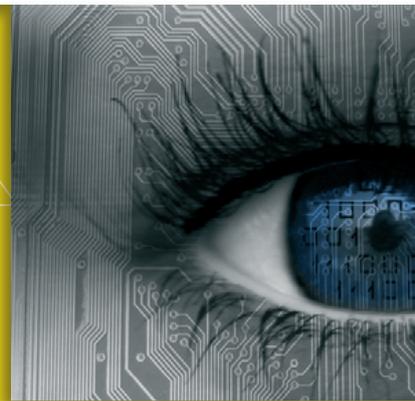


Andreas Oberweis ■ Christof Weinhardt
Henner Gimpel ■ Agnes Koschmider
Victor Pankratius ■ Björn Schnizler (Hrsg.)

eOrganisation: Service-, Prozess-, Market-Engineering 8. Internationale Tagung Wirtschaftsinformatik

Band 2



universitätsverlag karlsruhe

Andreas Oberweis / Christof Weinhardt / Henner Gimpel /
Agnes Koschmider / Victor Pankratius / Björn Schnizler (Hrsg.)

eOrganisation: Service-, Prozess-, Market-Engineering

8. Internationale Tagung Wirtschaftsinformatik

Band 2

eOrganisation: Service-, Prozess-, Market-Engineering

8. Internationale Tagung Wirtschaftsinformatik
Karlsruhe, 28. Februar – 2. März 2007

Band 2

Andreas Oberweis
Christof Weinhardt
Henner Gimpel
Agnes Koschmider
Victor Pankratius
Björn Schnizler
(Hrsg.)



universitätsverlag karlsruhe

Impressum

Universitätsverlag Karlsruhe
c/o Universitätsbibliothek
Straße am Forum 2
D-76131 Karlsruhe
www.uvka.de



Dieses Werk ist unter folgender Creative Commons-Lizenz
lizenziiert: <http://creativecommons.org/licenses/by-nc-nd/2.0/de/>

Universitätsverlag Karlsruhe 2007
Print on Demand

ISBN: 978-3-86644-095-1 (Band 2)
ISBN: 978-3-86644-094-4 (Band 1)
ISBN: 978-3-86644-093-7 (Set)

Inhaltsverzeichnis

Prozess-Engineering

eLearning: Anforderungen an das Content Management

Track-Chairs: Eric Schoop, Udo Winand, Martin Wessner

- Mass Customization und Serviceplattformstrategien im Blended Learning Engineering - konzeptionelle Grundlagen und evaluationsgestützte Erfahrungen* 3
Roland Gabriel, Martin Gersch, Peter Weber
- Generische Transformation von Learning-Content. Das MOCCA Projekt.* 21
Michael A. Herzog, Matthias Trier
- Reengineering der Content-Erstellungsprozesse in Industrieunternehmen durch Content-Modellierung: Fallbeispiel* 39
Pavlina Chikova, Katrina Leyking, Peter Loos, Eva-Maria Bruch, Lasse Lehmann
- Modularisierungskonzepte als Voraussetzung zur Gestaltung nachhaltiger Bildungsangebote am Beispiel der Aus-, Fort- und Weiterbildung von Ökonomielehrkräften* 57
Oliver Kamin, Karl Knispel

IS-Architekturen

Track-Chairs: Frank Leymann, Johannes Siedersleben, Elmar J. Sinz

- Eine Strukturvorlage zur effektiven Dokumentation von Software- und IT Architekturen* 77
Gernot Starke, Peter Hruschka
- Anwendungsarchitektur-Beurteilung unter Berücksichtigung zu erwartender Flexibilität* 89
Felix R. Müller
- Sizing Considerations for Enterprise Applications in Dynamic Data Centre Environments* 107
H. Thimm

Modellierung als Innovationsmotor

Track-Chairs: Ulrich Frank, Robert Winter

*Konfiguration von Informationsmodellen - Untersuchungen zu Bedarf und
Werkzeugunterstützung* 127
Patrick Delfmann, Ralf Knackstedt

*A Pattern based Approach for constructing Enterprise Architecture Management
Information Models* 145
**Sabine Buckl, Alexander M. Ernst, Josef Lankes, Kathrin Schneider,
Christian M. Schweda**

*Ein hierarchischer, architekturbasierter Ansatz zur Unterstützung des
IT-Business-Alignment* 163
Ronny Fischer, Robert Winter

*Fostering the Evaluation of Reference Models: Application and Extension of the
Concept of IS Design Theories* 181
Michael Schermann, Tilo Böhmann, Helmut Krcmar

*Modellbasierter Entwurf strukturanaloger Architekturen auf Basis der
Partitionierung von Graphen* 199
Stephan Aier, Marten Schönherr

*„Open Model“ – ein Vorschlag für ein Forschungsprogramm der
Wirtschaftsinformatik* 217
Ulrich Frank, Stefan Strecker, Stefan Koch

Softwareprozessverbesserung

Track-Chairs: Georg Herzwurm, Ralf Kneuper, Marcus A. Rothenberger

*From Communication to Implementation - A framework for understanding and
reducing unintentional complexity in software development processes* 237
Matthias Biggeleben

*Interaktive Entscheidungsunterstützung für die Auswahl von Software-Komponenten
bei mehrfachen Zielsetzungen* 255
Thomas Neubauer, Christian Stummer

*Ein Kennzahlensystem zur Optimierung von SAP-Systemen - Eine Fallstudie am
Beispiel der deutschen Automobilindustrie* 271
Eva Peggy Sekatzek, Helmut Krcmar

<i>Systematische Auswahl einer Methode zur Aufwandschätzung in der Softwareentwicklung</i>	289
Ulrike Dowie, Georg Herzwurm	
<i>Practical Application of SPICE in the Finance Sector</i>	307
Christian Völcker, Johann Leibig, Josef Maier, Holger Kühnl	
<i>Towards a Model-driven Development of Monitored Processes</i>	319
Christof Momm, Robert Malec, Sebastian Abeck	
<i>Management of Portal Evolution</i>	337
Hong Tuan Kiet Vo, Helmuth Elsner	

Supply Chain Management

Track-Chairs: Jens Gottlieb, Peter Mertens, Stefan Voß

<i>Regelbasierte Koordinierung von agentengestützten Transportprozessen</i>	355
Mathias Petsch, Dirk Pawlaszczyk, Hagen Schorcht	
<i>Distribution network design with postponement</i>	373
Frank Schwartz, Stefan Voß	
<i>Order-driven planning in build-to-order scenarios</i>	391
Thomas Volling, Thomas Spengler	

Techniken des Operation Research in der Wirtschaftsinformatik

Track-Chairs: Peter Korevaar, Dirk Mattfeld, Leena Suhl

<i>A Fast Look-ahead Heuristic for the Multi-depot Vehicle Routing Problem</i>	411
Juergen Branke, Christian Schmidt, Markus Withopf	
<i>Progress in solving large scale multi-depot multi-vehicle-type bus scheduling problems with integer programming</i>	429
Uwe H. Suhl, Swantje Friedrich, Veronika Waue	
<i>Optimierungssysteme für die Dienstplanung im ÖPNV</i>	447
Leena Suhl, Natalia Kliewer, Ingmar Steinzen	

<i>Barwertorientierte Projektplanung mit mehreren Akteuren mittels eines verhandlungsbasierten Koordinationsmechanismus</i>	465
Andreas Fink	
<i>Konzeption und Implementierung eines Planungssystems für Schnittstellentests</i>	483
Franz Rothlauf	
<i>Water Contamination Detection</i>	501
Armin Fügenschuh, Simone Göttlich, Michael Herty	

Wissensmanagement

Track-Chairs: Norbert Gronau, Marcus Spieß, Steffen Staab, Walter Waterfeld

<i>Third Generation Knowledge Management Systems - Towards an Augmented Technology Acceptance Model</i>	521
Kai Dingel, Sarah Spiekermann	
<i>Unterstützung wissensintensiver Prozesse im Produktlebenszyklus durch Suche in Produkt- und Prozessdaten</i>	539
Thomas Barth, Christian Lütke Entrup, Walter Schäfer	
<i>Mobile Knowledge Management Support in Fire Service Organisations</i>	557
Andrea Freßmann, Ralph Bergmann, Brian Taylor, Ben Diamond, Gary Carr-Smith	
<i>Kontextbasiertes Information Retrieval zur Unterstützung von Entwicklungsprozessen</i>	575
Andreas Henrich, Karlheinz Morgenroth	
<i>Towards Understanding the Dynamics of Communication Networks</i>	593
Matthias Trier, Annette Bobrik, Tilmann Bartels	
<i>Multiperspektivische Wissensvisualisierung für Wissensaustausch in heterogenen Community-Netzwerken</i>	611
Jasminko Novak	

Market-Engineering

Agenten- und Multiagenten-Technologien für betriebliche Anwendungen

Track-Chairs: Jörg P. Müller, Bernhard Bauer

Multi-agent Simulation for the transshipment problem with a non-negligible transfer lead time and a limited transportation mean capacity 631

Mohamed Hmiden, Lamjed Ben Said, Khaled Ghédira

eFinance

Track-Chairs: Hans Ulrich Buhl, Wolfgang Gaertner, Peter Gomber, Wolfgang König

Außerbörslicher Emittentenhandel: Handelsplatzwahl von Online-Investoren 651
Dennis Kundisch, Tobias Stuber

Travo-Projekt – Der Handel mit Aktionärsstimmen 669
Christoph Lattemann, Hartmut Klemens, Michael Durica

OR-basierte Persönliche Finanzplanung 687
Oliver Braun

An Order-Channel Management Framework for Institutional Investors 705
Bartholomäus Ende, Peter Gomber, Adrian Wranik

IT-Architekturmanagement in Banken - Ergebnisse einer leitfadengestützten Expertenbefragung 723
Christian Schmidt, Peter Buxmann, Zbynek Sokolovsky

MiFID-Readiness - Die Umsetzung der MiFID „Markets in Financial Instruments Directive“ in der deutschen Finanzindustrie 741
Peter Gomber, Markus Gsell, Claudia Reininger

eLearning Geschäftsmodelle

**Track-Chairs: Andrea Back, Michael H. Breitner, Uwe Hoppe,
Wolfgang Kraemer**

E-Learning-Geschäftsmodelle für Hochschulen. Entscheidungsunterstützung bei der strategischen Positionierung 761

Jan vom Brocke, Christian Buddendick

Ein Ansatz zur prozessorientierten Evaluation von Learning Management Systemen 779

Michael Luther, Matthias Trier

Modellbasiertes Curriculum-Design für Learning Management Systeme: Ein Integrationsansatz auf Basis von ARIS und IMS Learning Design 795

Guido Grohmann, Wolfgang Kraemer, Frank Milius, Volker Zimmermann

Einflussfaktoren für den Einsatz von Customer Focused E-Learning 813

Kai-Uwe Götzelt, Susanne Robra-Bissantz

Elektronische Märkte und elektronische Verhandlungen

Track-Chairs: Michael Rebstock, Mareike Schoop, Rudolf Vetschera

Strategisches Gebotsverhalten in Name-Your-Own-Price-Auktionen 833

Oliver Hinz

Design von Reputationssystemen in Online-Auktionen - Eine vergleichende Marktanalyse 851

Kai Riemer, Dorothee Korn

eMedia

Track-Chairs: Thomas Hess, Konrad Hilbers, Matthias Schumann

Intermediation in der TV-Branche: TV-Sender als Auslaufmodell? 871

Thomas Wilde, Konrad Hilbers, Thomas Hess

eServices: Anwendungen, Technologien und Management

Track-Chairs: Martin Bichler, Peter Buxmann, Lutz Heuser

Optimized Dynamic Allocation Management for ERP Systems and Enterprise Services 891

Valentin Nicolescu, Martin Wimmer, Raphael Geissler, Daniel Gmach, Matthias Mohr, Alfons Kemper, Helmut Krcmar

Analyse des Beitrages von Axiomatic Design zum Entwurf Serviceorientierter Architekturen 909

René Fiege, Dirk Stelzer

Services and resource profiles as metrics for the allocation of IT infrastructure costs 927

Reinhard Brandl

Informationsmanagement in der Energie- und Umweltinformatik

Track-Chairs: Wolf Fichtner, Wolfram Münch, Axel Tuma

TOP-Energy - Ein Framework für Softwarelösungen in der Energietechnik 947

Gregor Wrobel, Stefan Herbergs, Eckardt Augenstein

MDA-basierte Kopplung heterogener Informationssysteme im EVU-Sektor – ein Framework 955

Mathias Uslar, Niels Streekmann, Sven Abels

Rechtsfragen der Informationsgesellschaft

Track-Chairs: Thomas Dreier, Rupert Vogel, Andreas Wiebe

IT-Systeme in der Rechnungslegung und entsprechende Prüfungsanforderungen - Rechtliche Rahmenbedingungen: Aktueller Stand und Perspektiven 985

Elizaveta Kozlova, Ulrich Hasenkamp

Autorenverzeichnis 1003

Einführung in den Track

eLearning: Anforderungen an das Content Management

Prof. Dr. Eric Schoop

Technische Universität Dresden

Dr. Martin Wessner

Fraunhofer IPSI, Darmstadt

Prof. Dr. Udo Winand

Universität Kassel

Für das World Wide Web aufbereitete eLearning-Inhalte basieren heutzutage auf vielfältigen Technologien. Für die Erstellung, Pflege und Verwaltung von eLearning-Inhalten werden unterschiedlichste Methoden und IT-Systeme (Learning Content Management Systeme, Learning Management Systeme, WBT-Systeme, Autorensysteme, Wiki-Systeme) eingesetzt. Funktionalitäten von Content Management Systemen, die mittlerweile in unterschiedlichste für das eLearning einsetzbare IT-Systeme integriert sind, tragen erheblich zum wirtschaftlichen Umgang mit eLearning-Inhalten bei. Die Wirtschaftlichkeit des eLearnings sowie der Schutz von in eLearning-Content getätigten Investitionen wird von verschiedensten Faktoren beeinflusst, hierzu gehören u. a. Content-Engineering-Methoden, eLearning-Standards, der zweckgemäße Einsatz von IT-Systemen, die IT-Systeme selbst sowie Abhängigkeiten zwischen eLearning-Inhalten und IT-Systemen. Die Einhaltung internationaler eLearning-Standards kann die Interoperabilität von eLearning-Systemen bis zu einem bestimmten Grad ermöglichen, so dass eLearning-Inhalte in unterschiedlichen Systemumgebungen mehrfach genutzt werden können. Hinsichtlich der Produktion, Pflege und Weiterentwicklung von eLearning-Inhalten existieren hingegen in der Regel starke Abhängigkeiten zu IT-Systemen, so dass Migrationen hin zu anderen Produktionssystemen extrem aufwendig sein können. Für die Sicherstellung einer permanent hohen Qualität von eLearning-Produkten sind zudem IT-gestützte und organisatorische Qualitätsmanagement- und Qualitätssicherungsmaßnahmen bedeutsam.

Programmkomitee

Beate Bruns, time4you

Dr. Svenja Hagenhoff, Universität Göttingen

Prof. Dr. Thorsten Hampel, Universität Paderborn

Dr. Tilman Kuchler, IMC

Dr. Jan Pawlowski, Universität Duisburg-Essen

Prof. Dr. Eric Schoop, Technische Universität Dresden

Dr. Martin Wessner, Fraunhofer IPSI

Prof. Dr. Udo Winand, Universität Kassel

Mass Customization und Serviceplattformstrategien im Blended Learning Engineering

- konzeptionelle Grundlagen und evaluationsgestützte Erfahrungen

Roland Gabriel

Lehrstuhl für Wirtschaftsinformatik
Ruhr-Universität Bochum
44780 Bochum
rgabriel@winf.rub.de

Martin Gersch

Competence Center E-Commerce
Ruhr-Universität Bochum
44780 Bochum
martin.gersch@rub.de

Peter Weber

Lehrstuhl für Wirtschaftsinformatik
Ruhr-Universität Bochum
44780 Bochum
peter.weber@rub.de

Abstract

Wie können Universitäten Lern-Services anbieten, die bei einer konsequenten Zielgruppenorientierung (bis hin zur Lernindividualisierung) auch unter Kostengesichtspunkten wettbewerbsfähig realisierbar bleiben und damit didaktische und ökonomische Zielsetzungen erfolgreich zu integrieren vermögen? Der vorliegende Artikel zeigt mit dem Konzept der Mass Customization auf der Grundlage einer Serviceplattformstrategie eine auf hybride Lernarrangements ausgerichtete Möglichkeit hierzu auf und entwickelt entsprechende Standardisierungs- und Individualisierungsoptionen für das Blended Learning Engineering. Erste Evaluationsergebnisse entsprechend realisierter Lernarrangements bestätigen die konzeptionellen Überlegungen.

1 Eine sich transformierende Branche als Bezugsrahmen

Das Bildungswesen ist aktuell massiven Einflüssen ausgesetzt, die schon jetzt zu nachhaltigen Änderungen der Wertschöpfungs- und Wettbewerbsstrukturen führen. Zahlreiche neue (z.T. konsekutive) MBA-Studiengänge, die Einführung von Studiengebühren, die Vereinheitlichung von universitären Abschlüssen im Rahmen des Bologna-Prozesses, die Entwicklungen hin zu einem „lebenslangen Lernen“ sowie die Ausstattung der Universitäten mit Globalhaushalten veranschaulichen diesen Transformationsprozess für den Hochschulbereich eindringlich.¹ Unter anderem wird vor diesem Hintergrund, aus der Hoffnung heraus, eine größere Orts- und Zeitunabhängigkeit der Lehre sowie höhere Lernerfolge und mehr Lernzufriedenheit gewährleisten zu können, der Einsatz von E-Learning von zahlreichen Akteuren aktiv vorangetrieben, da sich diese hiervon einen Wettbewerbsvorteil versprechen bzw. hierin einen unverzichtbaren Schritt im Rahmen ihres Flexibilitätsmanagements sehen. Dem E-Learning fällt auch deshalb eine besondere Rolle zu, weil in Bezug auf die Leistungsdimensionen des Absatzobjektes Bildung² zum einen ein verbessertes Leistungsergebnis erreichbar erscheint (höherer Lernerfolg und größere Lernzufriedenheit), gleichzeitig aber auch ein unter Kostengesichtspunkten optimierter Leistungserstellungsprozess bzw. eine effizientere Bereitstellungsleistung diskutiert wird.

2 Mass Customization im Rahmen eines Blended Learning Engineering

Die in zahlreichen Veröffentlichungen und Konferenzdiskussionen angeführten Erfahrungen mit E-Learning und Blended Learning lassen vermuten, dass die für andere Branchen diskutierten generischen Wettbewerbsstrategien – entweder über einen nutzenbezogenen Differenzierungsvorteil oder über einen effizienzbezüglichen Kostenvorteil zu verfügen – auch im Bereich

¹ Branchentransformation bezeichnet den ergebnisoffenen Wandel einer Branche in einen veränderten Zustand und ist zu verstehen als eine Stufe im Rahmen einer branchenspezifischen E-Business-Entwicklung, die sich insbesondere durch tief greifende Änderungen der Markt-, Wettbewerbs- und Wertschöpfungsstrukturen auszeichnet [GeGo04, 1529ff.]; [Deis00, XVII].

² Die Unterscheidung der bei allen Leistungen identifizierbaren Leistungsdimensionen Bereitstellungsleistung, Leistungserstellungsprozess und Leistungsergebnis geht zurück auf Werner H. Engelhardt [Enge66] und hat insb. im Rahmen des Dienstleistungsmanagements eine starke Verbreitung gefunden. Die einzelnen Dimensionen sollen nach Engelhardt, Reckenfelderbäumer & Kleinaltenkamp [EnKR93] dabei wie folgt verstanden werden: Die Bereitstellungsleistung ist die Kombination der internen Potenzial- und Verbrauchsfaktoren, die eine Leistungserstellung ermöglicht. Der (finale) Leistungserstellungsprozess stellt die durch die Aktivierung dieser Bereitstellungsleistung ausgelöste Tätigkeit dar, bei der interne und gegebenenfalls externe Faktoren in einen Produktionsprozess integriert werden. Das Leistungsergebnis schließlich ist das Ergebnis des Leistungserstellungsprozesses, das geeignet ist, einen Nutzen für den Nachfrager zu stiften.

der Bildung relevant sind. Hieraus leitet sich vermeintlich das Erfordernis ab, als Bildungsanbieter entweder die eine oder die andere Strategierichtung für sich in Anspruch zu nehmen und die eigenen Lern-Services konsequent auf Kostenvorteile oder auf nutzenbezogene Differenzierungsvorteile auszurichten. Wie Frank Piller und Kathrin Möslin für den Bereich der Management Education zeigen, wird in der Praxis jedoch vielfach weder die eine noch die andere strategische Ausrichtung verfolgt, sondern statt dessen einem „Paradoxon“ Vorschub geleistet, welches sich durch auf formale Marktstandards ausgerichtete Leistungsergebnisse (z.B. in Form international vereinheitlichter Abschlüsse) in Verbindung mit zwar vermeintlich hoch flexiblen, aber auch kostenintensiven Leistungserstellungsprozessen auszeichnet [PiMö02]. Augenscheinlich werden die beiden Ausrichtungen in diesen Fällen in einer geradezu nachteiligen Weise kombiniert. Weder durch Standardisierung ermöglichte Kostenvorteile werden erzielt, noch werden Differenzierungsvorteile, die die aktuell aufwendig anmutenden Leistungserstellungsprozesse eigentlich vermuten lassen, nachhaltig realisiert und/oder kommuniziert. Statt jedoch der Fokussierung einer der generischen Strategiealternativen das Wort zu reden, wird nachfolgend eine komplementäre Kombination beider Strategierichtungen vorgestellt, die das aktuelle (Miss-)Verhältnis von Leistungserstellung und Leistungsergebnis konsequent neu interpretiert. Hierbei werden Lernangebote als Services interpretiert, Erkenntnisse der „integrativen Leistungslehre“ adaptiert sowie Erfahrungen aus anderen Servicebranchen genutzt. Eine aus Anbietersicht (teilweise) standardisierte Bereitstellungsleistung und Leistungserstellung wird mit einer zielgruppendifferenzierten Leistungsangebots(re)konfiguration auf der Grundlage so genannter „Serviceplattformen“³ und einem, soweit möglich und sinnvoll, individualisierten Leistungserstellungserlebnis durch den Nachfrager verbunden. Auch die in Abschnitt 3 dargestellten Evaluationsergebnisse entsprechend realisierter Lernarrangements stützen die bereits in anderen Serviceindustrien berichteten Erfahrungen, dass Standardisierung und Differenzierung/Individualisierung keineswegs unvereinbare Gegensätze darstellen. Weiterhin kann Standardisierung regelmäßig sogar mit einer – auch durch den Nachfrager empfundenen – Qualitäts-

³ Die Diskussion von „Plattformstrategien“ hat ihren Schwerpunkt im industriellen Bereich. Erste Übertragungen auf Services lassen aber ein großes Anwendungspotenzial auch im Servicebereich vermuten. Stauss charakterisiert in Anlehnung an Meyer und Lehnerd Serviceplattformen als entwickelte Sets von optionalen Teilelementen/-systemen und Schnittstellen, die eine mehrfach verwendbare Struktur bilden auf deren Grundlage immer wieder differenzierte Leistungsangebote effizient und effektiv entwickelt und realisiert werden können [Stau06, 322ff.]; [MeLe97]. In Abschnitt 2.2 wird die Plattformstrategie am Beispiel entwickelter Lehrveranstaltungstypen konkretisiert.

steigerung des Leistungsangebotes einhergehen.⁴ Eine zielgerichtete Standardisierung von Teilleistungen und Teilprozessen im Rahmen einer speziell auch didaktisch fundierten Modularisierungsstrategie offenbart derartige Möglichkeiten, zielgruppenspezifische oder gar individualisierte Lern-Services auch unter Kostengesichtspunkten wettbewerbsfähig zu realisieren und die generischen wettbewerbsstrategischen Ansätze auch im Bildungsbereich zu verbinden.

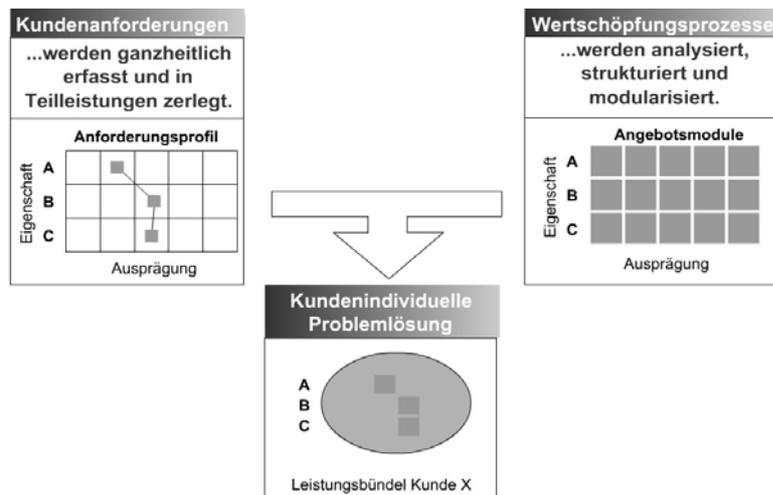


Abb. 1: Kundenindividualisierte Leistungen auf Basis modularisierter Wertschöpfungsprozesse [WeWe00, 480]

Der im Folgenden dargestellte Lösungsansatz baut auf einer standardisierten Vorgehensweise bei der Lernmodulproduktion auf, mit deren Hilfe aufeinander abgestimmte, strukturell ebenfalls standardisierte Lernobjekte erstellt werden, die im Prozess des didaktischen Designs⁵ mit weiteren Komponenten und insbesondere auch externen Faktoren⁶ zu zielgruppenspezifischen Leistungsbündeln verknüpft werden. Hierdurch wird mit Hilfe einer Modularisierung der Anteil vielfach verwendbarer Komponenten erhöht und zugleich über das didaktische Potenzial des Blended Learning eine negativ empfundene Standardisierung auf Seiten der Nachfrager ver-

⁴ Vgl. die in der Literatur dokumentierten Ergebnisse zum Beispiel im Bereich von Hotels, Restaurants sowie weiteren klassischen Dienstleistungen, z.B. bei: [BePa91]; [Hill88]; [Shos84]; [FiHH01]; [MeDo98]. Aktuell scheinen ähnliche Erfahrungen auch auf Seiten von deutschen Krankenhäusern gemacht zu werden, die nach einer Privatisierung mit anschließender Geschäftsprozessstandardisierung eine bis zu 30% Kostensenkung mit gleichzeitig deutlich verbesserten Kennziffern in Bezug auf Kunden- und Mitarbeiterzufriedenheit aufweisen können [Rhön05].

⁵ Didaktisches Design als didaktisch ausgerichteter Gestaltungsprozess berührt „den Lernprozess, die Lernzeit und den Lernort, die Lehrenden und Lernenden sowie die eingesetzten Medien. Diese Gesamtheit wird in der Regel als Lernarrangement bezeichnet.“ Didaktisches Design betrifft also in erster Linie Planung, Durchführung und Nachbereitung eines Lernangebotes“ [Trie05].

⁶ Externe Faktoren sind Faktoren, „... die zeitlich begrenzt in den Verfügungsbereich eines Dienstleistungsanbieters gelangen und mit den internen Produktionsfaktoren in einen Verarbeitungsprozess integriert werden. Mögliche externe Faktoren sind z.B. Personen (Nachfrager oder seine Mitarbeiter), Objekte, Tiere, Rechte, Nominalgüter und/oder Informationen. [EnKR93, 401].

mieden. Nachfolgend werden nun jeweils exemplarische Standardisierungs- und Individualisierungsansätze aufgezeigt. Abbildung 1 zeigt das Grundprinzip der Erstellung kundenindividueller Leistungen auf der Basis modularisierter Wertschöpfungsprozesse.

2.1 Standardisierung von Leistungserstellungsprozessen und Elementen der Bereitstellungsleistung

Eine Umsetzungsmöglichkeit soll am Beispiel einer konkreten universitären Veranstaltung aufgezeigt werden. Es handelt sich hierbei um eine Hauptstudiumsveranstaltung einer deutschen Universität in den Schwerpunktbereichen Marketing/Wirtschaftsinformatik, die sich inhaltlich den ökonomischen Konsequenzen von Digitalisierung und Vernetzung widmet und in Bezug auf das didaktische Konzept ein hybrides Setting aufweist, welches über eine Lernplattform koordiniert wird. Es werden unter anderem verschiedene Web-Based-Trainings (WBTs) verwendet, die in Abhängigkeit des Verwendungszeitpunktes und des thematischen Ankers unterschiedliche Einsatzzwecke verfolgen und daher in unterschiedliche Lernkontexte eingebettet sind.⁷ Die Veranstaltung ist dabei in einem größeren Gesamtzusammenhang zu sehen, da die verwendeten Lernmodule und Fallstudien als wichtige Teile der (E-Learning-) Bereitstellungsleistung im Rahmen eines durch das BMBF geförderten Projektes vorproduziert wurden.⁸

Die Erläuterung folgt der bereits angesprochenen Unterscheidung der Leistungsdimensionen Bereitstellungsleistung, Leistungserstellungsprozess und Leistungsergebnis, die sich bei jeder Art von Leistungserstellung identifizieren lassen und insbesondere im Bereich Services einen gut geeigneten Strukturierungsansatz repräsentieren. Es werden zudem explizit sowohl die vornehmlich auf Qualitätsdimensionen und somit indirekt auf mögliche Differenzierungsvorteile abzielende Nachfragerperspektive,⁹ als auch die Anbieterperspektive als eher kostenorientiertes

⁷ Dieser Aspekt wird im weiteren Verlauf des Beitrags noch näher erläutert.

⁸ In diesem Verbundprojekt mehrerer deutscher Universitäten wurde ein umfassender Pool an WBTs und Fallstudien erstellt, der nun von jedem Partner für eigene Lehrveranstaltungen genutzt werden kann. Die Spannbreite der Inhalte ist dabei so groß, dass ein äußerst flexibles Repertoire zur Anreicherung und Unterstützung von Lehrveranstaltungen besteht.

⁹ Als Ausgangspunkt der Qualitätsforschung wird häufig das Confirmation / Disconfirmation-Paradigma (CD-Paradigma) genannt, welches Qualität auf einen Vergleich der tatsächlichen Erfahrung eines Kunden bei Inanspruchnahme einer Leistung (Ist-Leistung) mit seinem Erwartungsniveau (Soll-Leistung) zurückführt. Somit ist der Abgleich von Standardisierungserwartung und erlebter Standardisierung durch den Bildungsnachfrager im vorliegenden Anwendungsfall als ein wichtiges qualitätsbildendes Element zu betrachten. Vgl. hierzu u.a. [Flie06, 136ff.]; [Gers95]; [HoSt05]; [StHe92]; [ZeBP88].

Pendant berücksichtigt. Durch diese integrierte Betrachtung der Anbieter- und Nachfragerperspektive sind die beiden Hauptansatzpunkte der Porterschen Strategiealternativen – Qualität und Kosten – interdependente Teile des Konzepts und Rückgrat der Mass Customization-Strategie.¹⁰ Tabelle 1 gibt einen Überblick über die verschiedenen Ebenen, an denen die nachfolgend exemplarisch dargestellten Standardisierungen und Individualisierungen ansetzen.

Standardisierung	Ebene	Individualisierung
Konzipierung von Veranstaltungsgrundtypen als Serviceplattformen	Veranstaltungen	Kombination der Veranstaltungsphasen; Individualisierung des Ablaufs, auch durch Einfluss der externen Faktoren
Standardisierte Abläufe und Vermittlungskomponenten innerhalb von Teilarrangements mit spezifischen Lernzielen	Veranstaltungsphasen	Zielgruppenorientierte Kombination der Komponenten; Ergänzung durch individuelle Leistungskomponenten; konstruktivistische Lernprozesse bei Fallstudienbearbeitung
Vereinheitlichung von Aufbau, Bedienung, Design und technischen Schnittstellen; Vorgabe eines didaktischen Modells	Komponenten (hier WBTs)	Eigenindividualisierung der Lernpfade; Eigenindividualisierung durch Wahl der Vermittlungsalternative bzw. der in Anspruch genommenen WBTs

Tab. 1: Ebenen der Standardisierung und Individualisierung

Ein wichtiger Ansatzpunkt der Standardisierung im Rahmen der Bereitstellungsleistung auf Komponentenebene bezieht sich auf die WBTs, somit auf einen wesentlichen Träger der Wissensvermittlung in den späteren Leistungsangeboten (Lernarrangements). Sowohl der Produktionsprozess als auch die Struktur sowie die enthaltenen Elemente der WBTs selbst wurden standardisiert, so dass konzeptionell sowohl eine Prozessstandardisierung des Erstellungsvorganges als auch eine Ergebnis(teil-)standardisierung vorgenommen wurde. Indirekt werden darüber hinaus zum Beispiel durch die Interfacegestaltung sowie die ermöglichten Lernpfade durch die WBTs auch spätere Nutzungsprozesse sowie die Eingriffs- und Gestaltungsmöglichkeiten des externen Faktors mehr oder minder stark standardisiert. In Bezug auf die Prozessstandardisierung ist hier die Unterscheidung von Bereitstellungsleistung und finaler Leistungserstellung von hoher Bedeutung, da sich die Leistungserstellung aufgrund der Beteiligung der Studierenden als

¹⁰ Das aus den Begriffen Mass Production und Customization zusammengesetzte Oxymoron Mass Customization [Pine93] bezeichnet ein – zumeist technologisch gestütztes – Konzept zur Auflösung der u.a. von Michael Porter postulierten und für viele Branchen über lange Zeit nicht angezweifelte Gegensätzlichkeit der generischen Wettbewerbsstrategien Kostenführerschaft oder Differenzierung [Port99]; [PiSc99, 1114f.]. Kunden(gruppen) differenzierte Leistungsangebote werden durch Mass Customization zu einem der Massenproduktion vergleichbarem Kostenniveau realisiert und angeboten. Für eine ausführliche Darstellung sei insbesondere verwiesen auf [Pill06].

hoch integrativer Prozess darstellt, so dass die Standardisierungspotenziale speziell auch von der Gleichförmigkeit der externen Faktoren abhängen.¹¹ So fällt auch im vorliegenden Fall die Prozessstandardisierung im Rahmen der autonomen Bereitstellungsleistung deutlich intensiver aus als die der späteren Durchführung des Lernarrangements beziehungsweise der ermöglichten Lern-/Nutzungspfade durch den Nutzer, die aber dennoch ebenfalls eine Determinierung erfahren. Der Lernmodulproduktionsprozess besteht im Wesentlichen aus den folgenden drei Schritten:

- Zunächst wird auf der Basis der verfügbaren Rohmaterialien eine *Inhaltsskizze* erstellt, die den Inhalt eines WBT kurz zusammenfasst, das WBT in den Gesamtkontext der verfügbaren Selbstlerneinheiten einordnet und grundlegende didaktische Optionen und Entscheidungen, wie z.B. Lernziele, die verwendeten Darstellungsformen, vorgesehene Abbildungen, Animationen oder Videos aufzeigt.
- Die *Feingliederung* als zweiter Schritt dient der Strukturierung der Inhalte entsprechend der Vorgaben des didaktischen Modells, welches im konkreten Anwendungsfall für alle WBT-Autoren verbindlich festgelegt wurde. Die Inhalte werden disaggregiert und jeweils einem bestimmten Komponententyp zugeordnet. Die Feingliederung dient gleichzeitig als Statuskontrolle im Rahmen kooperativer Produktionsprozesse, da z.B. die Produktion von Animationen, Abbildungen und Videos Spezialisten übertragen worden ist, die durch Drehbücher gesteuert werden.
- Auf der Basis der Feingliederung werden anschließend die *Komponenten* produziert, d.h. die Inhalte geschrieben, Multimediaelemente erstellt und die einzelnen Elemente zu Kernaussagen, Vertiefungen, Beispielen, Übungen oder Lernkontrollen zusammengefügt. Jede Komponente verfügt über eine spezifische, im didaktischen Modell festgelegte Position und Funktion, so dass grundsätzlich eine (Re-)Kombination der einzelnen Komponenten zu neuen WBTs möglich ist.¹² Um eine Einhaltung der Vorgaben zu erreichen, wird jede Stufe des Produktionsprozesses

¹¹ Diesbezüglich wird in der Literatur kontrovers diskutiert, ob autonome Prozesse einfacher zu standardisieren seien als integrative Prozesse. Für eine kritische Diskussion vgl. u.a. [Flie06, 170ff.]; [Gers95, 29ff.].

¹² Es sei jedoch betont, dass dieses Potenzial nur in Grenzen und nur bei einer ausreichenden didaktischen Kontrolle und ggf. Anpassung der Komponenten nutzbar ist, um nicht die Konsistenz der Selbstlerneinheiten zu gefährden.

durch einen zweiten Modulautor begleitet („Vier-Augen-Prinzip mit Coaching-Effekt“).

Im Prozess des didaktischen Designs werden die WBTs mit anderen Komponenten und Elementen der traditionellen Lehre zu Lernarrangements kombiniert. Die Standardisierung der WBTs bezieht sich vornehmlich auf die enthaltenen Komponenten, die Bedienung, das Design, die technischen Schnittstellen sowie das didaktische Modell (u.a. ermöglichte Lernpfade, Navigationsmöglichkeiten sowie zum Teil alternative Formen der Wissensaneignung) und betrifft sowohl die Anbieter- als auch die Nachfragerperspektive. Wie bereits dargelegt wurde, gilt es bei einer nachfrageorientierten Betrachtung die Standardisierungserwartungen sowie die tatsächlich erlebte Standardisierung der Nachfrager zu berücksichtigen, um die integrativen Leistungserstellungs- und Nutzungsprozesse sowie die angebotenen Leistungs(teil)ergebnisse hierauf abstellen zu können. Wie noch anhand der Evaluationsergebnisse gezeigt wird, empfinden die Studierenden beispielsweise die Standardisierung der WBT-Struktur als äußerst positiv. Ein wichtiger Grund hierfür dürfte in der kognitiven Belastung der Lernenden durch die neue Form des Lernens liegen. Die Nutzung der WBTs muss daher schnell zur Selbstverständlichkeit werden können, was durch Standardisierung gefördert wird.

Die Standardisierung auf Ebene der Veranstaltungsphasen bezieht sich auf eine idealtypische Vorkombination von Leistungskomponenten, die als Teilarrangements bestimmte Zielsetzungen und Abläufe repräsentieren. So ist beispielsweise bei Beginn insbesondere interdisziplinärer Veranstaltungen häufig eine Phase erforderlich, die es den Studierenden der verschiedenen Vertiefungsrichtungen erlaubt, das Basiswissen der einzelnen Teildisziplinen auszubauen. Hierzu bietet sich ein Teilarrangement mit schwerpunktmäßigem Selbststudium an, bei dem z.B. WBTs die Grundlageninhalte zeit- und ortsunabhängig bereitstellen und Diskussionsforen zur Klärung von Fragen verfügbar sind, die tutoriell aktiv betreut werden. Im weiteren Verlauf der Veranstaltung bzw. in Veranstaltungen mit anderen Zielsetzungen, bei denen eine problemorientierte Auseinandersetzung mit praxisnahen Fragestellungen in Gruppenarbeit im Vordergrund stehen soll, bietet sich die Verwendung einer speziell kollaborativen Veranstaltungsphase an, die mit Hilfe multimedialer Fallstudien eine aktive Wissenskonstruktion, losgelöst von objektiven Wissensbeständen fördert [Löbl06]. Im weiteren Verlauf des Beitrags wird ein Beispiel für ein dreigeteiltes Lernarrangement dargestellt – hierbei werden beispielhaft drei aufeinander aufbauenden Veranstaltungsphasen skizziert.

Auf Veranstaltungsebene schließlich stellen die Veranstaltungsgrundtypen als Serviceplattformen die Grundlage für verschiedene Leistungsangebote dar. Sie setzen sich aus idealtypischen Veranstaltungsphasen, Leistungspotenzialen (WBTs, Fallstudien, Vortragsaufzeichnungen, Betreuung, usw.), Prozessen und Schnittstellen zusammen, die gemeinsam die Grundlage zur Entwicklung und Realisierung immer wieder differenzierter Leistungsangebote darstellen.

2.2 Didaktisches Design und Individualisierung im Blended Learning Engineering auf der Grundlage von Serviceplattformstrategien

Blended Learning wird hier verstanden als ein Konzept, das die Möglichkeiten der Digitalisierung und Vernetzung u.a. über Internet oder Intranet mit klassischen Lernformen und -medien in einem zielgerichteten Lernarrangement verbindet [SaSB04, 68].¹³ Das anerkannte didaktische Potenzial von Blended Learning kennzeichnet im vorliegenden Kontext aber nicht nur die zielgerichtete Kombination von E-Learning und Präsenzlehre oder die Integration methodischer Ansätze wie etwa Selbststudium und gruppenbasiertes Lernen, sondern auch die Zusammenführung von standardisierten Teilleistungen mit – zum Teil auch persönlichen – Interaktionsprozessen im Rahmen der integrativen Leistungserstellung, die in Summe das Leistungsbündel des Lernarrangements ausmachen. Der hier verwendete Begriff des Blended Learning Engineering adaptiert ein Verständnis von Software Engineering, das von Helmut Balzert als „zielorientierte Bereitstellung und systematische Verwendung von Prinzipien, Methoden und Werkzeugen für die arbeitsteilige, ingenieurmäßige Entwicklung und Anwendung von umfangreichen Software-Systemen“ definiert wird [Balz00, 36].¹⁴ Der Begriff Blended Learning Engineering kann somit wie folgt konkretisiert werden:

- Im Mittelpunkt steht die zielorientierte Bereitstellung¹⁵ und systematische Verwendung von Prinzipien (hier u.a. Mass Customization und Serviceplattformstrate-

¹³ Der Begriff „hybrides Lernarrangement“ wird vorliegend als Synonym für den Begriff „Blended Learning-Arrangement“ verwendet. Der Begriffsbestandteil ‚hybrid‘ bringt dabei den Kerngedanken des Blended Learning zum Ausdruck.

¹⁴ Anzumerken ist, dass Balzert mit dem Zitat den deutschen Begriff des Software Engineering („Software Technik“) definiert.

¹⁵ Die Zielorientierung erfolgt im Hinblick auf die Dimensionen Kosten, Zeit, Qualität und/oder Flexibilität.

gien), Methoden und Werkzeugen, für die Entwicklung und Realisierung hybrider Lehr- und Lernangebote.

- Die Entwicklung und Durchführung von hybriden Lernarrangements können als systematische, arbeitsteilige Prozesse betrachtet werden, die eine streckenweise gleichförmige Realisierung zulassen.

Die Zusammenführung der Komponenten spielt eine zentrale Rolle, da hier Sorge dafür zu tragen ist, dass das zu konzipierende Lernarrangement – auf wettbewerbsfähigem Kostenniveau – den Qualitätsansprüchen der Studierenden gerecht wird und sich möglichst positiv von konkurrierenden Angeboten abhebt. Dieser Prozess kann aufgrund seiner Vielseitigkeit und Komplexität im Detail nicht Gegenstand dieses Beitrags sein. Es geht zunächst vielmehr um die Verdeutlichung von Optionen und um den Nachweis, dass bei dem gewählten didaktischen Design das dargestellte Standardisierungskonzept mit seinen skizzierten Möglichkeiten tatsächlich keine negativen Auswirkungen auf die Qualitätseinschätzung der Studierenden hat. Die Zusammenführung der modularisierten Teilleistungen stellt zudem nur ein Individualisierungsinstrument neben weiteren dar, die im Folgenden aufgezeigt werden. Abbildung 2 illustriert exemplarisch das Ergebnis eines Designprozesses in Form der Konzeptdarstellung einer hybriden Veranstaltung, um die vielfältigen Möglichkeiten zur Kombination der einzelnen Komponenten und Elemente sowie die daraus resultierenden Individualisierungspotenziale anzudeuten. Auch die bereits dargestellten Ansatzpunkte der Standardisierung können auf diese Weise noch einmal in einem konkreten Kontext nachvollzogen werden. Das realisierte Lernarrangement besteht aus insgesamt drei Phasen, die sich hinsichtlich der angestrebten Lernprozesse als „Individuelles Lernen“, „Tutoriell unterstütztes Lernen“ und „Kooperatives Lernen“ charakterisieren lassen. Jede der drei Phasen beinhaltet entsprechend des vorgesehenen Verwendungszwecks unterschiedliche Komponenten, die zudem eine zielgerichtete Gewichtung aufweisen. In der Phase des individuellen Lernens, die auf eine Angleichung der Wissensstände der aus verschiedenen Schwerpunktbereichen stammenden Studierenden abzielt, stehen beispielsweise, im Gegensatz zur Phase des tutoriell unterstützten Lernens, die WBTs im Vordergrund, während anschließend schwerpunktmäßig die Vorlesungen, ergänzt durch eine über eine Lernplattform (LP) tutoriell unterstützte Lernmodulbearbeitung, der Erarbeitung neuer Inhalte dienen. In der dritten Phase können die behandelten Inhalte der Veranstaltung durch die Studierenden anhand zusätzlicher WBTs zur Prüfungsvorbereitung vertieft werden und es werden mit Hilfe einer multimedialen Fallstudie kooperative Lernprozesse angestoßen, die im Sinne des konstruktivistischen Lernpa-

radigmas eine aktive Wissenskonstruktion fördern und damit für einen Teil des Lernarrangements die Lernprozesse und den Lerngegenstand individualisieren.¹⁶

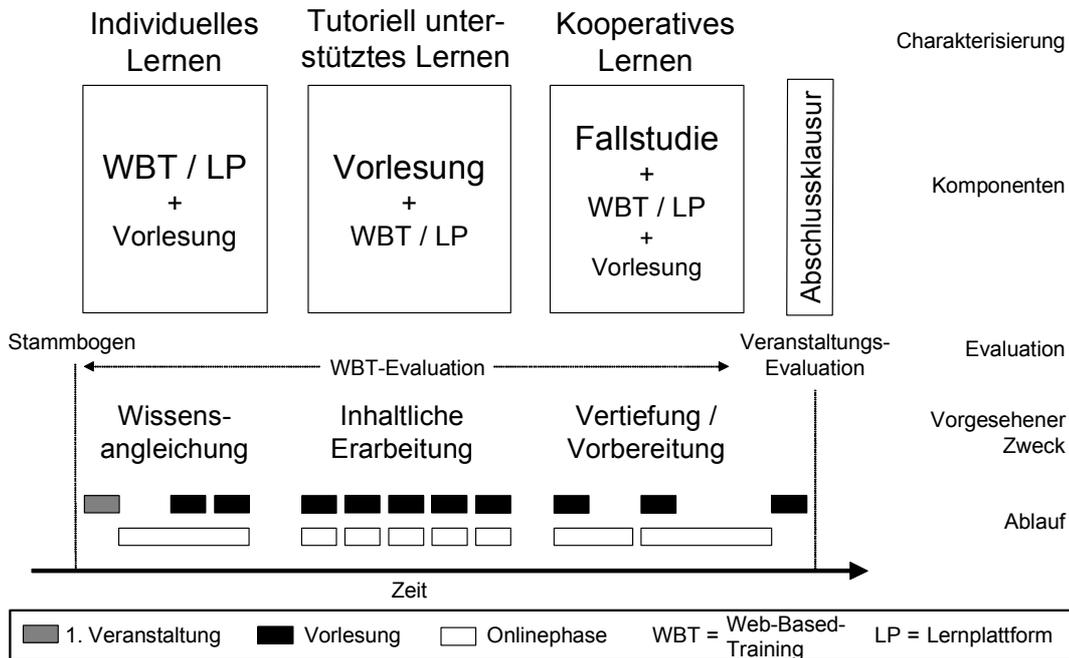


Abb. 2: Konzeptdarstellung Lernarrangement „Blended Learning/Lernaktive Vorlesung“

Die vornehmlich ökonomisch begründete Standardisierung kann bei dem vorgestellten Konzept der Mass Customization insbesondere mit didaktisch geprägten Individualisierungsmöglichkeiten kombiniert werden, die sowohl nachfragerseitig (Eigenindividualisierung) als auch anbieterseitig ansetzen:

Nachfragerseitige Eigenindividualisierung (Beispiele)

- Die Lernenden können die Lernpfade in Bezug auf die Auswahl und Reihenfolge der in Anspruch genommenen Informationsquellen weitgehend frei bestimmen. Vorgaben von Seiten der Lehrenden werden lediglich in Form von vorgegebenen Veranstaltungsphasen gegeben, die jeweils bestimmte inhaltliche Schwerpunkte

¹⁶ Im Gegensatz zum Objektivismus, der von einer objektiven, erfahrbaren Wirklichkeit als Lerngegenstand ausgeht wird beim Konstruktivismus eine entsprechend der gesonderten Wahrnehmung eines jeden Individuums individuell konstruierte und interpretierte Realität unterstellt. Damit steht bei konstruktivistischen Lernumgebungen die Initiierung von Lernprozessen und nicht die Vermittlung konkreten Wissens im Vordergrund. Nach [Bolz02, 27] lässt sich ein starker Bezug zwischen den Forderungen des Konstruktivismus und den Stärken der Fallstudienmethode feststellen.

setzen und in Bezug auf den Veranstaltungsablauf spezifische Funktionen übernehmen (s. o.).¹⁷

- Ebenfalls obliegt es an verschiedenen Stellen den Lernenden selbst, die subjektiv hilfreichste Form der Inhaltsdarstellung aus einem Kanon alternativer Aufbereitungsvarianten zu wählen. Bei der vorliegenden Veranstaltung stehen im Wesentlichen WBTs, Aufsätze, Vorlesungen und Fachvorträge von externen Referenten zur Verfügung.
- Auch innerhalb der WBTs, die nach einem speziellen didaktischen Modell aufgebaut sind, können individuelle Lernpfade gewählt werden.¹⁸
- Darüber hinaus individualisieren die Lernenden bei der Bearbeitung der konstruktivistisch ausgestaltete Fallstudien Teile der Lernprozesse und des Lerngegenstands.¹⁹

Anbieterseitige Individualisierung (Beispiele)

- Die teilstandardisierten idealtypischen Lernarrangementphasen können entsprechend der Zielsetzungen der Lehrveranstaltung und auf der Grundlage entwickelter Veranstaltungsgrundtypen (Serviceplattformen) kombiniert werden.²⁰ Sie bestimmen so die Rahmenstruktur der Veranstaltung.

¹⁷ So kann beispielsweise jeder Veranstaltungsteilnehmer in der Phase des individuellen Lernens eigenständig darüber entscheiden, welche Grundlagen er mit Hilfe der online zur Verfügung gestellten WBTs und Texte bearbeitet bzw. wiederholt. Die Präsenzveranstaltungen in dieser Veranstaltungsphase, in denen die Inhalte diskutiert und vertieft werden, helfen den Teilnehmern dabei, ihren eigenen Leistungsstand einzuschätzen und etwaige Lücken auszumachen. Bei ähnlichem Inhalt in verschiedener Aufbereitungsform (z.B. Text und WBT) ist dies kenntlich gemacht. Die Lernenden wählen die ihnen jeweils individuell geeignet erscheinende Lernform.

¹⁸ Es handelt sich bei dem didaktischen Modell um den so genannten Kernaussagenansatz, der auf einem Set möglicher Modul-Komponenten basiert (Kernaussagen, Vertiefungen, Übungen, Beispiele, ...). Die jeweils in den Kernaussagen enthaltenen Inhalte können vom Lerner entsprechend seiner individuellen Neigungen durch Bearbeitung der optionalen Zusatzkomponenten hinterfragt bzw. vertieft werden. Die Komponenten repräsentieren dabei auch unterschiedliche Aufbereitungsformen. Zum Kernaussagenansatz siehe [GeMa03].

¹⁹ Zu den Möglichkeiten konstruktivistischer Lernprozesse im Bereich der Managementlehre siehe [Löbl06].

²⁰ Vgl. zur Übertragbarkeit der „Plattformstrategien“ auf den Dienstleistungsbereich die Überlegungen in Abschnitt 2 sowie insb. [Stau06]. Bisher wurden von den Autoren drei Veranstaltungsgrundtypen im Detail entwickelt und eingesetzt [(1) „Lernaktive Vorlesung“; (2) „Fallstudienbasierte Übung“ und (3) „Medialisierte Vortrag“], die – gemeinsam mit einer Service-/Leistungsphilosophie – den Charakter von Serviceplattformen als systemische Grundlagen des Leistungsangebotes haben. Durch Modifikation der explizit vorgesehenen Gestaltungsdimensionen (z.B. einzel-/gruppenorientierte Aufgabenstellung; lokale, nationale oder internationale Zusammenarbeit; kollektiver oder kompetitiver Aufgabencharakter) lassen sich diese Serviceplattformen vielfach variieren. Das Konzept Mass Customization nutzt somit die Möglichkeiten der hiermit verbundenen Baukasten- / Modularisie-

- Innerhalb der Phasen eines Lernarrangements können wiederum die modularisierten Komponenten zielgruppenspezifisch kombiniert werden. Dabei kann nicht nur der Umfang und die inhaltliche Ausrichtung sondern auch die Auswahl der entsprechend der Zielgruppenanalyse adäquat erscheinenden Lernformen bestimmt werden.
- Auch auf der Ebene Leistungskomponenten wird durch eine Auswahl geeigneter Aufbereitungsformen eine Ausrichtung auf die Zielgruppe vorgenommen. So stehen die Fallstudien beispielsweise als reine Online-, aber auch als Hybridversionen zur Verfügung, wobei letztere vor allem auch Präsenzelemente in das Setting integrieren. Auf dieser Ebene findet somit auf Grundlage einer Zielgruppenanalyse eine Spezifizierung des Komponentenangebotes statt, wobei die Komponenten aber weiterhin – als wichtigen Konzeptbestandteil – zielgerichtet gestaltete Möglichkeiten für eine Eigenindividualisierung durch die Lernenden bieten.

3 Exemplarische Evaluationsergebnisse und Schlussfolgerungen

Für Veranstaltungen sowie einzelne Komponenten wurde ein dreischrittiges Evaluationsverfahren konzeptionell entwickelt und angewendet. Einer Erhebung der Stammdaten, Einstellungen und Erwartungen der Teilnehmer mit Hilfe eines Stammbogens zu Beginn des ersten Präsenztermins schließen sich Evaluationen der einzelnen verwendeten WBTs an. Abschließend wird nach Ablauf der Veranstaltung eine multiattributive merkmalsorientierte Qualitätsmessung des gesamten Lernarrangements und einzelner Leistungskomponenten durchgeführt. Die Ergebnisse aus dem Sommersemester 2005 zeigen deutlich das Potenzial der gewählten Vorgehensweise und sollen hier als erster Hinweis dafür dienen, dass das auf den ersten Blick effizienzorientierte Standardisierungskonzept zugleich auch einer nutzenbezogenen Wettbewerbsfähigkeit zuarbeitet bzw. dieser zumindest nicht im Wege steht.

Statement	Durchschnittswert*	Standardabw.
Ein Bezug zwischen Theorie und Praxis wurde hergestellt.	1,65	0,55
Ich lernte etwas Wichtiges.	1,53	0,73
Die Zusammenstellung der Themen war interessant.	1,61	0,72
Die Tatsache, dass Gastreferenten eingeladen wurden, finde ich gut.	1,35	0,55

rungslogik bei der (Re)Konfiguration und Realisierung konkreter Leistungsangebote auf der Grundlage derartiger „Serviceplattformen“.

Die Unterstützung der Veranstaltung durch Blackboard erleichtert das Lernen.	1,58	0,76
Ich würde Blackboard jederzeit wieder benutzen, wenn ich die Möglichkeit dazu hätte.	1,59	0,98
Die Vermittlung von Inhalten durch die bereitgestellten Lernmodule auf Blackboard fand ich gut.	1,52	0,63
Die Lernmodule konnten die Inhalte insgesamt nicht effektiv vermitteln.	3,19	0,70
Das eigenständige Erarbeiten von Inhalten mit Hilfe der Lernmodule hat mit gut gefallen und den Lernerfolg gefördert.	2,10	0,88
Den Anteil und die Bedeutung von Lernmodulen an der Veranstaltung war zu hoch.	2,90	0,84
Für jede Veranstaltung sollten zumindest ergänzend Lernmodule bereitgestellt werden.	1,87	0,86
Gesamtnote** für die Veranstaltung:	1,84	0,64
Gesamtnote** für den (die) Dozenten:	1,39	0,56
Die Veranstaltung hat mich inhaltlich interessiert.	1,48	0,57
Meine Einstellung zu E-Learning hat sich im Verlaufe der Veranstaltung positiv entwickelt.	1,87	0,85
Alter	25,71	2,30
Fachsemester	7,89	3,58
Lernmoduleevaluation (über alle bisherigen Veranstaltungen, Stand: Juli 2006)		
Gesamtnote**	2,35	(1243 Antw.)
Struktur des Lernmoduls**	2,15	(1227 Antw.)
Nachvollziehbarkeit der Inhalte**	2,20	(1242 Antw.)
* Zustimmung zum vorgegebenen Statement auf einer Skala von 1 bis 4 (1 = trifft voll zu ... 4 = trifft nicht zu), keine Antwort = Enthaltung		
** Gesamtnoten nach Schulnotensystem: 1 = sehr gut ... 6 = ungenügend		

Tab. 2: Evaluationsergebnisse des Lernarrangements aus dem SS 2005 (31 Teilnehmer)

Die Evaluationsergebnisse stützen die Ausgangsthese, Lern-Services durch einen zielgerichteten Standardisierungsansatz unter Nutzungsgesichtspunkten wettbewerbsfähig zu realisieren und zugleich durch die einhergehende Modularisierung ein Potenzial für eine nachhaltige Kostendegression aufzubauen. Moderne IT- und mediengestützte Lehre auf Seiten der Universitäten muss von Anfang an in ein Strategiekonzept eingekleidet werden, das den einzelnen Akteuren die notwendige Flexibilität erhält und zugleich auf die sich bereits deutlich abzeichnenden Konsequenzen des Transformationsprozesses der Branche Bildung abstellt. Mass Customization bietet, gerade in Kombination mit den Potenzialen von Blended Learning, große Möglichkeiten, ökonomische Vorteile unter didaktisch sinnvollen Prämissen zu erreichen. Der Forderung nach einem qualitätsorientierten Wettbewerb der Universitäten ist unbedingt zuzustimmen und es wurde mit dem Vorschlag einer Mass Customization bei Lern-Services auf der Basis von Serviceplattformen ein auf Standardisierungspotenziale ausgerichtetes Konzept hierfür vorgestellt. In einem nächsten Schritt beteiligen sich die Autoren auf der Basis der positiven Erfahrungen auf der Ebene der Einzelveranstaltungen an einem universitätsübergreifenden Projekt, welches die Grundidee des hier vorgestellten Ansatzes auf einen Studiengang überträgt und sich wiederum an konkreten Evaluationsergebnissen messen lassen wird.

Literaturverzeichnis

- [Balz00] Balzert, Helmut: Lehrbuch der Software-Technik: Software-Entwicklung. 2. Aufl., Spektrum Akademischer Verlag, Heidelberg; Berlin 2000.
- [BePa91] Berry, Leonard L.; Parasuraman, A.: Marketing Services: Competing through Quality. Free Pr., New York 1991.
- [Bolz02] Bolz, Andre: Multimedia-Fallstudien in der betriebswirtschaftlichen Aus- und Weiterbildung. Eul, Lohmar; Köln 2002.
- [BüLu97] Büttgen, Marion; Ludwig, Marc: Mass Customization von Dienstleistungen. In: Die Betriebswirtschaft, 57 (1997) 6, S. 857-858.
- [Deis00] Deise, Martin V. et al.: Executive's Guide to E-Business. Wiley, New York et al. 2000.
- [Enge66] Engelhardt, Werner H.: Grundprobleme der Leistungslehre, dargestellt am Beispiel der Warenhandelsbetriebe. In: Zeitschrift für betriebswirtschaftliche Forschung, 18 (1966), S. 158-178.
- [EnKR93] Engelhardt, Werner H.; Kleinaltenkamp, Michael; Reckenfelderbäumer, Martin: Leistungsbündel als Absatzobjekte. In: Zeitschrift für betriebswirtschaftliche Forschung, 45 (1993) 5, S. 395-426.
- [FiHH01] Fischer, Marc; Herrmann, Andreas; Huber, Frank: Return on Customer Satisfaction. Wie rentabel sind Maßnahmen zur Steigerung der Zufriedenheit? In: Zeitschrift für Betriebswirtschaft (ZfB), 71 (2001) 10, S. 1161-1190.
- [Flie06] Fließ, Sabine: Prozessorganisation in Dienstleistungsunternehmen, Kohlhammer, Stuttgart 2006.
- [Gers95] Gersch, Martin: Die Standardisierung integrativ erstellter Leistungen. Arbeitsbericht Nr. 57 des Instituts für Unternehmensführung und Unternehmensforschung, Bochum 1995.

- [GeGo04] Gersch, Martin; Goeke, Christian: Entwicklungsstufen des E-Business. In: Das Wirtschaftsstudium (wisu), 33 (2004) 12, S. 1529-1534.
- [GeMa03] Gersch, Martin; Malinowski, Tanja: Didaktische Aufbereitung zur Gestaltung von Lernsoftware – Dargestellt an einem konkreten WBT-Entwicklungsbeispiel, Arbeitsbericht Nr. 03-46 des Lehrstuhls für Wirtschaftsinformatik, Bochum 2003.
- [Hill88] Hill, Charles W.: Differentiation vs. low cost or differentiation and low cost. Academy of Management Review, 13 (1988) 3, S. 401-412.
- [HoSt05] Homburg, Christian; Stock, Ruth: Kundenzufriedenheit und Kundenbindung bei Dienstleistungen – eine theoretische und empirische Analyse. In: Corsten, Hans; Gössinger, Ralf (Hrsg.): Dienstleistungsökonomie. Duncker & Humblot, Berlin 2005, S. 301-327.
- [Löbl06] Löbler, Helge: Learning Entrepreneurship from a Constructivist Perspective. In: Technology Analysis & Strategic Management, 18 (2006) 1, S. 1-20.
- [MeDo98] Meyer, Anton; Dornach, Frank: Jahrbuch der Kundenzufriedenheit. FGM-Verlag, München 1998.
- [MeLe97] Meyer, Marc; Lehnerd, Alvin: The Power of Product Platforms. Free Press, New York 1997.
- [Pill06] Piller, Frank: Mass Customization – Ein wettbewerbsstrategisches Konzept im Informationszeitalter. 4. Aufl., DUV, Wiesbaden 2006.
- [PiMö02] Piller, Frank; Möslein, Kathrin: Are we practicing what we preach? - Strategic perspectives for the management education industry, Proceedings of the EURAM Conference 2002, Stockholm, 9-11 May, 2002.
- [PiSc99] Piller, Frank, Schoder, Detlef: Mass Customization und Electronic Commerce. In: Zeitschrift für Betriebswirtschaft (ZfB), 69 (1999) 10, S. 1111-1136.
- [Pine93] Pine, B. Joseph: Mass Customization. Harvard Business School Press, Boston 1993.

- [Port99] Porter, Michael: Wettbewerbsstrategien. 10. Aufl., Campus Fachbuch, Frankfurt a.M. 1999.
- [Rhön05] Rhön Klinikum: Geschäftsbericht der Rhön Klinikum AG 2004. Bad Neustadt, Saale. <http://www.rhoen-klinikum-ag.com/internal/download/04GBd.pdf>, Abruf am 03.01.2006.
- [SaSB04] Sauter, Annette M.; Sauter, Werner; Bender, Harald: Blended Learning: Effiziente Integration von E-Learning und Präsenztraining. 2. Aufl., Luchterhand, Unterschleißheim; München 2004.
- [Shos84] Shostack, Lynn G.: Planung effizienter Dienstleistungen. In: Harvard Manager, 6 (1984) 3, S. 93-99.
- [Stau06] Stauss, Bernd: Plattformstrategien im Service Engineering, in: Bullinger, Hans-Jörg; Scheer, August-Wilhelm: Service Engineering, 2. Aufl., Springer, Berlin et al. 2006, S. 321-340.
- [StHe92] Stauss, Bernd; Hentschel, Bert: Messung von Kundenzufriedenheit. Merkmals- oder ereignisorientierte Beurteilung von Dienstleistungsqualität. In: Marktforschung und Management, 36 (1992), S. 115-122.
- [Trieb05] Triebelhorn, Thomas: Didaktisches Design. http://www.crashkurs-elearning.ch/?adr=http://www.crashkurselearning.ch/html/didaktisches_design.htm, 2005, Abruf am 29.12.2005.
- [WeWe00] Weiber, Rolf; Weber, Markus: Customer Lifetime Value als Entscheidungsgröße im Customer Relationship Marketing. In: Weiber, Rolf (Hrsg.): Handbuch Electronic Business. Informationstechnologien – Electronic Commerce – Geschäftsprozesse. Gabler, Wiesbaden 2000, S. 333 – 357.
- [ZeBP88] Zeithaml, Valarie A.; Berry, Leonard L.; Parasuraman, A.: Communication and Control Processes in the Delivery of Service Quality. In: Journal of Marketing, 52 (1988), S. 35-48.

Generische Transformation von Learning-Content

Das MOCCA-Projekt

Michael A. Herzog

Forschungsgruppe INKA
Fachhochschule für Technik und Wirtschaft
10313 Berlin
herzog@fhtw-berlin.de

Matthias Trier

Institut für Wirtschaftsinformatik
Technische Universität Berlin
10587 Berlin
trier@syesdv.tu-berlin.de

Abstract

Die Produktion, Datenhaltung und Integration von Multimedia Content ist von zeitaufwändigen, kosten- und ressourcenintensiven Prozessen geprägt. Die Vielfalt der Datenrepräsentation in den Autorensystemen und die fehlende formatübergreifende Transparenz setzen dem Trend des Cross Media Authoring derzeit noch enge Grenzen. Mit dem hier diskutierten Konzept der Generischen Content Transformation und einem darauf basierenden Softwareprojekt werden für den Bereich der E- und M-Learning Content Produktion dramatische Reduktionen der Herstellungszeiten erreicht. In einer Fallstudie zur Podcast-Produktion wird die Transformation von Lerninhalten aus verschiedenen Autorenumgebungen in multivariante, multimodale Ausgaben mit dem MOCCA-Framework vorgestellt.

1 Ausgangssituation und Problemumfeld

Die Speicherung, Auszeichnung und Archivierung von multimedialem Content auf der Basis von Datenbanken und Repositories zielt heute mit Anwendungen aus den Bereichen Enterprise Content Management, Information Lifecycle Management und Data Mining auf die immer

stärkere Konzentration von Unternehmensdaten in einheitlichen Formaten. Für zahlen- und textbasierte Informationen einschließlich der Metadaten ist die Software-Evolution dank des Einsatzes von XML und weiterer Standards in den letzten Jahren deutlich fortgeschritten und wird mit der Entwicklung semantischer Technologien zunehmend attraktiver. Auch im Bereich von Bildern und zeitbasierten Medien schreitet der Konsolidierungsprozess voran. Mit dem Ausbau der Metadaten-Speicherung (MPEG7, MPEG21), der automatisierten Metadatenerkennung und deren Einbindung in semantische Systeme, mit der Verbesserung der Multimedia-Information-Retrieval Methoden [Schm05, LSDJ06], steigt auch die Verfügbarkeit der Medien, verkürzen und verbessern sich die Suchoperationen.

An diese technischen Entwicklungen ist die Erwartung geknüpft, die jährlich zwischen 60% und 200% wachsenden Contentmengen [Jupi05] über mehr Strukturierung, Zentralisierung und vollautomatisierte Verknüpfung effizienter erschließbar und beherrschbarer zu gestalten. Dieser Welle von Content-Konsolidierung und Zentralisierung stehen die zahlreichen Entwicklungen einzelner Softwarewerkzeuge zur Lösung sehr individueller Aufgaben und Probleme entgegen, die im Bereich von Multimedia-Autorensystemen häufig mit der Entwicklung von neuen Datenformaten und Codecs einher gehen. Allein die Menge der heute verbreiteten, von verschiedenen Anwendungen genutzten Bilddatenformate übersteigt die Zahl von 150 deutlich. Die Zahl der proprietären Projektdatenformate in Autorensystemen dürfte diese Summe nochmals deutlich übersteigen, insbesondere wenn auch die Versionshistorie der Formate einbezogen wird.

Die divergierenden Produktentwicklungen im Markt der Autorenwerkzeuge sorgen überdies für die beinahe grenzenlose Vielfalt an Projektdaten- und Distributionsformaten. Der digitale Medienbruch erschwert den Austausch der Ergebnisse gestalterischer Tätigkeiten und das Änderungsmanagement über die Grenzen einer Authoring-Anwendung hinweg. Die in den Einzellösungen angebotenen Schnittstellen zum Datenaustausch betreffen spezielle, meist herstellerbezogene Paketformate und eine Hand voll dedizierter Asset-Datenformate, die für den jeweiligen Zweck abgestimmte Zulieferungen von speziellen Dateitypen mit einer eingeschränkten Vielfalt von Codecs erfordern. Das bedingt in den Produktionsprozessen die zeitaufwändige, personalintensive Konvertierung von Medien und Extraktion von Inhalten in kompatible Formate. In den hochkomplexen Werkzeugen zur Bearbeitung zeitbasierter Medien wird dieser Problematik durch den Einsatz von Medienarchitekturen wie Windows Media, Quicktime u.a. entgegengewirkt. Meist sind die Schnittstellen überwiegend auf einzelne Assets beschränkt, selten jedoch für den Import von Projektdateien aus anderen Autorenumgebungen entwickelt.

In der Printindustrie wurden seit Mitte der 90er Jahre workflow-orientierte Lösungen mit teil- und vollautomatisierten Prozessanteilen auf der Basis des Adobe PDF Formats etabliert, die unter Berücksichtigung einer Formatvielfalt der Assets auf der Eingabeseite und mittels Pipelines standardisierte Prozessanteile realisieren. Diese PDF-Workflows verarbeiten Ergebnisse verschiedener Authoring-Lösungen für nicht-zeitbasierte Medien in integrativer Weise, sind jedoch fast ausschließlich am Printprodukt ausgerichtet und bieten kaum Cross Media Authoring mit multipler Ausgabefunktionalität.

Die produktionsunterstützenden Lösungen in der Medienindustrie, wie z.B. Medienlogistiksysteme [KrDr05], stellen überwiegend Funktionen zur strukturierten Datenhaltung in Repositories, zur Archivierung und zur Logistik im Rahmen von Workflows bereit, sind jedoch aufgrund der begrenzten Schnittstellen daran gehindert, Aggregationen von Content aus unterschiedlichen Autorenumgebungen durchzuführen.

Für den weitgehend proprietär geregelten Content-Austausch werden von den Desktop-Autorenwerkzeugen Schnittstellen für eine begrenzte Menge von Assetformaten bereitgestellt.

In einigen Werkzeugen werden auch dedizierte Ex- und Importfunktionen für die 1:1 Konvertierung von zusammengesetzten Inhalten eines spezifischen anderen Projektdatenformats angeboten, was jedoch eher die Ausnahme darstellt. Neben der Komplexität der Konvertierung und dem hohen Aufwand der Implementierung von integriertem Content bestand bei kommerziellen Anbietern von spezifischen Autorensystemen eher ein geringes wirtschaftliches Interesse an der formatunabhängigen Repräsentation der Daten wie z.B. in SMIL [BuRu04].

Die beschriebene Situation betrifft in ganz spezifischer Weise auch den sich schnell entwickelnden Markt des Multimedia Authoring für die E-Learning Produktion. Hier haben viele Firmen festgestellt, dass die Zeiten und Kosten der Erstellung von E-Learning Inhalten die Vorteile schnell wettmachen können [Vri04]. Daher motivierte sich die Etablierung eines kosten- und zeitsparenden Rapid E-Learning. Dieses ist definiert durch kurze Inhaltserstellungszeiten, durch die Elimination technischer Transformationsschritte, durch die dadurch ermöglichte Fokussierung auf den Inhaltsersteller als Hauptressource und die damit einhergehende Konzentration auf marktübliche Werkzeuge zur Inhaltserstellung (z.B. Powerpoint). Dabei kommen nur solche Medien zum Einsatz, die das Lernen unterstützen aber keine ‚Technologiebarriere‘ hervorrufen.

2 Forschungsziel

Vor dem Hintergrund der oben genannten Umfeldentwicklung wurde als Forschungsziel die Entwicklung eines allgemeinen Ansatzes sowie eines Produktmodells zur Datentransformation, Datenhaltung und Aggregation von Mediencontent aus verschiedenen Autorenumgebungen definiert. Damit soll sowohl eine drastische Verkürzung von Produktionsprozessen bei der Herstellung von singulären Medieninhalten vor allem aber Synergien in der Cross Media Produktion ermöglicht werden. Forschungstheoretisch kommt bei der Entwicklung die Methodologie der Designwissenschaften (Design Research) zum Einsatz, vgl. [VaKu04], da ein neues ‚IT-Artefakt‘ zu entwickeln ist. Dazu wird in den nächsten Abschnitten zunächst das unterliegende Konzept der generischen Content Transformation vorgestellt, um daraus den technischen Prototypen abzuleiten. Mit einer Nutzerstudie für die ausgewählte Einsatzdomäne E-Learning wird der Einsatz des entwickelten Prototypen evaluiert.

3 Entwicklung eines Konzepts zur generischen Content Transformation für containerbasierte Medieninhalte (GCT)

Mit der schrittweisen Zuwendung vieler Hersteller zu einer Repräsentation der Mediendaten in transparenten XML-Containern eröffnen sich im Umfeld der Mediensoftwaresysteme derzeit neue Möglichkeiten, die Konvertibilität von Inhalten um Dimensionen zu verbessern.

Dazu wurde ein Konzept entwickelt, das es erlaubt, Transcodierungen für Inhalte aus vollständigen Mediencontainern in alternative Repräsentationsformen vorzunehmen. Ein weiterer Schritt ist die Generalisierung dieser medienintegrierenden Containerformate mit dem Ziel, eine möglichst von speziellen Autorensystemen unabhängige Datenhaltung und Transformation zu gewährleisten. Als Vision wird der Ansatz eines generischen Transformationswerkzeuges im Sinne eines Content-HUB verfolgt, das Mediendaten aus verschiedensten Authoring Anwendungen (im weitesten Sinne) in einheitlichen, standardisierten Datenrepräsentationen zusammenführt und für verschiedene Distributionszwecke transformiert.

Es wird dabei davon ausgegangen, dass zahlreiche von Autorensystemen erzeugten Projektdateien gleichartigen Prinzipien unterliegen, indem sie beispielsweise

- Gleichartige Strukturen auf der Basis vektorbasierter Darstellungen aufweisen (vgl. Abbildung 1)

- Eine begrenzte Anzahl von Funktionen zur Navigation und Interaktion nutzen und
- Darüber hinaus gehende Softwarefunktionalität in codierter Textform beschreiben.

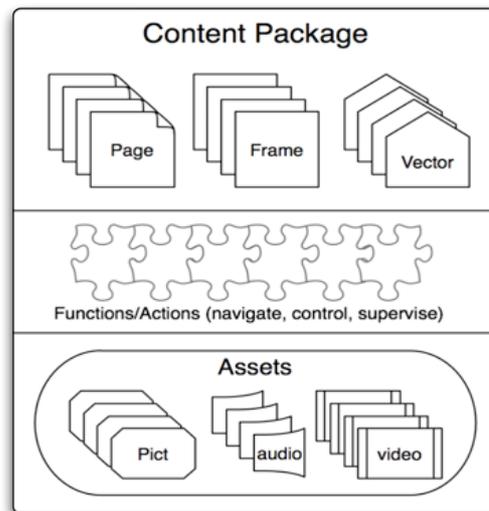


Abbildung 1: Struktur von Multimedia Content Packages

In einer technischen Voruntersuchung zu 22 ausgewählten Projektdatenformaten in Autorensystemen wurde ermittelt, dass sich in allen Fällen mehr als 80% aller Informationen im Format der für das Projekt entwickelten XML-Datenstruktur unaufwändig erhalten und weiterverarbeiten lassen. Weitere Bestandteile, die insbesondere Anwendungslogik und Steuerung proprietärer Mediencontainer betreffen, sind mit erheblich höherem Aufwand durchaus erfass- und konvertierfähig, wobei hier jeweils in spezifische Lösungsansätze investiert werden muss. Bei der Betrachtung des Aufwand-Nutzen-Verhältnis hilft insbesondere die Prozessperspektive bei der Abschätzung von erwarteten Einsparungen bei der Medienproduktion.

Im Detail konnte für die bisher untersuchten 2D-Einsatzgebiete auf dem Gebiet Interaktive Fernsehinhalte (MHP/SWF Transformation) und E-Learning (SCORM und Podcast Produktion) bei der vollautomatischen Transformation in das native XML-Format eine durchschnittliche Datenerhaltung von 97 % beim Layout, 100% bei Texten und Assets, sowie über 80 % bei der Funktionalität erreicht werden. Bei der vollautomatischen Ausgabe-Transformation sind durch die Beschränkungen der Distributionsformate sehr unterschiedlich hohe Raten der Datenerhaltung erreicht worden, da beispielsweise im Bereich der Ausgabe für mobile Geräte Anpassungen der Assets notwendig sind, wie Skalierung, Reduktion der Bildauflösung, Kürzung von

Texten, Wandlung von Vektor- in Bitmap-Informationen usw.. Besonders effektive Transformationsprozesse mit optimaler Qualität bis in die Distributionsformate konnten in den Bereichen erzielt werden, wo in den Ausgangsdaten eine hohe Transparenz und Assetqualität bestand oder wo die Ausgangsprojekte bereits an die Bedürfnisse der Zielformate angepasst wurden. Der GCT-Ansatz wird insbesondere deshalb als Erfolg versprechend angesehen, weil er berücksichtigt, dass nicht alle Funktionen und Assetformate zwingend in alle Distributionsformate transformierbar sein müssen. Ziel ist weitestgehende Erhaltung aller Komponenten in einem generischen Speicherformat mit einem vertretbaren Aufwand (80/20 Paradigma).

Die Transformation der über den Container verbundenen und arrangierten Medien in ein einziges offenes Datenformat ermöglicht verschiedene Synergien:

Datenhaltung und Information Retrieval: Die transparente Suche über verschiedene Content-Arten und Versionen in einheitlichen Repositories unter Berücksichtigung von Meta- und Strukturinformation bildet eine grundlegend breitere Basis für das Cross-Authoring und liefert im Zusammenhang mit semantischen Technologien eine wesentlich höhere Transparenz der Inhalte. Eine Suche wie beispielsweise »Zeige alle Bilder, die in Containern zum Thema ‚Fahrrad‘ vorkommen« wäre bereits mit schnellen und einfachen Algorithmen realisierbar und enthielte als Ergebnis Verweise auf alle Assets, die auf Seiten mit dem Textinhalt ‚Fahrrad‘ vorkommen. Durch die semantische Verknüpfung von Text und Bildinformation über das Layout im Container lassen sich so auch metadatenlose Assets in inhaltliche Zusammenhänge bringen und auffinden, wo sie bisher für Filesysteme oder Repositories verborgen geblieben sind. Die Ergänzung mit modernen Verfahren der Ähnlichkeitssuche in Bild- und Audioassets fördert die Erschließung dieser bislang nur über spezifische Werkzeuge zugänglichen Datenbestände im Unternehmen.

Persistenz: Konzepte für die langfristige Archivierung von Daten aus Containerformaten können auf der Meta-Ebene ansetzen und müssen nicht mehr jedes einzelne Projektdatenformat berücksichtigen. Damit kann die Content-Transformation integrativer Bestandteil des globalen Enterprise-Sicherungskonzeptes werden und bedient unmittelbar auch Compliance-Anforderungen.

Verteilte Produktion: Verteilt erstellte Inhalte, gerade wenn sie in verschiedenen Autorenumgebungen erstellt wurden, lassen sich auf der Basis des beschriebenen Konzepts erstmals zusammenführen und zentral verwalten. Dies dürfte eine erhebliche Verkürzung von Produktionszyklen durch Parallelisierung ermöglichen.

Paralldistribution: Mit der Bereitstellung mehrerer Distributionswege und -formate werden Autoren in die Lage versetzt, Content für dedizierte Ausgaben, gegebenenfalls als neu kombinierte Inhalte, zielgruppen- oder gerätespezifisch bereitzustellen (vgl. [ScWi04]). Technologien zur multivarianten Speicherung, Erzeugung und Distribution von XML-basierten Inhalten, wie sie in Frameworks wie Cocoon, OpenLaszlo oder Zope eingesetzt werden, bieten ein hohes Potential, wenn sie für die Verarbeitung von kombinierten Containerinhalten herangezogen werden.

Änderungsmanagement: Die Überführung von Inhalten aus proprietären Containerformaten in XML-basierte Strukturen und deren Verwaltung in Repositories führt zu einer neuen Qualität der Transparenz von Inhalten. Bei entsprechender Erweiterung und Anpassung der Versionierungssysteme wird eine hierarchische Dokumentaufteilung mit Zugriff bis in einzelne Assets bzw. Asset-Bestandteile oder -Spezifika möglich. Mittels der in Mediendatenbank- bzw. Medienlogistiksystemen bereits verfügbaren Techniken zur vollautomatischen Assetumwandlung, wie OPI, Farbraumwandlung oder Farbanpassung, werden neue Nutzungskonzepte vorstellbar, die sich erheblich auf die Produktivität von Cross-Publishing-Lösungen auswirken werden.

Neue Einsatzfelder: Das Konzept der generischen Content Transformation führt zu einem integrativen Cross-Authoring-Prozess, der es in einem Arbeitsgang ermöglicht, die Erstellung mehrerer Medien und das Bedienen verschiedener Distributionskanäle mit deutlich geringerem Zusatzaufwand als bisher zu verwirklichen. Durch die niedrige Eintrittsschwelle werden Autoren praktisch in die Lage versetzt, Content aus verschiedenen Quellen zu verschmelzen und zusätzliche Angebote zu kreieren.

Prozessoptimierung: Das skizzierte Konzept zur Content-Transformation aus Containern zielt im Kern auf die wesentliche Verbesserung von Prozesskennzahlen in der Medienproduktion. In Betracht kommen dafür Prozessvereinfachungen und die unmittelbare Verkürzung der Durchlaufzeiten in den Transformationsprozessen. Eine Vielzahl von mittelbaren Faktoren, wie etwa die Verkürzung von Suchoperationen unterstützen die Effektivierung der Medienproduktion. Als speziell entwickelte Komponente steht das Pipeline-Konzept zur vollautomatischen Abwicklung von definierten und parametrisierbaren Transformationsprozessen á la Workflows zur Verfügung, das insbesondere bei häufig vorkommenden Produktionsabläufen die Durchlaufzeiten nochmals drastisch verkürzt. Die vollautomatische Transformation führt hier insbesondere dann zu keinerlei manuellen Arbeiten, wenn das Ausgangsmaterial entsprechend vorbereitet ist.

Diese Anpassung wird von den Autoren in der gewohnten Autorenumgebung i.d.R. schnell und mit sehr geringer Einarbeitungszeit durchgeführt.

Das Konzept der generischen Content Transformation adressiert alle Projektdatenformate aus klassischen Media-Authoring-Anwendungen. Das zur Evaluierung des GCT-Konzeptes eingerichtete Softwareprojekt wird nachfolgend vorgestellt.

4 Technische Umsetzung des Prototypen - das MOCCA Projekt

Orientiert an der Vision des Content-HUB und am Konzept der Generischen Content Transformation für containerbasierte Medieninhalte wurde im Software-Entwicklungsprojekt MOCCA [Mocca] in Zusammenarbeit zweier Arbeitsgruppen an der FHTW Berlin und der TU Berlin ein kombiniertes Transformations- und Authoring Werkzeug entwickelt, das sich mit den Produktionsprozessen im Umfeld von interaktiven Fernsehinhalten und E-Learning befasst. Es wird um ein einheitliches Media-Repository ergänzt, das die Vorteile der im Format implizierten textbasierten Suchmöglichkeiten mit einigen aktuellen Methoden der inhaltsbasierten Ähnlichkeitssuche in Multimediadaten verbindet. Der für die Domäne E-Learning ausgebaute Prototyp ist dabei an den folgenden wesentlichen Merkmalen orientiert:

- Generische Speicherung von multimedialem Content aus verschiedenen Autorenumgebungen (z.B. PowerPoint, OpenOffice, Keynote) mittels einer speziellen XML-Repräsentation (Multimedia-Repository)
- Zusammenführung von Präsentationsfolien, Audio- und Videomaterial, Animationen etc., die für Präsenzvorlesungen oder für die Herstellung von E-Learning-Material entwickelt wurden (Import PlugIns)
- Automatisierte und manuelle Anreicherung mit allgemeinen Zusatzinformationen (Metadaten, Verarbeitungsmarker) und distributionsbezogenen Ergänzungen
- Begrenzte manuelle Bearbeitungsfunktionalität, wie Textänderung, Rahmenbearbeitung, Veränderung von Bildausschnitten und -Formaten, Setzen von Zeit-, Schnitt- und Sprungmarken, Einfügen und Ändern von Buttons und Links, manuelle Transformation von Medienelementen in andere Präsentationslayouts usw. (Authoring)

- Vollautomatisierte Verarbeitung von entsprechend mit Verarbeitungsmarken vorbereiteten Projektdaten (Pipeline)
- Gleichzeitige Bereitstellung von diversen Formaten, wie PDF, SCORM-Modulen und mobilen Lerninhalten (Distributoren)
- Semantische Suchmechanismen innerhalb des Repository (Spot)
- Unternehmensweite Erschließung von proprietären Projektdatenformaten durch automatisierte Überführung in das Multimedia-Repository (Crawler)

Das Authoring Framework und die PlugIn Architektur sind in der Programmiersprache Java5 realisiert. Dabei kommen im Wesentlichen die Eclipse Rich Client Platform (RCP) und das auf der Draw2D Bibliothek basierende Graphical Editing Framework (GEF) zum Einsatz. Für die Integration der Medienbearbeitung wurden die QuickTime for Java APIs verwendet.

Zur unabhängigen Beschreibung der Container für die generische Datenhaltung wurden im Vorfeld Untersuchungen und Experimente mit verschiedenen Standardformaten, wie z.B. ISO/IEC DIS 26300 durchgeführt. Resultierend wurde entschieden, die SVG-Spezifikation 1.2 basierend auf RelaxNG zu modifizieren, um ein schlankes, auf die Anforderungen der Transformation angepasstes Format zu definieren.

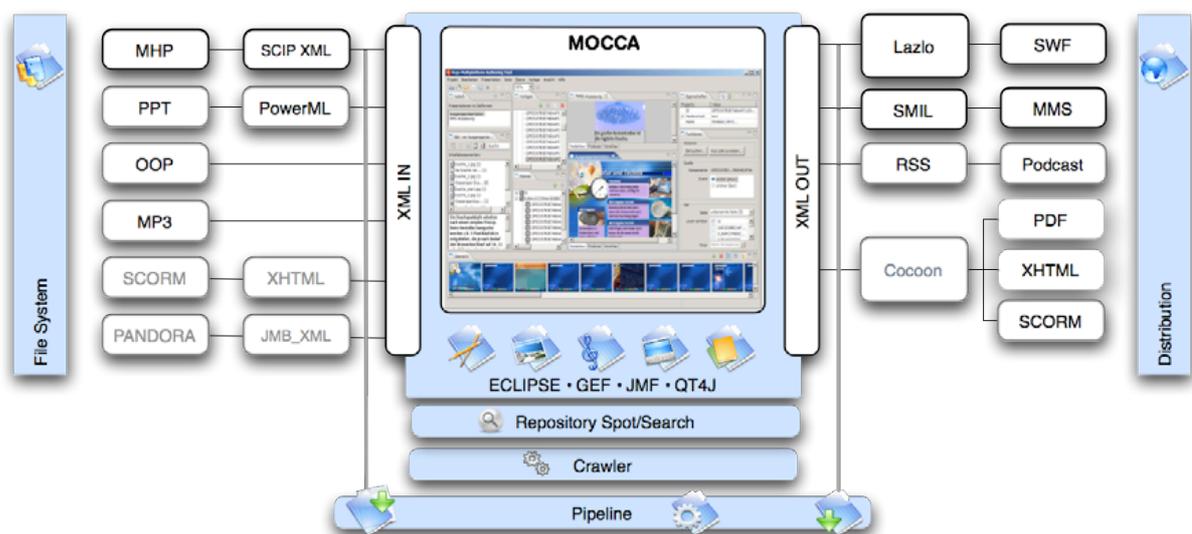


Abbildung 2: Architektur des MOCCA Prototypen für die E- und M-Learning-Produktion

Die Import- und Export-Komponenten bedienen sich neben XSL-Transformationsmechanismen wie SAX verschiedener Konstrukte und Bibliotheken zur Konvertierung proprietärer Dateiformate.

mate. Zur Distribution von Macromedia Flash Containern wurden hier beispielsweise Basis-
komponenten des OpenLaszlo Pakets integriert. Besondere Sorgfalt wurde auch auf die Kon-
zeption der Nutzer-Schnittstellen für die kontextsensitive Repository-Anbindung gelegt, worin
auch einige Anregungen aus den Projekten xSMART [ScBo05] und SampLe [FNOR04] einge-
flossen sind.

5 Fallstudie: Mobile-Learning Podcast-Produktion

Als ein Beispiel, wie der beschriebene Ansatz und die adäquate Werkzeugunterstützung den
Produktionsprozess für E-Learning-Materialien rationalisieren kann, wird hier die Herstellung
von Lehrmedien für mobile Geräte vorgestellt. Ausgangspunkt der Arbeiten waren die Erfah-
rungen aus einem bereits durchgeführten Projekt zur Prozessautomatisierung der SCORM-
Produktion für die Virtual Global University [MTH05]. Zur Ermittlung des Bedarfs wurde zu-
nächst in einer Feldstudie untersucht, ob und in welcher Weise ein lehrbegleitendes innovatives
Medium im mobilen Kontext sinnvolle didaktische oder organisatorische Effekte erzeugt, wel-
che Parameter für die Distribution von den Rezipienten akzeptiert werden und welche Möglich-
keiten technisch realisierbar sind. Als Rahmenbedingung wurde festgelegt, dass die Produktion
für die Lehrenden möglichst wenig technischen Aufwand verursachen sollte und das Medium
ergänzend zur Präsenzlehre eingesetzt werden kann. Als Präsentationskanal wurden RSS-
Podcasts ausgewählt, die sich unaufwändig abonnieren und auch im mobilen Kontext nutzen
lassen.



Abbildung 3: M-Learning Vodcast Vorlesung auf einem iPod

Die Studie wurde in der Lehrveranstaltung »Einführung in Multimedia« im Studiengang Angewandte Informatik an der FHTW Berlin im Wintersemester 2005/2006 durchgeführt.

Ausgangspunkt waren Audioaufnahmen, die überwiegend als Vorlesungsmitschnitte produziert wurden. Die Aufzeichnung erfolgte mit einem MP3-Recorder. Zusätzlich wurden die Präsentationsfolien des Dozenten (in der Regel PowerPoint-, auch OpenOffice- oder Keynote-Präsentationen) und Digitalfotos vom Whiteboard, von beteiligten Personen bzw. von Demonstrationobjekten verarbeitet. Aus diesen Medien wurde eine MPEG4-Audio Datei produziert, die neben der Tonspur die Folien der Vorlesung sowie Sprungmarken zu Beginn jeder neuen visuellen Information enthält. Der Herstellungsprozess für diese Medien – beginnend nach der Aufzeichnung – nahm zunächst das zwei bis vierfache der Länge der Vorlesung in Anspruch (vgl. Abbildung 9). Die Bildwechsel mussten nachträglich erkannt und markiert werden. Folien wurden parallel als Rasterbilder aus Powerpoint exportiert. Teilweise wurde vorher die Lesbarkeit der Folieninhalte für die mobile Präsentation angepasst. Und schliesslich wurden das Zielmedium als M4A-Datei sowie die entsprechende RSS-Datei zusammengesetzt.

Um eine größere Gerätevielfalt unterstützen zu können, wurden im nächsten Schritt zusätzlich mehrere Varianten der Podcasts hergestellt (vgl. Abbildung 4). Die Vorlesungsinhalte in Bild und Ton kombiniert mit ergänzenden Videoinhalten konnten in einem weiteren Schritt zu Vodcasts zusammengestellt und als MPEG4 Videodatei angeboten werden (vgl. Abbildung 3).

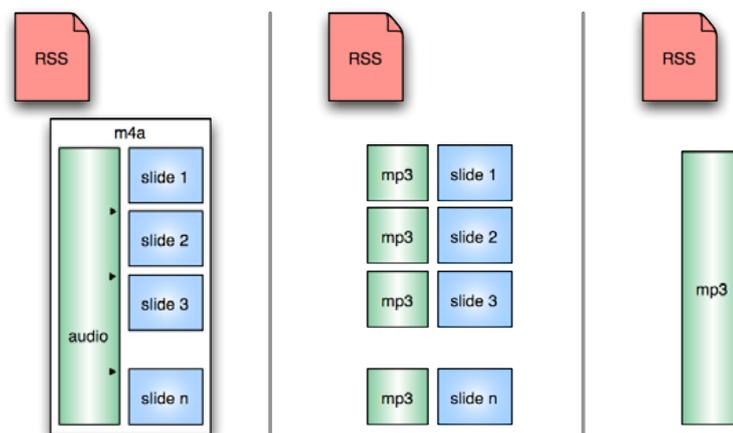


Abbildung 4: Podcast Varianten m4a, mp3 patches und mp3 plain

Die verwendeten Videoinhalte wurden nach einem in [HTS06] beschriebenen Konzept im vbc.studiolab ausserhalb des hier skizzierten Prozesses realisiert. Hier werden beispielsweise kurze Einführungen in den Kontext (Teaser) oder ergänzende Hinweise vom Lehrenden gege-

ben. Diese kurzen Videobeiträge unterstützen eine direkte und persönliche Ansprache des Rezipienten.

In einer Erhebung unter 39 Studierenden am Ende des Semesters wurde nach dem Nutzungskontext, der Bedeutung der Features, der Laufzeitumgebung und dem Zweck der Nutzung gefragt. Außerdem wurden die Struktur, Qualität, Dateigrößen, Formate und die Attraktivität des Angebots evaluiert.

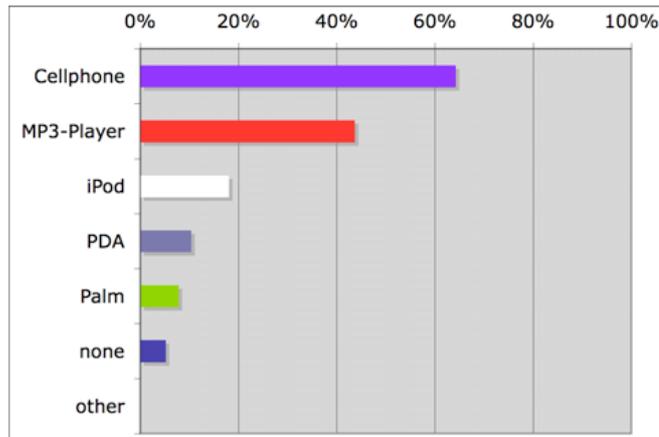


Abbildung 5: Verbreitung mobiler Geräte unter den Teilnehmern

Wie schon verschiedene Studien gezeigt haben, ist die Audioinformation ein wichtiger, wenn nicht der wichtigste Kanal für das Lernen. In dieser Erhebung haben 69% der befragten Teilnehmer Audiomitschnitte von Vorlesungen als wichtig bzw. sehr wichtig beurteilt. Ebenfalls 69% der Teilnehmer haben das Podcast-Angebot genutzt.

Bei der Erhebung zu den mobilen Endgeräten hat die hohe Verbreitung der mp3 Player unter den Teilnehmern überrascht, die zusammen mit den iPods 62% ausmachen, was in der Gruppe knapp der Verbreitung von Mobiltelefonen entsprach.

Sehr verschieden sehen die Nutzungsprofile der Teilnehmer aus, die sich – auch aufgrund der geringen Anzahl – nicht in Gruppen zusammenfassen lassen (z.B. Viel- oder Wenig-Nutzer, Stationär- oder Mobil-Nutzer etc.).

Im Durchschnitt wurde eine Vorlesung von den Podcast-Hörern bis zu 3 Mal gehört. Dabei wurden durchschnittlich 6 Stunden von insgesamt 16 Stunden des Materials genutzt. In der Regel wurden die Vorlesungen 1 Stunde hintereinander ohne nennenswerte Unterbrechungen gehört. 68% der Podcast-Nutzer gaben an, dass sie das Angebot auch mobil genutzt haben, wobei der größere Teil der Podcast-Nutzer (86%) die Vorlesungen am heimischen Computer verwendet hat.

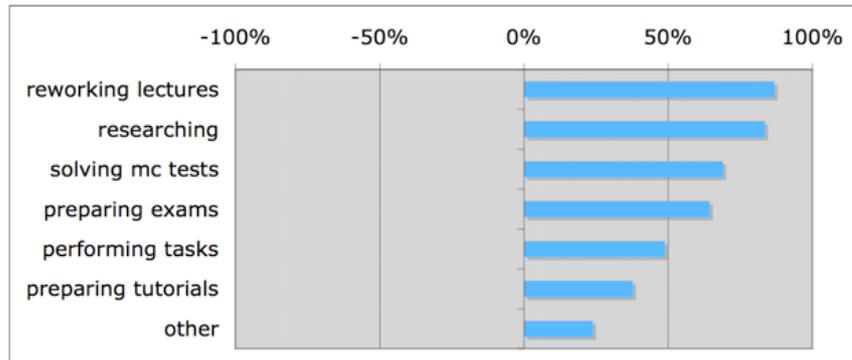


Abbildung 6: Nutzwert des Podcast-Angebots nach Zweck

Besonders interessant für die Lehrenden war die Befragung zum Zweck der Nutzung. Zur Klausurvorbereitung (86%) und zur Nachbereitung der Präsenzvorlesungen (54%) war das Medium besonders beliebt. Für alle nachgefragten Zwecke wurden die Podcasts als hilfreich bzw. sehr hilfreich bewertet (vgl. Abbildung 6), wobei der Nutzen zur Nachbereitung und für die allgemeine Recherche am größten beurteilt wurde. Damit ist nach unserer Einschätzung eine neue Qualität verbunden: Es ist zu vermuten, dass das Medium in vielen Fällen etwa einem Buch oder einer Online-Recherche vorgezogen wird. Die Studierenden hören lieber nochmals den sie interessierenden Teil der Vorlesung, als ein Text/Bildmedium zu nutzen. Auf spätere gezielte Nachfragen gaben mehrere Teilnehmer an, dass sie auch im weiteren Verlauf des Studiums das Medium für Recherchen wiederholt genutzt haben, weil der Bezug zum Kontext der Lehrveranstaltung und die mündliche Erklärung das Verständnis im Gegensatz zum Buch vereinfacht habe. Da sich ein Podcast nach dem hier beschriebenen Modell einfacher und schneller produzieren lässt als eine Niederschrift, dürfte das Aufwand-Nutzen-Verhältnis einer Buch- oder Skriptproduktion im Vergleich mit der Podcast-Produktion kaum vorteilhafter sein.

Bezüglich der Distribution konnte keine allgemeine Präferenz für ein bestimmtes Dateiformat ermittelt werden. Alle im Feldtest angebotenen Podcast-Varianten wurden von den Rezipienten genutzt. Die höchste Präferenz wurde für die MP3 plain Variante ermittelt, die keine Sprungmarken und keine visuellen Informationen enthält (vgl. Abbildung 7 links). Das steht im Widerspruch zu den Wünschen nach Sprungmarken (61%) und nach Bild-Informationen wie Folien oder Tafelbildern im Audiostream (71%).

Die Differenz lässt sich erklären, wenn man die verfügbaren mobilen Geräte betrachtet. Da für viele Player technisch nur die Verwendung der mp3 plain Variante infrage kam, weil bspw. kein ausreichendes Display vorhanden war, ist *das* optimale Distributionsformat derzeit nicht in

Sicht. Für die Produktionsunterstützung wurde hieraus gefolgert, dass es für die Erreichbarkeit der Rezipienten mit diesem Medium wesentlich ist, Formatvielfalt anzubieten.

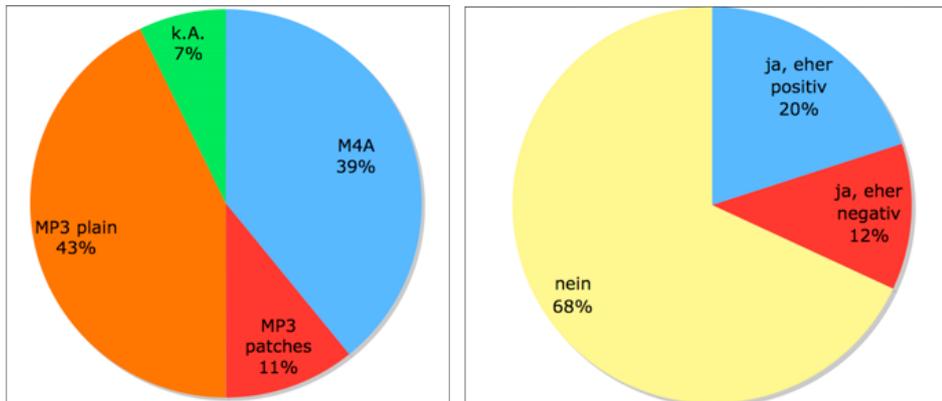


Abbildung 7: Genutzte Datenformate (links);
Hatte das Medienangebot Einfluss auf die Anwesenheit zu den Vorlesungen? (rechts)

Aus der fast ausnahmslos positiven Qualitätsbewertung der angebotenen Podcast Varianten wurde gefolgert, dass lediglich geringe technische Optimierungen zugunsten von Dateigröße und Qualität der Bildinformation in Abhängigkeit von den Distributionscodecs vorteilhaft wären. Eine wichtige organisatorische Frage war der Einfluss des Podcast-Angebots auf die Anwesenheit der Teilnehmer in der Präsenzveranstaltung (Abbildung 7 rechts). Die positive Tendenz bestätigt viele schon oft in der Literatur formulierten Thesen, dass Medien- und Materialvielfalt eher eine höhere Bindung an die Lehrveranstaltung erzeugt und keinesfalls die Präsenzlehre ersetzen wird.

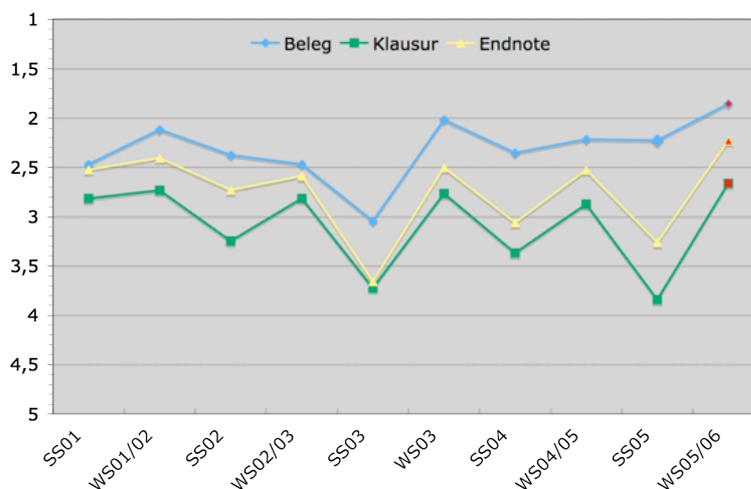


Abbildung 8: Benotung in der Lehrveranstaltung EMM

Als weiteres Indiz wurde die Leistungsbewertung der Teilnehmer über 10 Semester herangezogen (Abbildung 8). Die Veranstaltung wurde in dieser Zeit von zwei Lehrenden jeweils über mehrere Semester hintereinander betreut. In der einzig verfügbaren Stichprobe von einem Semester mit Podcast-Begleitung lässt sich feststellen, dass im theoretischen wie im praktischen Teil der Bewertung ein positiver Effekt vermutet werden darf.

Insgesamt lieferte der Feldtest ausreichende Belege, die den Aufwand der Prozessoptimierung mit Unterstützung durch neue Softwarekomponenten rechtfertigen. Im Mittelpunkt würde dabei eine Transformationslösung stehen, die es erlaubt, die produzierten Präsentationscontainer sowie weitere entstehende Assets, wie Audio, Video und Bilder zusammenzuführen und in einem automatisierten Prozess für mehrere Distributionswege aufzubereiten. Im Hinblick auf das dem MOCCA-Projekt zugrunde liegende Konzept von möglichst allgemeinen Transformationsmöglichkeiten lag es nah, den Bereich der Podcast-Produktion für das Mobile Learning in das Framework einzubeziehen. Ziel ist die wesentliche Vereinfachung und Verkürzung des in Abbildung 9 links beschriebenen Produktionsprozesses.

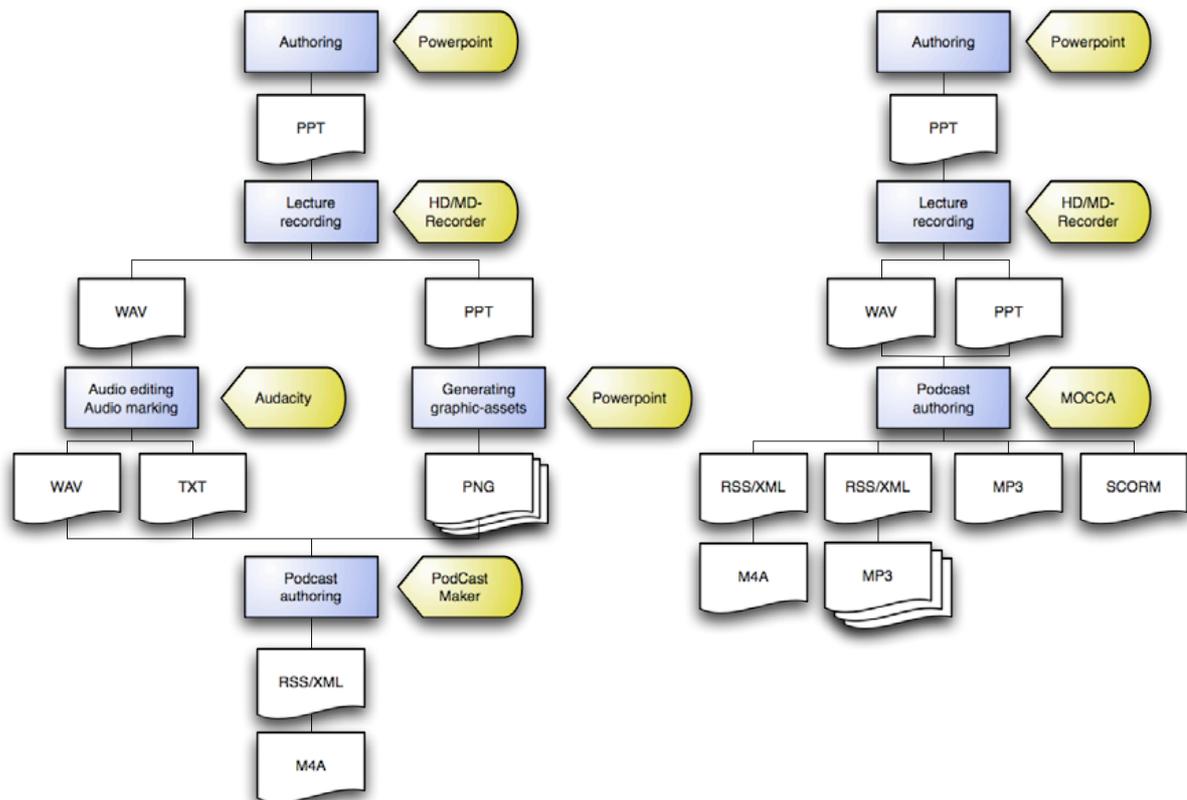


Abbildung 9: Podcast-Produktionsprozesse ohne (12/2005) und mit MOCCA (11/2006)

Dazu wurden beispielsweise die Medienarchitektur für die Herstellung von MPEG4 Audio erweitert und entsprechende Im- und Export-Plugins eingerichtet, die eine gleichzeitige Distribution der verschiedenen Distributionsformate ermöglicht. Daneben werden Präsentationsdetails wie Folienwechsel bereits während der Vorlesung in einer XML-Datei aufgezeichnet, um die Erstellung der Sprungmarken zu automatisieren [vgl. Gutb05]. Der Zielprozess ist in Abbildung 9 rechts modelliert.

Mit dem Release-Zeitpunkt Ende 2006 wird die Verkürzung der Durchlaufzeiten um 30 bis 50% erwartet. In einem weiteren Schritt ist die Erweiterung der Medienformate geplant: Auf der Eingabeseite für verschiedene weitere Autorensysteme und auf der Ausgabeseite für andere mobile Geräte, wie Mobiltelefone oder spezifische MP3 Player.

Die Nutzung des MOCCA-Frameworks bietet neben der Ausgabe von Podcast-Varianten auch andere Distributionsmöglichkeiten für E-Learning-Inhalte, beispielsweise die Erstellung von HTML-Präsentationen oder SCORM Paketen, was Gegenstand einer parallelen Entwicklung ist.

6 Fazit

In der Domäne E- und M-Learning Produktion wurde gezeigt, wie das Konzept der generischen Content Transformation zur erheblichen Aufwandsreduktion bei der Inhaltserstellung führen kann. Es können verschiedene proprietäre Projektdatenformate im Sinne des Cross Authoring in ein einheitliches Format integriert werden, was wiederum neue Qualitäten für den Content Lebenszyklus von Multimedia-Daten aus Authoring Systemen ermöglicht. Mit der Ausweitung des Modells auf weitere Domänen, wie bereits für das interaktive Fernsehen gezeigt, wird in Zusammenhang mit der konsequenten Prozessbetrachtung die Erschließung weiterer Rationalisierungsfelder und die Potenzierung der Synergieeffekte erwartet. Mit dem vorgestellten GCT-Konzept und dem MOCCA-Produktmodell wird damit ein signifikanter Beitrag zur effektiveren Datentransformation, Datenhaltung und Aggregation von Mediencontent aus verschiedenen Autorenumgebungen geleistet. Die damit verbunden Fortschritte bei der konsequenten Beseitigung digitaler Medienbrüche dürften auch zu universelleren Betrachtungen im Zusammenhang mit dem Enterprise Content Management führen und verschiedene Trends auf dem Gebiet der Komplexitätsreduktion von Pervasive Computing verstärken.

7 Danksagung

Die vorgestellten Arbeiten wurden von der Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF) im Rahmen des BMBF-Projekts IKAROS gefördert. Wir danken unseren Kollegen in Berlin, Dortmund und Görlitz für die bereichernden Diskussionen. Ein spezieller Dank gilt unseren Diplomanden Daniel Möller und Andreas Hohendorf, sowie Johannes Stein, Roman Frohn und Claudia Müller.

Literaturverzeichnis

- [BuHa05] Bulterman, Dick C. A., Hardman Lynda: Structured multimedia authoring, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), v.1 n.1, p.89-109, February 2005, revised Version of an article from 1993
- [BuRu04] Bulterman, Dick C.A.; Rutledge, Lloyd. SMIL 2.0 - Interactive Multimedia for Web and Mobile Devices. Springer, 2004
- [FNOR04] Falkovych, K., Nack, F., van Ossenbruggen, J., Rutledge, L.: SampLe: Towards a Framework for System-supported Multimedia Authoring. In Proceedings of the 10th International Multimedia Modelling Conference, p.362, January 5-7, 2004.
- [Gutbr05] Gutbrod, Martin et al.: Tutorntity-Projekt der Universität Braunschweig. <http://www.ibr.cs.tu-bs.de/users/gutbrod/tutorntity/>, Abruf am 12.6.2006
- [HTS06] Herzog, Michael A.; Trier, Matthias; Sieck, Jürgen: Production engineering for video based e- and m-learning content. In: UNESCO International Centre for Engineering Education. Proceedings of 10th Baltic Region Seminar, Szczecin, Poland, 4.-6. September, 2006. Accepted for publication.
- [Jupi05] Jupiter Research: Creating a valuable Enterprise Portal. <http://www.jupiterresearch.com> 11/2004, Abruf am 20.11.2005

- [KrDr05] Kretschmar, O., Dreyer, R.: Medien-Datenbank- und Medien-Logistik-Systeme. Anforderungen und praktischer Einsatz. R. Oldenbourg Verlag 2004
- [LSDJ06] Lew, Michael S.; Sebe, Nicu; Djeraba, Chabane; Jain Ramesh: Content-Based Multimedia Information Retrieval: State of the Art and Challenges. ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 2, No. 1, February 2006, Pages 1–19.
- [Mocca] Softwareprojekt Mocca der TU Berlin (Institut für Wirtschaftsinformatik, Forschungsgruppe IKM) und der FHTW Berlin (Forschungsgruppe INKA). <http://www.moccaonline.de>
- [MTH05] Müller, Claudia; Trier, Matthias; Herzog, Michael A.: Process-oriented production of learning units for sustainable e-learning offerings. In: Breitner, M., Hoppe, G. (Hrsg.), E-Learning. Einsatzkonzepte und Geschäftsmodelle, Physica/Springer 2005.
- [ScBo05] Scherp Ansgar, Boll, Susanne: Context-driven smart authoring of multimedia content with xSMART. In: Proceedings of the 13th annual ACM international conference on Multimedia. ACM Press 2005.
- [Schm05] Schmitt, Ingo: Ähnlichkeitssuche in Multimedia-Datenbanken. Retrieval, Suchalgorithmen und Anfragebehandlung. R. Oldenbourg Verlag 2005
- [ScWi04] Schmidt, A., Winterhalter, C.: User Context Aware Delivery of E-Learning Material: Approach and Architecture, Journal of Universal Computer Science (JUCS) vol.10, no.1, January 2004
- [Vri04] De Vries, Jennifer: Rapid E-Learning: Groundbreaking New Research. <http://www.elearningmag.com/ltimagazine/article/articleDetail.jsp?id=102399>. Abruf am 01.05.2006. (2004)
- [VaKu04] Vaishnavi, Vijay; Kuechler, William: Design Research in Information Systems. 20. Januar 2004, Update am 18. Januar 2006. <http://www.isworld.org/Researchdesign/drisISworld.htm>. Abruf am 01.06.2005

Reengineering der Content-Erstellungsprozesse in Industrieunternehmen durch Content-Modellierung

Fallbeispiel

Pavlina Chikova, Katrina Leyking, Peter Loos

Institut für Wirtschaftsinformatik (IWi) im
Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI)
66123 Saarbrücken
{pavlina.chikova,katrina.leyking,peter.loos}@iwi.dfki.de

Eva-Maria Bruch

Festo Lernzentrum Saar GmbH
66386 St. Ingbert
ena@de.festo.com

Lasse Lehmann

Technische Universität Darmstadt
Fachgebiet Multimedia Kommunikation (KOM)
64283 Darmstadt
llehmann@KOM.tu-darmstadt.de

Abstract

Der Prozess der Erstellung von eLearning-Content in Industrieunternehmen erweist sich derzeit als sehr komplex und ressourcenintensiv. Vor diesem Hintergrund verfolgt das Forschungsprojekt EXPLAIN den Ansatz, Unternehmen durch eine integrierte Authoring Management Plattform in die Lage zu versetzen, ihre Lerninhalte selbständig zu produzieren und so auf konkreten Lernbedarf zeitnah und kostengünstig reagieren zu können. Als zentraler Ansatz der EXPLAIN-Plattform wird eine Content-Modellierungsmethodik vorgestellt, mit deren Hilfe der gesamte Prozess zu vereinfachen und beschleunigen ist. Dass sich somit eLearning-Content in Industrieunternehmen effizienter und effektiver entwickeln lässt, wird anhand eines Fallbeispiels aus der Automatisierungsindustrie illustriert und aufgezeigt.

1 Einleitung und Motivation

Die effiziente Erstellung und der Einsatz von elektronischen Lerninhalten (Content) zur Schulung von Vertriebs- und Servicepersonal sowie Endnutzern entwickelt sich zunehmend zu einem wichtigen Wettbewerbsfaktor für produzierende Unternehmen [BaBS01]. Die konsequente Integration von eLearning in die Unternehmensprozesse ist kein neues Thema, jedoch entsteht die Frage, warum Training im Allgemeinen und Produkttraining im Speziellen heute in Deutschland und überwiegend in Europa zumeist noch in Form von klassischen Präsenzveranstaltungen durchgeführt wird [Attw03]. Aktuelle Studien sprechen von einem eLearning-Anteil in Unternehmen von circa 3 bis 5% im Verhältnis zum Präsenzanteil, obgleich steigende Tendenzen erkennbar sind. Die technologische Verfügbarkeit von Distributionswegen für Trainingsmedien ist signifikant gestiegen (DVD, CD-ROM, Internet, Mobile Geräte usw.) und auch der Reifegrad von Autorentools zur Unterstützung der Content-Erstellung ist weiter fortgeschritten. Insofern scheinen aus technischer und konzeptioneller Perspektive die Voraussetzungen für eine breite Anwendung gegeben. Die bisherigen Erkenntnisse in der Industrie zeigen aber, dass Unternehmen den Aufwand scheuen, eLearning-Content selbst zu produzieren und die personellen Ressourcen dafür zu intensivieren, da ihnen der Prozess zu kompliziert erscheint. Auch sind existierende Werkzeuge nicht intelligent und komfortabel genug, um das fehlende didaktische Wissen sowie Prozess-Know-how auszugleichen und optimal mit dem Expertenwissen in den Unternehmen zu verknüpfen [NHHA04]. Darüber hinaus wird der Anwendungsnutzen im Verhältnis zum wirtschaftlichen Aufwand insbesondere aus Managementsicht bislang nicht als adäquat angesehen.

Auch Unternehmen, die ihre Prozesse auf das Thema systematisch ausgerichtet haben und die Produktion multimedialer Inhalte für Produktschulungen etabliert haben, befassen sich ständig mit den oben stehenden Fragestellungen und passen ihre Strategie an Produktionsprozesse an, mit dem Ziel, Zeitaufwand und Produktionskosten für Trainingsmedien zu reduzieren.

Das konzeptionelle Design und die Produktion von eLearning-Content erweist sich derzeit als zu langwierig und kostenaufwändig angesichts schnelllebigter Produktportfolios. In der Regel sind in diesen Prozess mehrere Beteiligte wie Fachexperten, Autoren, Mediendesigner und Projektmanager involviert, da viele interdisziplinäre Detailkenntnisse (Technik, Tools, Projektmanagement, Medienproduktion, didaktische Expertise) benötigt werden. Für die Fachexperten beispielsweise, ist der Zeitaufwand vergleichsweise hoch, da ihr inhaltliches Know-how benötigt wird, die Explizierung ihres Wissens aber keinesfalls eine Routineaktivität

darstellt. Die Einbindung vieler Mitarbeiter aus verschiedenen Abteilungen erhöht den erforderlichen Aufwand zur Koordination und Kommunikation, besonders hinsichtlich des Dokumenten- und Datenaustauschs. Zudem unterstützen die bereits existierenden Werkzeuge nur singuläre Aspekte der Content-Produktion, bieten aber keine holistische Prozessintegration von Konzeption, Produktion und Management von Content. Auf der anderen Seite existieren aber auch Lösungen für das Management und den Austausch von Content, wie z.B. Content-Management-Systeme (CMS) und Repositories.

Diese Werkzeuge müssen, zusammen mit Editoren zur Konzept- und Drehbucheerstellung und anderen spezialisierten Tools zum Projektmanagement, manuell kombiniert werden. Was fehlt, ist ein übergreifendes, integratives System zur Kooperation und Ausführung dieser Tasks, das von allen Mitarbeitern verwendet wird.

Vor diesem Hintergrund wurde im Rahmen des vom BMWi-geförderten Forschungsprojekts EXPLAIN (Expertengestütztes Toolset zur entwicklungsbegleitenden Erstellung von Trainingsmedien im Product Life Cycle in Industrieunternehmen; <http://www.explain-project.de>) [ZBCH05] eine web-basierte Authoring Management Plattform zur Unterstützung des Erstellungsprozesses von eLearning-Content und des Managements unternehmensinterner Content-Projekte entworfen, die eine Integration bereits existierender Autorentools erlaubt (siehe Abb. 1). Nach einer Ist-Analyse der bestehenden Content-Erstellungsprozesse dreier Industrieunternehmen, wurde ein Reengineering der Prozesse durchgeführt [HaCh94], in dessen Rahmen ein idealer Soll-Prozess der Content-Produktion konzipiert und entworfen wurde [ChLM06]. Dieser Prozess soll in integrativer Art und Weise die Teilprozesse Content-Entwicklung und Content-Management durch Nutzung bestehender Schnittstellen miteinander verbinden und zusammen mit einfachen und günstigen Umsetzungstools die Akzeptanz und Nutzung in Unternehmen verbessern. Das Content-Modell gilt dabei als zentrales Element des gesamten Content-Erstellungsprozesses. Es ist das Ergebnis eines anfänglichen Projektplans, der die Lernziele und den Content beschreibt, wird in der Konzeptionsphase erstellt, und dient darüber hinaus als Basis für die Produktionsphase. Es ist vergleichbar mit einer Stückliste, die in Produktdesign und -entwicklung bzw. in der Produktionsplanung und -steuerung in Industrieunternehmen zum Einsatz kommt [Sche90].

Im Folgenden werden zunächst die Schwachstellen der Content-Erstellungsprozesse in produzierenden Unternehmen analysiert. Darauf bezugnehmend folgt eine Beschreibung der Methodik, die in der EXPLAIN Authoring Management Plattform zur Anwendung kommen

wird. Schließlich wird als Anwendungsfall aus der Industrie der Automatisierungstechnik das Unternehmen Festo vorgestellt, an dessen Beispiel die unternehmensspezifischen Ist-Prozesse der Content-Erstellung sowie die Erfahrungen mit der Methodik des Plattformszenarios bzgl. der reengineerten Prozesse und der Content-Modellierung erläutert und bewertet werden.

2 Schwachstellenanalyse der Content-Erstellung in Industrieunternehmen

Als praktische Grundlage der Betrachtung und Analyse der Schwachstellen bei der Content-Erstellung dient die gegenwärtige Praxis dreier Unternehmen in den Branchen Pharmazie, Elektrotechnik und Automatisierungstechnik. Ihre bestehende Affinität zum Thema eLearning, dessen unternehmensweite Umsetzung und dementsprechend etablierte Prozesse der Content-Erstellung zeichnen sie als geeignete Untersuchungsobjekte aus [ChLM06]. Im Rahmen der Analyse der Ist-Prozesse der Produktion von multimedialen Trainingsinhalten bei den drei EXPLAIN-Anwendungspartnern wurden nachfolgende Schwachstellen identifiziert:

- **Hoher Zeitbedarf bei „teuren“ Mitarbeitern:** Der Zeitbedarf von Fachexperten aus den Fachabteilungen bei der Produktion von Rich Media Content ist sehr hoch. Aufgrund ihrer inhaltlichen Kompetenz ist ihre Beteiligung aber essentiell. Um den Projekterfolg sicherzustellen, sind sie an der inhaltlichen Konzeption meist intensiv beteiligt. Sie erfüllen Aufgaben wie Bereitstellung von Basismaterial, Erklärung der Produkte (Briefing), Mitarbeit und Abstimmung bei der Entwicklung von Konzepten und Drehbüchern, Abnahme und Feedback von Entwicklungs- und Produktionsergebnissen sowie laufende Unterstützung bei der Produktion.
- **Fehlende Werkzeugintegration:** Über den gesamten Konzeptions- und Produktionsprozess kommt i.d.R. eine Vielzahl von Werkzeugen und Medientypen zum Einsatz und ein vermehrtes Auftreten von Medienbrüchen ist unvermeidlich. Die Gründe liegen zum Einen darin, dass das Ausgangsmaterial in verschieden gearteten Quellen vorliegt und zum Anderen in der heterogenen Infrastruktur in den Unternehmen existierender und eingesetzter Werkzeuge, die von Office-Applikationen über Drehbuecherstellungswerkzeuge bis hin zu Autorentools reichen. Hinzu kommen Spezialwerkzeuge für Grafik-, Video-, Audibearbeitung,

Animationen, Softwaresimulationen, CAD-Tools usw. Durch die fehlende Werkzeugintegration entstehen viele Aufwände an verschiedenen Stellen mehrfach.

- **Hoher persönlicher Abstimmungs-, Kommunikations- und Projektmanagementaufwand:** Meist sind in Projekten zur Content-Erstellung eine Vielzahl von Personen beteiligt (interne Stakeholder, externe Partner, Spezialisten, Didaktiker, Drehbuchautoren etc.), deren Abstimmung sehr aufwändig und kommunikationsintensiv ist. Oftmals wird die Abstimmung nicht durch elektronische Kommunikation unterstützt sondern erfolgt in Form von persönlichen Treffen, an denen auch die Fachexperten beteiligt sind, was die Laufzeit der Projekte allein aufgrund von Terminierungsschwierigkeiten verlängern kann. Dies betrifft sowohl die Abstimmung über generelle Vorgehensweisen, Projektstandards und Werkzeuge als auch die Abstimmung inhaltlicher Aspekte und Details.
- **Große Datenmenge und -redundanz:** Die im Prozess zu verarbeitende Datenmenge ist extrem hoch. Es fehlen integrierte Konzepte zur Verwaltung aller Teil- und Zwischenergebnisse, inhaltlichen Materialien und medialen Elemente. Dies führt in der Folge zu einem hohen Managementaufwand für Versionen, Releasestände, Produktvarianten, Fremdsprachensysteme usw. Aufgrund dessen, dass meist mehrere Personen im Prozess beteiligt sind, steigt der Datenhaltungsaufwand exponentiell. Eine hochgradige Redundanz der Datenhaltung für alle im Prozess entstehenden Ergebnisse ist zwangsläufig.
- **Zielgruppenspezifische Produktvarianten und aufwändige Lokalisierungsprozesse:** Meist sind die Trainingsmedien für Produkte in mehreren Sprachen zu erstellen. Trotz teilautomatisierter Prozesse für das Einspielen von Fremdsprachen (Texte, Audios) und die automatische Synchronisierung von Medien ist der Lokalisierungsaufwand sehr hoch.
- **Hohe Folgekosten durch Aktualisierungsaufwendungen:** Die Aktualisierung von Trainingsmedien bei Änderungen an den Produkten ist sehr aufwändig. Dies gilt zum Einen für die Analyse der Punkte, die geändert werden müssen und zum Anderen für die Durchführung der Änderungen in meist mehrfach existierenden Versionen, z.B. bedingt durch lokalisierte Inhalte in mehreren Sprachen.

- **Divergenz zwischen erwarteten Kosten und tatsächlich realisierten Gesamtkosten:** Die Komplexität der Content-Erstellung, die in der „Natur“ der Sache liegt und mit kreativen Prozessen einhergeht, wird meist unterschätzt. Selbst wenn umfangreiche Erfahrungen und Kompetenzen vorliegen, Projektteams eingespielt sind, Autorentools umfangreich eingesetzt und Prozesse standardisiert sind, werden die Kosten kontinuierlich unterschätzt.

Den identifizierten Schwachstellen liegen folgende Rahmenbedingungen zugrunde:

- Alle industriellen Partner verfügen bereits über Erfahrung in der Content-Erstellung und haben bereits mehrere Trainingsmedien produziert.
- Produktionsprozesse sind systematisch entwickelt und implementiert.
- Autorentools, definiert als Werkzeuge zur templategestützten Medienproduktion, wurden eingesetzt und sind in der Prozessanalyse berücksichtigt.
- Die Ist-Prozesse betrachten sowohl in-house-Produktionen als auch outgesourcte Content-Erstellungsprozesse durch unterschiedliche Produktionspartner.

Es kann somit weitestgehend ausgeschlossen werden, dass die obigen Problemstellungen aufgrund von individueller Kompetenz, Partnerstruktur, aktuellen Werkzeugen oder sonstigen Rahmenbedingungen als Einzelfälle zu betrachten sind. Vielmehr wird davon ausgegangen, dass die Schwachstellen grundsätzlich existieren und nur durch ein grundlegendes Überdenken des Content-Erstellungsprozesses, der Art der Produktion sowie der Integration der Werkzeuge entlang des Prozesses zu lösen sind. Hierin sieht EXPLAIN seine wesentlichen Forschungs- und Entwicklungsfragen.

3 EXPLAIN Authoring Management Plattform

Das Forschungsprojekt EXPLAIN verfolgt den Ansatz, Unternehmen durch eine integrierte Authoring Management Plattform in die Lage zu versetzen, ihre Lerninhalte selbständig zu produzieren und so auf konkreten Lernbedarf zeitnah und kostengünstig reagieren zu können. Dabei wäre es unrealistisch und ökonomisch wenig sinnvoll, wenn die Unternehmen alle bei der Produktion der eLearning Materialien benötigten Teilkomponenten völlig autark und eigenhändig erstellen. Stattdessen soll die EXPLAIN Authoring Management Plattform als

web-basierte Lösung eine breite Palette von Content-Erstellungsprozessen, Services und Tools auf „on-demand“-Basis anbieten (siehe Abb. 1) und auch externe Anbieter je nach Wunsch einbeziehen. Die hierbei verfolgte Idee ist, dass ein Unternehmen im Rahmen eines Content-Projektes auf genau die Werkzeuge zugreifen, genau die Unterstützung erhalten und genau die Dienstleistungen in Anspruch nehmen kann, die es gerade benötigt.

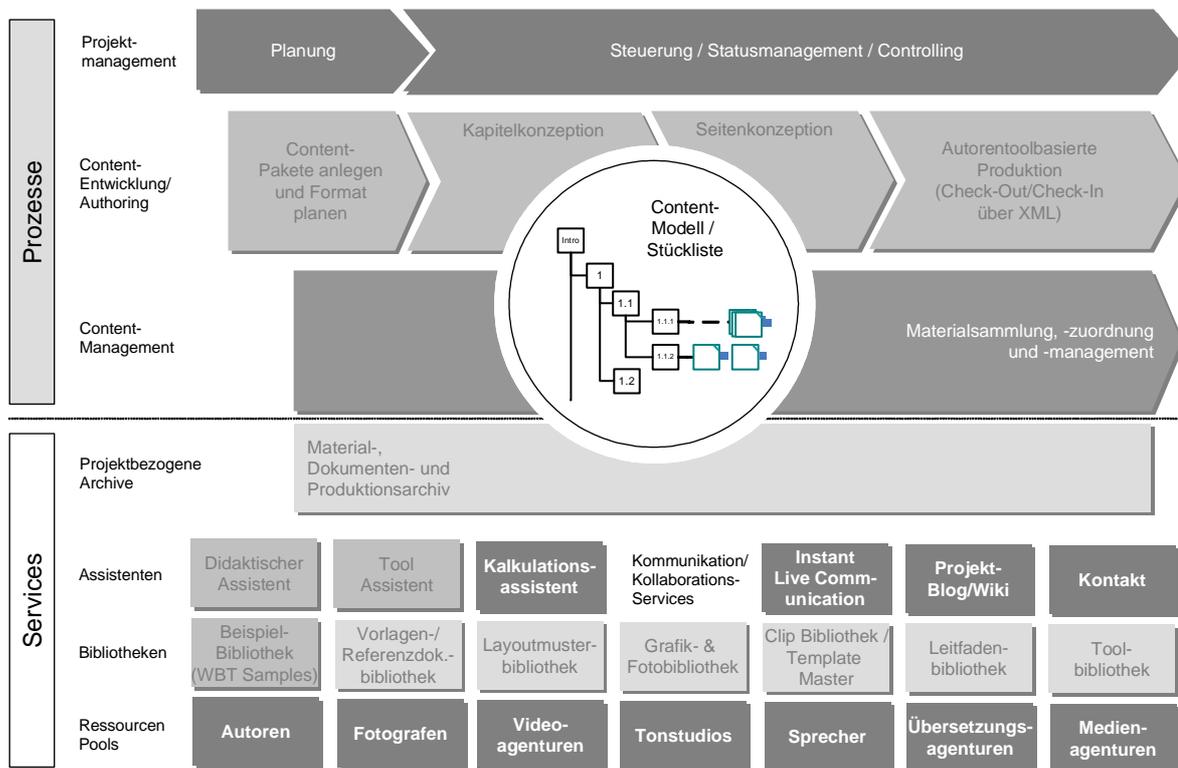


Abb. 1: Prozess- und Service-Landkarte der EXPLAIN Authoring Management Plattform

Zentraler Bearbeitungsgegenstand ist dabei das Content-Modell, welches alle anfallenden Aktivitäten entlang der Struktur des zu entwickelnden eLearning-Moduls abbildet. Es stellt somit die Schnittstelle zwischen den Prozessen des Content-Managements, der Content-Entwicklung und des Projektmanagements dar. Darüber hinaus bietet die Plattform Mehrwertdienste (Services) an, die Hilfestellung bei didaktischen Fragestellungen, bei der Werkzeugauswahl, bei der projektinternen und -übergreifenden Kommunikation und Kollaboration, bei der Suche nach Medienexperten sowie Bibliotheken vorgefertigter Templates und Medien-Assets bieten, und durch die sich die Plattform von CMS unterscheidet [ChLe06]. Für die weitere Betrachtung ist der Prozess der Content-Erstellung von besonderer Bedeutung, in dessen Rahmen viele Aspekte berücksichtigt werden müssen. Zusätzlich zur Medienproduktion und deren Zusammenstellung zu einem Lernobjekt, besteht dieser aus konzeptionellem Design, Material- und Ressourcenmanagement sowie Projektmanagement.

Das Ziel ist es, diesen komplexen kollaborativen Prozess zu beschleunigen und zu vereinfachen. Des Weiteren soll es Wissensträgern bzw. Fachexperten ermöglicht werden, die Zuständigkeit für viele Prozessschritte zu übernehmen, wobei dies durch automatisierte Führung entlang der einzelnen Schritte und durch kontextsensitive Tipps zur didaktischen Gestaltung vereinfacht werden soll [LARC06]. Die integrierte Betrachtung von Produkt- und Content-Entwicklung ermöglicht es, dass Content bereits während des Produktentwicklungsprozesses durch die Fachexperten erarbeitet werden kann, was insbesondere bei dieser am Content-Produktionsprozess beteiligten Personengruppe zu einer signifikanten Zeitersparnis, und damit verbunden auch zu einer Kostenreduktion für das jeweilige Unternehmen führt.

4 Die Content-Modellierungsmethodik

Da sich die EXPLAIN Authoring Management Plattform noch in der Entwicklung befindet, war diese für die Anwendungspartner im EXPLAIN Projekt für den Showcase noch nicht anwendbar. Nichtsdestotrotz sollten aber die Konzepte und Methoden, die einen großen Teil der Funktionalität der Plattform ausmachen zur Anwendung kommen. Hierzu zählt vor allem der Content-Modellierungsansatz der Plattform, der in diesem Kapitel beschrieben werden soll.

4.1 Content-Modell statt Drehbuch

Momentan herrscht bei der Content-Erstellung ein Prozess, wie er in Abb. 2 gezeigt ist vor.

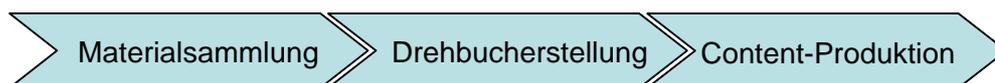


Abb. 2: Prozessschritte bei der Content-Erstellung

Der EXPLAIN-Plattform liegt die Annahme zu Grunde, dass die Content-Modellierung direkt an die Content-Produktion im Autorenwerkzeug anschließen kann, so dass die Prozesse der Materialsammlung und Drehbucheerstellung redundant werden. Diese Prozesse werden dabei durch die Content-Modellierung ersetzt, wodurch eine redundante Pflege von Materialien und anderen Dokumenten, die neben dem Drehbuch existieren, vermieden werden soll. Eine Grundvoraussetzung ist, dass das Content-Modell genügend Informationen enthält um als aussagekräftiges Modell des späteren Contents und so als Kommunikations- und Prozessgrundlage zu dienen. Dies wird noch durch die Modellierungsfähigkeiten vieler moderner Autorentools begünstigt. Somit ergibt sich eine Prozessabfolge gemäß Abb. 3.



Abb. 3: Content-Erstellungsprozess in EXPLAIN

Da die Prozessschritte „Materialsammlung“ und „Drehbucheerstellung“ aus Abb. 2 durch die „Content-Modell-Erstellung“ in Kombination mit den eingesetzten Autorenwerkzeugen ersetzt werden können, erfolgt eine Reduktion der Teilprozesse, wodurch z.B. Medienbrüche und redundante, nebenläufige Arbeitsschritte, aber auch weitere Prozessverluste der alten Teilprozesse vermieden werden. Zudem wird die Erstellung des Content-Modells als weniger komplex und zeitaufwendig angesehen als eine Drehbucheerstellung.

4.2 Das EXPLAIN Content-Modell

Das Content-Modell in der EXPLAIN-Plattform setzt sich aus drei Komponenten zusammen. Die *logische Struktur* wird mit Hilfe von Kapiteln und Unterkapiteln abgebildet, während die *physikalische Struktur* durch Seiten, bzw. falls die exakte Seitenzahl zum Modellierungszeitpunkt nicht bekannt ist, Seitengruppen abgebildet wird. Den Seiten können darüber hinaus *Materialien* zugeordnet werden, die später auf der Seite enthalten sein sollen. Zu Materialien zählen alle Medienobjekte, die in dem fertigen eLearning-Content auftreten können. Sollte ein Material zum Modellierungszeitpunkt noch nicht verfügbar sein, weil z.B. ein Bild erst noch produziert werden muss, kann es mit Hilfe einer so genannten Materialnotiz umschrieben werden. Diese wird dann zu einem späteren Zeitpunkt durch das echte Material ersetzt. Abb. 4 zeigt eine Instanz des beschriebenen Content-Modells.

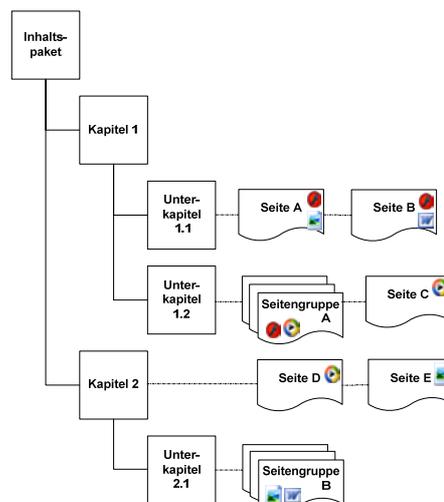


Abb. 4: Instanz des EXPLAIN Content-Modells

Neben der Anreicherung der Seiten durch Materialien können zu jedem Kapitel, jeder Seite und jedem Material zusätzliche Informationen festgehalten werden. Die folgende Tabelle zeigt einen Auszug der Informationen, durch die das Content-Modell angereichert wird. Die Anreicherung kann in der EXPLAIN-Plattform entweder manuell durch den Nutzer, in vielen Fällen jedoch auch voll- oder halbautomatisch durch die Plattform erfolgen. Dies ist z.B. beim Attribut „Status“, aber auch bei Kostenattributen und Vererbungen von Zuständigkeiten für untergeordnete Objekte möglich. Über diese inhaltlichen Informationen hinaus ist das Content-Modell im Kontext der EXPLAIN-Plattform noch in anderer Hinsicht angereichert. Es dient als Grundlage für die drei es umgebenden Prozesse (vgl. Abb. 1) und muss daher die notwendigen Informationen bezüglich des Projektmanagements und des Content-Managements enthalten. Diese werden, ebenso wie der für das Content-Modell der EXPLAIN-Plattform entwickelte Modellierungsansatz in [LARC06] beschrieben.

Attribut	Beschreibung
Inhaltsbeschreibung	Textuelle Kurzbeschreibung des Inhalts der Seite (oder des Kapitels)
Bemerkungen	Bemerkungen der Person, die die Modellierung durchführt
Screenertext	Text der auf der Seite erscheinen soll
Sprechertexte	Sprechertexte für evtl. auf der Seite vorhandene Video- oder Audioobjekte
Didaktischer Seitentyp	Beschreibt den Seitentyp unter Verwendung der Ergebnisse der didaktischen Hilfefunktion der Plattform
Offene Punkte	Konzeptionell noch offene Fragen
Navigationsanweisungen	Anweisungen, wie auf der Seite navigiert werden soll
Frage- und Antworttext	Text für evtl. auftretende Testfragen
...	...

Tab. 1: Anreichernde Informationen

4.3 Die Modellierung

Das beschriebene Content-Modell muss noch eine weitere Voraussetzung erfüllen. Es sollte leicht und schnell erstellbar, bearbeitbar und begreifbar sein; und dies insbesondere von den erwähnten Fachexperten. Aus diesem Grund wurde dem Content-Modell das bekannte Buchparadigma zugrunde gelegt (vgl. Abb. 4). Es ist auch von Fachexperten intuitiv und schnell erfassbar und kann deshalb ohne Probleme modelliert werden. In der EXPLAIN-Plattform wird die Modellierung mit einem leichtgewichtigen, einfach zu bedienenden Editor

umgesetzt, mit dessen Hilfe es möglich ist, innerhalb weniger Minuten die logische Struktur eines eLearning-Kurses zu erstellen und mit Seiten zu befüllen. Mit Hilfe der in die Plattform integrierten Repositories für Materialien können diese mit wenigen Klicks den Seiten zugeordnet werden. Einige der anreichernden Informationen können dabei von der Plattform erfasst und eingetragen werden.

5 Fallbeispiel aus der Automatisierungsindustrie

5.1 Der Festo-Anwendungsfall

Festo AG und Co. KG ist ein international agierender Anbieter von Automatisierungstechnik. Das global ausgerichtete, unabhängige Familienunternehmen mit Hauptsitz in Esslingen am Neckar hat sich in mehr als 40 Jahren in seiner Branche etabliert – durch Innovation und Problemlösungskompetenz im Bereich Pneumatik sowie mit einem großen Angebot an industriellen Aus- und Weiterbildungsprogrammen. Das Festo Lernzentrum als ein Unternehmen der Festo Gruppe ist anerkannter Weiterbildungsträger. Hier werden sowohl Führungskonzepte und betriebswirtschaftliches Know-how als auch globales Wissen und Erfahrungen weitergegeben. Für Unternehmen, öffentliche Institutionen und Privatpersonen bieten sich hier aussichtsreiche Rahmenbedingungen zur Weiterbildung. Das Leistungsspektrum des Festo Lernzentrums erstreckt sich von Coaching und Prozessbegleitung über Erst- und Verbundausbildung, Umschulung und Qualifizierung bis hin zu Weiterbildungsmaßnahmen für unterschiedliche Bereiche (Technik, EDV, Führung und Zusammenarbeit, Wirtschaft, Produktion, Logistik). Des Weiteren werden auch Forum-Veranstaltungen und Fördermittelberatung angeboten sowie eLearning-Aktivitäten durchgeführt.

Festo besitzt bereits langjährige Erfahrungen im Einsatz von CBT- und WBT-Lösungen. Seit der Einführung einer eLearning Plattform im Jahr 2000 werden die Trainingsinhalte den Mitarbeitern im Festo Konzern weltweit online angeboten. Hieraus ergibt sich die Notwendigkeit der Mehrsprachigkeit der angebotenen Inhalte. Das Angebot reicht von individuellen Produktschulungen für den Vertrieb bis hin zu standardisierten Lernprogrammen im Bereich der EDV-Schulungen. Weitere Themenschwerpunkte sind Fremdsprachentraining und Soft Skill Trainings. Der Einsatz der WBTs erfolgt in verschiedenen didaktischen

Szenarien von Selbstlernprozessen bis hin zu Blended Learning. Zur Flexibilisierung der Content-Prozesse für individuelle WBTs (z.B. Produktschulungen) und zur Kostensenkung beim Einsatz von eLearning werden diese WBTs seit 2003 weitgehend in-house produziert.

Im Rahmen des EXPLAIN-Projekts hat das Festo Lernzentrum im Showcase I erste Erfahrungen mit einem optimierten Content-Erstellungsprozess gemacht, der sich bereits in dieser frühen Phase von der bisherigen Praxis unterscheidet [Bruc05]. Als Thema des ersten Showcase wurde das Führungskräfte-Training „Auswahl von MitarbeiterInnen“ gewählt.

5.2 Bisherige Prozesse der Content-Erstellung bei Festo

Wie bereits erwähnt, wird Content bei Festo seit 2003 weitgehend in-house produziert. Zur Ausarbeitung der Inhalte werden Fachexperten herangezogen, die Produktion wird jedoch von speziell fortgebildeten Programmierern und Drehbuchautoren im Unternehmen durchgeführt. Dabei liefert der Fachexperte die Inhalte in von ihm bearbeitbaren Formaten (Word oder PowerPoint), welche dann durch den Drehbuchautor und Programmierer in die entsprechenden Autorentools eingefügt werden.

Der bisherige Prozess der Content-Erstellung verlief generell in folgenden Schritten:

- 1. Content-Planung: Fachexperte und Drehbuchautor definieren gemeinsam Zielgruppe und Lernziele des Inhaltes. Eine Grobstruktur des Lernprogramms wird in PowerPoint oder Word erstellt.
- 2. Drehbucherstellung: Anhand der Grobstruktur und auf Basis der durch den Fachexperten gelieferten Medien und Daten in Form von Word-Dokumenten oder PowerPoint-Präsentationen wird durch den Drehbuchautor ein Drehbuch in zwei Schritten erstellt: Screendesign des Lernprogramms in PowerPoint sowie das ausgearbeitete Drehbuch in Word. Das Drehbuch enthält die detaillierten Texte, Programmieranweisungen, Hinweise auf Medienquellen, Beschreibungen von noch zu erstellenden Medien, Beschreibungen zur Vertonung etc.
- 3. Abnahme des Drehbuchs: Der Fachexperte überprüft/korrigiert das Drehbuch.
- 4. Medienproduktion: Nach der Freigabe des Drehbuchs durch den Fachexperten beginnt die Medienproduktion: Bildersuche/Bildbearbeitung, Erstellung von Animationen, Sprachaufnahme, Musikauswahl, Videoauswahl/-produktion. Hierbei wird auf interne sowie externe Bilder- und Mediendatenbanken zurückgegriffen.

- 5. Content-Produktion: In diesem Schritt erfolgt die eigentliche Erstellung des Lernprogramms. Anhand der Vorlagen des Drehbuchs und des Screendesigns wird das Lernprogramm erstellt und die Medien zusammengeführt.
- 6. Abnahme des Lernprogramms / Korrekturen / Freigabe: Im letzten Schritt des Content-Erstellungsprozesses erfolgt die Kontrolle und Abnahme des Contents durch den Fachexperten. Im Rahmen dieses letzten Schrittes kann es zu einer oder mehreren Korrekturphasen kommen.

Im hier beschriebenen Prozess ergeben sich verschiedene Schnittstellen zwischen Drehbuchautor, Programmierer, Fachexperte und externen Dienstleistern (z.B. Tonstudio). Insbesondere die Schnittstelle zwischen Fachexperte und Drehbuchautor/Programmierer erweist sich häufig als problematisch. Die Schwierigkeit besteht darin, den Nicht-Spezialisten der Content-Erstellung (Fachexperte auf einem Gebiet) den Ablauf der Prozesse zu verdeutlichen und aufzuzeigen, in welchen Phasen die Unterstützung der Fachexperten unerlässlich ist. Aus den praktischen Erfahrungen hat sich gezeigt, dass der Fachexperte z.B. ein Drehbuch freigibt, ohne es gründlich geprüft zu haben, da die rein textbasierte Darstellung nicht geeignet ist, den späteren Content richtig zu repräsentieren. Diese Problematik hat bspw. Auswirkungen auf die Phase 6 der Content-Erstellung, da zusätzliche und zudem aufwändigere Korrekturphasen benötigt werden, was wiederum Auswirkungen auf Kosten und Zeitrahmen des Content-Projektes hat. Ein weiteres Problem der Content-Erstellung nach dem bisherigen Prozess ist die Beschaffung und Archivierung von Daten und Medienmaterial. Da mehrere Personen im Content-Erstellungsprozess arbeiten, werden z.B. Bilder oder Animationen häufig mehrfach archiviert. Derzeit gibt es keine zentrale Mediendatenbank, die einen gemeinsamen Zugriff ermöglichen würde. Dies hat eine redundante Speicherung von Daten und Medien zur Folge. Aufgrund der komplexen Prozesse und der für Nicht-Spezialisten ungeeigneten Technologien sind die Aktivitäten der Content-Erstellungsprozesse noch nicht systematisch mit den Geschäftsbereichen verbunden.

5.3 Reengineerte Content-Erstellung bei Festo

Im ersten Showcase von Festo wurde anhand der in EXPLAIN auf Basis der identifizierten Schwachstellen entwickelten neuen Prozesse und Methoden ein reengineerter Content-Erstellungsprozess angewandt, der nun folgendermaßen verläuft:

- 1. Content-Planung: In dieser Phase werden Lernziele, Zielgruppe und Grobstruktur festgelegt. Die Erstellung der Grobstruktur erfolgt auf der Basis der vom Fachexperten gelieferten Inhalte. Des Weiteren wird zur Grobkonzeption auf speziell entwickelte didaktische Muster [Nieg06] zurückgegriffen, um die geeignete Formatentscheidung für das Thema zu treffen. Die jetzt zur Verfügung stehenden didaktischen Muster und Anleitungen zur Formatentscheidung in Bezug auf Strukturierung, Gestaltung, Motivationsdesign und Instruktionsdesign sind in dieser Phase sehr hilfreich, da dieser Schritt im bisherigen Prozess nicht strukturiert war.
- 2. Erstellung des Content-Modells: In dieser Phase erfolgt nun die Erstellung des Content-Modells durch Darstellung der Kapitel- und Seitenstruktur des Contents. Zu den einzelnen Seiten werden zudem Hinweise zu Medien, Daten, Texten, bereits verfügbaren und zu erstellenden Materialien hinzugefügt. Das Content-Modell ersetzt nun das frühere Drehbuch.
- 3. Abnahme des Content-Modells durch den Fachexperten: Die Abstimmung und Abnahme mit dem Fachexperten erfolgt anhand des Content-Modells. Durch die visuelle Darstellung der Struktur des Lernprogramms und Kapitel- und Seitennamen, sowie Hinweise zu Materialien kann der Content wesentlich besser für den Fachexperten präsentiert werden.
- 4. Medienproduktion: In diesem Schritt ergaben sich keine Veränderungen.
- 5. Content-Produktion: Programmierung des Lernprogramms auf Basis des Content-Modells und Zusammenführung der Medien.
- 6. Abnahme des Lernprogramms / Korrekturen / Freigabe: Aufgrund der einfachen Zugänglichkeit des Content-Modells und dessen Verständlichkeit erfolgt der Abnahmeprozess durch den Fachexperten jetzt wesentlich schneller, Korrekturphasen können reduziert werden.

Abb. 5 zeigt einen kleinen Ausschnitt aus dem Festo-Content-Modell zum WBT „Auswahl von Mitarbeitern“. Die in grau markierten Seiten enthalten Festo-spezifischen Content.

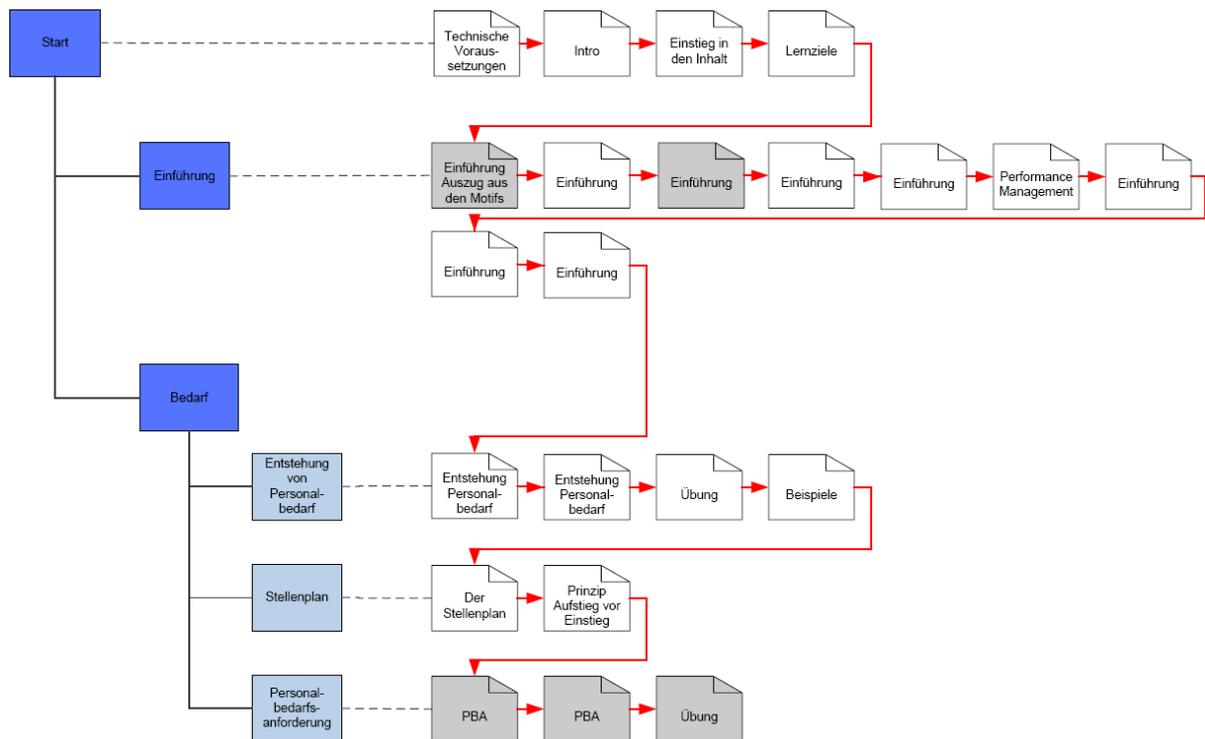


Abb. 5: Ausschnitt aus Festo-Content-Modell zum WBT „Auswahl von Mitarbeitern“

Insgesamt konnte der Content-Erstellungsprozess durch den Einsatz vom in EXPLAIN entwickelten Content-Modellierungsansatz vereinfacht und beschleunigt werden. Durch den Wegfall der Drehbucheerstellung konnte signifikant Zeit in der Content-Erstellung eingespart werden. Drehbücher sind aufgrund ihrer Textlastigkeit in der Handhabung oft komplex, werden von vielen Mitarbeitern zugleich gepflegt und enthalten dementsprechend viele Redundanzen bzw. sind daher schwer oder nur unter hohem Arbeitsaufwand anpassbar. Insgesamt stellt sich die Abnahme solcher unhandlichen Drehbücher daher in der Praxis als selten ernst genommenen Prozess dar. Durch den Einsatz des Content-Modellierungsansatzes wurde die Schnittstelle zum Fachexperten vereinfacht und somit deutlich verbessert. Das Content-Modell, welches das Lernprogramm visualisiert darstellt, erwies sich als eine wesentliche Verbesserung des gesamten Prozesses. Die Abnahme des Content-Modells, welches nun als Ersatz für das Drehbuch eingesetzt wird, war für den Fachexperten leichter und nachvollziehbarer, wodurch der Abnahmeprozess für die entsprechenden Verantwortlichen verbindlicher als bisher wurde. In Folge dessen hat sich der Bedarf an Korrekturphasen im letzten Schritt verringert, und damit auch der endgültige Abnahmeprozess verkürzt.

Die Content-Produktion insgesamt wurde beschleunigt und ist nun weniger aufwändig. Der sonst komplexe Prozess der Drehbucheerstellung mit z.B. der Ausformulierung von Texten entfiel bzw. wurde direkt in die Programmierung/Produktion übertragen.

Hierbei ist auch zu erwähnen, dass weniger Werkzeuge (nur Visio und Autorentool anstatt Word, PowerPoint, Autorentool) verwendet wurden und somit die Anzahl von Medienbrüchen während des Content-Erstellungsprozesses reduziert werden konnte. Dies bedeutete eine einfachere Handhabung des gesamten Prozesses.

Im Rahmen des Projektmanagements hat sich der Abstimmungs- und Kommunikationsaufwand deutlich reduziert. Da die Abstimmungsprozesse im Vergleich zur vorherigen Praxis nun schneller und übersichtlicher laufen, hat sich die Zusammenarbeit zwischen den einzelnen Akteuren verbessert. Aufgrund der visualisierten Darstellung der Content-Erstellung verfügen die einzelnen Prozessschritte über hohe Transparenz und Nachvollziehbarkeit für alle am Content-Projekt Beteiligten entlang des gesamten Prozesses. Das sehr gute Feedback der internen Kunden, die den Modellierungsansatz angewandt und evaluiert haben, spricht für dessen zukünftigen Einsatz im Unternehmen.

Der reengineerte Prozess bei Festo hat auch andere Vorteile für das Content-Management mit sich gebracht, wobei sich die Menge der im Prozess zu verarbeitenden Daten und Materialien niedriger gehalten werden konnte und eine formatübergreifende Referenzierung und Zuordnung von Materialien ermöglicht wurde. Zudem konnten Datenredundanzen vermieden werden, da die Verwaltung der Inhalte in einer transparenten und übersichtlichen Struktur erfolgte.

Insgesamt ist festzuhalten, dass allein durch den methodischen Veränderungsansatz der Content-Modellierung, Festo den Prozess der Content-Erstellung sichtbar optimieren konnte. Von einer Einbindung dieses Konzeptes in die EXPLAIN-Plattform verspricht man sich zusätzliche Verbesserungen hinsichtlich des reibungslosen Zusammenspiels unterschiedlicher Content-Bestandteile (Reduktion von Medienbrüchen), der in Einsatz befindlichen Systeme (Integration) und der an der Content-Modellierung Beteiligten (Kollaboration).

6 Zusammenfassung und Ausblick

Im Rahmen dieses Beitrages wurden die Herausforderungen der Content-Erstellung in Industrieunternehmen dargelegt. Der Ansatz durch Content-Modellierung den Prozess zu vereinfachen und in bestehende organisatorische und technologische Strukturen zu integrieren

ist ein zentrales Forschungsergebnis des EXPLAIN-Projektes auf dessen Basis die EXPLAIN Authoring Management Plattform entworfen und entwickelt wird. Die vorgestellte Modellierungsmethodik unterstützt den gesamten Prozess der Content-Erstellung, inklusive Projektmanagement, Medienproduktion und -sammlung, Materialmanagement sowie die Gestaltung und Produktion der Lerninhalte. Dass sich somit eLearning-Content in Industrieunternehmen effizienter und effektiver entwickeln lässt, wurde anhand der Fallstudie Festo illustriert. Dabei hat sich gezeigt, dass semi-formale Modellierung die bisher fehlende Brücke zwischen betriebswirtschaftlichen, pädagogischen und technischen Fachexperten schlägt. Wie schon in anderen Anwendungsszenarien bekannt, sind semi-formale Modelle wie das vorgestellte Content-Modell einerseits leicht zu verstehen und für jeden zugänglich. Andererseits bieten sie aber eine ausreichende Präzision, um als Anforderungsbasis für die weitere informationstechnische Verarbeitung, in unserem Fall der Content-Erstellung zu dienen. Der Fokus der nächsten Entwicklungsschritte liegt auf der Implementierung der Plattform und der Integration der Content-Modellierung. Angesichts eines prototypischen Systems können in folgenden Showcases mit den Anwendungspartnern detaillierte Urteilsschlüsse über die Qualität und Validität des Ansatzes gezogen und auf die Weiterentwicklung der Modellierungssprache – insbesondere hinsichtlich einer didaktischen Erweiterung – angewandt werden.

Literaturverzeichnis

- [Attw03] *Attwell, Graham: The challenge of e-learning in small enterprises. Issues for policy and practice in Europe. In: Cedefop Panorama series 82 (2003), Luxembourg 2003.*
- [BaBS01] *Back, Andrea; Bendel, Oliver; Stoller-Schai, Daniel: E-Learning im Unternehmen. Orell Füssli, Zürich 2001.*
- [Bruc05] *Bruch, Eva-Maria: e-learning: Wie bringt man e-learning in den Mittelstand? In: Bruch, E.-M. (Hrsg): Newsletter Festo Lernzentrum GmbH (2005) 9, St. Ingbert 2005, S. 1-2.*
- [ChLe06] *Chikova, Pavlina; Leyking, Katrina: EXPLAIN – eine web-basierte Authoring Management Plattform. In: LERNET-Report 19 (2006), S. 8-10.*

- [ChLM06] *Chikova, Pavlina; Leyking, Katrina; Martin, Gunnar*: Data and Process Integration of eLearning Content Development and Product Engineering in SMEs. In: *Remenyi, Dan (Hrsg.)*: Proceedings of the International Conference on eLearning ICEL 2006, Montreal, Canada, 22.-23. Juni 2006. ACL, Reading, England 2006, S. 57-68.
- [HaCh94] *Hammer, Michael; Champy, James*: Business Reengineering: die Radikalkur für das Unternehmen. 3. Aufl., Campus, Frankfurt am Main et al. 1994.
- [LARC06] *Lehmann, Lasse; Abdelhak, Aqqal; Rensing, Christoph; Chikova, Pavlina; Leyking, Katrina; Steinmetz, Ralf*: A Content Modeling Language as Basis for the Support of the Overall Content Creation Process. Accepted for the 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006), 5.-7. Juli 2006, Kerkrade, Niederlande 2006.
- [NHHA04] *Niegemann, Helmut M.; Hessel, Silvia; Hochscheid-Mauel, Dirk; Aslanski, Kristina; Deimann, Markus; Kreuzberger, Gunther*: Kompendium E-Learning, Springer, Berlin et al. 2004.
- [Nieg06] *Niegemann, Helmut M.*: Integrating and Conveying Knowledge on How to Design Multimedia Learning Environments - The Development of an eLearning Design Assistant. In: *Uskov, Victor (Hrsg.)*: Proceedings of the 9th IASTED International Conference on Computers and Advanced Technology in Education, Lima, Peru, 4.-6. Oktober 2006. Acta Press, Anaheim, CA 2006, S. 394-398.
- [Sche90] *Scheer, August-Wilhelm*: CIM – Der computergesteuerte Industriebetrieb. 4. Aufl., Springer, Berlin et al. 1990.
- [ZBCH05] *Zimmermann, Volker; Bergenthal, Kathrin; Chikova, Pavlina; Hinz, Didier; Lehmann, Lasse; Leyking, Katrina; Martin, Gunnar; Rensing, Christoph*: Authoring Management Platform EXPLAIN. A new learning technology approach for efficient content production integrating authoring tools through a web-based process and service platform, ARIADNE PROLEARN Workshop, 1. Dezember, TU Berlin 2005.

Modularisierungskonzepte als Basis zur Gestaltung von nachhaltigen internetgestützten Bildungsangeboten am Beispiel der Aus-, Fort und Weiterbildung von Ökonomielehrkräften

Dr. Oliver Kamin, M.A. Karl Knispel

Institut für Ökonomische Bildung
An-Institut der Carl von Ossietzky Universität Oldenburg
26122 Oldenburg
{kamin, knispel}@ioeb.de

Abstract

Am Beispiel des Vorhabens Ökonomische Bildung online wird gezeigt, in wiefern modularisierte Lehr-/Lernmaterialien Voraussetzung sind, um einen erfolgreichen Beitrag zur bildungsmaßnahmenübergreifenden Realisierung eines Fort- und Weiterbildungsprogrammes im Zuge eines lebenslangen Lehr-/Lernprozesses (am Beispiel von Lehrkräften an allgemein bildenden Schulen) mit E-Learning-Arrangements zu leisten. Zunächst werden theoretische Grundlagen und wesentliche Anforderungen aus insbesondere didaktischer Sicht an einem Produkt-Baukasten formuliert, um einen durchlässigen bzw. maßnahmenübergreifenden Einsatz einzelner Lehrveranstaltungen verschiedener Anbieter ermöglichen zu können, um diese anschließend sinnvoll zu Qualifizierungspaketen auf veranstaltungsübergreifender Sicht bündeln zu können.

1 Einleitung

Die Steigerung der Leistungsfähigkeit von E-Learning-Materialien geht zweifelsohne mit der allgemeinen IT-Entwicklung einher. Insbesondere haben sich die Technologien zur multimediale Gestaltung in ihrem Leistungs- und Funktionsumfang weiterentwickelt. Zumindest theoretisch können E-Learning-Materialien aufgrund der gestiegenen Adaptionfähigkeit und der hieraus resultierenden Individualisierbarkeit für eine breitere Zielgruppe und einer größeren Vielfalt von Anwendungsszenarien ausgelegt werden. Die praktische Umsetzung bleibt aber oft hinter den Erwartungen zurück. Auch vernetzte Wissensstrukturen können im Vergleich zu

sequenziell angeordneten Inhaltsatomen aufgrund ihrer komplexen Strukturen selten ökonomisch vertretbar realisiert werden. Zusätzlich sind die mediendidaktischen Entscheidungen der Entwickler und Betreiber von E-Learning-Materialien oder virtuellen Lernumgebungen kritisch zu betrachten. Häufig entsteht der Eindruck, dass aufgrund des reinen Vorhandenseins einer Technologie hervorgerufene Leitideen blind verfolgt werden. Im schlimmsten Fall stören solche Angebote den individuellen und kollektiven Lernprozess mit der Konsequenz, dass es zu höheren Abbrecherquoten als im Vergleich zu thematisch gleichen Präsenzveranstaltungen kommt. Zunächst wird auf Produktebene aufgezeigt, dass es für eine zielführende und effiziente Einbindung von E-Learning-Arrangements in Lernprozessen erforderlich ist, sowohl Materialien als auch Bildungsmaßnahmen mit Hilfe von Modularisierungskonzepten mehrfachverwendbar zu gestalten. Erst hierdurch kann eine ökonomische Rechtfertigung des hohen Entwicklungsaufwands erfolgen. Auch ist eine individuelle Konfigurationsmöglichkeit von E-Learning-Materialien (Bildungsprodukte) und von Aus-, Weiter- und Fortbildungsmaßnahmen (Bildungsangebote) aus inhaltlich-didaktischer Sicht zwingend erforderlich, um individuelle Qualifizierungsbedarfe im Rahmen von lebenslangen Lernprozessen befriedigen zu können. Erst wenn dieses Fundament steht, kann über ein durchlässiges Qualifizierungskonzept und Geschäftsmodell nachgedacht werden, um eine sehr heterogen vorgebildete oder regional verstreute Zielgruppe zu erreichen. Der Aufbau einer solchen Lern-Community auf Basis des oben beschriebenen Fundaments wird am Beispiel Ökonomische Bildung Online vorgestellt.

2 Grundlagen

2.1 Module und Bausteine

Die steigende Komplexität von Software-Anwendungen erfordert neue Techniken zur Entwicklung flexibler Bausteine, die sich leicht an sich ändernde Anforderungen anpassen lassen [ABNe03, 19]. Somit stellt die Wiederverwendung einen wichtigen Aspekt in der Software-Entwicklung dar. Modulare Systeme können im Vergleich zu nicht modularen Systemen wesentlich einfacher konfiguriert und ausgebaut werden [Kilb+93, 126] und demnach wieder verwendet werden. Unter Modularisierung wird hier die Bildung einzelner Teilsegmente aus einer Gesamtheit verstanden, die durch ihre Austauschbarkeit gekennzeichnet sind [BaBr96, 51]. Module lassen sich also als geschlossene Programmteile beschreiben, die unabhängig vonein-

ander implementiert werden und über definierte Schnittstellen mit ihrer Umwelt kommunizieren. So erwächst die Möglichkeit, komplexe Programme arbeitsteilig bzw. parallel voneinander zu entwickeln [Müll00, 23] und auftretende Fehler leichter zu identifizieren.

Ein verwandter Begriff zum Modul ist der Baustein. Dieser findet häufig in der industriellen Produktentwicklung Anwendung. Bei Bausteinen werden in systematischer Weise standardisierte Objekte entwickelt. Sie zeichnen sich durch Vielfalt an Kombinationsmöglichkeiten aus und besitzen normierte Eigenschaften, die innerhalb des Baukastenumfelds aufeinander abgestimmt sind. Bausteine sind in sich geschlossen, das heißt, sie werden bei der Konfiguration gemäß ihrer vorbestimmten Zwecke angeordnet, ohne ihre Gestalt zu verändern [Kohl97, 8]. Bausteine können nach vielfältigen Kriterien klassifiziert werden. Mussbausteine stellen Bausteine dar, deren Vorhandensein im Baukastenprodukt obligatorisch sind. Ein Kannbaustein verleiht dem Baukastenprodukt hingegen additive Eigenschaften [Kohl97, 25]. Des Weiteren lassen sich Bausteine nach ihrer Funktionsübernahme im Endprodukt unterscheiden. Grundbausteine entsprechen den vorgenannten Mussbausteinen und bilden die Basis des Baukastenprodukts. Anpassbausteine dienen dazu, Produkte zu adaptieren. Sie erfüllen Funktionen, die in jeder Produktvariante vorhanden sein müssen, aber unterschiedlich auftreten können.

Das Zusammenspiel von Grund- und Anpassbausteinen bei der Konfiguration eines Erzeugnisses zeigt die Abb. 1 [Ehr195, 629]. Die Metallkonstruktionen im inneren der Lehne und der Sitz-

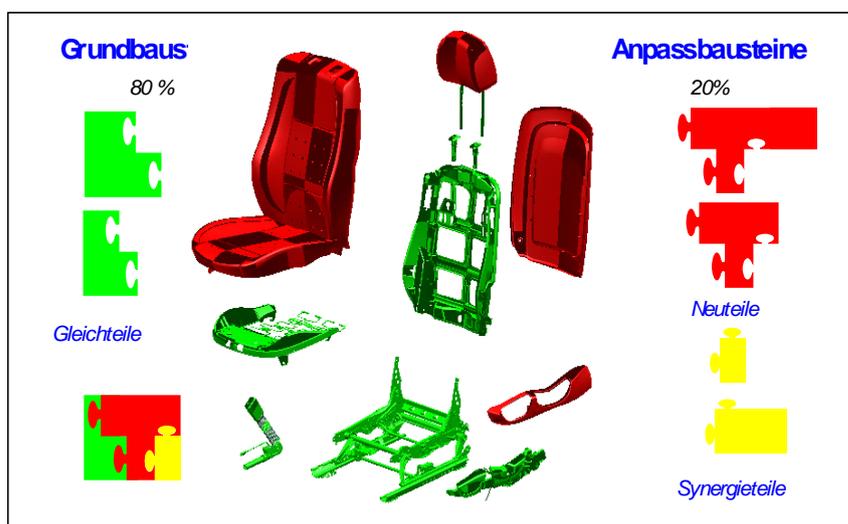


Abb. 1: Sitzbaukasten

fläche sowie die nach außen ragenden Befestigungseinrichtungen sind Grundbausteine Autositzes. Die Sitzbezüge der Lehnen können je nach Ausstattungsvariante aus Leder oder Stoff bestehen und verschiedene Formen aufweisen. Dies sind zum einen Eigen-

schaften von Anpassbausteinen, da sie in verschiedenen Form- und Funktionalitätsvarianten vorliegen können. Zum anderen können sie auch als Mussbausteine angesehen werden, da die Lehne ein obligatorisches Funktionsmodul innerhalb eines Autositzes darstellt.

2.2 Lebenslanges Lernen und individuelle Lernbiographien

Modularisierung wird auch in der Aus- und Weiterbildung aus vielerlei Hinsicht vorangetrieben. Neben Individualisierungsaspekten (Berücksichtigung von unterschiedlichen Bildungsvoraussetzungen des Lernenden) findet auch der Gedanke der Mehrfachverwendung einmal konzipierter Lernmodule (Kapselung von thematisch oder funktional zusammengehörigen Lerninhalten) eine hohe Beachtung, um beispielsweise die in Abb. 2 aufgeführten Aus- und Weiterbildungsmaßnahmen im Rahmen eines lebenslangen Lernprozesses miteinander zu verbinden. So können Module eines Ausbildungsberufs in anderen Berufsfeldern als Zusatzmodul in der Weiterbildung verwendet werden. Der Graph symbolisiert den Lernpfad einer Person, wobei in diesem Beispiel sechs Stationen durchlaufen werden [Kami04, 91f.]. Erfolgt nun eine Standardisie-

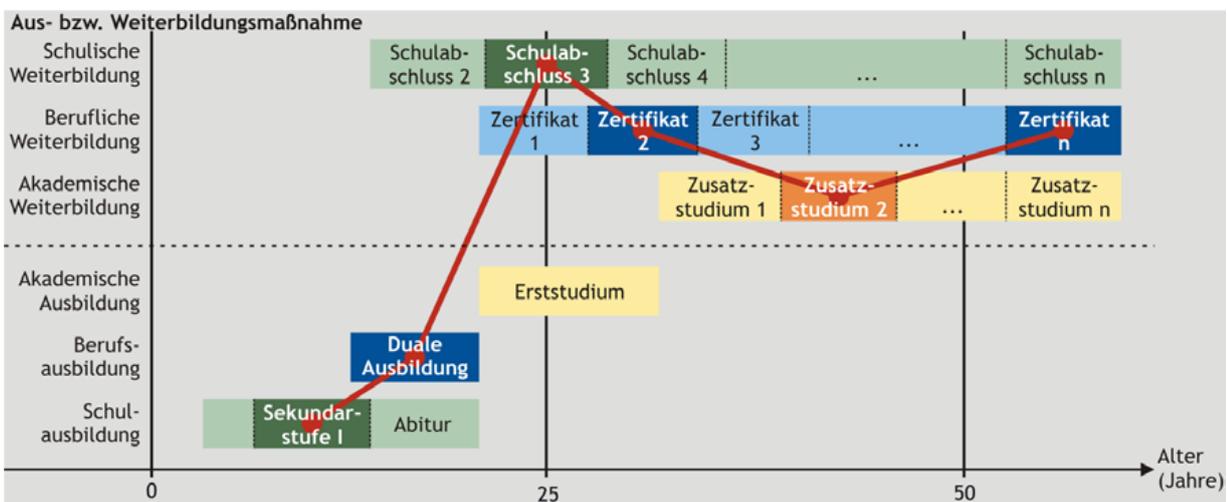


Abb. 2: Verlauf eines lebenslangen Lernprozesses

rung oder die Etablierung von Mindestnormen, können mehrfachverwendbare Lernmodule dezentral entwickelt und an verschiedenen Orten eingesetzt werden. Durch die Kapselung von Lerninhalten wird darüber hinaus ermöglicht, in Teilabschnitten erworbene Berufserfahrungen oder Wissensbestände separat zu zertifizieren [Kloa98, 203]. Eine solche Teilzertifizierung kann beispielsweise im Rahmen einer Leistungsanrechnung erfolgen. Dies ist eine der wichtigsten Voraussetzungen für die Durchlässigkeit und Flexibilität der beruflichen Aus- und Weiterbildung. Nur wenn entsprechende Lerninhalte ohne rechtliche und organisatorische Restriktionen auf andere Lehr-/Lernsituationen übertragen werden können, kann ein lebenslanger Lernprozess durchlaufen werden [Dubs03, 33].

Alle genannten Vorstellungen dienen als Grundlage, um eine schnellere und flexiblere Reaktion auf veränderte Anforderungen des Beschäftigungssystems zu ermöglichen. Bei modularen Systemen muss beispielsweise nicht das gesamte Berufsbild an neue Gegebenheiten angepasst

werden, sondern nur Teile. Ein weiterer Vorteil liegt in der risikoloserer Integration von neuen, noch nicht erprobten Teilqualifikationen, die bei einem Scheitern auf dem Arbeitsmarkt problemlos durch andere ersetzt werden können, ohne weitreichende Umstrukturierungen des Ausbildungsgangs vorzunehmen [Kloa98, 203]. Um dies einfacher zu ermöglichen, sollte auf zwei Ebenen modularisiert werden. Zunächst sind komplexe Bildungsmaßnahmen in Module zu strukturieren. Diese sind dann inhaltlich zu kapseln (zusammenhängende Themenstellungen als Lernmodule). Der Einsatz modularer Konzepte und die Auswahl der zu verwendenden Lerntechnologien haben konkrete Auswirkungen auf die Gestaltung von Lernprozessen. Dieser Umstand wirkt sich direkt auf den Aufbau von Lernmodulen, Unterrichtseinheiten und Curricula von Aus- und Weiterbildungsgängen aus [Dubs03, 32].

3 Entwicklung eines modularen Konzepts

3.1 Unterscheidung von Betrachtungsebenen

Durch die Modularität eines Bildungsproduktes können die wesentlichen Perspektiven von Autoren, Software-Entwickler, Bildungsprodukt-Anbieter, Aus- und Weiterbildungseinrichtungen als Bildungsprodukt-Abnehmer sowie den Lernenden berücksichtigt werden, die sich durch den gesamten Produktlebenszyklus ziehen in effizienter Weise unterstützt werden [KaHa03, 232f.]. Demnach können gleichartige Gestaltungskategorien sowohl bei der Planung von komplexen Bildungsangeboten als auch von einzelnen Bildungsprodukten herangezogen werden.

3.1.1 Mikroebene

Auf der Ebene der Unterrichtsführung (Mikroebene) geht es in der Unterrichtsplanung um die konkrete Ausgestaltung von Unterrichts- und Lernsituationen [Sloa00, S. 338]. Die Mikroebene repräsentiert also die Konfiguration von Bildungsprodukten und umfasst die Gestaltung der zu verwendenden Elemente, Komponenten und Lernmodule bis zum gesamten Bildungsprodukt. Die didaktische Beschreibung eines Bildungsprodukts bzw. einer Unterrichtsstunde kann durch

Ebene	Lerntheoretische Didaktik	Rahmenlehrplan	Lernziele	Modulkonzept
Makro	Didaktisches Konzept	Lehrplan/ Curriculum	Richtziele	Kombination von Bildungsprodukten zu Bildungsangeboten
Mikro	Strukturplanung Unterrichtsreihe	Lernfeld	Grobziele	Kursstruktur aus Lernmodulen (einzelnes Bildungsprodukt)
	Struktur- und Ablaufplanung Unterrichtsstunde	Themeneinheit	Feinziele	Lernmodul (Einheit aus Komponenten und Elementen)

Tab. 1: Analoge Strukturen bei der Planung und Beschreibung von Bildungsmaßnahmen

Feinziele erfolgen. Beispielsweise stellen einzelne Schwerpunktthemen eines Lernfelds und die dazugehörigen Zielformulierungen den Umfang eines Lernmoduls dar (vgl. Tab. 1). Module können zu einem Bildungsprodukt, welches den inhaltlichen Umfang eines Lernfelds abdeckt, gebündelt werden. Tab. 1 zeigt im mittleren Bereich, dass hier u.a. die betroffenen Feinziele zu einem Grobziel aggregiert werden [Kami04, 196].

3.1.2 Makroebene

Die Makroebene beinhaltet die Curriculumentwicklung, die sich in den spezifischen Lehrplänen für bspw. Ausbildungsberufe bzw. Bildungsangebote wiederfindet [Sloa00, 338]. Die Kombination verschiedener Bildungsprodukte zu Bildungsangeboten (bspw. Studien- und Ausbildungsgänge) spiegelt die Makroebene für den Einsatz von Bildungsprodukten wieder. Die wesentlichen Zielsetzungen eines Bildungsangebots lassen sich u.a. durch die Aggregation von Grobzielen zu Richtzielen formulieren (vgl. Tab. 2). Es kann auch eine Verknüpfung von Curricula und Lehr-/Lernprozessen ermöglicht werden, da durch die modulare Konzeption der Inhalte in den Bildungsprodukten die Möglichkeit besteht, die entwickelten und durch Lernziele, Inhaltsangaben und Vorwissensstrukturen dokumentierten Module ohne nennenswerte Schwierigkeiten mehrfach zu verwenden. Hier findet im Gegensatz zum klassischen Content-Management neben der inhaltlichen Zusammenstellung auch eine didaktische Modellierung statt. Die Dokumentation der Module und Komponenten ermöglicht ein einfaches Umkonfigurieren bestehender Bildungsprodukte für bspw. andere kaufmännische Ausbildungsberufe, sofern diese bspw. den Rahmenlehrplänen der Kultusministerkonferenz folgen.

3.2 Bildungsproduktentwicklung auf Mikroebene

3.2.1 Modulkonzept

Das hier zu entwickelnde Produktkonzept kann als Baukasten aufgefasst werden, wobei eine thematische Einheit bzw. Lektion stellt zunächst ein Modul darstellt. Abb. 3 [Kami04, 200] zeigt, dass jedes Modul aus mehreren Komponenten besteht. Die Komponenten beinhalten wiederum verschiedene Elemente. Elemente sind bspw. Textpassagen, Grafiken oder Animationen und stellen die kleinste bzw. atomare Einheit dar. Module ermöglichen einerseits systembezogene Verbindungen [KlSt02, 62], wie bspw. die Nutzung gemeinsamer Elementbibliotheken, die über ein Rahmenmodul zur Verfügung gestellt werden. Andererseits sind Schnittstellen vorzuhalten, um inhaltsbezogene Verknüpfungsstrukturen [KlSt02, 62] für die Navigation über das

Bildungsprodukt abbilden zu können [KaHa04]. Hiermit können verschiedene Gestaltungsformen von Bildungsprodukten mit einer entsprechenden Lernlogik abgebildet werden. Je nach Präferenz lassen sich Module entfernen und hinzufügen, wobei die Kernfunktionalitäten der bestehenden Module nicht, oder nur bedingt beeinflusst werden. Durch die Bündelung von zusammenhängenden Inhalten sowie die Trennung von technischen und inhaltlichen Bildungsprodukt-

Bestandteilen in den Modulen wird einerseits erreicht, dass insbesondere aus didaktischer Sicht sinnvolle Zusammenhänge nicht weiter zerlegt werden können. Darüber hinaus kann durch die Modellierung inhaltlicher Wirkungs-

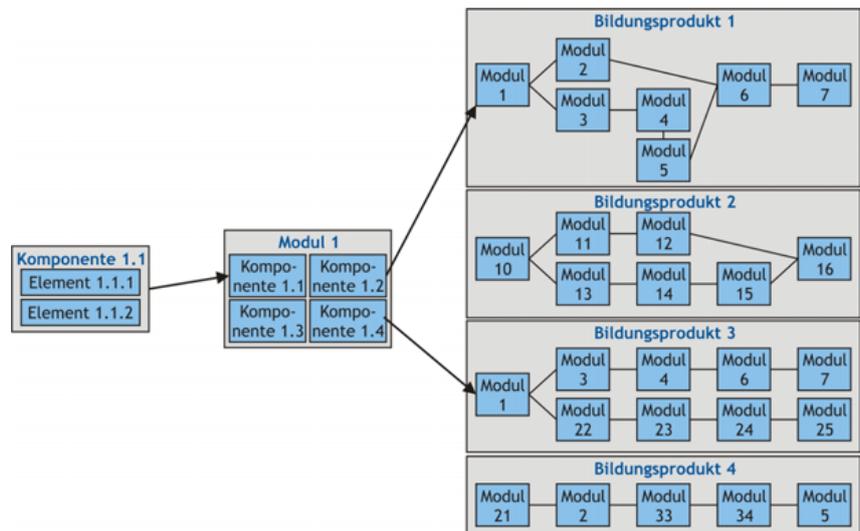


Abb. 3: Architektur eines Bildungsprodukts

beziehungen zwischen den Modulen in überschaubarer Weise eine Rekonfiguration des Produktes unter Einhaltung didaktischer Obliegenheiten vorgenommen werden. Die weitere Hierarchisierung der Bestandteile innerhalb der Module bis auf feingranularer Ebene ermöglicht, die Vorzüge des klassischen Content Managements auf die Produktarchitektur und zukünftigen Bildungsprodukten zu übertragen. Durch die auch hier vorgenommene Teilung zwischen inhaltsbezogenen bzw. systembezogenen Komponenten und Elementen kann mit geringem Aufwand die Anpassung an eine andere Designlinie vorgenommen werden [Kami04, 198ff.].

3.2.2 Bausteinkonzept

In Anlehnung an die Architekturen modularer Industrie- und Software-Produkte werden nun die zuvor definierten Module als Bausteine klassifiziert. Dies ist notwendig, da die besonderen Baukastenarchitekturen von Lernsoftware notwendige und nicht notwendige Teile vorsehen. Obligatorische Module werden als Mussbausteine bezeichnet. Diese Bausteine können im Gegensatz zu anderen Bausteinen nicht ausgeblendet werden. Mussbausteine lassen sich in inhalts- und systembezogene Bausteine einteilen und gewährleisten den Rudimentärbetrieb des Bildungsprodukts. Inhaltsbezogene Module tragen den fachlichen und didaktischen Anforderungen aus Sicht der Lehrenden und Lernenden Rechnung. So sind für die kognitivistische Ausrichtung

von Lerninhalten spezielle Kursmodule vorgesehen, die insbesondere den Lerntypen des Serialisten, der die aufeinander folgende Abarbeitung von Lerneinheiten bevorzugt, entgegenkommt [Brem02, 19]. Jedes dieser Module bildet eine abgeschlossene Lerneinheit mit Angabe des behandelten Themas, der Feinziele und des nötigen Vorwissens. Die Lernwegstruktur wird

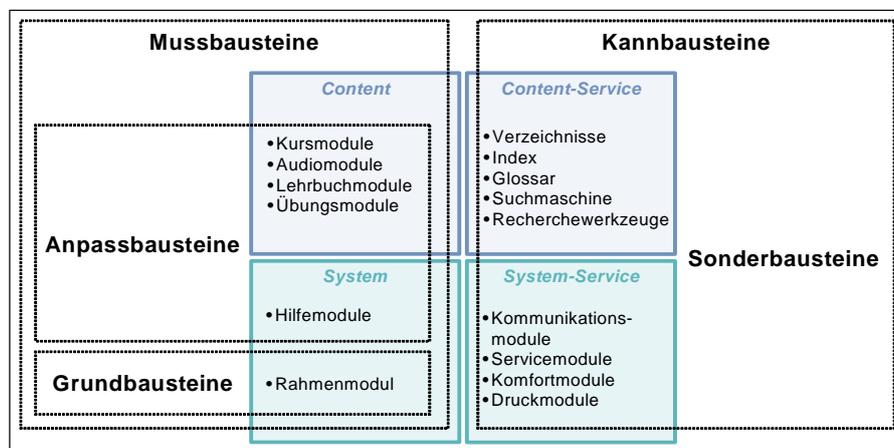


Abb. 4: Modulklassen eines Bildungsprodukts

mit Hilfe der bereits genannten Lernziel- und Vorwissensstrukturangaben, die die Funktion einer inhaltsbezogenen Schnittstelle wahrnehmen, realisiert. Diese Module lassen sich ähnlich wie das

Rahmenmodul als Mussbausteine charakterisieren (vgl. Abb. 4). Jedoch stellen sie trotz ihres Charakters keine obligatorischen Bausteine dar. Zwar müssen Inhaltsmodule im Bildungsprodukt vorhanden sein, jedoch ist im Gegensatz zu den obligatorischen Modulen nicht festgelegt, welche konkreten es sein sollen. Im weiteren Verlauf werden diese Mussbausteine als Anpassbausteine klassifiziert [Kami04, 211ff].

Auch das Hinzuschalten weiterführender Lernangebote zu den Kursmodulen kann mit Hilfe der inhaltlichen Schnittstellenbeschreibung erfolgen. Ein hierdurch entstandener Kurs lässt sich als gesamte Einheit oder modulweise ansteuern. Insbesondere konstruktivistisch geprägte Lernumgebungen können durch die Zuschaltung von Simulations- und Recherchemodulen bereichert werden, was dem sensorischen Lerntyp zu Gute kommt [Brem02, 19]. Auch in diesem Fall kann modulübergreifend oder -spezifisch verknüpft werden. Neben der Stoffpräsentation stehen weitere Bausteine zur Verfügung, die mittels Übersichten oder Verzeichnissen eine komprimierte Darstellung über die zu vermittelnden Lerninhalte geben. Dynamische Ausgaben werden durch Such- und Recherchebausteinen durch den Lernenden generiert. Durch die Verknüpfungsfähigkeit aller Module untereinander können verschiedene Lernstrategien miteinander kombiniert werden [Brem02, 19]. Im Zuge dieser Produktarchitektur werden diese Kann- bzw. Sonderbausteine als inhaltliche Servicemodule bezeichnet.

Voraussetzung für einen optimalen Lernprozess ist, dass der Lernende die Bedienung des Bildungsprodukts beherrscht. Deswegen sind Hilfsangebote unentbehrlicher Bestandteil bei der

Unterstützung des Lernenden. [Balz+04, 131]. Im hier vorgestellten Bausteinkonzept ist ein entsprechendes Hilfemodul vorgesehen, welches aufgrund seiner vorgenannten Eigenschaften insbesondere aus didaktischer Sicht zu den inhaltlichen Mussbausteinen zählt [KaHa04].

3.3 Bildungsangebotentwicklung auf Makroebene

3.3.1 Analoges Architekturkonzept

Viele Gestaltungsparadigmen zur Konfiguration von Bildungsprodukten können auch bei der Konzeption von Bildungsangeboten verwendet werden. Abb. 5 zeigt, dass Bildungsangebote aus einem Pool (die Schnittmenge der Mikro- und Makroebene in der Abbildung) verschiedener Bildungsprodukte und Dienstleistungen zusammengestellt werden. In diesem vereinfachten Beispiel nehmen die Bildungsprodukte eine ähnlich gelagerte Rolle ein, wie die inhaltsbezogenen Bausteine bei der Konfiguration innerhalb eines einzelnen Bildungsprodukts. Die

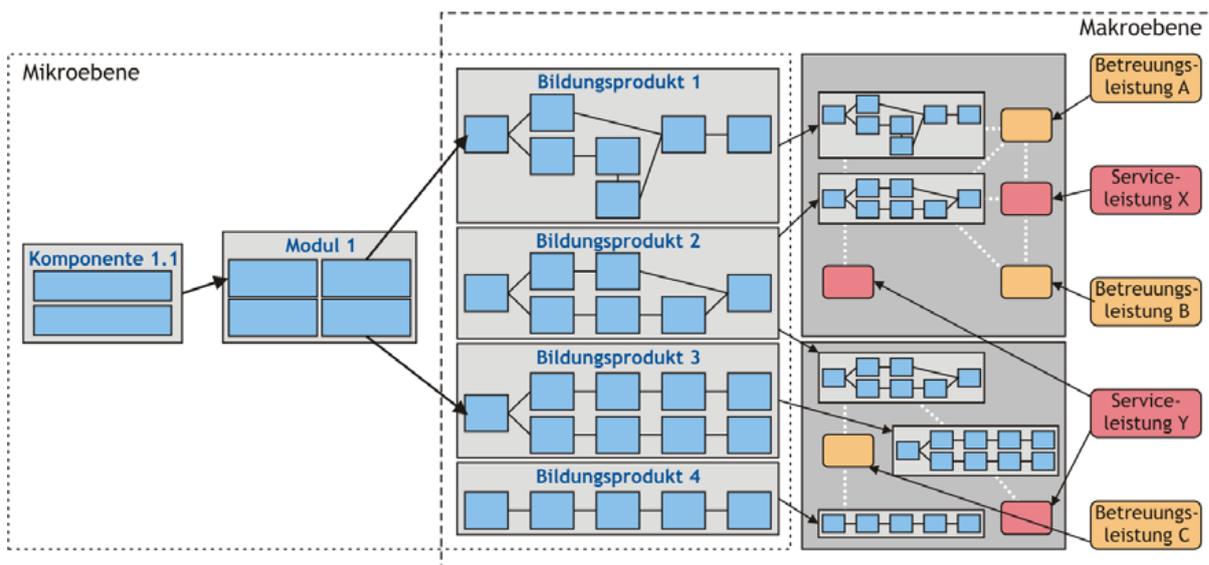


Abb. 5: Konfiguration von Bildungsangeboten

Betreuungs- und Serviceleistungen der Makroebene können nach den gleichen Prinzipien definiert und in das Angebot eingebunden werden, wie die inhalts- und systembezogenen Service-Module eines Bildungsprodukts. Die Bereitstellung von Mussbausteinen zur Sicherung des Rudimentärbetriebs des Bildungsangebots ist hierbei nicht zwingend erforderlich. Jedoch kann je nach Einsatzzweck oder Menge der zu integrierenden Bildungsprodukte bzw. Services eine Differenzierung nach obligatorischen und freiwillig zu absolvierenden Bestandteilen vorgenommen werden. Dienstleistungen und Bildungsprodukte können sich ebenfalls aufeinander beziehen bzw. einander ausschließen. Dies wird auf der Mikroebene durch entsprechende Schnittstellen realisiert, auf der Makroebene sind im organisatorisch/institutionellen Bereich

ähnliche Strukturen zu schaffen. Sofern es sich um komplexere Bildungsangebote handelt, kann eine weitere Zwischenhierarchie in der Bildungsangebotsarchitektur eingeschoben werden. Beispielsweise könnten die Bildungsangebote I und II als Teilangebote eines übergeordneten Bildungsangebots fungieren. In diesem Fall nehmen die Bildungsprodukte die Aufgabe von inhaltsbezogenen Komponenten wahr [Kami04, 260ff.].

3.3.2 *Adaption der Instrumente*

Die in der Mikroebene verwendeten Beschreibungsfelder und -elemente und deren Inhalte auf die Bedürfnisse der Makroebene müssen entsprechend angepasst werden. Im Folgenden wird beispielhaft mittels Beschreibung des Ausbildungsberufs Industriekaufmann/-kauffrau dieses Vorgehen verdeutlicht. In dem zugrunde liegenden Rahmenlehrplan wird mit Hilfe von Lernzielen der Umfang des Bildungsangebots beschrieben [KMK02]. Es handelt es sich um Richtziele, die dem Lehrenden viele Realisierungsalternativen bieten. Inhaltlich zusammenhängende Themenschwerpunkte des Lehrplans werden in entsprechenden Lernfeldern zusammengefasst

Ebene	Konzeption von Bildungsangeboten und Bildungsprodukten		Klassische Curriculumentwicklung und Unterrichtsplanung		Abstraktion Lernziele
	Makro	Bildungsangebot		Rahmenlehrplan	
BA-Modul			Lernfeld		Summe Grobziele
BA-Komponente		Bildungsprodukt	Themenschwerpunkt	Unterrichtssequenz	Grobziel(e)
Mikro		BP-Modul		Unterrichtsstunde	Summe Feinziele
		BP-Komponente			Feinziel(e)
		BP-Element			

Tab. 3: Betrachtungsebenen, Planungshorizonte und Lernzielabstraktion

(Ableitung Grobziele). Sie geben charakterisierende Inhalts- und Verhaltenskomponenten an und schaffen die Grundlage für die Entwicklung von Unterrichtssequenzen. Die konkrete Ausgestaltung einer einzelnen Unterrichtsstunde erfolgt mit Hilfe von konkreten Feinzielen. Dieses Vorgehen der klassischen Unterrichtsplanung lässt sich mit wenigen Modifikationen in die inhaltliche/didaktische Beschreibung modularer Bildungsprodukte und Bildungsangebote überführen (vgl. Tab. 3). Auf der Mikroebene werden für einzelne Lernmodule Feinziele beschrieben. Die Aggregation zu Grobzielen bzw. Lernfeldern beschreibt die wesentlichen Zielsetzungen eines Bildungsprodukts und stellt somit die höchste Abstraktionsstufe in der Mikroperspektive dar. Ein Bildungsprodukt kann als kompletter Kurs oder Kursbestandteil fungieren. Diese Einordnung repräsentiert die niedrigste Abstraktionsstufe der Makroperspektive. Die Aggregation sämtlicher Grobziele bzw. Kurse zu einem Ausbildungsgang bzw. Bildungsangebot bildet die Grundlage zum Formulieren von Richtzielen auf höchster Ebene.

3.4 Integrierte Betrachtung

Für ausgewählte Beschreibungsfelder (inhaltliche und didaktische) ist die Perspektivenerweiterung von der Bildungsproduktebene auf die Bildungsangebotsebene sehr sinnvoll. Es können die Wechselwirkungen zwischen der Planung von einzelnen Lehrveranstaltungen und komplexen Bildungsmaßnahmen erkannt und entsprechend berücksichtigt werden. Auch ist es möglich, dass im Rahmen des laufenden Lehrbetrieb bei der konkreten Bestückung von Aus- und Weiterbildungsangebotsmodulen mit Lehrveranstaltungen bzw. Bildungsprodukten externer Anbieter aufgrund der einheitlichen Beschreibungsweise schnell und effizient die einzufügenden curricularen Bausteine bewertet und eingeordnet werden können. Hierzu ist neben der einheitlichen Auszeichnungspraxis jedoch auch eine kompatible Bildungsangebotsstruktur erforderlich. Die Praxis hat jedoch gezeigt, dass eine kompatible Angebotsstruktur (durch Rahmenlehrpläne o.ä.) häufig schwierig in die Tat umzusetzen ist. Wird dieses Problem im Rahmen von Weiterbildungsvorhaben jedoch gelöst, können die entsprechenden Anrechnungs- und Integrationsproblematiken, die sich insbesondere bei der durchlässigen und offenen Bildungsangeboten ergeben entschärft werden, so dass die hier eintretenden Entlastungs-, Einspar- und Transparenzeffekte den oben genannten Mehraufwand rechtfertigen.

4 Ökonomische Bildung online als Realisierungsbeispiel

4.1 Gegenstandsbeschreibung

Bisher wurde aus theoretischer Sicht ein Konzept zur modularen Gestaltung von E-Learning-Materialien auf Bildungsproduktebene und E-Learning-Maßnahmen auf Bildungsangebotsebene beschrieben, wobei Bezüge auf die berufliche Aus- und Weiterbildung genommen wurde. Die Auswahl des Vorhabens Ökonomische Bildung online (ÖBO) aus dem Bereich Aus-, Fort- und Weiterbildung von Lehrkräften an allgemein bildenden Schulen soll zeigen, dass die Ideen aus dem zuvor beschriebenen Modulkonzept zum einen über die Grenzen einer Hochschule, Projekt oder Disziplin hinaus erfolgreich zum Einsatz kommen.

Das Vorhaben Ökonomische Bildung online wurde initiiert, da allgemein anerkannt ist, dass eine Implementierung ökonomischer Bildung in das Curriculum des allgemein bildenden Schulsystems stattfinden muss. Die Umsetzung ist allerdings als ein schwieriger Prozess, in dem zahlreiche Faktoren eine Rolle spielen. Unter anderem ist unverzichtbar, dass eine ausreichende

Qualifizierung von Lehrkräften stattfindet. Vor diesem Hintergrund hat sich eine heterogene Trägerschaft zusammengefunden, um mittels des ÖBO gemeinschaftlich ein bundesweit einsetzbares Qualifizierungsinstrument für das neue Feld ökonomische Bildung und der noch fehlenden Community auf der Basis neuer Technologien zu entwickeln, das im Hinblick auf unterschiedliche Zielgruppen wie auch auf die situationsspezifischen Bedürfnisse der teilnehmenden Bundesländer (in Abb. 6 dunkel gekennzeichnet) hin konzipiert ist. Es ist oberstes Ziel, eine seriöse und nachhaltige fachwissenschaftliche und fachdidaktische Verbesserung der Lehrerqualifikation für ein Fach Wirtschaft oder affine Fächer mit Anteilen ökonomischer Bildung zu etablieren. Der Bezug zur individuellen Lehrplansituation des jeweiligen Bundeslandes und die Unterstützung der Schulpraxis werden berücksichtigt.

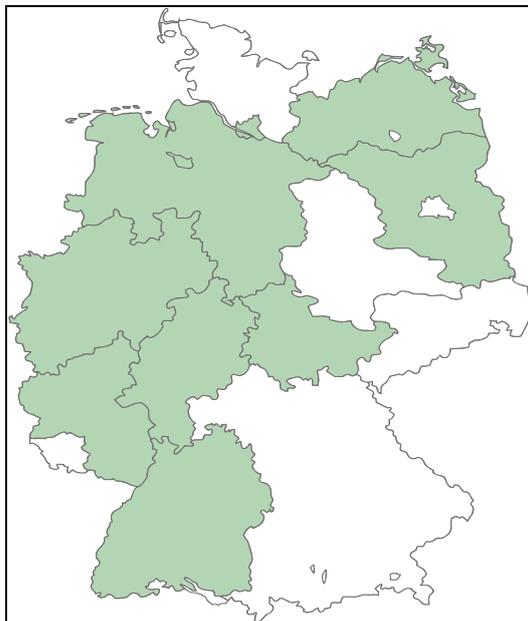


Abb. 6: Teilnehmende Bundesländer

ÖBO umfasst neben 75 elektronischen Qualifizierungs-Bausteinen zur ökonomischen Bildung (erstellt von namhaften Wissenschaftlern der deutschsprachigen Hochschullandschaft) die komplette technische Infrastruktur (Kurs-, Benutzer-, Contentverwaltung sowie Tutorenunterstützung). Diese werden zum einen im Rahmen der Ausbildung für Lehrkräfte an allgemein bildenden Schulen in zertifizierten, internetbasierten und gestuften Studiengängen (Bachelor/Master Ökonomische Bildung) eingesetzt. Zum anderen kommen sie auch in außeruniversitären zertifizierten, modularisierten Qualifizierungsangeboten für die Dauer von drei Monaten bis zwei Jahren für Fort- und Weiterbildung von Lehrern zum Einsatz, um den Bedarf an Ökonomielehrkräften sowie den Nachqualifizierungsbedarf an allgemein bildenden Schulen zu decken. Die Betreuung der Lernenden wird in allen Bildungsmaßnahmen durch fachlich, methodisch und technisch ausgebildete Tutoren durchgeführt. Neben dem Online-Lernen finden regelmäßig Präsenzphasen statt, in denen landesspezifische Themen sowie fachdidaktische und methodische Inhalte im Vordergrund stehen. Ergänzt wird dieses Angebot durch eine internetbasierte Datenbank mit Unterrichtsmaterialien, fachdidaktischen Hilfen, Praktikumsbörsen, Praxiskonkernpartnern, Medientipps usw., die auf die Aus-, Fort- und Weiterbildungsmodulen und mit den Qualifizierungsinhalten der Lehrpläne in den Bundesländern abgestimmt sind.

4.2 Inhaltlich-didaktische und technische Konzeption

In ÖBO sind Forschungsergebnisse sowohl aus mediendidaktischer wie auch aus technischer Sicht umgesetzt worden. Als Ausgangslage ist die aktuelle Problemstellung einer Neudefinition der Elemente Lernobjekt und Informationsobjekt innerhalb eines Lern- und Erkenntnisprozesses zu verzeichnen. Die Forderung nach Wiederverwendbarkeit, Modularisierung sowie Kontextfreiheit wirft Widersprüche auf, die als Reusability Object and Instruction Paradox (ROI Paradox) [BaKa05] bezeichnet werden können. Beispielsweise führt in diesem Zusammenhang die Forderung nach möglichst umfassender Wiederverwendbarkeit von Lernobjekten aus didaktischer Sicht in ein Dilemma der Kontextfreiheit eines Lerninhaltes und den ursprünglich intendierten didaktischen Zielstellungen. Aus didaktischer Sicht ist ja gerade eine effektive Lernziel-erreichung nur möglich in optimaler Anpasstheit des Inhalts an den Kontext.

Dieses Paradoxon soll auf Mikroebene durch die Definition des Lernobjekts in Anlehnung an Baumgartner [BaKa05, 97ff] entschärft werden, wobei auf das Autodesk Content Model [EMC02, 25] zurückgegriffen wird. Hier wird auf Ebene zwischen Lernobjekten und Informationsobjekten (Medienobjekten) unterschieden. Nach dem verfeinerten Modell von Baumgartner

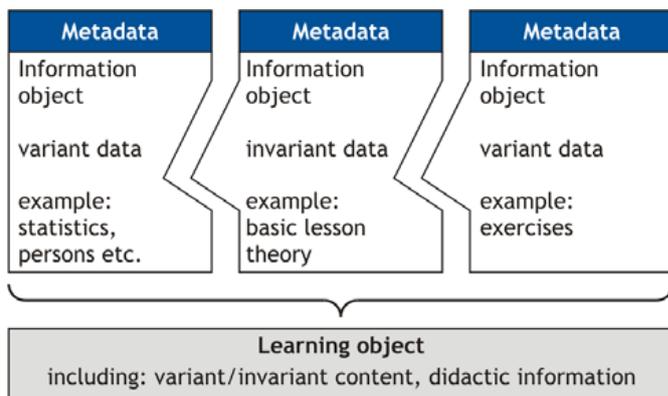


Abb. 7: Lernobjekt aus Informationsobjekten

werden Lerninhalte auf Elemente bzw. Informationsobjekte umgesetzt, die erst in der Laufzeitumgebung zu einer Komponente bzw. einem Lernobjekt zusammengestellt werden, dem wiederum ein dedizierter didaktischer Kontext zugrunde liegt. Ein Informationsobjekt stellt hierbei (vgl.

Abb.7) die Zusammenfassung von (mehreren) Medienobjekten dar, die exakt ein Lernziel unterstützen. Hier kommt zum Ausdruck, dass die Forderung der Wiederverwendbarkeit (der Elemente und Komponenten) auf verschiedenen Ebenen möglich ist und auch in unterschiedlichen Kontexten realisiert werden kann. Neben dem Wissensinput mittels des Objekts (ÖBO-Bausteins) zum Fachwissen ökonomischer Bildung werden parallel dazu Informationsobjekte wie z.B. einsatzfähige Unterrichtsmaterialien, Unterrichtsbeispiele, aktuelle Statistiken für den Unterricht, aufbereitete Handelsblattartikel (Wirtschaft aktuell im Unterricht), Begriffsklärungen (1x1 der Wirtschaft), interaktive Übungen, Lehrfilme (visueller Umsetzung erklärungs-aufwendiger Begrifflichkeiten) oder Spezialbeispiele (Fußball WM, Energiedebatte etc.) einge-

setzt. Sie werden je nach didaktischem Kontext über ein Metadatensystem gesteuert in der Laufzeitumgebung zu einer Komponente zusammen geschaltet werden und in einer so genannten „Marginalien-

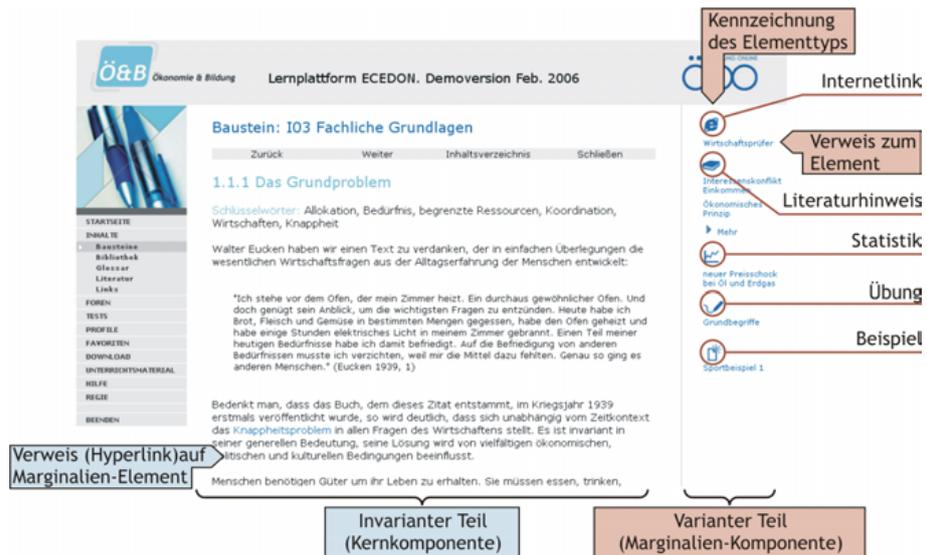


Abb. 8: Screenshot ÖBO-Baustein I03 Fachliche Grundlagen

spalte“ angezeigt. Auf einen Klick werden alle Elemente im jeweils relevanten Kontext des fachlichen Bausteins eingebettet (vgl. Abb. 8).

Auf Makroebene werden dann verschiedene ÖBO-Bausteine zusätzlich mit Betreuungs-, Kommunikations- und weiteren inhaltlichen Unterstützungangeboten versehen zu entsprechenden Aus-, Fort- und Weiterbildungsmaßnahmen formiert, die dann nicht nur über das ÖBO-Portal (www.oebo.de), sondern auch über andere Bildungsportale betrieben werden können.

Ziel von ÖBO ist es, die einzelnen Phasen der Aus-, Fort- und Weiterbildung von Ökonomelehrern zu verzahnen. Wie bereits angesprochen, ist sowohl derzeit keine Community von Ökonomelehrern an allgemein bildenden Schulen als auch keine flächendeckende und integrierte Ausbildung (Phase 1 und 2) sowie Fort- und Weiterbildungsinfrastruktur (Phase 3) zur ökonomischen Bildung vorhanden. Dies ist eine gute Chance, bisherige Konzeptionsfehler bei der Ausgestaltung eines Qualifizierungssystems für Lehrkräfte zum umgehen, indem eine integrierte Bildungsplattform geschaffen wird, die alle drei Phasen abdecken kann. Abb. 9 zeigt mittels der dunklen großen Fläche einen möglichen (auch fachübergreifenden) Einsatzraum für ÖBO. Da ÖBO sowohl auf Mikro- als auch auf Makroebene modular aufgebaut ist und zwischen inhalts- und systembezogenen Modulen unterscheidet, kann es an die individuellen Anforderungen einer Qualifizierungsphase (mittels Vorwissens- und Lernzielstrukturen) oder einer Bildungseinrichtung angepasst werden, da es speziell auf die Anforderungen von Lehrkräften zugeschnitten und auf deren Prozessunterstützung hin optimiert ist.

Mit Hilfe der individuellen lebenslangen Lernbiographie eines Lehrers lassen sich diese Potenziale leicht ableiten. Im Zuge der Erstausbildung an der Universität werden die zukünftigen Lehrkräfte erstmals mit den Inhalten und Funktionen von ÖBO konfrontiert. Gleichzeitig erhalten sie Zugang zum System und systemseitig wird die Stammdatenbasis des Studierenden angelegt. Nach Abschluss des 1. Staatsexamens (bzw. Masterabschluss) treten die Lehrer in die zweite Ausbildungsphase ein, die an die Studienseminaren und den Schulen abläuft. Hier kann der angehende Lehrer sich selbstgesteuert in ÖBO weiterbilden oder für seine Unterrichtsgestaltung Inhalte und Materialien besorgen. Gleichzeitig wird er über interessante Qualifizierungsmöglichkeiten für die Vorbereitung auf das 2. Staatsexamen informiert, wobei die Informationen auf Basis seiner Stammdaten generiert werden. Auch kann er sich mit anderen (gleichge-

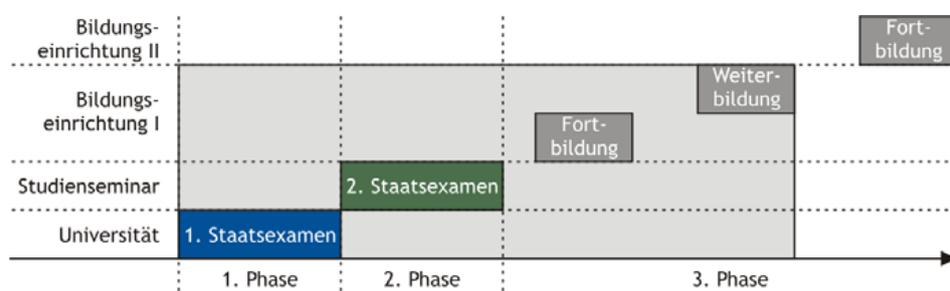


Abb. 9: Einsatzfeld ÖBO

sinnten) Lehrkräften austauschen (internetbasierte Reflexion über Unterricht, Projekte

etc.). Dieser Austausch ist in der dritten Phase der Lehrerqualifizierung (Förderung der wechselseitigen Analyse von Unterrichtsentwicklung) besonders wichtig. Lehrer sind angehalten, ihre Wissensbasis stets auf den aktuellsten Stand zu halten. Auch müssen sie zur Erteilung von Unterrichtsfächern Befähigungen vorweisen, die in ÖBO erworben und zertifiziert werden können. Jedoch ist diese Phase dadurch gekennzeichnet, dass viele Lehrkräfte aufgrund der Verweilzeiten zwischen einzelnen Qualifizierungsgängen den Bezug zur Community verlieren können. ÖBO kann aufgrund der gesammelten Profildaten des Lehrers, seiner offene Schnittstellen zu anderen Systemen, seines Bausteinangebots sowie weiterer Vorgaben für Lehrkräfte individuell zugeschnittene Bildungsangebote schnüren und als Blended Learning Szenario durchführen lassen. Die Lehrkräfte sind mit ÖBO in der Community verwurzelt, die Infrastruktur ist bekannt und die individuellen Qualifizierungswünsche können befriedigt werden. Dies ist ein seit fünf Jahren erfolgreich beschrittener Weg für die Qualifizierung von Lehrern.

4.3 Verzahnung von Lehreraus- und -fortbildung

Ziel von ÖBO ist es, die einzelnen Phasen der Aus-, Fort- und Weiterbildung von Ökonomielehrern zu verzahnen. Wie bereits angesprochen, ist keine flächendeckende und integrierte Ausbildung (Phase 1 und 2) sowie Fort- und Weiterbildungsinfrastruktur (Phase 3) zur ökonomi-

schen Bildung vorhanden. Dies ist eine gute Chance, bisherige Konzeptionsfehler bei der Ausgestaltung eines Qualifizierungssystems für Lehrkräfte zum umgehen, indem eine integrierte Bildungsplattform geschaffen wird, die alle drei Phasen abdecken kann. Abb. 9 zeigt mittels der dunklen großen Fläche einen möglichen (auch fachübergreifenden) Einsatzraum für ÖBO. Da ÖBO sowohl auf Mikro- als auch auf Makroebene modular aufgebaut ist und zwischen inhalts- und systembezogenen Modulen unterscheidet, kann es an die individuellen Anforderungen einer Qualifizierungsphase (mittels Vorwissens- und Lernzielstrukturen) oder einer Bildungseinrichtung angepasst werden, da es speziell auf die Anforderungen von Lehrkräften zugeschnitten und auf deren Prozessunterstützung hin optimiert ist.

Mit Hilfe der individuellen lebenslangen Lernbiographie eines Lehrers lassen sich diese Potenziale leicht ableiten. Im Zuge der Erstausbildung an der Universität werden die zukünftigen Lehrkräfte erstmals mit den Inhalten und Funktionen von ÖBO konfrontiert. Gleichzeitig erhalten sie Zugang zum System und systemseitig wird die Stammdatenbasis des Studierenden angelegt. Nach Abschluss des 1. Staatsexamens (bzw. Masterabschluss) treten die Lehrer in die zweite Ausbildungsphase ein, die an die Studienseminaren und den Schulen abläuft (vgl. Abb. 9). Hier kann der angehende Lehrer sich selbstgesteuert in ÖBO weiterbilden oder für seine Unterrichtsgestaltung Inhalte und Materialien besorgen. Gleichzeitig wird er über interessante Qualifizierungsmöglichkeiten für die Vorbereitung auf das 2. Staatsexamen informiert, wobei die Informationen auf Basis seiner Stammdaten generiert werden. Auch kann er sich mit anderen Lehrkräften austauschen (internetbasierte Reflexion über Unterricht, Projekte etc.). Dieser Austausch ist in der dritten Phase der Lehrerqualifizierung (Förderung der wechselseitigen Analyse von Unterrichtsentwicklung) besonders wichtig. Lehrer sind angehalten, ihre Wissensbasis stets auf den aktuellsten Stand zu halten. Auch müssen sie zur Erteilung von Unterrichtsfächern Befähigungen vorweisen, die in ÖBO erworben und zertifiziert werden können. Jedoch ist diese Phase dadurch gekennzeichnet, dass viele Lehrkräfte aufgrund der Verweilzeiten zwischen einzelnen Qualifizierungsgängen den Bezug zur Community verlieren können. ÖBO kann aufgrund der gesammelten Profildaten des Lehrers, der Schnittstellen zu anderen Systemen sowie seines Bausteinangebots für Lehrkräfte individuell zugeschnittene Bildungsangebote schnüren und als Blended Learning Szenario durchführen lassen. Dies ist ein seit fünf Jahren erfolgreich beschrittener Weg für die Qualifizierung von Lehrern.

5 Fazit

Der Artikel zeigt auf, dass bei konsequenter Anwendung von definierten Modularisierungskonzepten bei der Produktentwicklung von E-Learning-Materialien sowie internetgestützten Bildungsangeboten nachhaltig tragfähige Qualifizierungsangebote entstehen. Wurde zunächst auf Produktebene das Konzept von Modul und Baustein im Sinne einer Komponentenentwicklung verdeutlicht, wurde in einem zweiten Schritt der Bogen zur Übertragung dieses Modells auf Bildungsprodukte geschlagen und anhand der konkreten Realisierung in der Aus-, Fort-, und Weiterbildung für Ökonomelehrer dargestellt. Im Sinne der Forderung einer Umsetzung des lebenslangen Lernens beinhaltet dieses Modularisierungsmodell richtungweisendes Potenzial.

Literaturverzeichnis

- [AßNe03] Aßmann, U., Neumann, R.: Quo vadis Komponentensysteme – Von Modulen zu grauen Komponenten. In: Heilmann, H., Strahinger, S. (Hrsg.): Neue Konzepte in der Software-Entwicklung, HMD – Praxis der Wirtschaftsinformatik, Heft Nr. 231, Heidelberg 2003, 19-28
- [BaBr96] Ballin, D., Brater, M.: Handlungsorientiert Lernen mit Multimedia - Lernarrangements planen, entwickeln und einsetzen, Nürnberg 1996
- [Balz+04] Balzert, H., Balzert, H., Zwintzsch, O.: Die E-Learning-Plattform W3L, Anforderungen - Didaktik - Ergonomie - Architektur - Entwicklung - Einsatz. In: Wirtschaftsinformatik, Nr. 2/2004, Wiesbaden 2004, 129-138
- [BaKa05] Baumgartner, P., Kalz, M.: Wiederverwendung von Lernobjekten aus didaktischer Sicht. In: Auf zu neuen Ufern! E-Learning heute und morgen. Hrsg.: D. Tavangarian und K. Nölting. New York - München - Berlin, Waxmann 2005. Medien in der Wissenschaft: Bd. 34: 97-106
- [Brem02] Bremer, C.: Online lehren leicht gemacht, Leitfaden für die Planung und Gestaltung von virtuellen Hochschulveranstaltungen. In: Berendt, B., Voss, H., Wildt, J. (Hrsg.): Neues Handbuch Hochschullehre, Lehren und Lernen effizient gestalten, Berlin 2002, Beitrag D 3.1
- [Dubs03] Dubs, R.: IT-Weiterbildung: Schlüsselfaktor Qualifikation. In: Mattausch, W., Caumanns, J. (Hrsg.): Innovationen der IT-Weiterbildung, Bielefeld 2003, 30-43

- [Ehr195] Ehrlenspiel, K.: Integrierte Produktentwicklung, Methoden für Prozessorganisation, Produkterstellung und Konstruktion, München 1995
- [KaHa03] Kamin, O., Hagenhoff, S.: Considerations for the Conceptual Design of Modular Web Based Teaching and Learning Environments - Avoiding Existing Deficits of E-Learning. In: Abramovicz, W., Klein, G. (Hrsg.): Business Information Systems, Proceedings of BIS 2003, Colorado Springs (USA) 2003, 293-299
- [KaHa04] Kamin, O., Hagenhoff, S.: Modular Web Based Teaching and Learning Environments as a Way to improve E-Learning. In: Howard, C., Schenk, K., Discenza, R. (Hrsg.): Distance Learning and University Effectiveness: Changing Educational Paradigms for Online Learning, Hershey (USA), 2004, 190-211
- [Kami04] Kamin, O.: Mehrfachverwendbare elektronische Lehr-/Lernarrangements, Köln/Lohmar 2004
- [Kilb+93] Kilberth, K., Grycznan, G., Züllinghoven, H.: Objektorientierte Anwendungsentwicklung – Konzepte, Strategien, Erfahrungen, Braunschweig 1993
- [Kloa98] Kloas, P.: Modularisierung unter Beibehaltung des Berufskonzepts. In: Davids, S. (Hrsg.): Modul für Modul zum Berufsabschluss, Berufsbegleitende Nachqualifizierung zwischen Flexibilität und Qualitätssicherung, Bielefeld 1998
- [KlSt02] Klein, M., Stucky, W.: Erstellung von hypermedialen Kursen – Ein Vorgehensmodell. In: Eicker, S. (Hrsg.): E-Learning: Modelle, Instrumente und Erfahrungen, Multi-Konferenz Wirtschaftsinformatik 2002, Nürnberg 2002, 53-70
- [KMK02] Kultusministerkonferenz (Hrsg.): Rahmenlehrplan für den Ausbildungsberuf Industriekaufmann/Industriekauffrau, <http://www.kmk.org/beruf/rlpl/rlpindustriekfm.pdf>, 2002, Abruf am 2006-11-13
- [Kohl97] Kohlhase, N.: Strukturieren und Beurteilen von Baukastensystemen, Strategien, Methoden, Instrumente. In: Verein Deutscher Ingenieure (Hrsg.): Fortschrittsberichte VDI, Reihe 1, Konstruktionstechnik, Maschinenelemente, Nr. 275, Düsseldorf 1997
- [Müll00] Müller, M.: Modularisierung von Produkten, Entwicklungszeiten und -kosten reduzieren, München, 2000
- [Sloa00] Sloane, P.: Das Lernfeldkonzept in der Ausbildung von Bankkaufleuten, Didaktische Grundanliegen und Implementierungsprobleme. In: Riekeberg, M., Stenke, K., Banking 2000, Perspektiven und Projekte, Wiesbaden 2000, 328-346

Einführung in den Track

IS-Architekturen

Prof. Dr. Frank Leymann

Universität Stuttgart

Prof. Dr. Elmar J. Sinz

Universität Bamberg

Prof. Dr. Johannes Siedersleben

T-Systems Enterprise Services GmbH

Moderne betriebliche Informationssysteme (IS) zeichnen sich durch eine feine Granularität ihrer Komponenten und Dienste, einen hohen Verteilungsgrad und durch strukturelle Dynamik aus. Um die damit einhergehende Komplexität bewältigen zu können, wird die Architektur immer mehr zum zentralen Bezugspunkt für die Analyse, die Gestaltung und das Management betrieblicher Informationssysteme. Ziel des Tracks ist es, aktuelle Entwicklungen auf allen Ebenen von IS-Architekturen (Prozess-, Anwendungs- und Infrastrukturebene) sowie einschlägige Praxiserfahrungen vorzustellen und zu diskutieren. Es werden Beiträge aus Wissenschaft und Praxis zu relevanten Themen auf dem Gebiet der IS-Architekturen erbeten.

Programmkomitee:

Dr. Uwe Dumsclaff, sd&m AG

Prof. Dr. Stefan Eicker, Universität Duisburg-Essen

Prof. Dr. Wilhelm Hasselbring, Universität Oldenburg

Prof. Dr. Susanne Leist, Universität Regensburg

Prof. Dr. Heinrich C. Mayr, Universität Klagenfurt

Prof. Dr. Martin Schader, Universität Mannheim

Prof. Dr. Klaus Turowski, Universität Augsburg

Prof. Dr. Robert Winter, Universität St. Gallen

Eine Strukturvorlage zur effektiven Dokumentation von Software- und IT Architekturen

Dr. Gernot Starke

D-50858 Köln
gs@gernotstarke.de

Dr. Peter Hruschka

Atlantic Systems Guild
D-52080 Aachen
hruschka@b-agile.de

Abstract

IT-Entwicklungsprojekte verwenden heute immer noch ungebührlich viel Zeit zur Entwicklung projektspezifischer Strukturen für die Dokumentation von Software- und IT-Architekturen. Durch die Verwendung von *Strukturvorlagen* lässt sich einerseits dieser Aufwand erheblich reduzieren, andererseits die Qualität von Architekturdokumentation deutlich steigern. Der Beitrag stellt die arc42-Schablone zur Architekturdokumentation vor, die sich in vielen kommerziellen, industriellen und Open-Source Projekten bewährt hat.

1 IT-Praxis benötigt Strukturvorlagen

In vielen Projekten, in denen wir als Berater involviert werden, verwenden Mitarbeiter ungebührlich viel Zeit für die Entwicklung von Strukturen und Prozessen für die Dokumentation der jeweiligen Software-Architektur. Dabei besitzen gerade die Strukturen von Architekturdokumentation erhebliches Wiederverwendungspotenzial über die Grenzen einzelner Projekte hinweg: Theorie und Praxis sind sich über viele notwendige Merkmale und Bestandteile von Architekturdokumentation einig (siehe dazu insbesondere [Clements+02] sowie [Starke02]). Unter anderem gehören dazu:

- Verwendung verschiedener Sichten (etwa: Kontextsicht, statische & dynamische Sicht, Verteilungs-/Deploymentsicht), um die unterschiedlichen Strukturebenen von IT-Systemen adäquat repräsentieren zu können.
- Dokumentation wesentlicher Entwurfsentscheidungen, beispielsweise zu Ablaufsteuerung, Persistenz, Benutzungsoberfläche, Fehler-/Ausnahmebehandlung, Protokollierung, Tracing, Monitoring, Systemverteilung/Middleware und anderen.
- Dokumentation wesentlicher Architektur- und Entwurfsziele, um Architekturdokumentation autark von anderen Dokumenten verständlich zu halten (Kruchten hat in [Kru95] hierfür eigens die Use-Case-View propagiert).
- Weitmögliche Verwendung standardisierter Modellierungssprachen: Zur strukturellen Beschreibung von Softwaresystemen hat sich beispielsweise die vielen Unternehmen die UML (Unified Modeling Language) etabliert.

Weiterhin erhöhen bekannte Strukturen *per se* die Verständlichkeit und Wiederverwendbarkeit der darin dokumentierten Inhalte – was sich für Unternehmen als zusätzlicher Effizienzgewinn darstellt.

Aus diesem Grund sollten IT-Entwicklungsprojekte in hohem Maße vorhandene (und erprobte!) Strukturen von Architekturdokumenten (hier genannt Strukturvorlagen) wiederverwenden, anstatt solche Strukturen projektspezifisch neu zu entwerfen.

2 Eine Strukturvorlage für Software- und IT-Architekturen

Ein nach unserer Erfahrung geeignetes Vorbild einer solchen Strukturvorlage für Software- und IT-Architekturen ist das frei verfügbare arc42-Template. Es ist seit 2004 frei verfügbar ([arc42]) und steht unter der sehr flexiblen Creative-Commons Lizenz, die dieses Template für beliebigen Gebrauch freigibt. Seine Struktur resultiert aus der Erfahrung vieler Entwicklungsprojekte in unterschiedlichen Branchen. Das arc42-Template ist wahlweise in MS-Word Format, pdf oder als Mindmap erhältlich. Abbildung 1 zeigt die oberste Gliederungsebene als Mindmap, die für Textdokumente (je nach Umfang) durchaus der Kapitel- oder Dateistruktur entsprechen kann. Eine detaillierte Gliederung finden Sie im Anhang.

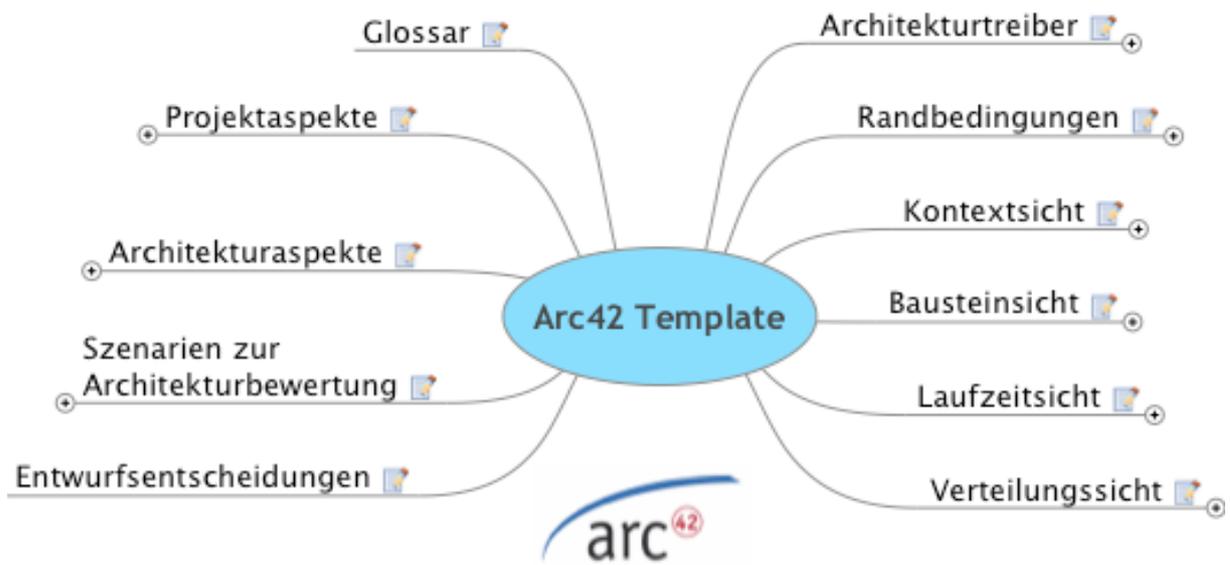


Abb. 1: Oberste Gliederungsebene des arc42 Strukturtemplates (Mindmap-Version)

Architekturdokumentation gemäß dieser Struktur soll als *Informationsrepository* interpretiert werden, nicht als einzelnes Dokument. Aus einem solchen Repository kann stakeholder- oder lesergerechte Information extrahiert werden, ohne die bei rein dokumentenbasierten Ansätzen übliche Redundanz. Als ideale Ergänzung hierfür erweisen sich mehrbenutzerfähige Modellierungswerkzeuge mit zentraler Datenhaltung (wie etwa Rational Rose, Rational Software Architect, MagicDraw oder Enterprise-Architect).

3 Der Kern: Architektursichten

Schon in den 70er-Jahren hat David Parnas gefordert, dass man mindestens zwei Strukturen für Software-Architekturen dokumentieren sollte: einerseits die vorhandenen Moduln und deren Zusammensetzung und andererseits eine dynamische Ablaufsicht, d.h. das Verhalten des Softwaresystems zur Laufzeit. Im Lauf der Jahre sind viele Vorschläge für die Anzahl der notwendigen Sichten und deren Wahl entstanden (Mehr dazu siehe unten im Abschnitt: Was sagen die Anderen).

Unsere Erfahrung zeigt, dass drei wesentliche Sichten in der Praxis genügen - unter vereinfachten Randbedingungen sogar noch weniger.

3.1 Bausteinsicht

Unvermeidbar ist die Bausteinsicht. Sie zeigt die (statischen) Bausteine des Systems und deren Zusammenhänge. Wir haben für arc42 das neutrale Wort „Baustein“ als Oberbegriff für die Vielzahl von Namen gewählt, die im Projektalltag dafür verwendet werden, wie z.B. Module, Komponente, Klasse, Subsystem, Interface, Paket, Bibliothek, Framework, Schicht, Partition, Tier, Funktion, Makro, Operation, Tabelle, Datenbank oder Datei. Ein Baustein ist also im Endeffekt jedes kleine oder große Stück Quellcode, unabhängig davon mit welcher Technologie es entwickelt wird. Wir wollen für jeden Baustein in der Architekturdokumentation wissen, wie er heißt, welche Verantwortung er im Rahmen des Gesamtsystems übernimmt, welche Leistung er erbringt, über welche Schnittstellen er verfügt, u.v.a. mehr. Das arc42-Template stellt dafür sowohl Muster für seine Black-Box-Charakterisierung zur Verfügung, als auch ein Muster für die Beschreibung des Innenlebens in Form einer White-Box-Beschreibung. Die abstrakteste Form eines Bausteins ist das Gesamtsystem. Im logischen Kontextdiagramm wird es in Bezug zur Systemumgebung gesetzt.

3.2 Verteilungssicht

Mehr und mehr Systeme sind heute auf mehrere Standorte oder auf mehrere Rechner verteilt. Diese Dimension decken wir in der Verteilungssicht ab. Alles, worauf Software laufen kann, wird als Knoten in den Verteilungsdiagrammen dokumentiert. Neben diesen Knoten brauchen wir auch die "Kanäle", über die Informationen zwischen den Knoten ausgetauscht werden kann. Den Knoten werden dann die Bausteine, die darauf gespeichert werden oder laufen, zugeordnet. Zu den Kanälen hält man die Art und Klassifikation der Informationen fest, die darüber fließen. Sie brauchen eine derartige Sicht nicht zu erstellen, wenn Sie Software für einen einzigen Standardrechner schreiben, also kurz: wenn Verteilung für Sie keine Rolle spielt. Aber Vorsicht: selbst in einfacheren Fällen hat uns die Praxis gezeigt, dass Verteilungsdiagramme nützlich sind, insbesondere das Verteilungskontextdiagramm. In dieser Vogelperspektive dokumentieren Sie neben Ihrem System als einen Knoten wenigstens die Kanäle (oder Medien), über die Ihr System mit Nachsystemen Informationen austauscht. Insbesondere erleichtert die Verteilungssicht die Abstimmung zwischen Entwicklungs- und Betriebsorganisationen (Rechenzentren).

3.3 Laufzeitsicht

Als dritte Sicht schlagen wir vor, ausgewählte Laufzeitszenarien zu dokumentieren. Diese Laufzeitsicht veranschaulicht das gewünschte dynamische Verhalten Ihres Softwaresystems. Gehen Sie an diese Herausforderungen ähnlich wie Software-Tester heran: Wählen Sie zunächst Szenarien für den normalen Ablauf aus. Ergänzen Sie diese dann um Grenz- und Stress-Szenarien für wichtige Ausnahme- oder Sonderfälle. Laufzeitsichten sind ein Hilfsmittel, um dynamischen Verhalten zu finden oder zu bewerten. Beispielsweise erachten wir den Systemstart (Bootstrapping) oder wichtige Anwendungsfälle als geeignete Kandidaten für Laufzeitszenarien.

Für alle diese Sichten bietet beispielsweise die UML geeignete grafische Dokumentationsmittel. Eine ausführliche Darstellung, welche Ausdrucksmittel für welchen Zweck finden Sie in [ESA 05] oder unter [arc42].

4 Die Kapitel der arc42 Strukturvorlage

Lassen Sie uns einige wichtige Punkte aus der arc42-Strukturvorlage herausgreifen, ohne im Rahmen dieses Artikels vollständig auf die einzelnen Kapitel einzugehen (für weitergehende Erläuterungen und Beispiele siehe [arc42]).

- Formulieren Sie grundsätzlich explizite Architekturtreiber und Architekturziele (im Abschnitt „Architekturtreiber“)! Danach beurteilen Sie am Projektende, ob Ihre Architektur gut genug ist. Sie werden diese Ziele auch als Ausgangsbasis brauchen, wenn Sie Ihre Architekturen systematisch bewerten wollen.
- Legen Sie Randbedingungen, Einschränkungen und Annahmen schriftlich fest (im Abschnitt „Randbedingungen“), damit Sie wissen, wo Sie Freiheitsgrade für Entwurfsentscheidungen haben und wo nicht.
- Grenzen Sie Ihr System gegen die Systemumgebung systematisch ab und dokumentieren es in der „Kontextsicht“. Aus logischer Sicht hat das hoffentlich schon ein Requirements Engineer gemacht. Sie sollen die physischen Schnittstellen und die Kanäle bzw. die Medien von und zu Nachbarsystemen dokumentieren. An diesen Schnittstellen müssen Sie mit anderen Projekten oder Systemen verhandeln.

- An der „Bausteinsicht“ geht kein Weg vorbei. Sie entspricht in etwa dem Grundrissplan eines Hauses. Wenn Sie sonst gar nichts von Ihrer Architektur dokumentieren - diese Sicht muss (bis zum benötigten Niveau) auf jeden Fall dokumentiert sein.
- Entwurfsentscheidungen müssen über die Lebensdauer von Systemen nachvollziehbar bleiben. Deshalb sieht das arc42-Template dieses Kapitel vor, um alle wichtigen Entscheidungen, die Sie als Architekt im Laufe eines Projekts getroffen haben, rasch wieder auffindbar zu machen.
- Durch die szenariobasierte Architekturbewertung (beispielsweise ATAM, siehe [Clements02]) haben wir Mittel und Wege, über gezielt ausgewählte Szenarien bestimmte wünschenswerte Eigenschaften von Softwarearchitekturen, wie Flexibilität, Robustheit, Zuverlässigkeit, ... nachprüfen und bewerten zu können. Wenn Sie solche Bewertungen vornehmen, dann dokumentieren Sie die Grundlagen dafür im Abschnitt „Szenarien zur Architekturbewertung“.
- Das Kapitel der „Architektur Aspekte“ ist in der detaillierten Gliederung (siehe Anhang) sehr ausführlich. Nicht alle Aspekte, die für Software- Architekturen relevant sind, lassen sich direkt in Baustein-, Laufzeit- oder Verteilungssichten darstellen. Zu diesen „Architektur Aspekten“ zählen unter anderem Persistenz, Ablaufsteuerung, Fehler- & Ausnahmebehandlung, Monitoring, Migration und andere.
- Unter der Überschrift „Projektaspekte“ sind viele Punkte gesammelt, die die Schnittstellen zwischen Architekten und anderen Projektbeteiligten betreffen. So z. B. Risikolisten als Grundlage für die Diskussion mit Projektleitern oder Technologieexperten, Change Requests zur Diskussion mit Auftraggebern, Endanwendern und Requirements Engineers oder Auswirkungen zur rechtzeitigen Diskussion mit der Inbetriebnahme, den Organisatoren oder Migrationsspezialisten.

5 Praktische Erfahrungen

Wir haben in den letzten Jahren bei zahlreichen Kunden unterschiedlicher Branchen (etwa: Telekommunikation, Medizintechnik, Logistik, öffentliche Verwaltung) die arc42 Schablone in

Projekten verschiedener Größe und Ausrichtung erfolgreich eingesetzt. Auf Basis dieser Erfahrungen haben wir Struktur des Templates iterativ angepasst – jedoch seit 2004 keine signifikanten strukturellen Änderungen mehr vorgenommen. Die meisten Ergänzungen oder Erweiterungen betrafen den Abschnitt „Architektur Aspekte“, zu dem unsere Kunden zahlreiche Ergänzungsvorschläge unterbreitet haben. Einige konkrete Erfahrungen zur Einführung:

- Die Einarbeitung in die Struktur des Templates bedeutet, wie die Einführung jeglicher Methodik oder Technologie, eine Lernkurve für die Beteiligten. Unserer Erfahrung nach lässt sich diese Lernkurve durch geschickte Wahl der betreffenden Pilotprojekte deutlich reduzieren: Projekte mit ohnehin strukturell durchdachter und ausformulierter Dokumentation sind nach unserer Erfahrung innerhalb kürzester Zeit in der Lage, die vorhandenen Dokumente in der arc42-Struktur abzubilden.
- Projekte können das arc42-Template als eine Art „Checkliste“ verwenden, um ihre eigene Architekturdokumentation auf Vollständigkeit hin zu prüfen. Dies kann etwa in Form halb- bis ganztägiger Architekturreviews geschehen.
- Mit der Einführung einer projektübergreifenden Strukturvorlage zur Dokumentation geht oftmals die Überarbeitung oder Optimierung der betroffenen Architekturprozesse einher: Mehrmals wurde durch die thematisch breit angelegte Struktur des arc42-Templates den betroffenen IT-Architekten erst die Breite ihres eigenen Aufgabenspektrums vor Augen geführt. Das wiederum führte für betroffene Unternehmen zu einer klar umrissenen Rollendefinition¹ der Architekten.
- In regelmäßigen Abständen von 3-6 Monaten sollten Architekten aus verschiedenen Projekten eines Unternehmens in kurzen Workshops ihre Erfahrungen mit dem Einsatz der Strukturvorlage gegenseitig abgleichen. Ein einfaches Mittel ist die *Architekturpräsentation* der jeweiligen Systeme, bei der in Form eines Vortrags von maximal 60 Minuten Dauer eine Kurzfassung jedes Kapitels der arc42-Dokumentation präsentiert wird. Insbesondere müssen dabei spezifische Erweiterungen und Ergänzungen an der Strukturvorlage abgestimmt werden.

¹ Arc42 postuliert folgende sechs Kernaktivitäten von IT-Architekten: 1.: Einflussfaktoren klären (d.h. Anforderungen und Randbedingungen prüfen), 2.: Systeme strukturieren, 3.: Architektur Aspekte konzipieren, 4.: Umsetzung überwachen, 5.: Architekturen bewerten und 6.: Architekturen kommunizieren.

Insgesamt empfehlen wir Organisationen und Unternehmen, so schnell wie möglich eine einheitliche Struktur zur Dokumentation von IT-Architekturen einzuführen – der Produktivitäts- und Qualitätsgewinn in Projekten und Systemen wiegt die (in der Regel ohnehin nur kurze) Lernkurve schnell wieder auf.

6 Alternative Ansätze von Strukturvorlagen für IT-Architekturen

In der Literatur und im Internet gibt es einige weitere Ansätze von Strukturvorlagen zur Dokumentation von Software- und IT-Architekturen. Einige davon möchten wir kurz vorstellen.

Philippe Kruchten hat mit seinem Artikel über „4+1 Views“ [Kruchten95] das Konzept der Architektursichten hoffähig gemacht. Dieses Modell hat für die zentralen Ideen in arc42 Pate gestanden, ebenso für die Ansätze aus [Starke05].

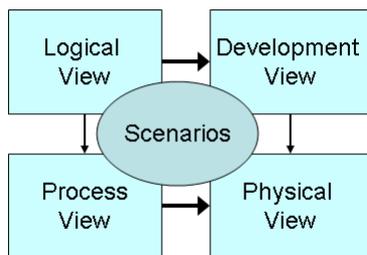


Abb. 3: 4+1 Views von Philippe Kruchten

Als Bestandteil des Rational Unified Process (RUP) wird der 4+1 Ansatz heute in der Praxis eingesetzt, außerhalb den RUP unseres Wissens nach nicht. Die Konsistenz zwischen den einzelnen Sichten ist schwer herstellbar. Positiv an Kruchtens Ansatz ist der Einbezug der fachlichen Sicht, die sowohl in den Scenarios (in Form von Use-Cases) als auch im Logical View Eingang finden.

Zu den Urvätern von Architektursichten gehört der ehemalige IBM-Forscher John A. Zachmann (siehe [zifa]). Er schuf ab 1970 den nach ihm benannten Architekturframework [zifa]. Dieser Framework geht über die reine Beschreibung von IT-Systemen deutlich hinaus, und bezieht sich im Ursprung auf Unternehmen als Ganzes. Der Zachman-Framework postuliert 30 (!) Sichten. Das empfinden wir als deutlich zu viel für den alltäglichen Gebrauch, obwohl Zachmann sich insbesondere im Umfeld US-amerikanischer Großunternehmen verbreiten konnte.

In den USA werden viele (hauptsächlich militärische) Systeme nach dem C4ISR Architecture Framework (siehe [C4ISR]) entwickelt. Er teilt Architekturen in 5 Sichten: Funktionen, Informationen, Organisation, Infrastruktur und Netzwerk. Jede Sicht wird weiter in statische Modelle, so genannte Produkte, zerlegt. Es gibt dazu nur wenig frei zugängliche Dokumentation, jedoch basieren andere Ansätze wie TOGAF (siehe unten) und DODAF auf dem C4ISR-AF.

Deutlich weiter verbreitet ist das ISO Referenzmodell für Open Distributed Processing (RM-ODP, [rmodp]). Es geht nach einem objektorientierten Ansatz vor und enthält neben 5 Sichten (genannt Viewpoints: Enterprise, Information, Computational, Engineering, Technology) noch 8 unterschiedliche Transparencies. Diese beschreiben die übergreifenden Aspekte access, failure, location, migration, replication und transaction. Zu diesem Modell gibt es eine ganze Reihe ergänzender Dokumentation, jedoch kaum veröffentlichte Beispieldokumentation. Unserer Erfahrung nach ist RM-ODP recht schwerfällig und eignet sich eher für langlaufende Großprojekte (jedoch würde einigen von denen auch ein schlankerer Dokumentationsansatz unserer Ansicht nach gut stehen :-).

Das Architekturframework der Open Group, genannt TOGAF (siehe [togaf]), kann man für 90 Tage kostenfrei evaluieren, die kommerzielle Nutzung ist teilweise eingeschränkt. Zur Dokumentation von Architekturen schlägt TOGAF ein *Enterprise Continuum* vor, eine Sammlung sämtlicher architekturelevanten Ergebnisse. Diese sind untergliedert in Geschäfts-, Daten-, Applikations- und Technologiearchitektur. Die Dokumentation von TOGAF enthält kaum konkrete Hinweise, wie Architekturdokumentation praktisch aussehen soll.

Dana Bredemeyer, geistiger Vater des HP-internen Software-Architekturprozesses, schlägt auf [bredemeyer] vor, Architekturbeschreibungen mit den Architekturtreibern zu beginnen, unterteilt in Architekturvision, Architektur Anforderungen, Annahmen, Einflussfaktoren sowie Schlüsselkonzepte und Architekturstil. Anschließend dienen auch bei Bredemeyer eine Reihe von Sichten der Beschreibung der detaillierten Architektur. Er gruppiert Sichten in folgende Matrix:

	Verhaltenssicht	Struktursicht
Konzeptionelle Architektur (abstrakt)	Trace einzelner "Kollaborationen", d.h. Szenarien	Architekturdiagramm, informelle Komponentenspezifikation
Logische Architektur (detailliert)	Kollaborationsdiagramme	Architekturdiagramm mit Schnittstellen, Schnittstellenspezifikationen
Ausführungs-	Kollaborationsdiagramme,	Architekturdiagramme, die

architektur (Prozess- und Verteilungssicht)	die Prozesse ausdrücken	aktive Komponenten zeigen
---	-------------------------	---------------------------

Tab. 1: Sichten nach Dana Bredemeyer [Bredemeyer]

Damit enthält sein Vorschlag einiges von den Dingen, die wir für Architekturbeschreibungen als notwendig erachten, leider fehlt in seinem Modell Platz für die übergreifenden Lösungsaspekte. Positiv an seinem Ansatz erachten wir, dass er die notwendigen Sichten (Bausteine, Laufzeit, Verteilung sowie Kontextsicht) immerhin mit seiner Begrifflichkeit abdecken kann - wenn uns auch die klare terminologische Trennung von Sichten und Architekturebenen fehlt.

7 Fazit

Die Verwendung einer Strukturvorlage zur Dokumentation (und letztlich Kommunikation) von IT- und Softwarearchitekturen erweist sich in der Praxis als förderlich für die Effektivität und Effizienz von Entwicklungsprojekten und auch der Qualität der betreffenden IT-Systeme. Die arc42-Vorlage bietet eine erprobte und praxistaugliche Struktur an, die sich in vielen Fällen bewährt hat. Diese Struktur lässt sich bei Bedarf unternehmens- oder auch systemspezifisch anpassen. Sie ist leicht erlern- und vermittelbar und von daher mit geringem Aufwand in Unternehmen und Organisationen einzuführen.

8 Danksagung

Die Autoren möchten sich bei den zahlreichen Anwendern und Reviewern des arc42 Templates für die konstruktiven Diskussionen sowie Verbesserungsvorschläge bedanken, die diese Strukturvorlage letztendlich erst praxistauglich gemacht haben.

9 Anhang: Detaillierte Gliederung des arc42-Architekturtemplates

Folgende Tabelle zeigt die Verfeinerung der arc42-Struktur auf die in der Praxis relevanten Gliederungsebenen.

Nr.	Titel	Erläuterung
1	Einleitung	Kurzbeschreibung des Systems, Formalia: Änderungshistorie, Versionsstand, Release-Notes
2 2.1 2.2 2.3	Architekturtreiber Architekturziele Aufgabenstellung Stakeholder	Dieses Kapitel ist ein Extrakt aus den Systemanforderungen, die im Idealfall als eigenständige Dokumentation (Requirements Model, Systemanalyse) vorliegen.
3 3.1 3.2 3.3	Randbedingungen Technische Randbed. Organisatorische Randbed. Konventionen	Dieses Kapitel beschreibt die Einschränkungen, denen Architekten beim Systementwurf unterliegen. Thematisch sind sie mit den Anforderungen verwandt, stammen in der Regel jedoch aus anderen Quellen.
4	Kontextsicht	Hier wird das System im Kontext aller Nachbarsysteme und externen Schnittstellen dokumentiert.
5 5.1 5.2 5.3	Bausteinsicht Gesamtsystem Ebene 2 Ebene 3	Hier wird die Bausteinstruktur des Systems beschrieben, wechselweise in White- und Blackboxbeschreibungen.
6 6.1 6.2 6.3	Laufzeitsicht Laufszenario 1 Laufszenario 2 Laufszenario 3	Hier wird das Zusammenspiel der Systembausteine zur Laufzeit dokumentiert. Dabei können spezifische Szenarien (etwa: Bootstrapping) gesondert herausgehoben werden.
7	Verteilungssicht	Hier wird die Verteilung des Systems auf Hardware dokumentiert – (Deployment-Sicht).
8	Entwurfsentscheidungen	
9	Szenarien zur Architekturbewertung	Grundlagen für die szenariobasierte Architekturbewertung.
10 10.1 10.2 etc.	Architektur Aspekte: <ul style="list-style-type: none"> • Persistenz • Benutzungsoberfläche • Ergonomie • Ablaufsteuerung • Transaktionsbehandlung • Sessionbehandlung • Sicherheitsaspekte • Verteilung • Ausnahmebehandlung • Parallelisierung • Konfiguration • Internationalisierung • Migration • Monitoring • Protokollierung 	Hier dokumentieren Sie sichtenübergreifende (i.d.R. technische) Aspekte. Die Aufzählung ist als Beispiel zu betrachten und i.d.R. system- oder unternehmensspezifisch zu ergänzen.
11	Projektaspekte	Informationen und Risiken von oder für das Management des Projektes.
12	Glossar	Projekt- oder systemspezifische Begriffe.

Tab. 2: Verfeinerte Gliederung des arc42 Architekturtemplates

Literaturverzeichnis

- [arc42] Peter Hruschka, Gernot Starke: arc42 – Das Internet-Portal für Software-Architekten. Online: www.arc42.de
- [Bredemeyer] Bredemeyer, Dana: Architecture Documentation Action Guide. Online: http://www.bredemeyer.com/architecture_documentation_action_guides.htm (Stand: Juli 2006).
- [C4ISR] C4ISR Architecture Framework. Online: www.software.org/pub/afc/c4isr-dodaf.asp, Stand Juli 2006.
- [Clements01] Clements, Paul, Rick Kazman & Mark Klein: Evaluating Software Architectures: Methods and case Studies. Addison Wesley, 2001.
- [Clements03] Clements, Paul et. al: Documenting Software Architectures. Addison Wesley, 2003.
- [Kruchten95] Kruchten, Philippe: The 4+1 View Model of Architecture. IEEE Software Vol. 12, Nr. 6, 1995.
- [rmodp] Online: www.dstc.edu.au/Research/Projects/ODP/standards.html (Stand: Juli 2006).
- [Starke05] Starke, Gernot: Effektive Software-Architekturen – Ein praktischer Leitfaden. Carl Hanser Verlag, München, 2. Auflage, 2005.
- [togaf] The OpenGroup: Architecture Framework (Version 8.1). Nach kostenfreier Registrierung online erhältlich unter www.opengroup.org. (Stand Juli 2006)
- [zifa] Zachman Institute for Framework Advancement. Online: www.zifa.com (Stand: Juli 2006)

Anwendungsarchitektur-Beurteilung unter Berücksichtigung zu erwartender Flexibilität

Felix R. Müller

Lehrstuhl für Controlling und Logistik
Universität Regensburg
93053 Regensburg
felix.mueller@wiwi.uni-regensburg.de

Abstract

Für die Gestaltung einer Anwendungsarchitektur existieren Richtlinien, die zu einer „guten“ Abgrenzung von Anwendungssystemen führen sollen. Es ist jedoch noch unklar, wie die Güte einer Aufgabenzerlegung und -zuordnung beurteilt werden kann. In diesem Beitrag wird argumentiert, dass die Flexibilität einer Anwendungsarchitektur ein wichtiges Gestaltungsziel darstellt. Dieses Ziel ist im Software Engineering unter dem Namen *Wartbarkeit* bekannt. Ausgehend von der Annahme, dass komplexe Systeme bei sonst gleichem Leistungsumfang und bei gleichem Anpassungsbedarf einen größeren Wartungsaufwand verursachen, wird versucht, die Größen Kohäsion und Kopplung – für die im Software Engineering anerkannte Metriken existieren – für den Bereich der Gestaltung von Informationssystemen auf Anwendungssystemebene zu adaptieren.

1 Motivation und Fragestellung

Für die Bedienung von Informationsverarbeitungsbedarfen in einem Unternehmen stehen im Rahmen der Gestaltung der Anwendungsarchitektur mehrere Alternativen zur Verfügung. Die Auswahl einer architektonischen Alternative aus dem Lösungsraum hat Einfluss auf den Entwicklungs- und Einführungsaufwand, den Aufwand für den Betrieb der Anwendungssysteme und auf den Aufwand, der für Wartungstätigkeiten zu erwarten ist. Es ist deshalb notwendig, architektonisch verschiedene Lösungen für ein Anforderungsprofil beurteilen zu können, um

eine sinnvolle Auswahl zu treffen. In diesem Beitrag soll ein Werkzeug entwickelt werden, das eine solche Beurteilung ermöglicht.

2 Begrifflicher Hintergrund

Eine Anwendungsarchitektur ist ein Modell eines Informationssystems, das betriebliche IV-Anforderungen (Aufgaben) sowie (Anwendungs-)Softwaresysteme¹ beinhaltet, die jene realisieren. Ein Softwaresystem ist eine Zusammenfassung von Softwaremodulen, die zusammen mindestens eine Funktion erfüllen. Hier wird der Begriff (*Anwendungs-*)*Softwaresystem* verwendet, da der Begriff *Anwendungssystem* auch für das Ensemble mehrerer Softwaresysteme, der Aufgabenrealisierung und ggfs. menschlicher Aufgabenträger verwendet wird. Zwischen diesen Softwaresystemen können Integrationsbeziehungen bestehen. Ein Dienst wird durch funktionale und nicht-funktionale Anforderungen spezifiziert. Funktionale Anforderungen legen fest, welche zu automatisierenden Aufgaben ein Dienst beinhaltet, nicht-funktionale Anforderungen stellen eine Einschränkung des Lösungsraums dar, da sie festlegen, unter welchen Bedingungen die Leistung erbracht werden muss. Es lassen sich nun Mengen von Software-Systemen bilden, die diese Anforderungen erfüllen (Sachziel der „Automatisierungs- und Integrationsaufgabe“ [FeSi01,220]). Architekturentscheidungen beziehen sich auf die Größe der Softwaresysteme² und die Art der Integrationsbeziehungen zwischen den Softwaresystemen [Mert04,20].

3 Anwendungsbereich

Die Beurteilung alternativer gültiger architektonischer Lösungen für ein Anforderungsprofil auf Anwendungssystemebene lässt sich als Teilaufgabe des Integrationsmanagements betrachten, das die „Zusammenfassung aller mit der Integration verbundenen Aufgaben“ [Rose99,5] darstellt. Das Problem lässt sich der Gestaltung eines Informationssystems im Rahmen der Aufgabenzerlegung und der Zuweisung zu automatisierender Aufgabenbestandteile auf Aufgabenträger zuordnen [FeSi01,215-217]. Ist ein „Neuaufwurf“ bestehender Anwendungssysteme und deren Integrationsbeziehungen untereinander zulässig, ist eine derartige Beurteilung jedoch

¹ Da nur Anwendungs-Softwaresysteme betrachtet werden, wird im Folgenden vereinfachend das Präfix weggelassen.

² Die Größe eines Softwaresystems lässt sich zunächst über die Anzahl der durch dieses automatisierten Aufgabenbestandteile messen; für eine präzisere Betrachtung s. u.

auch sinnvoll, da durch architektonische Entscheidungen auf Anwendungsebene mitunter eine bessere Zielerreichung realisiert werden kann.

4 Ziele

Zu Zielen der Gestaltung der Anwendungsarchitektur existiert kein universell anerkannter Katalog. Als wichtiges Teilziel im Rahmen des Integrationsmanagements wird der „optimale Integrationsgrad“ [Rose99,13] behandelt. Ferstl entwickelt – aufbauend auf einer groben Formalziel-Klassifizierung der „Gestaltungsaufgabe Automatisierung von IS“ – Einzelziele der Integration [Fers92,9–14], die jedoch sowohl innerhalb der Klasse als auch klassenübergreifend von Ausprägungen anderer Einzelziele abhängig sein können. Schwinn argumentiert, dass ein optimaler Integrationsgrad schwer zu ermitteln sei und gruppiert fünf Gestaltungsziele, die diesen beeinflussen. Als Oberziel dieser Gestaltungsziele führt er „Agilität“ ein [ScWi05; HaSc06]. Es bleibt jedoch offen, ob die jeweiligen „Gruppen“ vollständig und abgeschlossen sind und wie die Zielerreichung sinnvoll gemessen werden kann. Huber et al. erwähnen „Faktoren“, denen zukünftige Anwendungsarchitekturen genügen müssen; diese sind Flexibilität, Integrationsfähigkeit und Standardisierung [HABÖ00,173–174].

Sowohl Flexibilität als auch Agilität lassen sich unter dem Begriff Anpassungsfähigkeit subsumieren. Flexibilität ist die Fähigkeit eines Systems, auf system- oder umweltinduzierte Veränderungen sinnvoll zu reagieren [Kalu93; Steg04]. Für die Anwendungsebene eines Informationssystems lässt sich Flexibilität als die Fähigkeit beschreiben, geänderte Informationsverarbeitungsbedarfe zu befriedigen (vgl. [Dunc95; GeLe05,44]) oder sich sonstigen Umweltveränderungen (z. B. der Infrastruktur) anzupassen.³ Für den Bereich der Applikationsintegration ist Agilität „die ‚Leichtigkeit‘(...), mit der sich Änderungen oder Erweiterungen [an der Anwendungslandschaft] durchführen lassen“ [HaSc06,277].

Für den Bereich des Software-Designs ist Wartbarkeit (engl. *maintainability* [IEEE90,46]) ein besonders relevantes Qualitätskriterium, da sich Architektur-Entscheidungen darauf auswirken, in welchem Umfang mögliche Änderungen an einer Komponente Anpassungen anderer Systemelemente bedingen. Synonym wird der Begriff Änderbarkeit (engl. *modifiability* [BCKB98; BBKM78]) verwendet. Beide Begriffe beschreiben, wie einfach sich Änderungen an einem System vollziehen lassen, um Fehler zu beseitigen, die Leistungsfähigkeit zu erweitern oder um

³ Für die Infrastruktur können ähnliche Betrachtungen angestellt werden, s. z. B. [Pfau97,73].

das System an eine veränderte Umwelt anzupassen. Diese Anwendungsbereiche korrespondieren mit einer gängigen Klassifizierung von Wartungstätigkeiten (*corrective, perfective, adaptive* [Scha05,480]).

Es verhält sich also so, dass für die Gestaltung von Software-Systemen und die eines Informationssystems auf Anwendungsebene ein ähnliches Ziel formuliert ist: Durch eine sinnvolle Gestaltung des jeweiligen Systems sollen potentielle Anpassungen möglichst günstig durchgeführt werden können.⁴ Maßnahmen, die im Lebenszyklus eines Informationssystems auf Anwendungssystemebene u. U. notwendig sind, lassen sich als Wartungstätigkeiten⁵ in eine der oben vorgestellten Kategorien einordnen; also lässt sich dieses Qualitätskriterium für die Gestaltung eines Informationssystems verwenden (s. a. [Ande02,26]).

Es ließe sich argumentieren, dass diese Reduktion auf Wartbarkeit als einziges Qualitätsziel eine unstatthafte Vereinfachung darstelle. Bezugnehmend auf den Katalog in [Fers92,11–14] ist festzustellen:

- Das Formalziel *Korrektheit* ist nicht weiter skaliert: Eine architektonische Variante, die die funktionalen Anforderungen an den Dienst nicht korrekt erfüllt, ist ungültig. Es sollen hier ausschließlich gültige architektonische Lösungen beurteilt werden.
- Das *Echtzeitverhalten* kann – analog zur Softwareentwicklung [Scha05,277f.] – als nicht-funktionale Anforderung verstanden werden, die den Lösungsraum bereits vorab einschränkt.
- *Integration* kann als Formalziel nicht unabhängig von Flexibilität behandelt werden, da die Ausprägungen der Zielerreichung für Einzelziele der Integration maßgeblich die Flexibilitätseigenschaften einer Lösung beeinflussen. So sind beispielsweise anpassende Wartungstätigkeiten bei minimaler Funktionsredundanz c. p. mit minimalem Aufwand durchführbar (s. auch [HaSc06,281]).

Das Wirtschaftlichkeitsziel für Umsetzung und Durchführung⁶ ist nicht vernachlässigbar. Systeme, die jeweils eine andere Anwendungsarchitektur aufweisen, können sich in den Kosten für Implementierung und Betrieb – Aktivitäten, die *nicht* Wartungstätigkeiten sind – unterscheiden:

⁴ Für die Quantifizierung des zu erwartenden Aufwands für Anpassungen stehen verschiedene Ansätze zur Verfügung. Vereinfachend werden hier z. B. Personen-Stunden angenommen.

⁵ Genauer: Wartungs-, Systemverbesserungs- und Systemanpassungsmaßnahmen [Kore72,200].

⁶ Fünftes Formalziel im o. g. Katalog, ähnlich [ScWi05]: „Minimale Projektaufwände für die Integration.“

Eine Alternative mag z. B. den Betrieb spezieller Infrastrukturkomponenten erforderlich machen, wohingegen eine andere nur durch den Einsatz versierter Experten realisiert werden kann. Dadurch, dass mögliche durchzuführende Anpassungen nicht vollständig im Voraus bekannt und damit nicht bewertbar sind, können diese den Entwicklungs- und Betriebskosten nicht gegenübergestellt werden. Nichtsdestotrotz existieren Beiträge, die die Eintrittswahrscheinlichkeiten aller denkbaren Änderungen der Informationsverarbeitungsbedarfe als bekannt voraussetzen und diese in einem Entscheidungsmodell verarbeiten (z. B. [GeSc06]) oder abstrakte Turbulenzmaße verwenden und qualitative Betrachtungen dokumentieren (z. B. [HiSH05]).

5 Beurteilung anhand zu erwartender Wartbarkeit

Der Aufwand, der für zukünftige mögliche Wartungstätigkeiten anfällt, lässt sich nur schwer abschätzen, da die Anpassungsbedarfe nicht vollständig antizipierbar sind. Deshalb muss auf Indikatoren zurückgegriffen werden, die für Wartbarkeit signifikant sind.

5.1 Indikatoren zur Abschätzung der Wartbarkeitseigenschaften eines Systems

Im Software Engineering werden Wartbarkeit und Einfachheit als proportionale Größen angesehen [StMC74,115]. Um die steigende Komplexität von Softwareprodukten beherrschbar zu machen, wurden in den 1970er Jahren Prinzipien, Methoden und Werkzeuge entwickelt, mit denen sich eine Strukturierung des zu gestaltenden Systems durchführen lässt, u. a. die Modularisierung, also die Zusammenfassung abgeschlossener funktionaler Einheiten mit definierten Schnittstellen (vgl. [Balz98,571–574]). Komplexe Systeme durch Modularisierung und Abstraktion verständlicher zu machen ist jedoch kein Instrument, das spezifisch für die Softwareentwicklung ist; vielmehr handelt es sich um eine Vorgehensweise, die in vielen Disziplinen thematisiert wird [Alex64; Simo65]. Auf den allgemeinen Fall übertragen handelt es sich hierbei um das Problem der Abgrenzung von Subsystemen [Ulri68,107–109].

Für die Modularisierung der Anwendungslandschaft stehen Methoden zur Verfügung, die aus einer Zuordnung von Funktionen oder Prozessen zu Informationsobjekten Softwaresystemkandidaten und deren Datenaustauschbeziehungen untereinander ableiten und somit automatisch zu einer gültigen Zerlegung und Zuordnung führen (z. B. [Gags71] oder als Teil des Business Systems Planning [IBM 84]). Der Verwendung einer derartigen Vorgehensweise für den Vergleich von architektonischen Lösungen stehen jedoch zwei Probleme entgegen: Erstens ist nicht nach-

gewiesen, dass eine Aufgabenzuordnung, die Resultat der Anwendung der Methode auf ein Anforderungsprofil ist, zwingend eine maximale Wartbarkeit gewährleistet. Zweitens bleibt offen, wie alternative architektonische Lösungen verglichen werden können, da keine Abstandsmaße für verschiedene Zerlegungen und Beziehungen definiert sind.

Einfache – und dadurch günstig wartbare – Systeme zeichnen sich durch lose gekoppelte Subsysteme aus; das System ist „beinahe zerlegbar“ [Simo65,69–76].⁷ Änderungen eines Subsystems wirken sich minimal auf andere Subsysteme aus [WaWe90]. Hohe Kohäsion innerhalb der Subsysteme fördert sowohl eine lose Kopplung [StMC74] als auch eine gute Wartbarkeit der Subsysteme selbst. Diese Eigenschaften, Kopplung und Kohäsion, können dazu verwendet werden, die Ergebnisse alternativer Modularisierung zu beurteilen. Im Software Engineering werden zusätzlich zu Metriken, die sich auf Kohäsion und Kopplung beziehen, weitere Maße zur Bestimmung struktureller Komplexität vorgeschlagen⁸, Kohäsion und Kopplung nehmen jedoch eine zentrale Rolle ein.⁹

5.2 Messung von Kohäsion und Kopplung

Es ist nun zu untersuchen, wie Kohäsion und Kopplung in einem Informationssystem auf Anwendungsebene gemessen werden können. Dazu soll erfasst werden, welche Mittel im Software Engineering Anwendung finden. Als Grundlage werden typische Einteilungen verwendet, die sich auf Softwareentwicklung nach dem prozeduralen Programmierparadigma beziehen.

5.2.1 Kohäsion

Die Eigenschaft *Kohäsion* beschreibt den Bezug der Elemente innerhalb eines Moduls. Die Größe gibt an, wie stark die Aufgaben, die ein Modul realisiert, zusammenhängen.¹⁰ Die Eigenschaft ist ordinal skaliert [StMC74,121], es existieren jedoch Metriken zur Bestimmung verschiedener Kohäsionsgrade für prozedural organisierte Programme (z. B. [BiOt94]). Die folgende Darstellung orientiert sich an einer gängigen Einteilung verschiedener Kohäsionsgrade (vgl. [Myer78;Scha05,170–175]); diese sind nach steigender Intensität geordnet (*informational cohesion* ist also die erstrebenswerteste Form der Kohäsion).

⁷ „nearly decomposable“, Übersetzung übernommen aus der deutschen Ausgabe [Simo94].

⁸ Für eine Übersicht s. [Keme95].

⁹ „Any study of past research on structural complexity of software leads to the conclusion that coupling and cohesion are fundamental underlying dimensions“ [DKST05,983].

¹⁰ „The manner and degree to which the tasks performed by a single software module are related to one another.“ [IEEE90,17].

Operationen in Modulen entsprechen betriebliche Aufgabenbestandteile in Softwaresystemen: Beide repräsentieren aus Anforderungssicht bestimmte sinnvolle Funktionen. Zwar ist die Abgrenzung von Operationen bereits Bestandteil des Softwareentwicklungsprozesses, die Ableitung zu automatisierender Aufgabenbestandteile ist jedoch auch zentral für die Gestaltung von Informationssystemen.

Zufällige Kohäsion liegt demnach vor, wenn in einem Softwaresystem Aufgaben zusammengefasst sind, die in keinem fachlichen Zusammenhang stehen (z. B. *Lohnnebenkosten ermitteln* und *Stapler disponieren*). Analog zur Argumentation im Software Engineering ist zufällige Kohäsion als schlechteste Art des Zusammenhalts in einem Softwaresystem zu betrachten: Sind Änderungen an einer Aufgabenrealisierung im Softwaresystem, ist davon eine fachlich unabhängige Funktion betroffen. Ferner ist zufällige Kohäsion ein Indikator für die Kopplung zwischen Softwaresystemen, sofern nicht davon ausgegangen werden muss, dass sämtliche Aufgaben isoliert voneinander, also unabhängige Prozesse sind. Es kann dann geschlossen werden, dass mindestens eine Kopplungsbeziehung zu einem anderen Modul existieren muss.

In einem Softwaresystem liegt logische Kohäsion dann vor, wenn die Aufgabenbestandteile zwar logisch zusammenhängen (z. B. *Belege erstellen* als Modul, das für sämtliche Arten von Belegen Funktionalität bereitstellt), innerhalb des Moduls jedoch starke Abhängigkeiten existieren. Diese Form der Kohäsion ist problematisch, da Funktionen im Softwaresystem starke Abhängigkeiten aufweisen, so dass bspw. bei Änderungen in der Druckausgabe auch Anpassungen der Bildschirmanzeige erforderlich sind. Außerdem lässt sich wie beim Vorliegen zufälliger Kohäsion auf eine ungünstige Kopplung zwischen den Softwaresystemen schließen (unnötige Parameter, Kontrollkopplung).

Werden Aufgaben in einem Softwaresystem zusammengefasst, deren Bearbeitung zeitlich zusammenfällt, liegt zeitliche Kohäsion vor (z. B. *Wareneingang buchen* und *Stapler disponieren*). Diese Zusammenfassung ist als ungünstig zu betrachten, da das zeitliche Zusammentreffen selbst Änderungen unterworfen ist. Werden Geschäftsprozesse umgestaltet, ist der zeitliche Bezug zweier Aufgaben möglicherweise nicht mehr gegeben. Schließt man die Möglichkeit aus, neue Softwaresysteme zu bilden, die dem neuen Ablauf Rechnung tragen, so können Softwaresysteme, die ursprünglich zeitliche kohärent gestaltet sind zu solchen mit zufälliger Kohäsion degenerieren. Eine analoge Argumentation lässt sich für Systeme mit ablauforientierter Kohäsion anführen, in denen Aufgaben zusammengefasst sind, deren sequentielle Durchfüh-

rung in einem Teilprozess festgelegt ist. Deshalb soll diese Unterscheidung im Folgenden vernachlässigt werden.

Im Unterschied zur ablauforientierten Kohäsion gilt für kommunikative Kohäsion (*communicational cohesion*) zusätzlich die Einschränkung, dass alle Operationen auf gemeinsamen Ein- oder Ausgabedaten operieren (z. B. *Sendung erfassen* und *Packstück-Labels erstellen*). Systeme mit kommunikativer Kohäsion sind einfacher als Systeme, in denen ablauforientierte Kohäsion vorherrscht; die Wartungsfreundlichkeit (des Moduls) ist – ohne weitere Beeinflussung der Kopplung – größer, da das System c. p. kleiner ist, jedoch keinen zusätzlichen komplexen Kommunikationsaufwand verursacht. Bei Änderungen am Datenmodell sind – verglichen mit zeitlicher und prozeduraler Kohäsion – weniger Funktionen betroffen. Deshalb kann diese Kohäsionsart auch auf Anwendungsebene nicht mit den beiden vorher genannten zusammengefasst werden.

Realisiert ein Subsystem genau eine Aufgabe, so liegt funktionale Kohäsion vor (z. B. *Stapler disponieren*). Diese zählt für die klassische Softwareentwicklung zu den besten Kohäsionsarten, da neben der Wartbarkeit durch erleichterte Fehlersuche und einfache Austauschbarkeit auch die Wiederverwendung positiv beeinflusst wird. Die positive Bewertung lässt sich auf Anwendungssystemebene jedoch nicht ohne weiteres übernehmen, da noch keine Aussagen über die Ansteuerung der Einzelfunktionen oder über Datenhaltungsaspekte getroffen werden.

Schließlich können Aufgaben, die voneinander unabhängig auf demselben Datum operieren, in Softwaresystemen zusammengefasst werden (*informational cohesion*). Diese Art der Kohäsion liegt beispielsweise abstrakten Datentypen und Objekten im Rahmen des objektorientierten Programmierparadigmas zu Grunde, wenn diese gut gestaltet sind. Bei der Übertragung dieses Konzepts auf die Anwendungsarchitektur entsteht pro „Informationsobjekt“ oder „Resource Business Element“ [MSJL06,85–88] ein Softwaresystemkandidat, in dem sämtliche Aufgaben, die vorrangig auf dieser Information operieren, zusammengefasst sind (z. B. *Lieferscheineintrag erstellen* und *Lieferschein abschließen*). Auch für diese Art der Kohäsion gilt, dass die Einteilung ohne eine Betrachtung der Ablauflogik nicht beurteilt werden kann. Diese kann über alle Subsysteme verteilt oder in einem oder mehreren separaten Modulen implementiert sein; je nach Realisierung ist ein derartiges System mehr oder weniger komplex.

In Tabelle 1 sind die behandelten Kohäsionsarten übersichtsartig dargestellt. In der Spalte *Art (SW)* sind die Arten aus der Software-Entwicklung aufgeführt; in der Spalte *Art (IS)* die zu unterscheidenden Arten für die Beurteilung von Anwendungssystemen.

Art (SW)	Beschreibung	Art (IS)
Coincidental	Modul beinhaltet unzusammenhängende Operationen.	Coincidental
Logical	Modul beinhaltet zusammenhängende Operationen, die vom aufrufenden Modul angesteuert werden.	Logical
Temporal	Modul beinhaltet Operationen, deren Aufruf zeitlich zusammenfällt.	} Temporal
Procedural	Modul beinhaltet Operationen, deren Aufruf in einer bestimmten Reihenfolge zusammenfällt.	
Communicational	Modul beinhaltet Operationen, deren Aufruf in einer bestimmten Abfolge zusammenfällt und die auf den selben Daten operieren.	Communicational
Functional	Modul führt genau eine Operation aus.	Functional
Informational	Modul beinhaltet voneinander unabhängige Operationen, die auf denselben Daten operieren.	Informational

Tabelle 1 Kohäsionsarten

5.2.2 Kopplung

Für den Bereich der Softwareentwicklung ist Kopplung definiert als ein Maß der Verknüpfungsstärke, die aus einer Verbindung zweier Module erwächst.¹¹ Diese Definition bezieht sich auf Entwicklung nach dem prozeduralen Programmierparadigma. Ein Modul im weiteren Sinne ist eine Darstellung einer funktionalen Einheit oder einer semantisch zusammengehörenden Funktionsgruppe, die abgeschlossen ist, definierte Schnittstellen für Externbezüge besitzt und „im qualitativen und quantitativen Umfang handlich, überschaubar und verständlich“ ist [Balz98,572]. Der Kopplungsgrad zwischen Modulen hängt von der Komplexität der Schnittstelle, des Verbindungsbezugs (Modul als solches oder Element innerhalb des Moduls) sowie der Art der Nachricht ab [StMC74,117–121]. Im Folgenden wird eine gängige Einteilung von Kopplungsarten (nach abnehmendem Kopplungsgrad sortiert) aufgegriffen (vgl. [Your79]) – *content coupling* ist also als stärkste Form der Kopplung am wenigsten erstrebenswert.

Es sollen nun Beziehungen *zwischen* Softwaresystemen untersucht werden. Auch hier wird angenommen, dass diese mit den Modulen des Software Engineering korrespondieren: Ein Softwaresystem beinhaltet also die Implementierung mindestens eines Aufgabenbestandteils und kann mit anderen Softwaresystemen in Beziehung stehen. Die Stärke der Beziehung wird durch den Grad der Kopplung angegeben.

Inhaltliche Kopplung liegt vor, wenn ein Softwaresystem direkt ein Datum oder Funktionalität eines anderen Softwaresystems ändert. Dies ist zum Beispiel der Fall, wenn ein Softwaresystem *Konstruktion* eine interne Variable im Softwaresystem *Behälterplanung* überschreibt. Dieses Design macht selbst bei internen Änderungen im referenzierten Softwaresystem, die also

¹¹ „The measure of strength of association established by a connection from one module to another“ [StMC74,117].

nicht die externen Schnittstellen der Software betreffen, Anpassungen der referenzierenden Softwaresysteme erforderlich.

Für Anwendungsarchitektur besonders relevant ist die Kopplung über gemeinsame Datenbanken (*common coupling*), wobei mehrere Softwaresysteme die enthaltenen Daten ändern können (z. B. „Datenintegration“ [Fers92,19–21]). Aus Sicht des Software Engineering wird diese Kopplungsart als Indikator für ein schlechtes Design angesehen, da bei Änderung der Deklaration einer globalen Variable alle Funktionen, die diese Variable verwenden, angepasst werden müssen, da Sicherheitsprobleme auftreten können und da sich die Anzahl derartiger Kopplungsbeziehungen eines Moduls ohne Änderungen am Modul selbst ändern kann („clandestine common coupling“ [SJWH03]). Diese Probleme gelten für Anwendungsarchitekturen analog: Änderungen am Datenschema können Wartung gekoppelter Softwaresysteme erforderlich machen und die Anzahl an Softwaresystemen, die an einer solchen Beziehung teilnehmen, kann sich ändern, ohne dass diese Veränderung für ein bestimmtes Softwaresystem offensichtlich ist. Rückverfolgbarkeits- und Sicherheitsprobleme können ebenso auftreten.

Kontrollkopplung (*control coupling*) liegt vor, wenn ein Softwaresystem über einen Parameter beim Aufruf eines anderen Softwaresystems Einfluss auf den Kontrollfluss nimmt. Dies ist z. B. dann der Fall wenn ein Softwaresystem *Produktionssteuerung* das Softwaresystem *Auftragsfreigabe* aufruft und dieses für einen Auftrag eine Meldung zurückgibt, die dem aufrufendem System die Änderung der Auftragsgröße vorschreibt. Dazu muss das aufgerufene System Kenntnis von der Funktionsweise des aufrufenden Moduls haben. Ändert sich die Logik im aufrufenden Subsystem, sind Anpassungen des aufgerufenen Systems erforderlich.

Werden beim Aufruf eines Softwaresystems als Parameter unnötigerweise komplexe Informationsobjekte übergeben, liegt Datenbereichskopplung (*stamp coupling*) vor. In der Softwareentwicklung ist dies zu vermeiden, da die Schnittstellenkomplexität steigt, die Wiederverwendbarkeit sinkt und Sicherheitsprobleme auftreten können. Allerdings ist anzumerken, dass die schlechte Bewertung einer Datenbereichskopplung dadurch verstärkt wird, dass in einigen verbreiteten Programmiersprachen (z. B. Java) komplexe Strukturen normalerweise als Referenz und einfache Datentypen als Wert übergeben werden, so dass zusätzliche Komplexität dadurch entsteht, dass eine Rückverfolgbarkeit der Änderungen an einer Struktur oder einem Objekt durch häufige „Weitergabe“ erschwert wird. Diese Unterscheidung muss also auch auf Anwendungssystemebene getroffen werden, um eine Beurteilung dieses Musters durchführen zu können. Wird beispielsweise einem Softwaresystem *Versand* zur Erstellung der Versandbelege als

Argument ein Informationsobjekt *Kunde* übergeben, so sind Statistiken zum Bestellverhalten des Kunden für die Ausführung der Aufgabe zunächst unerheblich. Dennoch hat das Softwaresystem in diesem Falle Zugriff auf den kompletten Kundenstammdatensatz. Müssen Änderungen der Kennzahlen zum Bestellverhalten am Informationsobjekt nachvollzogen werden, so muss in diesem Fall auch das Softwaresystem *Versand* untersucht werden.

Erhält ein Softwaresystem beim Aufruf einer Funktion genau die Informationsobjekte, die zur Erfüllung der Aufgabe erforderlich sind, unterhält dies mit dem aufrufenden System eine Beziehung vom Typ Datenkopplung (*data coupling*). Hierbei bleibt zunächst unberücksichtigt, ob ein System zur Durchführung der Aufgabe auf andere Datenbestände zurückgreift, die jedoch vom aufrufenden System nicht verändert werden. Datenkopplung wird im Software Engineering als Indikator für ein gutes Design betrachtet.¹² In Tabelle 2 sind die Kopplungsarten zusammengefasst (Darstellung analog zu Tabelle 1).

Art (SW)	Beschreibung	Art (IS)
Content	A referenziert direkt den Inhalt von B.	Content
Common	A und B haben Zugriff auf ein globales Datum.	} External
External	A und B kommunizieren über eine gemeinsame Datei.	
Control	A ruft B mit einer Nachricht auf, um den Kontrollfluss in B zu beeinflussen.	Control
Stamp	A übergibt B unnötigerweise eine komplexe Datenstruktur	Stamp
Data	A übergibt B eine einfache Variable oder sinnvollerweise eine komplexe Datenstruktur	Data

Tabelle 2 Kopplungsarten

6 Beispiel

Als Anwendungsbeispiel dient ein Geschäftsprozess, der im Rahmen des Forschungsprojekts „Supra-adaptives Logistiknetzwerk Automobilwirtschaft in Bayern“ erhoben wurde. Aus diesem Prozess wurden zu automatisierende Aufgabenbestandteile abgeleitet, die verschieden realisiert werden können. Im Folgenden werden zwei Alternativen dargestellt und der oben entwickelten Beurteilung unterworfen.

6.1 Problem

Einige Logistikdienstleister (LDL), die an der Schnittstelle zwischen 1-st-Tier-Zulieferer und Hersteller (OEM) in der Automobilindustrie agieren, sehen sich mit der Situation konfrontiert, dass Zulieferer Packstücke (PS) mit VDA-Warenanhänger [VDA 94] übergeben, auf denen Barcodes in einem proprietären Format aufgedruckt sind. Für die Abwicklung beim LDL ist es

¹² „Data coupling is a desirable goal.“ [Scha05,180].

notwendig, im Wareneingang jedes PS zusätzlich mit einem eigenen Label zu versehen. Es ist also sinnvoll, das PS-Label bereits initial mit einer vom LDL automatisiert lesbaren PS-ID und mit einer Routing-Angabe zu versehen. Der Prozess ist in Abbildung 1 zusammengefasst.

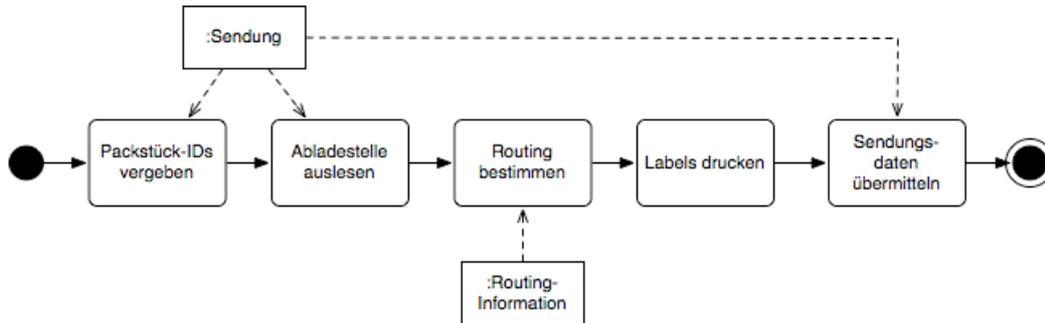


Abbildung 1 Prozess Packstückbelabelung

6.2 Alternative 1

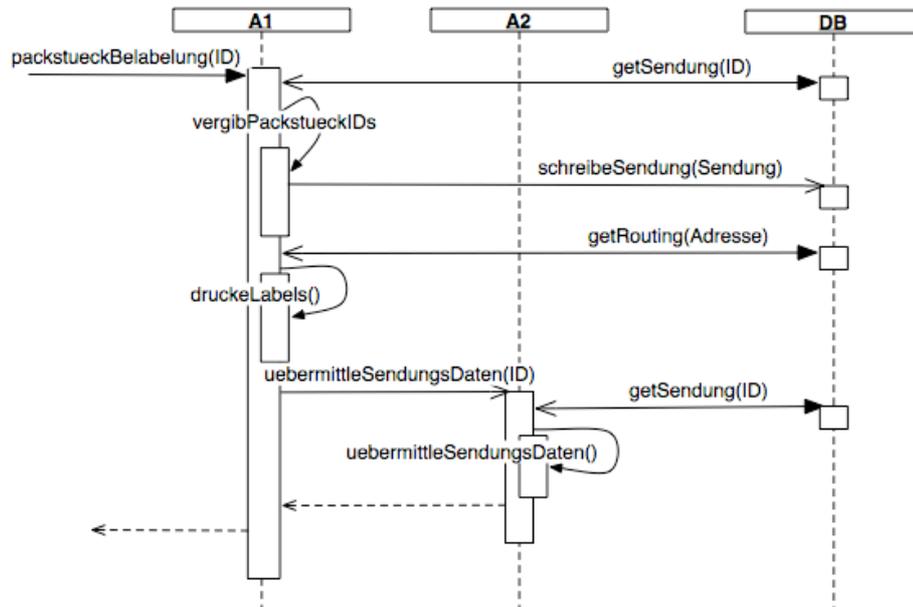


Abbildung 2 Umsetzung Packstückbelabelung – Alternative 1

Der erste Realisierungsvorschlag umfasst zwei Softwaresysteme (vgl. Abbildung 2): In A1 sind sämtliche Funktionen außer der Übermittlung der Sendungsdaten (A2) zusammengefasst. Für A1 liegt zeitliche Kohäsion vor: Die enthaltenen Funktionen sind aus einer Aufrufabfolge heraus betrachtet zusammenhängend, operieren jedoch auf unterschiedlichen Ein- bzw. Ausgabedaten. A2 beinhaltet genau eine Funktion, ist also funktional kohärent. A1 speichert die erstellten PS-IDs im Sendungs-Objekt ab. A2 liest diese Daten aus und übermittelt sie an ein Softwaresystem beim Logistikdienstleister. Es liegt also externe Kopplung über eine Datenbank vor.

6.3 Alternative 2

In der zweiten Variante werden drei Softwaresysteme gebildet (vgl. Abbildung 3): A1 für die Vergabe der PS-IDs, A3 für die Bestimmung der Zielniederlassung und A2 für die Durchführung der sonstigen Aufgaben. A1 und A3 sind funktional kohärent (genau eine Aufgabe wird durchgeführt), für A2 liegt maximal zeitliche Kohäsion vor (Aufgaben fallen im Prozessmodell zusammen); da A2 eine Aufgabe im Ablauf übernimmt und somit den zeitlichen Zusammenhang für A3 *unterbricht*, ließe sich auch zufällige Kohäsion unterstellen. Es wird jedoch angenommen, dass A3 keine weiteren Aufgaben realisiert; der Zusammenhang ist somit über die Realisierung von Aufgaben für *einen* Geschäftsprozess – PS-Belabelung nämlich – hergestellt. Dem Softwaresystem A1 wird eine Sendung übergeben, für die PS-IDs generiert werden. Diese werden zurückgegeben. A3 liest das Sendungsobjekt übergibt die Abladestelle an das Softwaresystem A2, das Labels erstellt und die Sendungsdaten übermittelt. A1 erhält als Eingabedatum ein Sendungsobjekt, aus dem es jedoch nur bestimmte Daten ausliest; es liegt also Datenbereichskopplung vor. Als Rückgabewerte werden PS-IDs geliefert (Datenkopplung). A3 wird mit der Abladestelle als Argument aufgerufen; auch dieses Datum ist minimal für die Durchführung der Aufgabe. Liefert A3 lediglich das Kürzel der Empfangsniederlassung, liegt auch für die Rückgabe Datenkopplung vor.

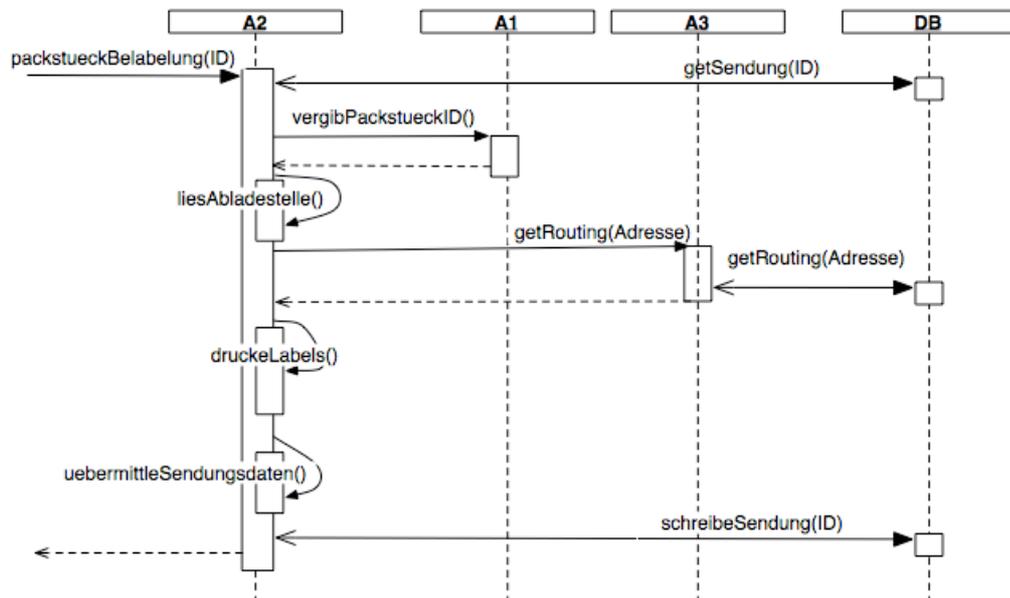


Abbildung 3 Umsetzung Packstückbelabelung – Alternative 2

Diese architektonische Alternative ist gemäß einem realistischen Anwendungsfall konstruiert worden: Ein LDL stellt seinen Kunden zur Abwicklung von Transporten über sein Netzwerk bisweilen ein eigenes Softwaresystem zur Verfügung, das über eine Schnittstelle, die vom Ver-

sandsystem angesprochen werden kann, für eine Adresse die entsprechende Empfangsniederlassung zurückliefert.

Dadurch, dass sich kein Softwaresystem durch eine schlechtere Kohäsion als ein Softwaresystem in der ersten Variante auszeichnet, und einige Instanzen externer Kopplung vermieden werden, ist die zweite Variante aus architektonischer Sicht vorzuziehen.

7 Zusammenfassung und weitere Schritte

Aufbauend auf bestehenden Ansätzen zur Beurteilung architektonischer Alternativen wurde argumentiert, dass Flexibilität mit dem Konzept der Wartbarkeit aus der Domäne der Softwareentwicklung korrespondiert. Es wird dort davon ausgegangen, dass Wartbarkeit mit Strukturkomplexität korreliert ist. Deshalb wurde versucht, Metriken zur Messung von Kohäsion und Kopplung auf die Ebene der Anwendungsarchitektur zu übertragen. Einige Stufen der jeweiligen Eigenschaften konnten dabei zusammengefasst werden, da eine weitere Unterscheidung für die Gestaltung von Anwendungssystemen nicht sinnvoll ist.

Die Beurteilung erfolgt bislang qualitativ und bedarf einer Einschätzung beteiligter Interessenträger. Das vorliegende Anwendungsbeispiel ist sehr einfach gehalten: Es gibt keinerlei Überschneidung, die eine Abwägung erforderlich machen würde. Es ist mit den vorgestellten Metriken also bislang nicht möglich, nicht-triviale Anwendungsfälle zu vergleichen. Um diese Probleme bearbeiten zu können, bietet sich beispielsweise der Einsatz des Analytic Hierarchy Process' (AHP, [Saat80]) als Instrument zur Entscheidungsunterstützung an.

Da die Anwendung von Integrationskonzepten einen Einfluss auf die Aufgabenzuordnung und die Verknüpfung betrachteter Softwaresysteme hat, soll abgeschätzt werden, welche Ausprägungen für die Eigenschaften Kohäsion und Kopplung des Ergebnisses einer Gestaltung gemäß dem jeweiligen Konzept möglich sind. Des Weiteren soll ermittelt werden, welchen Einfluss die Anwendung von Referenzmodellen und Instrumenten der strategischen Informationssystem-Planung auf die oben vorgestellten Eigenschaften eines zu erwartenden „Endprodukts“ hat.

Literatur

- [Alex64] *Alexander, Christopher*: Notes on the Synthesis of Form, Harvard University Press, Cambridge (USA) 1964.
- [Ande02] *Andersson, Jonas*: Enterprise Information Systems Management: An Engineering Perspective Focusing on the Aspects of Time and Modifiability. Dissertation, Department of Electrical Engineering, KTH, Stockholm (Schweden), 2002.
- [Balz98] *Balzert, Helmut*: Lehrbuch der Software-Technik, Spektrum, Heidelberg 1998.
- [BBKM78] *Boehm, Barry W.; Brown, John R.; Kaspar, Hans; Myron, Lipow; MacLeod, Gordon J.; Merritt, Michael J.*: Characteristics of Software Quality, North-Holland, Amsterdam (NL) 1978.
- [BCKB98] *Bas, Len; Clements, Paul; Kazman, Rick; Bass, Ken*: Software Architecture in Practice, Addison-Wesley, Boston (USA) 1998.
- [BiOt94] *Bieman, James M.; Ott, Linda M.*: Measuring Functional Cohesion. In: IEEE Trans. Softw. Eng. 20 (1994) 8, S. 644–657.
- [DKST05] *Darcy, David P.; Kemerer, Chris F.; Slaughter, Sandra A.; Tomayko, James E.*: The Structural Complexity of Software: Testing the Interaction of Coupling and Cohesion. In: IEEE Trans. Softw. Eng. 31 (2005) 11, S. 982–995.
- [Dunc95] *Duncan, Nancy Bogucki*: Capturing Flexibility of Information Technology Infrastructure: A Study of Resource Characteristics and their Measure. In: JMIS 12 (1995) 2, S. 37–57.
- [Fers92] *Ferstl, Otto K.*: Integrationskonzepte betrieblicher Anwendungssysteme (Fachbericht Informatik 1/1992), Koblenz 1992 .
- [FeSi01] *Ferstl, Otto K.; Sinz, Elmar J.*: Grundlagen der Wirtschaftsinformatik. 4. Aufl., Oldenbourg, München 2001.

- [Forl06] Bayerischer Forschungsverbund supra-adaptive Logistiksysteme (Forlog): Supra-adaptives Logistiknetzwerk Automobilwirtschaft in Bayern, <http://www.forlog.de>, Abruf 2006-10-24.
- [Gags71] *Gagsch, Siegfried*: Probleme der Partition und Subsystembildung in betrieblichen Informationssystemen. In: *Grochla, Erwin; Szyperski, Norbert (Hrsg.)*: Management-Informationssysteme – Eine Herausforderung an Forschung und Entwicklung. Gabler, Wiesbaden 1971, S. 623–652.
- [GeLe05] *Gebauer, Judith; Lee, Fei*: Towards an Optimal Level of Information System Flexibility – A Conceptual Model. In: *Bartmann, Dieter et al. (Hrsg.)*: Proceedings of the 13th ECIS. Regensburg 2005. CD-ROM Edition.
- [GeSc06] *Gebauer, Judith; Schober, Franz*: Information System Flexibility and the Cost Efficiency of Business Processes. In: JAIS 7 (2006) 3, S. 122–147.
- [HABÖ00] *Huber, Thomas; Alt, Rainer; Barak, Vladimir; Österle, Hubert*: Future Application Architecture for the Pharmaceutical Industry. In: *Österle, Hubert; Fleisch, Elgar; Alt, Rainer (Hrsg.)*: Business Networking. Springer, Berlin 2000, S. 163–183.
- [HaSc06] *Hagen, Claus; Schwinn, Alexander*: Measured Integration – Metriken für die Integrationsarchitektur. In: *Schelp und Winter [ScWi06]*, S. 267–292.
- [HiKL95] *Hirschheim, Rudy; Klein, Heinz K.; Lyytinen, Kalle*: Information Systems Development and Data Modeling – Conceptual and Philosophical Foundations, Cambridge University Press, Cambridge (UK) 1995.
- [HiSH05] *Hilhorst, Cocky; Smits, Martin; van Heck, Eric*: Strategic Flexibility and IT Infrastructure Investments – Empirical Evidence in two Case Studies. In: *Bartmann, Dieter et al. (Hrsg.)*: Proceedings of the 13th ECIS. Regensburg 2005. CD-ROM Edition.
- [IBM 84] IBM (Hrsg.): Business Systems Planning – Information Systems Planning Guide (GE20-0527-4). 4. Aufl., IBM, Armonk (USA) 1984.

- [IEEE90] *IEEE Institute of Electrical and Electronics Engineers: IEEE Standard Glossary of Software Engineering Terminology*, IEEE Std 610.12-1990, 02/1991.
- [Kalu93] *Kaluza, Bernd: Flexibilität, betriebliche*. In: *Wittmann, Waldemar; et al. (Hrsg.): Handwörterbuch der Betriebswirtschaft*. Schäffer-Poeschel, Stuttgart 1993, Bd. 1, S. 1173–1183. 5. Aufl.
- [Keme95] *Kemerer, Chris F.: Software complexity and software maintenance: A survey of empirical research*. In: *Annals of Software Engineering* 1 (1995) 1, S. 1–22.
- [Kore72] *Koreimann, Dieter S.: Systemanalyse*, de Gruyter, Berlin 1972.
- [Mert04] *Mertens, Peter: Integrierte Informationsverarbeitung*, Bd. 1. 14. Aufl., Gabler, Wiesbaden 2004.
- [MSJL06] *McGovern, James; Sims, Oliver; Jain, Ashish; Little, Mark: Enterprise Service Oriented Architectures – Concepts, Challenges, Recommendations*, Springer, Berlin 2006.
- [Myer78] *Myers, Glenford J.: Composite, structured design*, Van Nostrand, New York (USA) 1978.
- [Pfau97] *Pfau, Wolfgang: Betriebliches Informationsmanagement – Flexibilisierung der Informationsinfrastruktur, Markt- und Unternehmensentwicklung*. DUV Gabler, Wiesbaden 1997. zugl. Habil. Universität Freiburg.
- [Rose99] *Rosemann, Michael: Gegenstand und Aufgaben des Integrationsmanagements*. In: *Scheer, August-Wilhelm; Rosemann, Michael; Schütte, Reinhard (Hrsg.): Integrationsmanagement (Arbeitsbericht Nr. 65)*. Institut für Wirtschaftsinformatik, Münster 1999, S. 5–18.
- [Saat80] *Saaty, Thomas L.: The analytic hierarchy process – Planning, priority setting, resource allocation*. McGraw-Hill, New York (USA) 1980.
- [Scha05] *Schach, Stephen R.: Object-oriented and Classical Software Engineering*. 6. Aufl., McGraw-Hill, New York (USA) 2005.

- [ScWi05] Schwinn, Alexander; Winter, Robert: Entwicklung von Zielen und Messgrößen zur Steuerung der Applikationsintegration. In: Ferstl, Otto K.; Sinz, Elmar J.; Eckert, Sven; Isselhorst, Tilman (Hrsg.): Wirtschaftsinformatik. Physica-Verlag, Heidelberg 2005, S. 587–606.
- [ScWi06] Schelp, Joachim; Winter, Robert (Hrsg.): Integrationsmanagement, Springer, Berlin 2006.
- [Simo65] Simon, Herbert A.: The Architecture of Complexity. In: van Bertalanffy, Ludwig; Rapoport, Anatol (Hrsg.): General Systems. Society for General Systems Research, Bedford (USA) 1965, Bd. 10, S. 63–76.
- [Simo94] Simon, Herbert A.: Die Wissenschaft vom Künstlichen. 2. Aufl., Springer, Wien 1994.
- [SJWH03] Schach, Stephen R.; Jin, Bo; Wright, David R.; Heller, Gillian Z.; Offutt, Jeff: Quality Impacts of Clandestine Common Coupling. In: Software Quality Journal 11 (2003) 3, S. 211–218.
- [Steg04] Steger, Hans-Diego: Systemflexibilität – Konzeption und Gestaltungsansätze einer systemorientierten Unternehmensflexibilität, Eberhard, München 2004.
- [StMC74] Stevens, Wayne P.; Myers, Glenford J.; Constantine, Larry L.: Structured Design. In: IBM Systems Journal 13 (1974) 2, S. 115–139.
- [Ulri68] Ulrich, Hans: Die Unternehmung als produktives soziales System – Grundlagen der allgemeinen Unternehmungslehre, Paul Haupt, Bern 1968.
- [VDA 94] VDA Verband der Automobilindustrie: Warenanhänger (barcodefähig); VDA-Empfehlung 4902 V.4. <http://www.vda.de/de/service/bestellung/downloads/4902.pdf>, Abruf 2006-07-06.
- [WaWe90] Wand, Yair; Weber, Ron: An Ontological Model of an Information System. In: IEEE Trans. Softw. Eng. 16 (1990) 11, S. 1282–1292.
- [Your79] Yourdon, Edward: Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design, Prentice Hall, Englewood Cliffs (USA) 1979.

Sizing Considerations for Enterprise Applications in Dynamic Data Centre Environments

Heiko Thimm

Institute of Business Information Technology
University of Applied Sciences Kiel
24149 Kiel
heiko.thimm@fh-kiel.de

Abstract

Dynamic data centres are regarded as promising approach to achieve a high degree of resource utilization. In such data centres, applications are dynamically allocated to servers under consideration of the actual resource demand and the actual load states of the servers. However, as a result of this flexible deployment model, it is difficult to size the initial servers and other hardware equipment. We present considerations for this specific problem and a corresponding sizing method that is based on a heuristic algorithm. For a given set of applications, our algorithm determines the number of required servers, the subset of applications to be allocated to each server, and the corresponding runtime characteristics of each server. The maximum load that needs to be considered for each of these servers is optimized from a global data centre perspective. This is achieved through a smart rule based orchestration of the individual applications' load profiles.

1 Introduction

The performance of an enterprise application depends substantially on the performance capacity of the underlying IT infrastructure. Henceforth, application performance requirements of the business, such as the application response time, can only be satisfied if the application is deployed on a well sized IT infrastructure.

As the experience with enterprise computing during the last decade shows, the backend servers of enterprise applications, such as database servers, have to be regarded as single most critical

IT infrastructure component from a hardware sizing point of view. Normally, the corresponding hosts on which these backend servers are being deployed belong to the most expensive data centre components. Therefore, whenever an existing application landscape is modified, e.g. extended by new applications, a corresponding sizing project needs to be completed. Such sizing projects usually follow the general approach to determine the required performance capacity of the hosts mainly based on workload assumptions. It is common business practise for traditional data centre environments to derive the required performance capacity from the expected peak load of the application(s). The resulting performance capacity usually includes a safety charge to compensate the fuzziness of this sizing method.

It is well known that this sizing practise results into poor server utilization, a fact that conflicts with the recent trend of increasingly shrinking corporate IT budgets. In the search for solutions to this problem of underutilized hosts, new technologies such as virtualization and capacity management techniques have been developed. These new technologies enable the implementation of data centres which are not based anymore on a persistent 1:1-allocation of applications to hosts. Under the notion of *dynamic data centres* we broadly subsume this new type of data centre. In dynamic data centres, applications may be dynamically allocated to hosts that belong to a pool of shared servers. This approach enables very flexible deployment options under consideration of the applications' actual resource demand and the actual server performance capacity at runtime. Therefore, available capacity management solutions for dynamic data centres monitor the system load and gather corresponding load data. Through an analysis of this load data, it is possible to obtain insights that are helpful to achieve a high degree of server utilization by dynamic resource allocation actions.

In our opinion, specialized sizing methods for dynamic data centres are required that already in the hardware planning phase take the "built-in" deployment dynamics of dynamic data centres into consideration. We could not find existing work that is geared at such sizing methods. In our research, we focus on this gap of knowledge. We strive on the investigation of effective sizing approaches and, in the long run, on the development of corresponding sizing tools for dynamic data centres. In this paper, we present our initial considerations for this research agenda that are largely based on our experience with sizing traditional data centre servers. In addition to that, we propose a sizing method that makes use of a heuristic algorithm. First, this algorithm obtains a base allocation of applications to hosts by grouping applications with similar characteristics together. An own dedicated host is allocated to each of these groups of applications and also to

those applications that definitely need to be deployed on a dedicated server. Each host's runtime properties are determined from the characteristics of the applications being allocated to the host. Then, in a next step, performed are optimization operations across the host-specific application sets. These operations re-orchestrate the sets of applications to make use of complementary load patterns among the applications. The performance capacity being required for a server may be minimized through such operations. In the near future, we will study the effectiveness of our algorithm through a simulation study.

The rest of the paper is organized as follows. Section 2 presents main aspects of traditional hardware server sizing for enterprise applications. Section 3 contains key characteristics of dynamic data centres from a sizing point of view. In Section 4.1, we first discuss considerations for the problem of server sizing for dynamic data centres. Then, in Section 4.2, we present various methods to predict the required performance capacity that are specialized to the different deployment modes found in dynamic data centres. Section 4.3 contains an informal description of our sizing method. Related work is discussed in Section 5 and our conclusions are given in Section 6.

2 Traditional Hardware Server Sizing for Enterprise Applications

Observations about traditional sizing practice. In several years of experience in sizing many different enterprise applications such as SAP R/3, Siebel CRM, and Oracle Applications, we observed that, in principle, sizing methods are usually composed of three steps.

First, the load profile of the targeted production system is determined. Typically, this is done through an assessment of the various activities to be performed by the system such as user activities, background jobs, and the system's own bookkeeping activities. In rare cases only, the load profile is derived through performance tests with a real system because of the large amount of efforts required for such experiments. Given the load profile, the peak load of the production system is determined and further considered in the next step.

In the second step, the performance capacity required by the hardware server on which the application will be deployed is predicted from the peak workload of the system. This predicted performance capacity is expressed either in application specific terms such as the number of SAPS in case of SAP R/3 [LoMa03] or in other terms, that are related to standard performance

benchmarks. In case of OLTP applications, it is often referred to TPC-C which is the OLTP benchmark of the *Transaction Processing Performance Council (TPC)*.

In the third step, the performance relevant server configuration is obtained from the performance capacity resulting from the previous step. Usually system engineers make use of hardware specific performance specification data and corresponding configuration recommendations. For example, such a guideline may describe that for a specific application the RAM size should be 2 GB per CPU of a particular CPU type.

Sizing risk. Traditional sizing methods as described above may be regarded as “fuzzy methods” because they mainly work with assumptions. By the concept of *sizing risk* we address this “fuzzy nature” of these methods. We use the notion of sizing risk to express the probability that the server may be overloaded at some point in time in the future during the production usage phase. In overload situations, the actual performance capacity of a server is lower than the performance capacity needed to satisfy the given application specific service levels such as response time for interactive users. It is possible to lower the sizing risk by extensive studies of the future deployment characteristics so that very accurate sizing assumptions are made available. However, such extensive studies are usually limited by cost and time constraints, respectively. In practice, often it is dealt with the sizing risk by the consideration of a safety charge that is added to the required performance capacity. It is assumed that this extra safety charge will compensate load generating activities ignored by the sizing method otherwise. Obviously, also the consideration of a safety charge is limited by given cost constraints.

Types of traditional sizing projects. It is possible to classify sizing projects into different types depending on the broader project context.

By *initial sizing project*, we refer to the case where the targeted application is deployed for the *first time at all* within the organization (hence the term *initial*). That is, there does not exist any experience in the deployment of the new application so that the sizing assumptions are relatively vague. As a consequence, a relatively high sizing risk needs to be considered for this type of sizing projects. For risk mitigation, it has been recommended to complete performance experiments with a corresponding test system.

In some cases, it is possible to leverage pre-existing deployment experience in initial sizing projects. For example, consider a hardware platform switch for an existing application landscape. Due to more accurate sizing assumptions, usually, such sizing projects need to deal

with only a minor sizing risk as compared to sizing projects without pre-existing deployment experience.

In so-called *upgrade sizing projects*, it is necessary to determine the additional performance capacity needed by an application system already running in production mode. Typical causes for such projects include an increase of the number of application users, the implementation of functional extensions, and release upgrades. Such projects are typically a subject of capacity management for which dedicated tools are available. For example, such tools allow to monitor, simulate, and analyse the work load based on actual load data and to predict the extra performance capacity needed. Henceforth, usually only a low sizing risk is to be considered in upgrade sizing projects.

Discussion. As a result of the above described sizing practice, traditional data centres suffer from a low degree of server utilization. For the decision makers this situation presents a dilemma because they only have a choice between a high sizing risk and a low degree of server utilization. Increasingly more attention to this problem has been paid for the last several years due to the cost pressure that IT departments need to deal with. In the search for solutions to this problem, approaches have been developed that different technology providers call “dynamic IT”, “dynamic infrastructure”, “adaptive infrastructure”, “dynamic data centre”, or “adaptive computing”. These initiatives share all the same idea to enable a flexible dynamic resource management and some self-management capabilities in order to provide a cost-effective and adaptable IT infrastructure. In this work, we broadly subsume these approaches under the notion of *dynamic data centre*. In the next section, we describe the general principles of dynamic data centres from a sizing point of view.

3 Key Characteristics of Dynamic Data Centres from a Sizing Perspective

In traditional data centres, enterprise applications are deployed typically on exclusive hosts. That is, the hardware servers are considered as exclusive computing resources for only a single application. They are not regarded as shared resources that may run multiple enterprise applications at the same time in a shared mode as it has been the case in mainframe computing environments. The consideration of an exclusive host usually leads to an installation procedure where the application software is combined with the server in a relatively radical way. For example, often the IP address of the host is hard-coded in the configuration files of an

application. As a consequence, a separation of the application software from the server at a later point in time is very hard to accomplish. Thus, usually it requires a lot of efforts to move such an application from one server to a different one.

Through the use of virtualization techniques, dynamic data centres are capable to work without such a persistent assignment of application software to underlying hardware servers. The servers are presented to the applications as pooled resources that may be deployed flexibly by the applications according to different deployment models. This enables to allocate applications to a given server only for a certain period of time and to re-allocate the application later to a different server. Furthermore, some applications may even be forced into a planned downtime mode. For example, during a high load phase, the application may be deployed on a server with a high performance capacity. From this server, the application may be moved to a less powerful server for a different time period where the application load is only low.

This flexibility allows for dynamic data centres to allocate applications to available hardware servers dynamically (hence the term *dynamic* data centre) under consideration of the actual resource demands of the applications and the actual load states of the servers. In several initiatives (e.g. [GSWK05]) concepts are investigated for a central management instance that is capable to automatically schedule and manage such dynamic re-allocation actions.

The re-allocation of applications, however, leads to some negative effects such as extra costs, an increasing risk for system failures, and application down time. Therefore, re-allocation actions should not occur with a too high frequency.

From a hardware server sizing view it is necessary to differ between different kinds of resource sharing models that may occur in dynamic data centres. In the following we present three different models.

Exclusive resource sharing deployment mode. If applications are deployed according to this mode, only one application may run on a given host at a time. That is, the complete performance capacity of the host is available for exclusive usage by only one application at a given point in time. However, it may occur that the application is moved to another server or put on hold in order to allow another application to be (exclusively) deployed on the same server.

Non-exclusive resource sharing deployment mode. In the non-exclusive sharing mode, the data centre servers are shared by multiple applications at a time. Each of these applications consumes a certain share of the servers total performance capacity.

Mixed resource sharing deployment mode. This type of deployment mode presents a combination of the previously two mentioned modes. For some periods of time, the server is deployed exclusively by only a single application, while during other periods of time multiple applications are deployed in parallel. Note that this may include the case where an application that has been deployed exclusively on a server from some point in time on will be accompanied (at the same server) by further applications. That is, at this mentioned point in time, it is switched from an exclusive deployment mode into a non-exclusive deployment mode.

4 Sizing Considerations for Dynamic Data Centre Environments

It is envisioned that dynamic data centres may be capable to allocate and re-allocate applications to computing resources autonomously without any participation of human system administrators, in the future. However, in today's available solutions, the allocation task is still controlled by the data centre personal. These solutions mainly build on the existence of load data gathered in the production usage phase. This load data is analyzed and the results are used to derive allocation plans.

For the task of sizing initial data centre servers, however, such load data obviously is not available. Among other reasons, this has lead to the fact that today it is still searched for an effective approach for sizing initial servers for dynamic data centre environments. The inherent property of such environments, that the allocation of applications to servers are dynamically changing over the time, presents one of the crucial problems for this effort.

Our research strives on the investigation of such sizing approaches and on the development of corresponding sizing tools in the long run. As starting point for the development of a first approach, we identified the general considerations presented in Section 4.1. Given this basis, we devised a set of methods to predict the required performance capacity and a first heuristic sizing method presented in Section 4.2 and 4.3, respectively.

4.1 General Considerations and Requirements

Optimization towards a high server utilization. It needs to be addressed that dynamic data centres are designed specifically to allow for a high utilization of the computing resources. Therefore, it is required to reflect the different deployment modes of Section 3. For example,

consider a given number of applications that are to be deployed on a single server in non-exclusive mode. The demanded performance capacity should not be obtained by simply adding together the peak loads of the individual applications' load profiles. In reality, the resulting total peak load will only occur in the very rare worst case situation where all individual applications' peak loads occur at the same point in time in parallel. Usually, relatively simple optimizations, e.g. by orchestrating the individual load profiles on a common time scale in an interlocking mode will lead to better sizing results. In order to achieve this optimization, it is necessary to explore thoroughly the expected load patterns of all applications.

Frequency of re-allocation operations. It is no question that the dynamic allocation of applications to servers also provides some drawbacks which has also been described in [GSWK05]. Each time when an application is re-allocated from one server to another one or put on hold, respectively, some performance capacity is bound to these extra operations. The extra load of these operations may lead to distortions and, possibly, even into an instable state of the data centre. Furthermore, each dynamic re-allocation involves the risk of application service failures, even if the same operation was completed successfully many times in the past¹. In order to prevent these drawbacks, it is necessary to limit the re-allocation frequency. A thorough analysis of the available re-allocation options is necessary which will include a careful prediction of the short term and long term effects of the re-allocation operations. For sizing projects this calls for a starting allocation that does not need to be revised through re-allocation operations in an early stage (i.e. shortly after production start).

Cross application specific aspects. As presented in Section 2, for sizing projects, a sizing risk needs to be considered. For traditional data centres, this sizing risk may be viewed separately for each single server. Due to the fact that in dynamic data centres applications are flexibly deployed on different servers, it is recommended to also look at the data centre as a whole from a risk investigation point of view. That is, for dynamic data centres, the sizing risk needs to include the single-server specific risks but also the cross-servers specific risks. For example, consider the fact that if a server is not sized properly, a high re-allocation frequency is likely to occur. For reasons described above, such a high re-allocation frequency will affect the inappropriately sized server, but the other servers, too.

¹ Consider in this context one of the system administrators' golden rule "Never change a running system".

4.2 Predicting the Required Performance Capacity Based on Load Schedules

The required performance capacity usually presents the key constraint for sizing IT infrastructure components. For dynamic data centres, it is necessary to predict this required performance capacity for multiple interdependent servers. In the following, we present straightforward prediction methods for each of the deployment modes described in Section 3. In our description, by A , we denote the set of applications that are to be deployed on a single host with $A = \{A_1, A_2, \dots, A_n\}$ and A_k denoting a single application with $A_k \in A$ and $k \leq n$. By $S_j(A, t)$, we refer to the aggregated total load profile resulting from a particular orchestration j of individual application load profiles. Considered in an individual orchestration are the load profiles of the applications in A for the time interval t with $t = [t_s, t_e]$. In the following, we refer to such an orchestration by the notion of *schedule*. By $S_{A,t}$, we denote the set of $j=1, \dots, m$ alternative schedules $S_j(A, t)$ that may be orchestrated with respect to A for time interval t . We assume that this orchestration problem may be solved by function $SCHEd(A, t)$ that takes as input the set of corresponding applications A and the time interval t , respectively, and yields the corresponding alternative schedules $S_{A,t}$. In addition to that, we define a function $OPT-SCHEd(S_{A,t})$ that finds within the set of alternative schedules $S_{A,t}$ that schedule $optS_{A,t} \in S_{A,t}$ which leads to the lowest total peak load. The further functions considered in our framework are as follows:

- $L_{\max_app}(A_k, t)$: function that computes the peak load of application A_k wrt. t
- $L_{glob_max_app}(A, t)$: function that finds the max. peak load among the set of applications A and wrt. t
- $L_{\max_sched}(S_j(A, t))$: function that computes the max. peak load of schedule $S_j(A, t)$
- $L_{glob_max_sched}(S_{A,t})$: function that computes the max. total peak load wrt. set of schedules $S_{A,t}$
- $P(L)$: function that computes the performance capacity required to satisfy load L

The three prediction methods described below share a common initial step where a time interval $t=[t_s, t_e]$ is determined. The size of interval t is defined so that the load profiles of all applications given by A are included in t .

Exclusive sharing deployment mode. Recall that in this mode, the set of applications given by A are deployed on the same server but the server is running only one of these applications at a time. It is possible to predict the required performance capacity for the given server in two steps. First, with respect to the $i=1, \dots, m$ applications, the maximum peak load is obtained by $L_{glob_max_app}(A, t)$. For this maximum peak load, the corresponding required performance capacity is determined through $P(L_{glob_max_app}(A, t))$.

Non-exclusive sharing deployment mode. In this mode, the server is shared by several applications at a time. The different long term load profiles of the applications may be orchestrated together so that the peak load of the resulting schedule will be minimal. Based on this general idea, we propose a prediction method that consists of the following steps. First, the set of alternative schedules $S_{A,t}$ is obtained by $SCHED(A, t)$. Then, the schedule $^{opt}S_{A,t}$ is found that leads to the lowest peak load through $OPT-SCHED(S_{A,t})$. In turn, the maximum peak load of schedule $^{opt}S_{A,t}$ is determined by $L_{glob_max_sched}(^{opt}S_{A,t})$ and the corresponding required performance capacity is obtained by $P(L_{glob_max_sched}(^{opt}S_{A,t}))$.

Mixed-mode deployment mode. A server that runs applications in mixed-mode deployment mode, at predefined points in time, will switch from exclusive sharing into non-exclusive sharing and vice versa, respectively. Therefore, for the sizing task both of these deployment modes need to be addressed, for example as follows. First, the set of all applications A is divided into two subsets A_{ex} and A_{ne} , respectively. By A_{ex} , we denote that subset of applications that are to be deployed *exclusively* during a set of time intervals t_{ex} . By A_{ne} , we refer to those applications that are to be deployed *non-exclusively* during a set of time intervals t_{ne} with $t_{ne} = t - t_{ex}$. Then, the prediction method for the exclusive deployment mode is applied to A_{ex} and the prediction method for the non-exclusive deployment mode is applied to A_{ne} . From the resulting two numbers that each express required performance capacity, the larger

value is to be considered as the final performance capacity necessary for the deployment of all mixed-mode applications.

4.3 Towards an Algorithm for Initial Sizing Projects for Dynamic Data Centres

Based on the above described considerations, we devised a first pragmatic sizing approach for dynamic data centres. To present this approach, the definitions given in Section 4.2 are extended as follows:

- α : set of all applications to be deployed with $\alpha = \bigcup_{i=1}^l \alpha_{e_i} \cup \alpha_{se} \cup \alpha_{ne} \cup \alpha_m$ and α_{e_i} the $i=1, \dots, l$ applications that need to be deployed definitely in *exclusive mode*, α_{se} the set of applications that may be deployed in a *special exclusive mode*, α_{ne} the set of applications that may be deployed *non-exclusively*, and α_m the set of applications that may be deployed in *mixed-mode*
- B : base allocation with $B = \{ \langle H_{e_i}, \alpha_{e_i} \rangle, \langle H_{se}, \alpha_{se} \rangle, \langle H_{ne}, \alpha_{ne} \rangle, \langle H_m, \alpha_m \rangle \}$ and $\langle H_{e_i}, \alpha_{e_i} \rangle$ the $i=1, \dots, l$ hosts H_{e_i} on which the $i=1, \dots, l$ applications α_{e_i} are deployed in *exclusive mode*, $\langle H_{se}, \alpha_{se} \rangle$ the single host H_{se} on which the set of applications α_{se} are deployed in a special “*semi-exclusive mode*”, $\langle H_{ne}, \alpha_{ne} \rangle$ the single host H_{ne} on which the set of applications α_{ne} are deployed in *non-exclusive mode*, $\langle H_m, \alpha_m \rangle$ the single host H_m on which the set of applications α_m are deployed in *mixed-mode*
- $S_{\alpha_{se}, t}, S_{\alpha_{ne}, t}, S_{\alpha_m, t}$: sets of alternative schedules for each of the application sets $\alpha_{se}, \alpha_{ne}, \alpha_m$, respectively, computed by function $SCHED(A, t)$
- $^{opt} S_{\alpha_{se}, t} \in S_{\alpha_{se}, t}, ^{opt} S_{\alpha_{ne}, t} \in S_{\alpha_{ne}, t}, ^{opt} S_{\alpha_m, t} \in S_{\alpha_m, t}$: single schedules - computed by function $OPT-SCHED(S_{A, t})$ - where the peak load is minimal

Our approach, that takes the interdependencies between the applications into account, consists of the following steps:

1. The short and long term deployment characteristics of all applications given by α are explored and described in a sizing information repository which will include per

application the corresponding load profile, deployment constraints, and service level requirements.

2. From the sizing information repository, the base allocation B is determined under consideration of all relevant constraints.
3. The servers given in B are sized separately.

For steps 2 and 3, we developed a first version for a heuristic algorithm that looks as follows:

1. Select for each single application in α a proper deployment mode by an evaluation of the sizing information repository under consideration of the following rules. *R1.1*: Consider the *exclusive deployment mode* as a pre-selection for applications that need to run on an own dedicated server. If for such an application the peak load occurs frequently such as shown in the example of Figure 1 (left diagram) select the (strict) exclusive deployment mode. For the converse case, where the peak load occurs only rarely and where substantial periods of idle time exist such as in the other example of Figure 1, select the special exclusive deployment mode. *R1.2*: Select the *mixed-mode deployment mode* for applications that only at some specific points in time need to run on an own dedicated server. *R1.3*: Select the *non-exclusive deployment mode* for applications which may all the time run on a host together with multiple other applications.

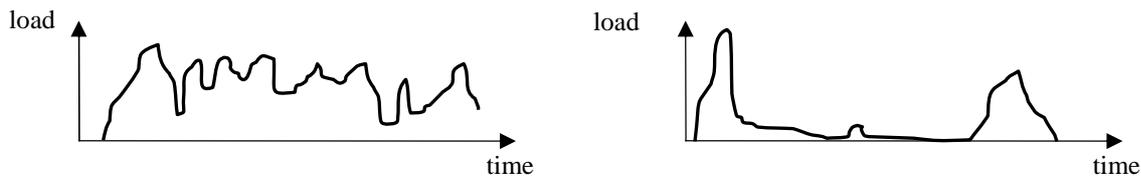


Figure 1: Load profile of an application to be considered for strict exclusive deployment (left side) and special exclusive deployment (right side).

2. Obtain base allocation $B' = \{ \langle H_{e_i}, \alpha_{e_i} \rangle, \langle H_{se}, \alpha_{se} \rangle, \langle H_{ne}, \alpha_{ne} \rangle, \langle H_m, \alpha_m \rangle \}$ being an initial allocation according to the following rules. *R2.1*: Consider a separate dedicated host H_{e_i} for each of the $i=1, \dots, l$ applications that are to be deployed in (strict) exclusive mode. *R2.2*: Consider a single common host H_{se} for all applications together that are

to be deployed in the special exclusive mode. *R2.3*: Choose a single common host H_{ne} for all those applications together that may be deployed non-exclusively. *R2.4*: Choose a single common sever H_m for all those applications together that may be deployed in mixed-mode.

3. Select a time interval t , s.t. the load profiles of all applications given by α_{se}, α_{ne} , and α_m in B' are included in $t=[t_s, t_e]$.
4. Compute the sets of alternative schedules $S_{se,t}, S_{ne,t}, S_{m,t}$ for the application sets α_{se}, α_{ne} , and α_m through the use of function $SCHED(A,t)$. Next, obtain the schedules $^{opt}S_{se,t} \in S_{se,t}, ^{opt}S_{ne,t} \in S_{ne,t}, ^{opt}S_{m,t} \in S_{m,t}$ through the optimization function $OPT-SCHED(S_{A,t})$.
5. Obtain optimized base allocation B from B' by modifying the application sets and schedules. The principles of this optimization approach are shown in Figure 2. It is attempted to close potential gaps in the schedule $^{opt}S_{ne,t}$ and $^{opt}S_{se,t}$, respectively. By “filling such gaps” with fitting load profiles that belong to applications in $^{opt}S_{m,t}$ the peak load of schedule $^{opt}S_{m,t}$ may be reduced. To formulate this optimization principle, we introduce the notion of *transfer operation* denoted by T_i with $T_i = T_i(A_k, \alpha_m, \alpha_{ne}, \alpha_{se})$. We define a transfer operation to move an application $A_k \in \alpha_m$ from its current source host H_m into either α_{ne} or α_{se} of the destination host H_{ne} or H_{se} if the above described optimization criterion is met. This criterion may be formulated as two post conditions for transfer operations as follows:

$$(1) L_{glob_max_sched}(SCHED((\alpha_m - A_k), t)) < L_{glob_max_sched}(SCHED(\alpha_m, t))$$

$$(2) L_{glob_max_sched}(SCHED((\alpha_{ne} \cup A_k), t)) \leq L_{glob_max_sched}(SCHED(\alpha_{ne}, t))$$

or $L_{glob_max_sched}(SCHED((\alpha_{se} \cup A_k), t)) \leq L_{glob_max_sched}(SCHED(\alpha_{se}, t))$.

It is possible that several alternative transfer operations T_i exist, i.e. $i=1, \dots, q$. For these cases, we propose to select that T_i among the q different alternatives where the resulting peak load reduction of schedule $^{opt}S_{m,t}$ in relation to the peak load of the

application A_k reaches the maximum value. That is, that T_i is chosen where $A_k \in \alpha_m$ yields the maximum value among all the alternatives for:

$$\frac{L_{glob_max_sched}(SCHED(\alpha_m, t)) - L_{glob_max_sched}(SCHED((\alpha_m - A_k), t))}{L_{max_app}(A_k, t)}$$

Note that after a transfer operation is completed, update operations are required for the application sets that have been modified. That is, it is necessary to re-compute the sets of alternative schedules and the schedule that leads to the minimal peak load for α_m and also either α_{ne} or α_{se} .

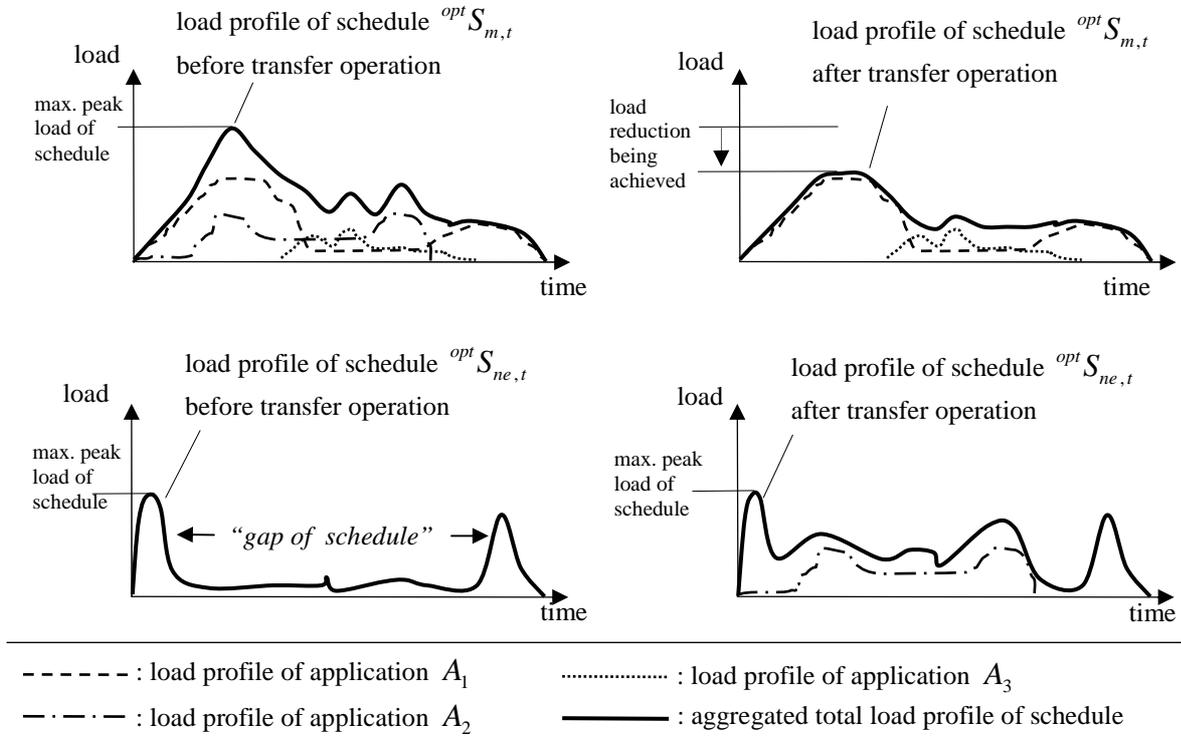


Figure 2: Optimization by transfer operations. The left side shows the global load schedule $S_{m,t}^{opt}$ (upper corner) and $S_{ne,t}^{opt}$ (lower corner), respectively, prior to a transfer operation. The transfer operation will move application A_2 into the gap of the global schedule $S_{ne,t}^{opt}$. As a result, the maximum global load of $S_{m,t}^{opt}$ will be reduced while the maximum load of $S_{ne,t}^{opt}$ will remain unchanged.

6. Predict the required performance capacity of all servers given by the base allocation B through the methods presented in Section 4.2. For the $i=1, \dots, l$ servers H_{e_i} use the

prediction method of the exclusive sharing deployment model. For the servers H_{se} and H_m , respectively, use the method of the mixed-mode deployment model. For the server H_{ne} use the method of the non-exclusive sharing deployment model.

7. Configure each server according to the predicted performance capacity.

5 Related Work

The problem of how to deal from a resource management point of view with the dynamics of a set of applications has been addressed previously. In the AutoGlobe Project of the Technical University Munich [GSWK05], concepts for the static and dynamic allocation of computing services are studied. This project aims at an adaptive computing infrastructure that includes advanced self-management services. In the AutoGlobe approach, a static allocation optimization is proposed that makes use of aggregated historic load data. Similar to our approach, in this optimization it is attempted to allocate services with complementary resource requirements on a common server. However, most research on the problem of resource allocation that can be found in the literature is focused on dynamic allocation techniques, e.g. to deal with overload situations or system errors. In the AutoGlobe Project, a fuzzy controller is proposed that handles such situations by corrective actions that are deduced through a rule based approach under consideration of the actual load situation. Online load measurements for dynamic resource allocation are also considered in [ChGS02]. The load measurements are combined with different prediction and resource allocation techniques in order to dynamically vary the resource shares in shared data centres to the changing workloads of applications. A so called predictive controller and various prediction algorithms for dynamic resource allocation in enterprise data centres are proposed in [XZSW06].

In [ThiK196] an adaptation mechanism is proposed for distributed Multimedia Database Systems that may dynamically adapt concurrent multimedia presentations to fluctuating network bandwidth. This mechanism makes use of the simplex method to globally optimize the adaptations so that the maximum presentation Quality of Service (QoS) is achieved under consideration of the individual user QoS. The difference of our project to these research projects is that we look at the allocation issue for multiple interdependent servers from a sizing perspective.

Our work is also related to system configuration and performance modelling research. In [AbRW01], a systematic method to find a satisfactory hardware and software configuration of a distributed message converter system is presented. Using layered queuing network models, a solution is described that distributes different jobs to different hosts and also configures the processes on the hosts. In addition to this general task that is related to the initial sizing task in our work, we also need to consider the aspect of dynamic resource allocations which was not necessary in this project. A mathematically based method for configuring distributed workflow management systems is proposed in [GWWK00]. This method is targeted at meeting the application's demands in terms of performance and availability while aiming to minimize the total system costs. Similar to our work, it is considered that it may be necessary to adapt the configuration over time due to changes of the workflows. The mathematical core of the proposed method consists of Markov-chain models that are derived from the application's workflow specifications. From these models the overall system's performance is derived. In contrast, we predict the required system performance mainly from the load profiles of the applications. The proposal for a large-scale network parameter configuration method presented in [YeTK02] shares with our approach that efficient parameter state space search techniques are required in order to optimize the allocation of applications to servers. In this related work and also in [XLRX04], finding an optimal configuration is formulated as a black-box optimization problem. For our long term research goal, which is the development of innovative sizing tools for dynamic data centres, we will also evaluate if techniques may be applied to our sizing problem that have been originally developed for configuring mechanical and electronic products [KrHG02].

6 Conclusions

In this paper, we have presented the current status of our research on effective sizing methods for dynamic data centres. We are investigating such approaches in order to develop effective and reliable sizing tools for such data centres in the long run. For the near future, we expect a growing demand for such sizing tools because dynamic data centres are becoming more and more popular.

We presented the main result of our current work status, which is a method for effective initial server sizing. This method leads to a set of hosts with specific deployment characteristics and a

corresponding set of applications that fit to these characteristics. The maximum load that may occur at each of these servers is optimized from a global data centre perspective. Due to the fact that our method makes use of a heuristic algorithm there is no guarantee that our method will yield the global optimum.

In a next step, we will evaluate our algorithm through a simulation study. Based on the simulation results, we will further develop and refine our sizing method. This will include more concrete definitions for the rule-based selection of proper deployment modes. These definitions will also address the concept of service levels qualities. Through the simulation study, we also expect to get insights about the proper dimension of the time interval considered in our algorithm. Our future work will also include a concrete specification of the scheduling functions applied in our algorithm and the functions for the various search tasks such as the identification of gaps in load schedules. We expect that for these issues standard algorithms are readily available or may be adapted to our specific purpose. Furthermore, in our future research, we will extend our sizing method to allow users to guide and to influence the sizing proposal generation. This will include the concept of costs, e.g. for performance capacity. This will also include lower and upper bounds for the number of servers per application class and the performance capacity of the servers. For example, one may use this option to guide the sizing proposal generation towards specific needs and preferences, respectively, defined for the data centre equipment. Moreover, we want to allow that users may influence the destination server that is considered in transfer operations.

References

- [AbRW01] Risse, T., Aberer, K., Wombacher, A., Surridge, M., Taylor, S.: *Configuration of Distributed Message Converter Systems*, Performance Evaluation, Vol. 58, Issue 1, Oct. 2004, Elsevier, pp. 43-80

- [ChGS02] Chandra, A., Gong, W., Shenoy, P.: *Dynamic Resource Allocation for Shared Data Centres Using Online Measurements*, Proc. Quality of Service - IWQoS 2003, 11th Int. Workshop, Berkeley, CA, USA, in Springer LNCS 2707, pp. 381-400

- [GWWK00] Gillmann, M., Weissenfels, J., Weikum, G., Kraiss, A.: *Performance and Availability Assessment for the Configuration of Distributed Workflow Management Systems*, Proc. of the 7th Int. Conf. on Extending Database Technology 2000, Konstanz, Germany, in Springer LNCS 1777, pp. 183-201
- [GSWK05] Gmach, D., Seltzsaam, S., Wimmer, M., Kemper, A.: *AutoGlobe: Automatische Administration von dienstbasierten Datenbankanwendungen*, 17th Int. GI Conf. on Database Systems for Business, Technology, and Web, Karlsruhe, Germany, February 2005
- [KrHG02] Krebs, T., Hotz, L., Günter, A.: *Knowledge-based Configuration of Configuring Combined Hardware/Software Systems*, Proc. 16 Workshop Planen, Scheduling und Konfigurieren (PuK 2002), Freiburg
- [LoMa03] Lober, B., Marquard, U.: *Anwendungs- und Datenbank-Benchmarking im Hochleistungsbereich von ERT-Systemen am Beispiel von SAP*, Datenbank-Spektrum 7/2003, dpunkt.verlag Heidelberg, S. 6-12
- [ThiK196] Thimm, H., Klas, W.: *Delta-Sets for Optimized Reactive Adaptive Playout Management in Distributed Multimedia Database Systems*, Proc. 12th Int. IEEE Conf. on Data Engineering, March 1996, New Orleans, LO, USA, IEEE Computer Society Press, pp. 584-592
- [XLRX04] Xi, B., Liu, Z., Raghavachari, M., Xia, C., Zhang, L. : *A Smart Hill-Climbing Algorithm for Application Server Configuration*, Proc. ACM WWW 2004, May 2004, New York, NY, USA, ACM Press, pp. 287-296
- [XZSW06] Xu, W., Zhu, X., Singhal, S., Wang, Z.: *Predictive Control for Dynamic Resource Allocation in Enterprise Data Centres*, Proc. of the 10th IEEE/IFIP Network Operations & Management Symp. (NOMS 2006), April, 2006, Vancouver, Canada
- [YeTK02] Ye, T., Kaur, H., Kalyanaramann, S.: *Large-scale network parameter configuration using online simulation*. Tech. Rep., Rensselaer Polytechnic Institute, 2002, currently also under review in IEEE Transactions on Networking

Einführung in den Track

Modellierung als Innovationsmotor

Prof. Dr. Ulrich Frank

Universität Duisburg-Essen

Prof. Dr. Robert Winter

Universität St. Gallen

Die Gestaltung und Verwaltung komplexer Systeme erfordert geeignete Abstraktionen. In den Ingenieurwissenschaften ist dies seit langem bekannt. Aber auch in der Betriebswirtschaftslehre werden vielfältige Modelle von Unternehmen eingesetzt, um Gestaltungs- bzw. Veränderungsentscheidungen zu unterstützen. In der Wirtschaftsinformatik kommt Modellen insofern eine besondere Bedeutung zu, als sie nicht nur eine Grundlage für den Entwurf betrieblicher Informationssysteme darstellen, sondern darüber hinaus ein Medium schaffen, um eine zielgerichtete Zusammenarbeit zwischen IT-Experten, Domänenexperten und Anwendern zu unterstützen. Modelle der Unternehmensstrategie, der Geschäfts- und Produktionsprozesse sowie des unterstützenden Informationssystems sind damit wesentliche Voraussetzung für ein effektives IT-Management sowie für die Planung und Realisierung innovativer Formen des IT-Einsatzes.

Der Track ist darauf gerichtet, die zentrale Rolle der Modellierung zu verdeutlichen und ihren angemessenen Einsatz in der Praxis zu fördern. Dazu sollen nicht nur Modellierungskonzepte berücksichtigt werden, sondern auch kritische Erfolgsfaktoren für deren wirtschaftliche Anwendung in der Praxis.

Programmkomitee:

Prof. Dr. Hans-Jürgen Appelrath, Universität Oldenburg

Prof. Dr. Jörg Becker, Universität Münster

Prof. Dr. Werner Esswein, Technische Universität Dresden

Prof. Dr. Ulrich Frank, Universität Duisburg-Essen

Prof. Dr. Norbert Gronau, Universität Potsdam

Prof. Dr. Dimitris Karagiannis, Universität Wien

Prof. Dr. Gerhard Knolmayer, Universität Bern

Prof. Dr. Susanne Leist, Universität Regensburg

Prof. Dr. Heinrich C. Mayr, Universität Klagenfurt

Prof. Dr. Markus Nüttgens, Universität Hamburg

Prof. Dr. Erich Ortner, Technische Universität Darmstadt

Prof. Dr. Michael Rebstock, Fachhochschule Darmstadt

Prof. Dr. Elmar J. Sinz, Universität Bamberg

Prof. Dr. Klaus Turowski, Universität Augsburg

Prof. Dr. Robert Winter, Universität St. Gallen

Konfiguration von Informationsmodellen

Untersuchungen zu Bedarf und Werkzeugunterstützung

Patrick Delfmann, Ralf Knackstedt

European Research Center for Information Systems
Westfälische Wilhelms-Universität Münster
Leonardo-Campus 3, 48149 Münster
{delfmann | knackstedt}@ercis.uni-muenster.de

Abstract

Die Unterstützung der Informationssystemmodellierung durch Modellierungswerkzeuge stellt ein etabliertes Gestaltungsziel der Wirtschaftsinformatik dar. Der vorliegende Beitrag belegt, dass die Anpassung von Informationsmodellen abhängig von ihren Anwendungskontexten eine aufwändige Aufgabe gegenwärtiger Modellierungspraxis darstellt und daher Bedarf zur Anpassungsunterstützung durch Modellierungswerkzeuge besteht. Auf der Basis empirischer Untersuchungen wird eine Klassifikation von Modellvariantenbildungsformen abgeleitet, auf deren Basis eine Evaluation bestehender Modellierungswerkzeuge vorgenommen wird. Die Untersuchung zeigt weit reichende Entwicklungspotenziale bei der Werkzeugunterstützung des Modellvariantenmanagements auf.

1 Einleitung

In nahezu jedem Modellierungsprojekt ergibt sich die Notwendigkeit, Informationsmodelle an die Bedürfnisse verschiedener Nutzergruppen anzupassen. Exemplarisch seien hier verschiedene Unternehmensbereiche genannt, die sich durch unterschiedliche betriebswirtschaftliche Merkmale unterscheiden, verschiedene Einsatzzwecke der Modelle wie z. B. Anwendungssystem- oder Organisationsgestaltung sowie Nutzergruppen, die sich hinsichtlich ihrer Präferenz bzgl. der Modelldarstellung unterscheiden. Die Anpassung dieser Modelle und die sich daran anschließende Pflege der nutzergruppenspezifischen Modelle kann Kosten verursachen, zu de-

ren weitgehender Vermeidung eine methodische Unterstützung durch Modellierungstools wünschenswert erscheint.

Ziel dieses Beitrags ist es, Bedarf für entsprechende Anpassungsmechanismen zu identifizieren sowie auf deren Grundlage eine Klassifikation für Anpassungsmechanismen zu formulieren und zu evaluieren. Weiterhin sind auf dieser Basis existente Toolunterstützungen zu evaluieren und ggf. die Notwendigkeit für die Entwicklung entsprechender Tools zu identifizieren.

Im Folgenden wird zunächst die hierfür verwendete Forschungsmethode vorgestellt (Kapitel 2) und auf verwandte Arbeiten verwiesen (Kapitel 3). Im Anschluss wird eine Klassifikation von Anpassungsmechanismen vorgestellt, die aus Praxiserfahrungen der Autoren abgeleitet wurde (Kapitel 4). Zur Evaluierung dieser Klassifikation werden die Ergebnisse einer fragebogenbasierten Umfrage zur Relevanz der innerhalb der Klassifikation formulierten Anpassungsmechanismen präsentiert (Kapitel 5). Mittels einer Toolevaluation wird überprüft, ob die gemäß Praxiserfahrungen und Umfrageergebnissen relevanten Anpassungsmechanismen durch momentan erhältliche Tools umgesetzt werden (Kapitel 6). Abschließend werden aus den Ergebnissen Forschungsperspektiven abgeleitet (Kapitel 7).

2 Forschungsmethode

In Anlehnung an eine wissenschaftstheoretisch interpretivistische Grundposition [BuMo79; KIMy99] wird für das Vorgehen zur Erkenntnisgewinnung im Rahmen dieses Beitrags eine kritische hermeneutische Forschungsmethode verwendet [Wimm95, S. 883f.].

Hermeneutik ist die Lehre des Verstehens von Sachverhalten durch humane Subjekte [AlSk00, S. 52ff.]. Sie geht stets von einem vorhandenen Vorwissen eines Subjekts über einen Sachverhalt aus. Auf Grundlage des Vorwissens ist das Subjekt in der Lage, durch Analyse des Sachverhalts diesen neu zu interpretieren [KIMy99, S. 71]. Interpretationen führen dazu, das Wissen über den Sachverhalt sukzessive den Ergebnissen der Interpretationen anzupassen. Ein geändertes Wissen bzgl. eines Sachverhalts erzeugt neue (Forschungs-)Fragen und führt wiederum zu Diskussionen und weiteren Analysen des Sachverhalts. Es entsteht ein Kreislauf, der als *hermeneutischer Zirkel* bezeichnet wird [AlSk00, S. 50ff.].

Hermeneutische Forschungsmethoden sehen aufbauend auf dem für die Hermeneutik charakteristischen Verstehensprozess im hermeneutischen Zirkel ein „durch theoretische und praktische Argumentationen gesichertes, systematisch-kritisches Verstehen von sprachlichen und nicht-

sprachlichen Handlungen und der ihnen zugrunde liegenden Zwecke, Regeln und Normen sowie deren Endzwecke“ vor [Wimm95, S. 883]. Eine hermeneutische Methode ist kritisch, wenn sie „sowohl die von ihr selbst vorgeschlagenen Argumentations- und Interpretationsnormen und -verfahren für Kritik, vor allem auch von Seiten des zu deutenden Sachverhalts, offen hält, als auch die Meinungen und Meinungssysteme, die theoretischen Aussagen und die Zwecke und Zwecksysteme, die praktischen Orientierungen zu Grunde liegen, einer kritischen Beurteilung unterzieht“ [Wimm95, S. 883f.].

Die Anforderungen, die im Rahmen einer kritischen hermeneutischen Forschungsmethode gestellt werden, werden im Folgenden auf den konkreten Anwendungsfall dieses Beitrags bezogen (vgl. auch Abbildung 1).

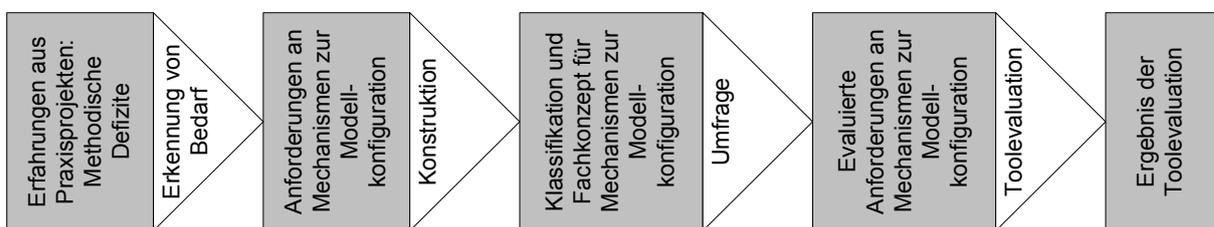


Abbildung 1: Forschungsmethode

- Ausgehend von Beobachtungen des praktischen Umfelds sowie dem Vorwissen der Autoren werden *Defizite* im Rahmen der methodischen Unterstützung der Modellkonfiguration aufgedeckt.
- Der erste Lösungsvorschlag in Form einer *Klassifikation* und eines *Fachkonzepts* ist die Folge der Auswertung der *Anforderungen*, die in Forschung und Praxis gestellt werden, sowie deren Kombination mit Ideen der Autoren, die sich aus dem Vorverständnis der Disziplin der Informationsmodellierung und den Anforderungen ergeben.
- Die anschließende kritische Evaluierung wird durch die Durchführung einer *fragebogenbasierten Umfrage* vollzogen, in der die innerhalb der Klassifikation formulierten Mechanismen zur Modellkonfiguration einem kritischen Publikum aus Forschung und Praxis zur Verfügung gestellt werden und im Hinblick auf ihre Relevanz und ihren Nutzen hinterfragt werden. Als Ergebnis der Umfrage liegt eine *evaluierte Menge von Anforderungen* vor, auf deren Grundlage ggf. eine Rekonstruktion des Fachkonzepts vollzogen werden kann.

- Mit dem Ziel, zu ermitteln, ob eine Unterstützung der Konfigurationsmechanismen bereits innerhalb von Modellierungstools vorliegt, wird eine *Toolevaluation* durchgeführt. Als *Ergebnis* kann erstens beurteilt werden, ob die Konstruktion eines Modellierungstools, das die Modellkonfiguration beherrscht, noch aussteht. Zweitens kann gezeigt werden, ob evtl. vorhandene Toolunterstützungen von den befragten Personen bereits wahrgenommen bzw. als praktikabel beurteilt werden.

3 Verwandte Arbeiten

Die methodische Unterstützung der Pflege, Verwaltung und Generierung von Modellvarianten wird mit unterschiedlichen Schwerpunktsetzungen in mehreren Arbeiten untersucht. SOFFER, GOLANY und DORI stellen konfigurierbare Referenzmodelle mit dem Ziel vor, ein modellbasiertes Customizing von ERP-Systemen zu ermöglichen. Im Rahmen des Customizing wird dabei der jeweilige Anwendungskontext anhand von Attributausprägungen spezifiziert, worauf die Modelle gemäß den Attributausprägungen zugeordneten Anwendungsszenarien angepasst werden [SoGD03].

Im Gegensatz zu dem Ansatz von SOFFER, GOLANY und DORI sehen ROSEMANN und VAN DER AALST in ihrer Arbeit zur Erweiterung Ereignisgesteuerter Prozessketten (EPK, vgl. [KeNS92]) nicht die Formulierung von Konfigurationsregeln vor, die in Abhängigkeit betriebswirtschaftlicher Kriterienausprägungen festlegen, welche Modellvariante zu wählen ist. Stattdessen unterstützen sie die Modifikation der Prozessmodelle durch die Spezifikation von Regeln, mit welchen sich die Abhängigkeiten zwischen den Prozessmodellelementen explizieren lassen. Trifft ein Modellersteller z. B. die Entscheidung, bestimmte Funktionen aus einem Modell zu entfernen, leitet ihn die Regelbasis dabei an, auch von diesen Funktionen abhängige Modellbereiche korrekt anzupassen [RoAa07]. Gemeinsam ist beiden Beiträgen, dass das Customizing von ERP-Systemen als Einsatzgebiet adressiert wird.

Der Ansatz von BECKER ET AL. zur konfigurativen Referenzmodellierung zeichnet sich im Vergleich zu den anderen genannten Ansätzen dadurch aus, dass er nicht an eine bestimmte Modellierungstechnik gebunden ist. Stattdessen wird ein Framework zur Unterscheidung verschiedener Konfigurationsmechanismen präsentiert, die eine regelbasierte Anpassung von Modellen und deren Modellierungssprachen an durch Konfigurationsparameter spezifizierte Anwendungskontexte erlauben. Die regelbasierte Konfiguration stellt eine Gemeinsamkeit mit dem

Ansatz von SOFFER, GOLANY und DORI dar. Die Konfigurationsparameter werden allerdings sowohl sprachunabhängig spezifiziert als auch exemplarisch auf verschiedene Modellierungstechniken angewendet [BDKK02; Knac06; Delf06].

Alle vorgestellten Konfigurationsansätze haben gemeinsam, dass die unterstützten Modellvarianten in einem Gesamtmodell integriert werden. Bis auf den Beitrag von ROSEMANN und VAN DER AALST verfügen alle Ansätze über eine Regelbasis, mit deren Hilfe für einen gegebenen Anwendungskontext die am besten passende, im Gesamtmodell enthaltene Variante identifiziert werden kann.

Für die Klassifikation von Konfigurationsmechanismen erfolgt in diesem Beitrag eine Anlehnung an den Ansatz von BECKER ET AL., da dieser sich durch eine besonders umfassende und detaillierte Unterscheidung von Konfigurationsmechanismen für unsere Zwecke besonders eignet.

4 Anforderungen aus der Praxis an die Informationsmodellkonfiguration

In mehreren Praxisprojekten, die u. a. insbesondere bei der *DeTe Immobilien GmbH* [BeKR05], in *Kommunalverwaltungen des Landes Nordrhein-Westfalen* [BADN03], der *Bayer Business Services GmbH* [BJDF06], bei der *Deutschen Bundeswehr* sowie in Zusammenarbeit mit der *itemis GmbH & Co. KG* durchgeführt worden sind, ist im Rahmen von Modellierungsaktivitäten von Modellnutzern die Notwendigkeit geäußert worden, Modelle nutzergruppenspezifisch anzupassen. Die Anpassungen bezogen sich dabei auf die Modellierungssprache, den Modellinhalt sowie die grafische Aufbereitung der Modelle.

Als Ursache der Anpassungsnotwendigkeit wurden einerseits betriebswirtschaftliche Merkmale von Unternehmen(sbereichen), so genannte *Unternehmensmerkmale*, identifiziert. Andererseits führten modellierungszweckbezogene, organisatorisch rollenbezogene sowie persönliche Präferenzen der Nutzergruppen zur Notwendigkeit der Modellanpassung. Letztere drei Aspekte lassen sich unter dem Begriff *Perspektive* zusammenfassen (vgl. hierzu ausführlich [BDKK02, S. 27ff.; Delf06, S. 53ff.; Knac06, S. 213ff.]).

Das folgende Anpassungsbeispiel soll verdeutlichen, welche Anpassungsarten in den Praxisprojekten gefordert waren. Es wird ein exemplarischer stufenweiser Anpassungsprozess dargestellt, der die verschiedenen Aspekte der Modellanpassung vereint (vgl. Abbildungen 2 und 3).

Der als EPK dargestellte Prozess der Rechnungsprüfung ist für ein Handelsunternehmen geeignet, das verschiedene Geschäftsarten (Lager-, Strecken- und Zentralregulierungsgeschäft) durchführt. Soll dieses Prozessmodell Unternehmensbereichen zur Verfügung gestellt werden, in denen ausschließlich Strecken- und Lagergeschäft durchgeführt wird (z. B. in einer Niederlassung), sind Aspekte des Zentralregulierungsgeschäfts aus dem Prozessmodell zu entfernen (Anpassungsschritt 1→2, vgl. grau schattierte Elemente in 1).

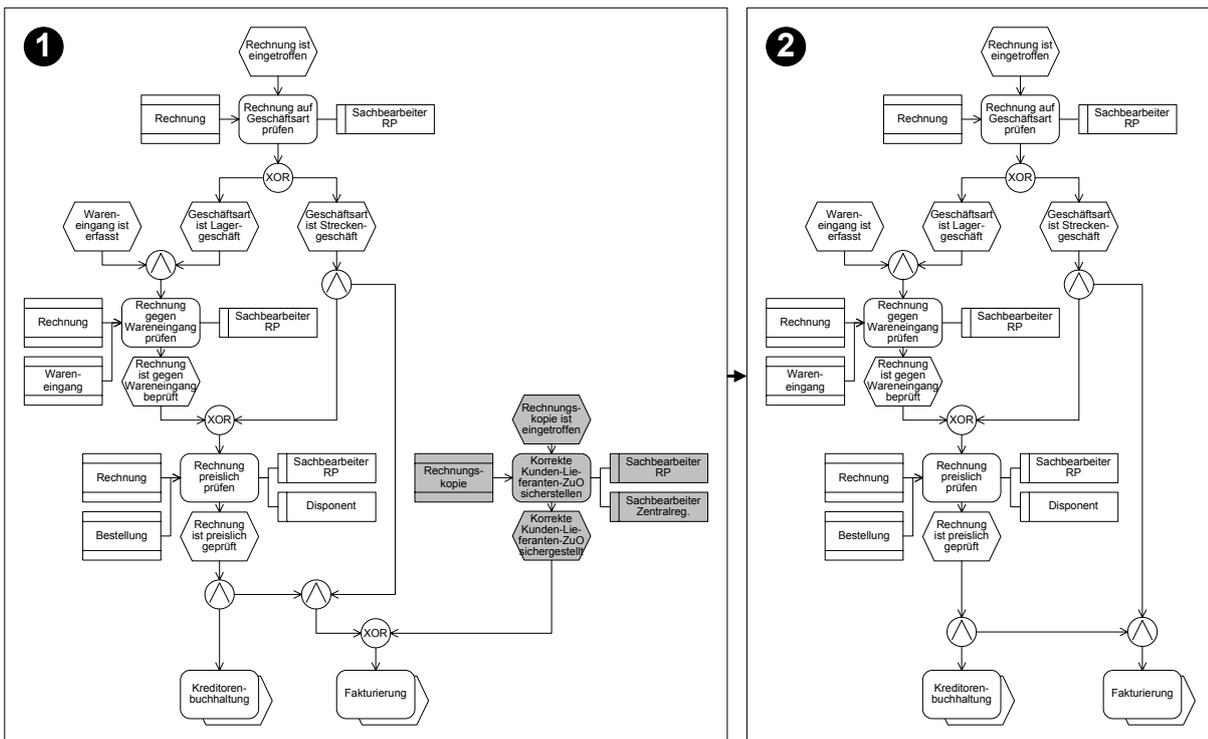


Abbildung 2: Exemplarischer Anpassungsprozess – Teil 1

Neben betriebswirtschaftlich-inhaltlichen Ursachen für die Anpassung von Informationsmodellen wird häufig gefordert, bspw. zu Schulungszwecken, Übersichtsmodelle zur Verfügung zu stellen, womit eine besonders gute Lesbarkeit angestrebt wird. Bspw. kann sich dies im Entfernen von sämtlichen an Funktionen annotierten Ressourcenobjekten äußern (Anpassungsschritt 2→3, vgl. grau schattierte Elemente in 2).

Ein weiteres vielfach gefordertes Instrument zur Übersichtsbildung ist das Entfernen von so genannten *Trivialereignissen* aus EPKs. Als *Trivialereignisse* werden solche Ereignisse bezeichnet, die nicht an Prozessverzweigungen beteiligt bzw. keine Start- oder Endereignisse sind (Anpassungsschritt 3→4, vgl. grau schattierte Elemente in 3).

Nutzergruppenspezifische Änderungen von Begriffen innerhalb von Modellelementbezeichnungen werden dann gefordert, wenn sich z. B. Begriffskonventionen in unterschiedlichen Unternehmensbereichen herausgebildet haben. Bspw. wird ein Sachbearbeiter aus der Rechnungs-

prüfung hinter der Bezeichnung „Rechnung“ stets eine Kreditorenrechnung vermuten. Umgekehrt ist für fachfremde Modellnutzer – bspw. Methodenexperten – zu explizieren, um welchen Typ von Rechnung es sich handelt (Anpassungsschritt 4→5, vgl. grau schattierten Text in 4). Die grafische Aufbereitung von Informationsmodellen kann dazu beitragen, Akzeptanzschwierigkeiten bzgl. formaler Modelle zu überwinden. Piktogramme anstatt Polygone als Modellelemente werden bspw. von Nicht-Methodenexperten als besonders anschaulich empfunden (Anpassungsschritt 5→6).

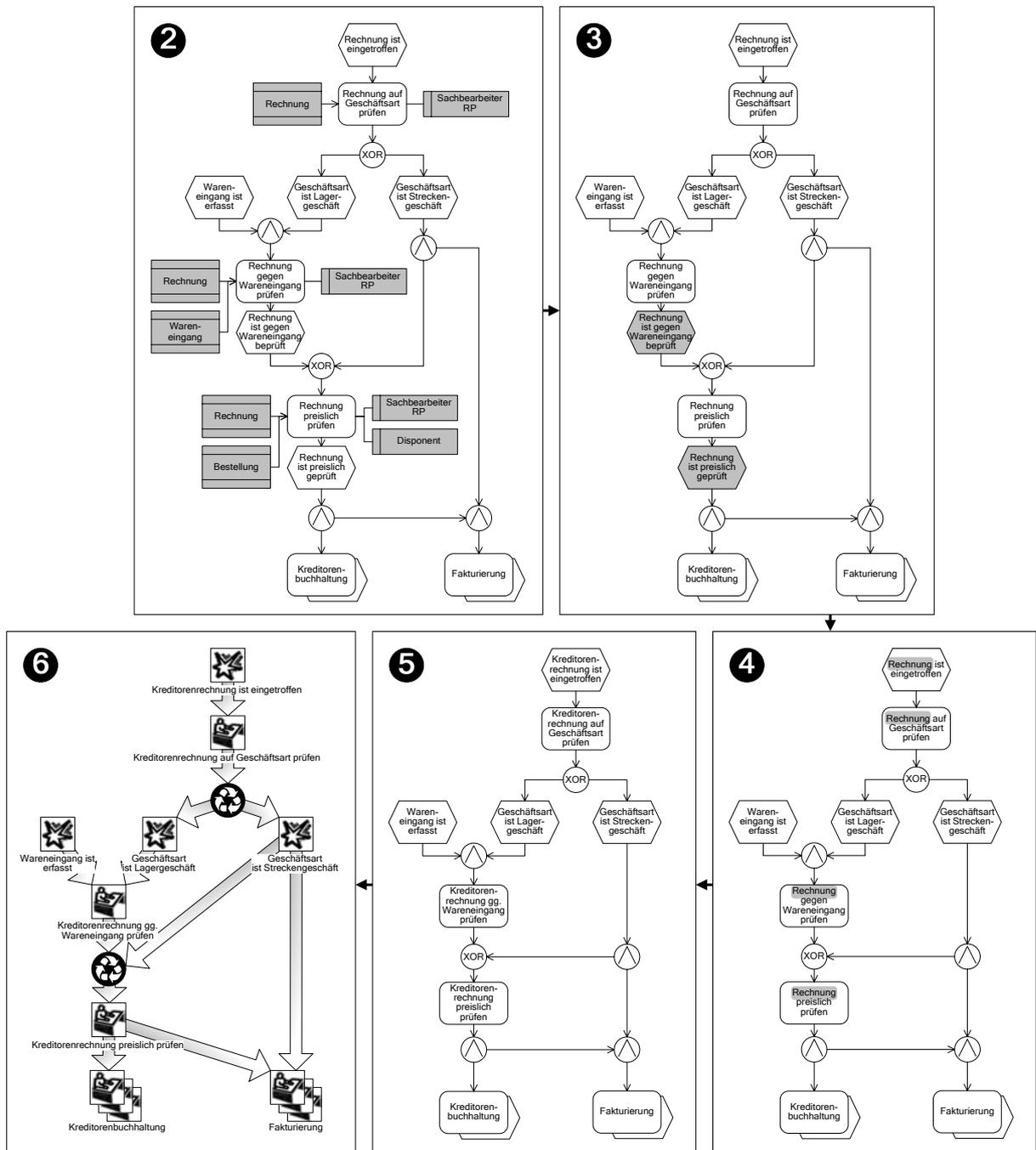


Abbildung 3: Exemplarischer Anpassungsprozess – Teil 2

Eine konventionelle Anpassung von Informationsmodellen an die Anforderungen unterschiedlicher Nutzergruppen – wie im Beispiel gezeigt – erforderte die Erstellung separater Modelle und zöge damit erhöhte Konstruktions- und Wartungskosten nach sich. Um diese Kosten zu verhindern bietet es sich an, die für sämtliche zu unterstützenden Nutzergruppen relevanten Informationen in einem Gesamtmodell zu integrieren und die einzelnen nutzergruppenspezifischen Modelle durch Ausblendung nicht relevanter Sachverhalte bzw. durch standardisierten Austausch von grafischen oder textuellen Aspekten automatisiert zu erzeugen. Ausgehend von den Anforderungen werden fünf Kategorien von Anpassungsmechanismen, so genannte Konfigurationsmechanismen, eingeführt (Die formale Spezifikation dieser Mechanismen findet sich in mehreren diesem Beitrag vorangegangenen Arbeiten. Vgl. exemplarisch [BDKK02; Delf06; Knac06]):

- *Modelltypselektion*: Modelltypen repräsentieren Ergebnistypen spezieller Modellierungssprachen. Die Relevanz der Modelltypen kann von der jeweiligen Nutzergruppe abhängig sein. Bspw. werden im Rahmen der Organisationsgestaltung zur Repräsentation von Daten Fachbegriffsmodelle gegenüber Entity-Relationship-Modellen bevorzugt. Umgekehrte Präferenzen sind bei Anwendungssystemgestaltern zu beobachten. Daher unterstützt der Konfigurationsmechanismus die Zuordnung von Modelltypen zu Nutzergruppen und damit eine grobgranulare, sprachbezogene Konfiguration des Modellsystems.
- *Elementtypselektion*: Elementtypselektionen ermöglichen es, zu Modelltypen Varianten zu bilden, die sich in der Menge der verwendbaren Modellelementtypen unterscheiden. Je nach Nutzergruppe kann es erforderlich sein, Modellelemente, die zu einem bestimmten Elementtyp gehören, auszublenden (vgl. Anpassungsschritt **2→3**). Elementtypselektionen ermöglichen solche Ausblendungen und stellen wie Modelltypselektionen sprachbezogene Konfigurationen des Modellsystems dar.
- *Elementselektion*: Elementselektionen äußern sich in der Selektion von einzelnen Instanzen von Modellelementtypen, bspw. einer einzelnen Prozessmodellfunktion „Korrekte Kunden-Lieferanten-Zuordnung sicherstellen“ (vgl. Anpassungsschritt **1→2**) bzw. von einzelnen Trivialereignissen (vgl. Anpassungsschritt **3→4**). Auf Basis dieser Selektion können einzelne Elemente ausgeblendet werden, die für die entsprechende Nutzergruppe nicht relevant sind. Elementselektionen sind damit im

Gegensatz zu Modelltyp- und Elementtypselektionen nicht sprach- sondern modellbezogene Konfigurationsmechanismen.

- *Bezeichnungsvariation*: Der Mechanismus Bezeichnungsvariation berücksichtigt, dass es erforderlich sein kann, die Bezeichnungen von Modellelementen in Abhängigkeit von Nutzergruppen auszutauschen (vgl. Anpassungsschritt 4→5).
- *Darstellungsvariation*: Die Darstellungsvariation ermöglicht die Zuordnung unterschiedlicher Repräsentationsformen zu Modellaspekten. Zum einen wird hiermit der Austausch von Modellelementsymbolen ermöglicht (vgl. Anpassungsschritt 5→6). Zum anderen werden zudem von unterschiedlichen Nutzergruppen unterschiedliche Modelltopologien bevorzugt. Beispielsweise werden teils Entity-Relationship-Modelle bevorzugt, deren Elemente nach dem Grad der Existenzabhängigkeit angeordnet sind, teils solche, deren Kantenüberschneidungsgrad minimal ist. Die Darstellungsvariation berücksichtigt auch diesen Aspekt.

Die nutzergruppenabhängigen Ursachen, die zur Konfiguration führen und die in *Unternehmensmerkmale* und *Perspektiven* unterschieden werden (s. o.), dienen den Konfigurationsmechanismen als Input und werden unter dem Begriff *Konfigurationsparameter* zusammengefasst.

5 Umfrage

5.1 Untersuchungsdesign

Die aus Praxiserfahrungen abgeleitete Klassifizierung von Konfigurationsmechanismen wurde zum Gegenstand einer fragebogenbasierten Umfrage gemacht. Untersucht werden sollte,

- ob Vertreter aus Praxis und Forschung das Management von Modellvarianten und die vorgestellten Mechanismen zur Modellkonfiguration als relevant erachten und
- für wie aufwändig die Pflege von Modellvarianten vor dem Hintergrund der derzeitigen Werkzeugunterstützung eingeschätzt wird.

Mit der Umfrage wurde kein Anspruch auf Repräsentativität hinsichtlich einer Grundgesamtheit verfolgt. Zur Teilnahme an der Untersuchung wurden 177 Wissenschaftler eingeladen, die auf den für die Wirtschaftsinformatik führenden Konferenzen zu einschlägigen Themen publiziert

haben. Außerdem wurden 170 Mitglieder des Networking-Portals www.openbc.de, die entsprechende Interessensgebiete genannt haben, gefragt, ob sie an der Studie teilnehmen möchten. Nach entsprechender Interessenäußerung wurden Zugangsdaten für die Befragung an insgesamt 28 Portalnutzer versandt. Insgesamt haben letztlich 14 der direkt eingeladenen Wissenschaftler und 14 Mitglieder des Networking-Portals an der Studie teilgenommen (wobei sich einer der Portalnutzer als Wissenschaftler einordnete) und den über das Umfrageportal 2ask bereitgestellten und auf geschlossenen Fragen basierenden Fragebogen bearbeitet. Die Teilnehmer verteilen sich entsprechend der gemischten Akquisitionstrategie auf unterschiedliche Berufszweige der Forschung und Praxis (14 x Hochschule, 1 x sonstige Forschungsorganisation, 4 x Industrieunternehmen, 7 x Beratung, 1 x sonstiges Dienstleistungsunternehmen, 1 x Sonstige).

5.2 Ausgewählte Untersuchungsergebnisse

Die Ergebnisse der Studie zeigen, dass sowohl Unternehmensmerkmale als auch Perspektiven überwiegend als relevante Auslöser für die Erstellung von Modellvarianten angesehen werden. Der Aufwand zur Erstellung entsprechender unternehmensmerkmals- bzw. perspektivenspezifischer Modellvarianten wird überwiegend als hoch angesehen. Die Teilnehmer aus Forschung und Praxis kommen dabei zu annähernd gleichen Einschätzungen (vgl. Abbildung 4).

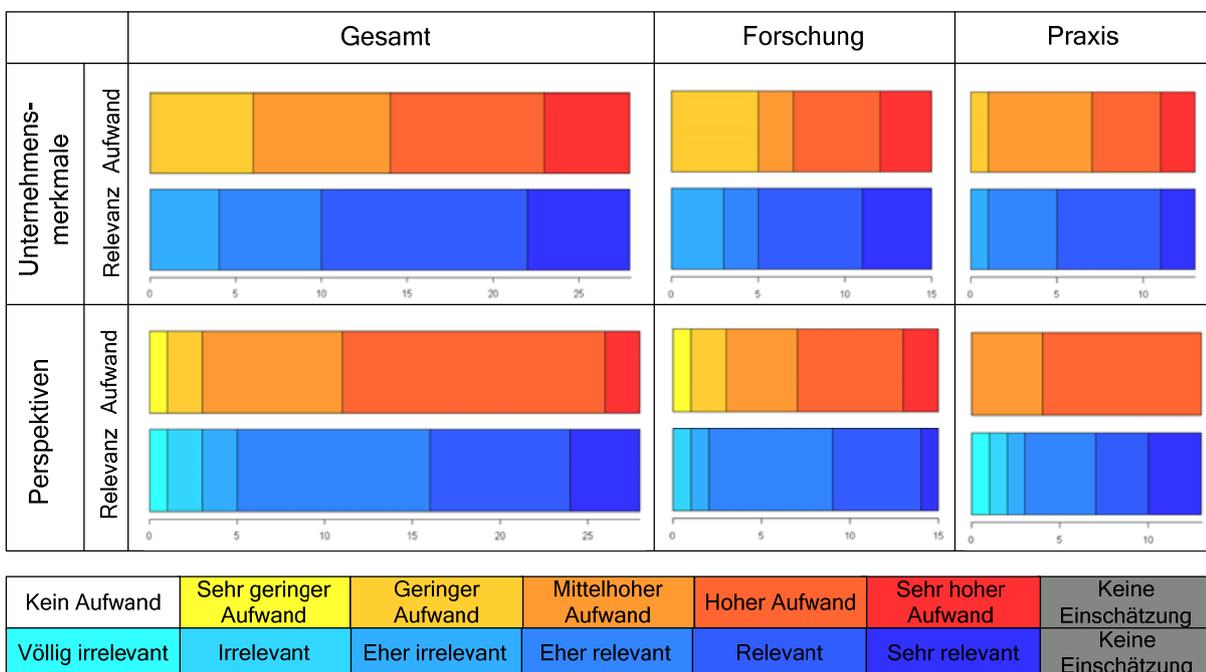


Abbildung 4: Aufwand und Relevanz unternehmensmerkmals- und perspektivenspezifischer Modellvarianten

Die verschiedenen in den Vorarbeiten identifizierten Variantenbildungsformen Modelltypselektion bis Darstellungsvariation werden von den Befragten alle überwiegend als relevant zur Bil-

derung unternehmensmerkmals- und perspektivenspezifischer Modellvarianten angesehen. In dieser Kernaussage stimmen die Einschätzungen der Teilnehmer aus der Praxis und Forschung überein (vgl. Abbildung 5).

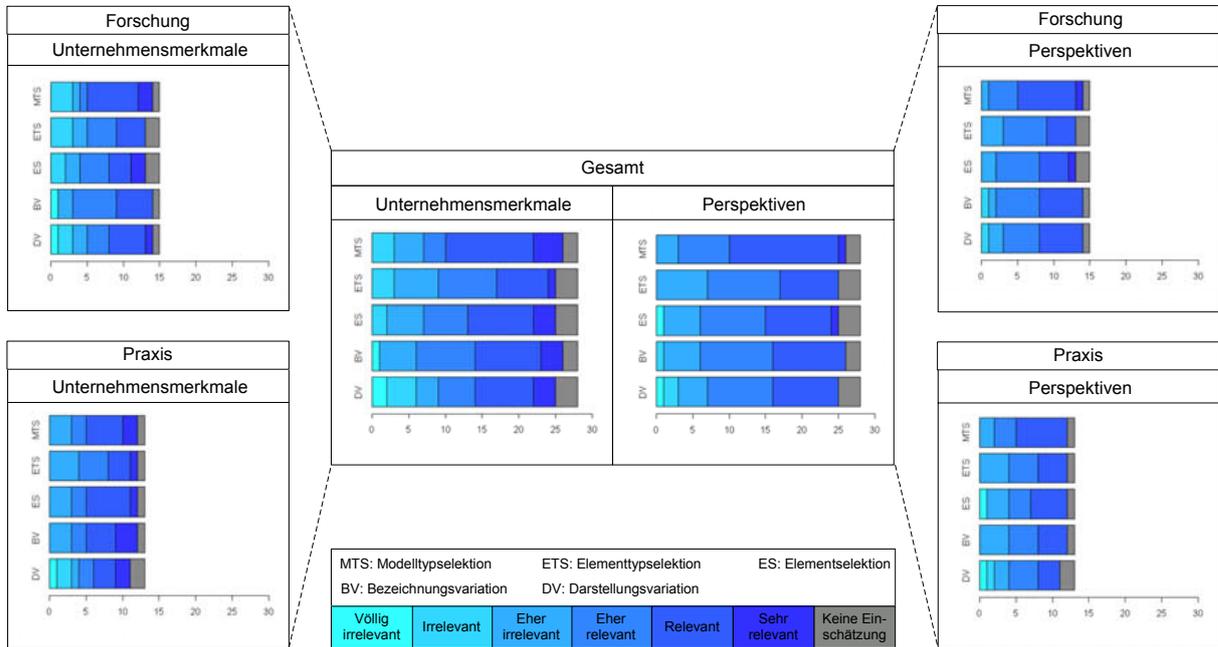


Abbildung 5: Relevanz der Variantenformen

Während die Teilnehmer den Stellenwert der Variantenbildungsformen bestätigen, schätzen sie gleichzeitig den Aufwand, der vor dem Hintergrund der derzeit zur Verfügung stehenden Modellierungswerkzeuge bei der Pflege von Modellvarianten betrieben werden muss, verhältnismäßig hoch ein. Auch hierbei sind die Einschätzungen von Forschung und Praxis nicht sehr unterschiedlich (vgl. Abbildung 6).

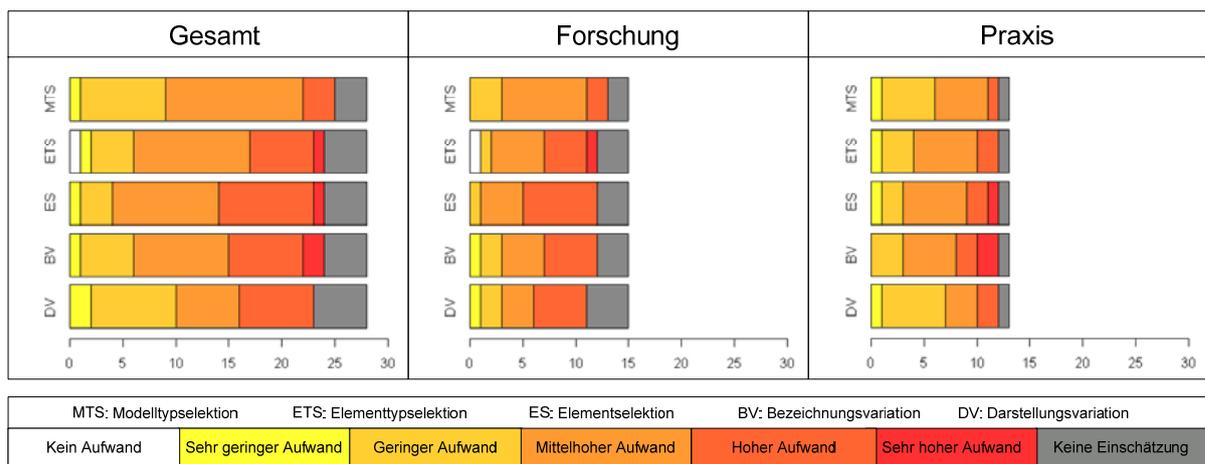


Abbildung 6: Einschätzung des Aufwandes zur Pflege unterschiedlicher Variantenformen

Hinsichtlich der verschiedenen Variationsformen von Modellen wurden die Teilnehmer auch nach ihrer Einschätzung gefragt, mit welchem Modellierungswerkzeug sich die Pflege entspre-

chender Modellvarianten zurzeit am besten realisieren lässt. Das Modellierungswerkzeug ARIS TOOLSET wurde dabei hinsichtlich aller Variantentypen am häufigsten genannt. Auffällig an dem Ergebnis ist allerdings auch, dass den Teilnehmern sehr häufig gar kein geeignetes Modellierungswerkzeug bekannt war (vgl. Abbildung 7).

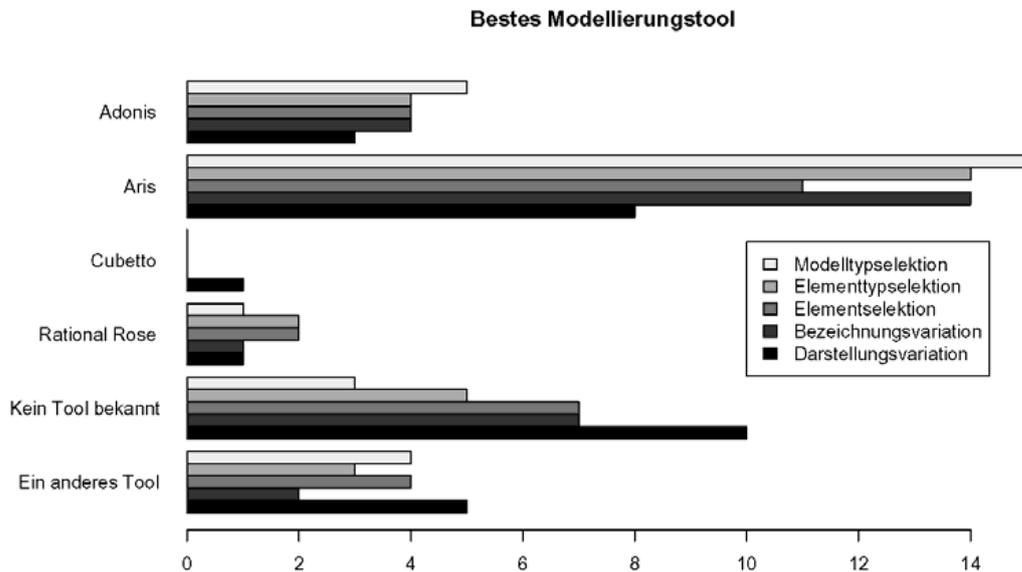


Abbildung 7: Von den Studienteilnehmern als beste Unterstützung der jeweiligen Modellvariantentypen genannte Modellierungswerkzeuge

6 Toolevaluation

6.1 Untersuchungsdesign

Ausgehend von den Ergebnissen der Befragung, in deren Rahmen sämtliche vorgeschlagene Konfigurationsmechanismen als relevant eingestuft wurden, sind unterschiedliche Modellierungstools (ARIS, CASEWISE, CONCEPTBASE, CUBETTO, GENGRAPH, METIS und SEMTALK) hinsichtlich ihrer Konfigurationsunterstützung untersucht worden. Neben der Unterstützung der vorgeschlagenen Konfigurationsmechanismen wurden Möglichkeiten und Komfort der Spezifikation und Abfrage von Konfigurationsparametern untersucht.

Es wurden insbesondere *Metamodellierungstools* betrachtet, da die in diesen Tools zur Verfügung gestellte Metamodellierungsumgebung eine viel versprechende Spezifikationsmöglichkeit für Konfigurationsmechanismen darstellt. Neben dieser Toolgruppe wurde das Produkt des Marktführers auf dem Gebiet der Modellierungstools, das ARIS TOOLSET, mit in die Untersuchung einbezogen, da dieses Tool laut Befragung als für die Konfigurationsunterstützung am besten geeignet empfunden wird. Generell sind nur solche Tools evaluiert worden, für die eine

kostenlose Testversion zur Verfügung stand bzw. die den Autoren durch bereits vorhandene Universitätslizenzen ohnehin zur Verfügung standen.

6.2 Untersuchungsergebnisse

Die Ergebnisse der Toolevaluation sind in Tabelle 1 zusammengefasst. Unter den evaluierten Tools existiert keines, das die Modellkonfiguration laut Anforderungen aus der Praxis sowie der Befragung unterstützt.

Möglichkeiten der Spezifikation von Konfigurationsparametern	Aris	Casewise	ConceptBase	Cubetto	GenGraph	Metis	SemTalk
Spezifikation von Unternehmensmerkmalen	●	○	●	●	○	○	○
Spezifikation von Perspektiven (Modellierungszweck)	●	○	●	●	○	○	○
Spezifikation von Perspektiven (organisatorische Rollen)	●	○	●	●	○	○	○
Spezifikation von Perspektiven (persönliche Präferenzen und sonstige Einflüsse)	●	○	●	●	○	○	○
Definition ausgewählter Kombinationen von Merkmalsausprägungen zur Konfigurationsparameterspezifikation	●	○	●	●	○	○	○
Möglichkeiten der Abfrage gültiger Konfigurationsparameterausprägungen	Aris	Casewise	ConceptBase	Cubetto	GenGraph	Metis	SemTalk
Abfrage der Gültigkeit einer einzelnen Merkmalsausprägung (z. B. <i>Perspektive</i> = „ <i>Methodenexperte</i> “)	○	○	●	●	○	○	○
Abfrage der Gültigkeit mehrerer UND-verknüpfter Merkmalsausprägungen	○	○	●	●	○	○	○
Abfrage der Gültigkeit mehrerer beliebig logisch verknüpfter Merkmalsausprägungen	○	○	●	●	○	○	○
Übersichtsdarstellung über Merkmale und ihre Ausprägungen (z. B. in Form von morphologischen Kästen)	○	○	●	●	○	○	○
Spezifikation von Beziehungen zwischen verschiedenen Merkmalsausprägungen (z. B. gegenseitiger Ausschluss)	○	○	●	●	○	○	○
Unterstützte Konfigurationsmechanismen	Aris	Casewise	ConceptBase	Cubetto	GenGraph	Metis	SemTalk
Modelltypselektion	●	●	●	●	●	○	○
Elementtypselektion	●	○	●	●	●	○	●
Elementselektion	○	○	●	●	○	○	○
Bezeichnungsvariation	○	○	●	●	○	●	○
Darstellungsvariation der Symbole	●	●	●	●	○	●	○
Darstellungsvariation der Modelltopologie	●	○	●	●	○	●	○
Legende:							
● Unterstützung vorhanden							
● Unterstützung durch Erstellung eigener Makros in der Skriptsprache des Tools (teilweise) möglich							
○ Unterstützung nicht möglich							

Tabelle 1: Ergebnisse der Toolevaluation

Im Einzelnen weisen die untersuchten Tools hinsichtlich ihrer Konfigurationsunterstützung folgende Besonderheiten auf:

- *ARIS*: Das ARIS TOOLSET ist grundsätzlich nicht auf die Modifikation von Eigenschaften der bereitgestellten Modellierungstechniken ausgelegt, so auch nicht für die Erweiterung um Konfigurationsmechanismen. Eine Metamodellierungsumgebung zur Unterstützung einer solchen Modifikation ist in ARIS nicht vorgesehen. Die in ARIS mitgelieferte Skriptsprache erlaubt rudimentäre Modifikationen von geöffneten Modellen, so auch die grafische Aufbereitung durch unterschiedliche Zuweisung von Symbolen zu Modellelementen sowie die Zuweisung von beliebig komplexen Konfigurationsparametern. Sie ist jedoch zur Abfrage von Konfigurationsparametern mit dem Ziel der Konfiguration nicht geeignet. Basiskonfigurationen werden in ARIS über den so genannten Methodenfilter zugelassen, der benutzerspezifisch die Bereitstellung von Modell- und Elementtypen verwaltet. Topologievariationen werden in ARIS durch unterschiedliche Modelltypen zugelassen, die durch Transformation ineinander überführt werden können.
- *casewise*: Dieses Metamodellierungstool erlaubt die Spezifikation von Konfigurationsattributen und deren Ausprägung, jedoch reicht die Skriptsprachenfunktionalität zur Automatisierung dieser Aufgaben nicht aus. Gleiches gilt für die Abfrage von Konfigurationsparametern sowie die Spezifikation von Konfigurationsmechanismen. Modelltypselektionen werden ähnlich wie in ARIS durch einen Filter zugelassen. Die Modifikation von Symbolrepräsentationen ist möglich, jedoch nicht im Rahmen eines automatisierten Konfigurationsmechanismus.
- *ConceptBase*: CONCEPTBASE ist ein sehr generisch gehaltenes Metamodellierungstool, das Spezifikationen von Modellierungssprachen mithilfe einer mitgelieferten Programmierumgebung vorsieht. Aufgrund der Ausdrucksstärke der eingesetzten Sprache lässt sich in CONCEPTBASE nahezu jede Modellmodifikation vorsehen. D. h., dass sämtliche Kriterien der Konfigurationsunterstützung grundsätzlich von CONCEPTBASE erfüllt werden können, sofern die Mechanismen zur Spezifikation und Abfrage von Konfigurationsparametern sowie die Konfigurationsmechanismen programmiert werden. Bei der Evaluation ist aufgefallen, dass die geringe Benutzerfreundlichkeit von CONCEPTBASE einen praktischen Einsatz erheblich erschwert.
- *Cubetto*: Ein ähnliches Evaluationsergebnis wie CONCEPTBASE liefert die Untersuchung von CUBETTO. Ebenso wie CONCEPTBASE ist CUBETTO nicht auf die Konfi-

guration von Modellen ausgerichtet. Jedoch ist Erfüllung nahezu sämtlicher Konfigurationsanforderungen durch Eigenprogrammierung möglich, da eine entsprechend mächtige Skriptsprache mitgeliefert wird. Im Vergleich zu CONCEPTBASE bietet CUBETTO eine in geringem Maße erhöhte Nutzerfreundlichkeit.

- *GenGraph*: Das noch relativ junge Metamodellierungstool GENGRAPH befindet sich im prototypischen Entwicklungsstadium und ist aus diesem Grund lediglich mit rudimentären Funktionalitäten ausgestattet – so z. B. mit einer Modelltyp- und Elementtypselektion, die ähnlich wie im ARIS TOOLSET gehandhabt wird.
- *Metis*: In METIS ist durch die Verwendung von Skripten die Realisierung der Darstellungsvariation der Symbole sowie rudimentär die der Topologie möglich. Darüber hinaus lassen sich ebenfalls per Skript Bezeichnungsaustauschtabellen definieren, die eine Bezeichnungsvariation erlauben. Weitere Modellmodifikationen sowie Möglichkeiten zur Spezifikation und Abfrage von Konfigurationsparametern sind in METIS nicht bzw. nur mit erheblichem Aufwand zu realisieren.
- *SemTalk*: SEMTALK ist ein Modellierungsaufsatz für MICROSOFT VISIO. Dieser ist grundsätzlich nicht auf die Konfiguration von Modellen ausgerichtet. Laut Dokumentation ist in SEMTALK die Spezifikation von Makros auf Visual-Basic-Basis möglich. Die nutzbaren Funktionalitäten sind jedoch nicht dokumentiert, so dass nicht beurteilt werden kann, ob sich diese Skriptsprache zur Erstellung von Konfigurationsfunktionalität eignet.

7 Fazit und Ausblick

Anforderungen von Praxisvertretern sowie die Ergebnisse der empirischen Studie zeigen, dass Bedarf zur toolunterstützten Verwaltung von Informationsmodellen besteht, da die nutzergruppenspezifische Anpassung von Informationsmodellen in nahezu jedem Modellierungsprojekt notwendig ist. Diese empirische Arbeit ergänzt hiermit bisherige Beiträge, die eine theoretische Begründung der Konfigurationsnotwendigkeit liefern (vgl. zu solchen Beiträgen Abschnitt 3).

Die Konfigurationsunterstützung vorhandener Tools ist zum momentanen Zeitpunkt demgegenüber nur bedingt gegeben. Tools mit umfangreicher Skriptsprachenfunktionalität wie CONCEPTBASE oder CUBETTO stellen viel versprechende Plattformen dar, die für den Zweck der

Konfiguration erweitert werden können. Hierbei ist allerdings die Frage zu stellen, ob sich solche Tools, die sich gegen etablierte, kommerzielle Tools wie bspw. ARIS behaupten müssen, durchsetzen können. Die Ergebnisse der Studie zeigen, dass sich der Bekanntheits- und damit Verbreitungsgrad solcher Tools in Grenzen hält, was diese Vermutung untermauert. Grundsätzlich viel versprechend erscheint ebenfalls der Ansatz, ein Modellierungstool auf Basis eines weit verbreiteten Zeichenprogramms zu entwickeln, wie er von SEMTALK verfolgt wird. Hier fehlt allerdings die Möglichkeit der Spezifikation von Konfigurationsfunktionalität.

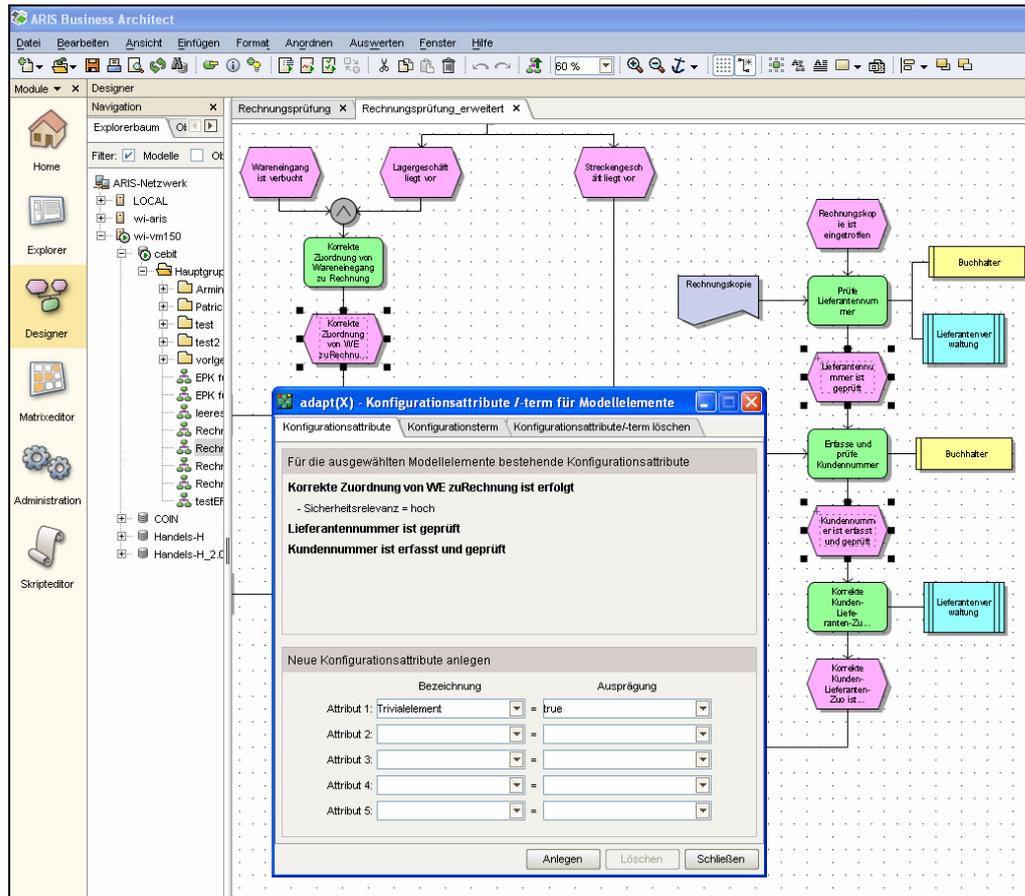


Abbildung 8: ADAPT(X)-unterstützte Zuweisung von Konfigurationsattributen zu Modellelementen in ARIS

Ausgehend von diesen Überlegungen werden auf der Basis des Fachkonzepts zur Konfiguration von Informationsmodellen zwei Modellierungstoolaufsätze für die am weitesten verbreiteten Tools entwickelt. Zum einen wird ein Konfigurationsaufsatz ADAPT(X) für das Modellierungstool ARIS entwickelt (vgl. Abbildung 8), der in der zum momentanen Zeitpunkt vorhandenen Betaversion die vorgestellten Konfigurationsmechanismen unterstützt (vgl. zu einer ausführlichen Beschreibung dieses Tools sowie eines weiteren konfigurativen Tools H2C zur Verwaltung von hierarchischen Modellen, die insbesondere in der OLAP-Berichtsspezifikation Verwendung finden [DJKR06]).

Der zweite entwickelte Modellierungstoolaufsatz trägt dem Umstand Rechnung, dass zahlreiche Unternehmen das Tool VISIO bereits einsetzen und nach Erfahrung der Autoren die hohen Anschaffungskosten von ARIS scheuen. Das Modellierungstool INDAPTA wird deshalb momentan als Aufsatz für VISIO entwickelt und wird neben einer Metamodellierungsumgebung die vorgestellten Konfigurationsmechanismen umsetzen.

Danksagung

Für die finanzielle Unterstützung des Forschungsprojekts „Komplexitätsmanagement für die Entwicklung und Anwendung adaptiver Referenzmodelle (KompadRe)“ (Geschäftszeichen: BE 1422/9-1), in dessen Rahmen dieser Beitrag entstanden ist, gilt unser Dank der Deutschen Forschungsgemeinschaft (DFG). Weiterhin danken wir Herrn Tobias Kley für die engagierte Unterstützung bei der Durchführung der Online-Befragung und ihrer Auswertung recht herzlich.

Literaturverzeichnis

- [AlSk00] Alvesson, M.; Sköldbberg, K.: Reflexive Methodology: New Vistas for Qualitative Research. London 2000.
- [BADN03] Becker, J.; Algermissen, L.; Delfmann, P.; Niehaves, B.: Konstruktion konfigurierbarer Referenzmodelle für die öffentliche Verwaltung. In: Dittrich, K.; König, W.; Oberweis, A.; Rannenber, K.; Wahlster, W. (Hrsg.): Informatik 2003. Innovative Informatikanwendungen. Frankfurt. 2003, S. 238-242.
- [BDKK02] Becker, J.; Delfmann, P.; Knackstedt, K.; Kuropka, K.: Konfigurative Referenzmodellierung. In: Becker, J.; Knackstedt, R. (Hrsg.): Wissensmanagement mit Referenzmodellen. Konzepte für die Anwendungssystem- und Organisationsgestaltung. Heidelberg 2002, S. 25-144.
- [BeKR05] Becker, J.; Kugeler, M.; Rosemann, M. (Hrsg.): Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. 5. Auflage, Berlin et al. 2005.

- [BJDF06] Becker, J.; Janiesch, C.; Delfmann, P.; Fuhr, W.: Perspectives on Process Documentation – A Case Study. In: Chen, C.-S.; Filipe, J.; Cordeiro, J. (Hrsg.): Enterprise Information Systems VII. 2006.
- [BuMo79] Burrell, G.; Morgan, G.: Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life. Brookfield 1979.
- [Delf06] Delfmann, P.: Adaptive Referenzmodellierung. Methodische Konzepte zur Konstruktion und Anwendung wiederverwendungsorientierter Informationsmodelle. Berlin 2006.
- [DJKR06] Delfmann, P.; Janiesch, C.; Knackstedt, R.; Rieke, T.; Seidel, S.: Towards Tool Support for Configurative Reference Modeling – Experiences from a Meta Modeling Teaching Case. In: Brockmans, S.; Jung, J.; Sure, Y. (Hrsg.): Meta-Modelling and Ontologies. Bonn 2006.
- [KeNS92] Keller, G.; Nüttgens, M.; Scheer, A.-W.: Semantische Prozeßmodellierung auf der Grundlage „Ereignisgesteuerter Prozeßketten (EPK)“. In: Scheer, A.-W. (Hrsg.): Veröffentlichungen des Instituts für Wirtschaftsinformatik, Heft 89. Saarbrücken 1992.
- [KIMy99] Klein, H. K.; Myers, M. D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MISQ 23 (1999) 1, S. 67-93.
- [Knac06] Knackstedt, R.: Fachkonzeptionelle Referenzmodellierung einer Managementunterstützung mit quantitativen und qualitativen Daten. Methodische Konzepte zur Konstruktion und Anwendung. Berlin 2006.
- [RoAa07] Rosemann, M.; van der Aalst, W. M. P.: A Configurable Reference Modelling Language. Information Systems 23 (2007) 1, S. 1-23.
- [SoGD03] Soffer, P.; Golany, B.; Dori, D.: ERP modeling: a comprehensive approach. Information Systems 28 (2003) 9, S. 673-690.
- [Wimm95] Wimmer, R.: Methode, hermeneutische. In: Mittelstraß, J. (Hrsg.): Enzyklopädie Philosophie und Wissenschaftstheorie 2. Stuttgart 1999, S. 883-884.

A Pattern based Approach for constructing Enterprise Architecture Management Information Models

Sabine Buckl, Alexander M. Ernst, Josef Lankes, Kathrin Schneider, Christian M. Schweda

Software Engineering betrieblicher Informationssysteme
Institut für Informatik
Technische Universität München
Boltzmannstraße 3, 85748 Garching
{buckls, ernst, lankes, schweda}@in.tum.de
{kathrin.schneider}@bmw.de

Abstract:

This paper sketches an approach to designing organization-specific information models for Enterprise Architecture (EA) Management based on patterns. Thus we intend to support the construction of *EA Management* information models in research and practice, a field we view in need of approaches to manage the complexity of such models.

Contributing to this complexity are the wide spread domains (e.g. processes, technical architecture, strategic issues) that are involved in *EA Management*. Moreover, this complexity burdens in practice both using existing information models and creating a new information model.

This necessitates approaches to manage this complexity. Our contribution in this respect lies in the introduction of patterns into the field of *EA Management* information models. Thereby, other approaches, as e.g. structuring the information models into layers, are complemented by the possibility to reuse pre-existing solutions to address *EA Management* issues.

1 Motivation

The following article identifies problems in approaches currently pursued in designing information models supporting *EA Management* and presents a solution based on composing such information models from patterns. Subsequently, we rely on the following definition of *EA Management*, which is in accordance to [ELSW06]:

”*EA Management* is a continuous and iterative process controlling and improving the existing and planned information technology (IT) support for an organization. The process not only considers the IT of the enterprise, but also business processes, business goals, strategies, etc.

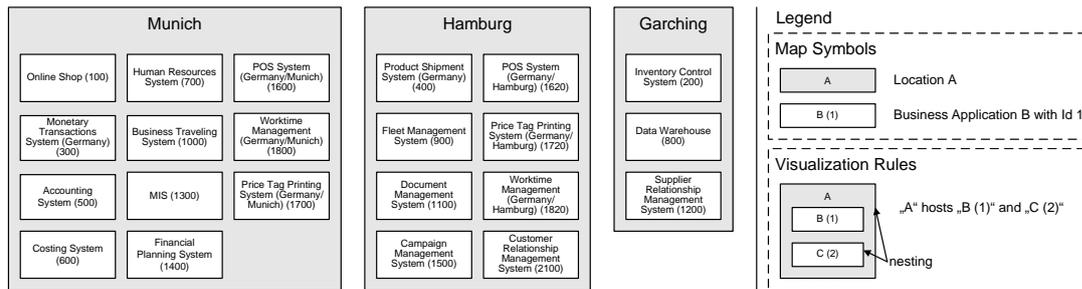


Figure 1: Exemplary software map showing which business application is hosted at which location

are considered in order to build a holistic and integrated view on the enterprise.

The goal is a common vision regarding the status quo of business and IT as well as of opportunities and problems arising from these fields, used as a basis for a continually aligned steering of IT and business.”

In order to effectively manage the EA, which includes planning and designing the EA, a systematic documentation of the elements contained therein is often regarded as an important prerequisite. Such documentations are the focus of our research project software cartography (see e.g. [LaMW05a]), in which we develop, in co-operation with our project partners (among others BMW, HVB Systems, T-Com), methods and models for documenting, planning, and designing application landscapes¹, which make up an essential part of an EA. During our research, we discovered a large number of different visualizations for application landscapes, which we call software maps. In practice, besides different kinds of software maps, as described in [LaMW05a], also other visualizations, e.g. portfolio matrices used in the context of project portfolio management, or textual descriptions, as e.g. tabular reports, can be used to support *EA Management*. The exemplary software map in Figure 1 shows via nesting rectangles, which business application is hosted at which location. Thus, it shows information conforming to the information model fragment shown in Figure 2.

Hereby, we define an *EA Management* information model as a model which specifies, which information about the enterprise architecture, its elements and their relationships should be documented, and how the respective information should be structured. This is usually achieved via a model expressed in a language suitable for conceptual modeling, as e.g. UML, enriched with descriptions detailing the exact meaning of the concepts.

Considering the different aspects of *EA Management* in an information model in practice leads to information models, that are according to [Sebis05] likely to contain easily over 100 different classes. Due to the complexity and amount of relevant information, the employment of tools for documenting the EA is according to e.g. [Fran02] required. A central part of such tools is the

¹In our research project we define application landscape as ”the entirety of all business applications and their relationships in an organization”.

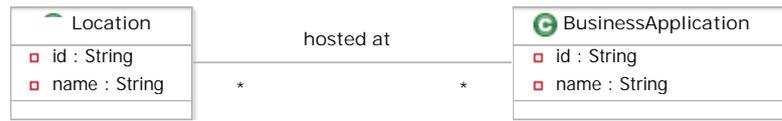


Figure 2: Information model corresponding to the software map shown in Figure 1

information model. The information models provided by tools for supporting *EA Management* were examined as part of the *EA Management* Tool Survey [Sebis05] conducted in the research project software cartography, helping us to identify several problems typically connected to creating and using information models.

Based on our experiences from the *EA Management* Tool Survey and insights gained from existing information models used by our project partners, we propose, as mentioned above, a concept based on information model fragments called *patterns*, which are developed particularly for specific concerns and recurring problems. These patterns can then be used as building blocks to compose an organization-specific information model.

The remaining article is structured as follows: Section 2 gives an overview of literature and approaches in practice concerning *EA Management*. Section 3 outlines problems of common information models and the approaches pursued in their creation. Our pattern based approach for information modeling is described in section 4 and illustrated by an example in section 5. The last section 6 summarizes the paper and provides a short outlook on planned research activities, which further pursue our pattern based approach.

2 Related Work

According to [LaWe04], who conducted an extensive literature review about *EA Management*, first publications in this field go back to [Zach87], but only recently the number of articles published about the subject has increased, indicating that the topic is entering mainstream interest. Thus, different organizations have created frameworks providing guidance for *EA Management*, as for example *The Open Group* (TOGAF [TOG05]), *Meta Group* (Enterprise Architecture Desk Reference [META02]), or the *US Department of Defense* (DoDAF [DoD04]). While those frameworks provide high-level guidance and some of them add a process model for *EA Management*, they do not detail how specific tasks should be carried out and do not provide detailed information models suitable for supporting these tasks.

Such details are of course implemented in *EA Management* tools, which may support one or more of the frameworks mentioned above. Vendors taking their *EA Management* approaches to standardization, as e.g. *Adaptive, Ltd.*, which is a major contributor to OMG's *IT Portfolio*

Management Facility (ITPMF) [OMG06], add to published approaches created by academia. The *ArchiMate Project*, for example, proposes a notation and viewpoints for *EA Management* in [Lank05]. [BrWi05] offers an EA metamodel with over 50 classes, managing the arising complexity by structuring the model in layers. The metamodel for *EA Management* supplied by [Fran02], which supports *planning, designing, and maintaining corporate information systems*, is structured via a language architecture integrating partial models for special purposes as e.g. *strategic modeling* or *modeling organizational structures*. While these languages might be seen as similar to patterns in the sense of our approach, [Fran02] is more focused on providing an integrated metamodel by putting together these languages. Subsequently, we complement such approaches by describing a way for documenting information model patterns, which can then be integrated to a comprehensive, organization-specific information model.

3 Problems of *EA Management* Information Models and their Approaches

As described above, we are trying to complement existing approaches for structuring the complexity of *EA Management* information models by utilizing patterns, as the benefit of existing models seems to be reduced by the following tendencies:

- Research projects seem to prefer developing new models instead of improving existing ones. The same is true for visualization techniques for *EA Management* specific information and methodologies prescribing procedures for working with this information and its visualizations. There are hardly any well-known results in the field that are further developed and verified by research or other practitioners, at least not by a larger community. Each project seems to start developing its own approach.

This tendency does not universally exist in other areas. In software engineering for example, there are publications that build on UML as a visualization of software structures and try to improve the readability of such diagrams, e.g. by finding advantageous layout criteria [PMCC01]. Contrary, the state of the art regarding enterprise modeling is described as *hardly coherent* by [Fran02].

- Projects in practice executing *EA Management* activities seem to neglect *EA Management* information models made available by research, unless a joint project with academia is executed or an information model is incorporated in the *EA Management* tool used. While such tools contain information models, visualizations, and methodologies, practitioners seem to experience problems therewith, as e.g. shown by the fact, that a major share of the project partners we have talked to in our research project *software cartography* uses *Microsoft Visio* or *Microsoft PowerPoint*.

based visualizations [LaMW05a] developed and used after self-defined guidelines² in their *EA Management* activities. Another practitioner states his experiences with *EA Management* information models and tools as follows [Riih05]:

”[...] we have tried out several tools in this area, without much success. [...] Their metamodels are rather complex, but not integrated within themselves.”

These tendencies describe a situation in which *reinventing the wheel* seems to be a common practice in the creation of *EA Management* information models, both in practice and in academia. This is surprising, as we experienced, detailed in [Sebis05], that some questions, e.g. ”which business application system supports which process at which organizational unit” are common to *EA Management* and should therefore also be covered by information models supporting *EA Management*. Additionally, the tendencies hamper the development of best practices, standards, and significant research results regarding *EA Management* information models, leaving the state of art in this field unable to satisfy the *need to use proven models*, which is e.g. indicated by [Bern03].

Sections 3.1 to 3.3 discuss issues in common *EA Management* information models and approaches guiding their creation, which possibly contribute to the above tendencies.

3.1 Model Size: Giant vs. Midget Models

According to [ELSW06] *EA Management* information models seem to be in danger of falling into one of two traps. On the one hand, small information models with brief documentation hardly deliver benefit to an *EA Management* project. On the other hand, all-embracing information models that cover almost all kinds of information that can be relevant to *EA Management* are also not without their problems. Table 1 gives an overview of the number of classes in the shipped metamodels of the tools covered by [Sebis05].

Number of classes	Number of tools
not known	2
up to 50	2
50 – 350	3
350 or more	2

Table 1: Overview of the information model sizes encountered in [Sebis05]

In most organizations, not all features of an all-embracing information model might be of relevance, due to the focus of a given *EA Management* project on specific issues or the sheer fact that an organization is too small to profit from explicitly managing certain facets of its EA. In

²In these cases, the information model can be seen as not being made explicit, giving the visualizations a drawing-like quality, as e.g. described in [ELSW06].

such cases, the disadvantages connected to the size of the information model are likely to outweigh its benefit. Such disadvantages center around understanding and using the model, as well as making organization-specific modifications to it. [Bern03], for example, states this complexity issue of *EA Management* information models. These disadvantages prevail, even if some parts of the information model can be hidden by the *EA Management* tool used (cf. [Sebis05]), as someone has to understand a big model at first in order to determine what has to be hidden and what has to be modified to fit the needs of the organization under consideration.

Additionally, giant and midget models seem to be no popular objects of research, contributing to the tendency of *EA Management*-specific research to build new information models instead of improving existing ones. While midget models often barely contain enough substance, which can be targeted by research and be used as a basis for proposing research questions or hypotheses, the situation also easily gets troublesome with the giant models. Subjecting certain aspects of an *EA Management* information model to research can be difficult in the complex web of an all-embracing information model.

3.2 Giving explicit Account for Model Utility: Stockpiling useless Information

Regarding architectural descriptions, conventions as e.g. the IEEE Std. 1471-2000 (Recommended practice for the architectural description of software intensive system) [IEEE00] state that the creation of architectural views has to be justified by the existence of stakeholders, whose concerns can be addressed by the respective views. This paradigm makes even more sense in the context of EAs, as the collection of the respective information, which often has to be organized in a decentralized fashion, is more difficult in this field than in the documentation of single application systems [LaMW05b]. Introducing certain data structures into an *EA Management* information model, thus creating the demand to collect the respective information, should therefore be justifiable by adequate concerns being addressed using this information.

Thus, failing to provide the information due to which concerns certain data structures are used in an information model, which is not an unrealistic danger in practice according to [Sebis05], leaves difficulties to decide which concepts are really relevant for addressing specific problems in a specific organization. This could contribute to the tendency that the respective information model is only hesitantly used in practice, especially, when this intransparency occurs in giant-like information models.

Failing to provide the concerns also constitutes an obstacle to research regarding *EA Management* information models. When evaluating the suitability of a certain model in order to use this as a basis for reasonable proposals for improvement, it seems beneficial to know what exact problems are to be addressed by this model.

3.3 Missing Methodologies

Even if it is known or somehow guessed from insufficiently explicated concerns, which data structures are intended to be taken into consideration when addressing a specific concern, it may not be clear how these structures should be used. While many information models seem to intend the information to serve for the creation of visualizations or reports that are meant to make the EA more transparent, also other use cases are possible, as e.g. metrics calculation or automated checking for violation of certain constraints.

The practical usage of an information model is made more difficult, if the methodologies for addressing concerns based on this information are left out. [Bern03] reinforces the importance and benefit of knowing the usage context of *EA Management* models. In case they are missing, methodologies have to be added by the users in the project introducing the information model, leading to additional effort. Also, descriptions of the intended data usage could often help to clarify what is meant by the abstract concepts³ forming the classes and relationships in an *EA Management* information model.

The advantage explicit methodologies could give to research regarding EA information models directly connects to the point made in section 3.2, as the evaluation whether an information model is useful for addressing certain concerns also has to take into consideration, how these concerns are meant to be addressed using the respective information model. Hereby, it becomes clear that an *EA Management* information model should not be seen in isolation. It is only complete together with *viewpoint definitions*, offering a problem-adequate way of visualizing data from the information model, and *methodologies*, which describe procedures how information model and viewpoint can be used to address specific concerns.

Especially the importance of methodologies can be substantiated via the methodology definition according to [KrSH93], which states that

”Within an engineering discipline, a method describes a way to conduct a process. In the context of systems engineering, a method is defined as consisting of:

- An underlying model (which refers to the classes of objects represented, manipulated and analysed by the method)
- A language (referring the concrete means of describing the products of the method)
- Defined steps and ordering of these steps (which refers to activities performed by the user of the method)
- Guidance for applying the method”

³The frequently and incoherently used term ”service” can serve as an example here. While there is hardly a common definition of this term, mutual understanding about it can increase significantly, if it is explicitly stated what is about to be done with services. Knowing that e.g. application development is the subject can add to the understanding regarding the term.

Of course the tendencies discussed above may be impacted by the fact that *EA Management* is a relatively young discipline, with mature standards and best practices therefore yet to come. But we view that the issues outlined in sections 3.1 to 3.3 might well obstruct the development of exactly these standards and best practices.

4 Pattern based Approach

Addressing the issues encountered in *EA Management* information models, which we regard to partially be arising from the approaches taken during information modeling, we propose an approach based on *patterns*. This concept is already used in various domains, as e.g. in architecture or in software design [GHJV94], where a pattern is defined as follows:

”A (design) pattern names, abstracts, and identifies the key aspects of a common design structure that make it useful for creating a [...] design. The design pattern identifies the participating classes and instances, their roles and collaborations and the distribution of responsibilities. Each design pattern focuses on a particular [...] design problem or issue. It describes, when it applies [...], and the consequences and trade-offs of its use.”

Defined like this, a pattern describes a basic idea, that has turned out to be useful for addressing specific concerns in a practical context and in a generalized way. Due to this generalization, a pattern can be adapted and reused in related contexts. Besides this, patterns are defined implementation-independent, constituting a blueprint for a solution actually to be implemented. We regard *EA Management* patterns as building blocks for the concept of an organization-specific support for *EA Management*, consisting of a conceptual information model, viewpoints and methodologies for using the respective information. Thereby, the *EA Management* patterns themselves are organization independent, based on best practices developed by academia and practice and hence are possibly of benefit to a wide range of users, offering the possibility to further improve existing patterns in an environment without organization-specific restrictions. *EA Management* patterns can furthermore be combined to one or more catalogues, which can be maintained by a community in order to control and foster the further development, like e.g. improvements of existing or the introduction of new patterns. This community can additionally make sure that the advantages of this approach can better be communicated than in the case of solutions created by single organizations or research groups.

Figure 3 sketches the ideal type of project⁴ concerned with *EA Management* support utilizing *EA Management* patterns. Such a project uses one or more catalogues of *EA Management* patterns, supplied by *pattern designers*, as a basis. From these catalogues, the *developers* for *EA Management* support choose patterns, that are perceived as adequate for addressing specific concerns

⁴Figure 3 sketches a waterfall-like approach, but of course also iterative approaches can be taken.

of the respective organization, preferably under participation of the prospective *users*. After integrating these *EA Management* patterns, thereby creating a coherent organization-specific conceptual model, the respective concepts can be implemented, e.g. in an *EA Management* tool or a suite of tools, that fit the requirements of the organization under consideration.

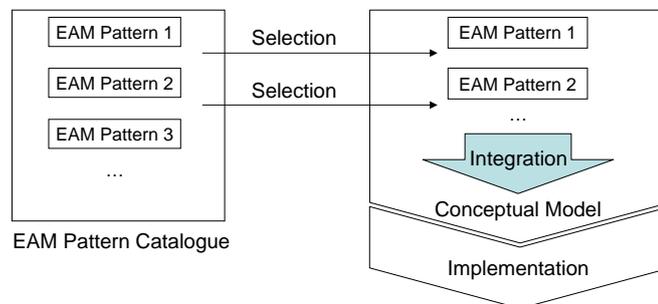


Figure 3: Implementing an *EA Management* approach based on *EA Management* patterns

Thus, the pattern based approach tries to avoid the pitfalls described in sections 3.1 to 3.3, as the *EA Management* patterns form small, reusable units preferably based on established practices that are used, when suitable for addressing specific concerns.

4.1 Basic Structure: A three-tiered Approach

In order to not solely describe the information under consideration in a specific pattern, but also to make the methodologies for using the respective data explicit and therefore avoiding pitfalls as described in section 3.3, we view it important to have descriptions with the information model patterns that delineate, how the data stored should be presented and used to address specific concerns, similar to the concepts of *language* and *steps for applying the method* in the methodology definition according to [KrSH93] in section 3.

Hence, we decided to build an approach on three different kinds of *EA Management* patterns:

Methodologies *defining steps* to be taken in order to address given concerns. Furthermore, as a guidance for applying the method, statements about its intended usage context are provided, which include the concerns to which the methodology can be applied. These concerns are addressed by procedures defined by the methodology, which can be very different, ranging from e.g. visualizations and group discussions to more formal techniques as e.g. metrics calculation.

Viewpoints providing the *languages* used by methodologies. A viewpoint proposes a way to present data stored according to one or more information model patterns.

Information model patterns supplying *underlying models* for the data visualized in one or

more viewpoints. An information model pattern conveys an information model fragment including the definitions and descriptions of the used information objects.

The three kinds of *EA Management* patterns are closely tied to each other, via relationships as shown in Figure 4. Templates, which serve to provide exemplary outlines for documenting and describing the patterns in an in-depth and structured way, can guide the compilation of actual patterns. An detailed description of the template regarding information model patterns can be found in the next section, while this article, focusing on information models, does not further elaborate on methodologies and viewpoints. Each *EA Management* pattern has to be adapted to the organization-specific context and combined with other *EA Management* patterns in order to make contributions to a specific concept for *EA Management* support.



Figure 4: Methodology, viewpoint, and information model pattern

4.2 Documentation and Usage of Information Model Patterns

According to the three-tiered structure of our *EA Management* pattern approach, described in the previous section, the *information model pattern*, its structure, and its usage is detailed below. An exemplary outline for documenting information model patterns can be seen in the template in Table 2. It is organized in three main parts, an *overview section*, a *solution section*, and a *consequence section*, similar to the basic elements of a pattern description stated by [GHJV94].

Overview section	
Id	An unique alphanumerical identifier
Name	A short and expressive name for the pattern
Alias	Names this pattern is also known as (optional)
Summary	A short summary of the pattern of about 100 words
Solution Section	
Information Model	An <i>information model fragment</i> in a certain language (see section 4.2.1), together with additional documentation
Consequence Section	
Appliance	Guidance on how to use the information model pattern
Consequence	Consequences resulting from the usage of the pattern

Table 2: Template structure of the Information Model Pattern

4.2.1 *Languages for Information Model Fragments*

Documenting an information model fragment always relies on a certain language, in which the elements of the fragment are expressed. Basically a lot of different languages suitable for that kind of conceptual modeling are known. The following list is intended to show some of the more prominent examples, outlining both advantages and disadvantages of the languages:

Textual Description in Natural Language: The model elements and their relationships are described in natural language. This seems to be a good choice as it produces easily understandable and adaptable descriptions. The main disadvantage is that this kind of documentation easily leads to mistakable constructs, insufficient for exactly defining information model fragments.

Meta Object Facility (MOF) and Unified Modeling Language (UML) Class diagrams: The model elements and their relationships are described via object-oriented concepts, captured e.g. in the UML or MOF 2.0 metamodel. This description can rely on UML class diagrams, which should be usable by most developers. A disadvantage of these languages is that they lack a formal basis and hence possibly limit the domain in which these kind of information model fragments can be used.

Ontology Languages: The model elements and their relationships are described in terms of an ontology language. Such a language might provide more expressiveness than the approach of MOF or UML and is based on a formal foundation including the possibility to use automated reasoning in the model. Disadvantageous is, that ontology languages are not so widespread in the area of conceptual modeling.

Mathematical Formalization: The model elements and their relationships are described in mathematical or logical terms, providing a strong formal background. The main disadvantage is, that mathematical or logical models are usually difficult to use.

Of course, the possibility to combine two or more of the variants described above could unleash the advantages of the languages combined, possibly without having to consider the individual disadvantages. Nevertheless, one disadvantage seems inevitably related to such a combined approach, namely the effort for describing an information model fragment in both languages and keeping these descriptions consistent.

4.2.2 *Language Choice and Documentation Guidelines*

With different possible languages for describing information model fragments and considering their individual advantages and disadvantages, some languages may be more or less adequate for a specific pattern. This is especially obvious, if the information model fragment under

consideration is only used for creating a visualization of the application landscape or a tabular report, where an object-oriented description should be sufficient. For other usecases, e.g. the calculation of metrics or the simulation of processes, the situation looks different, as this may only be possible in a reasonable way if the information model has a more formal basis.

Therefore, we propose using a language adequate to the problem addressed, strongly suggesting to use UML as the language of choice⁵, as this language is widely understood and has been found by us as problem-adequate in many situations. For example, this language is commonly used in the information models we found in our research project *software cartography*.

There is also a second reason for suggesting *one* language. Utilizing the pattern based approach for information modeling, a crucial point is the *integration* of information model patterns to form a complete information model. This integration can be regarded more simple, if the different patterns are specified using one language, which cuts the effort of translating the models between languages.

Nevertheless, there might be situations in which it seems advantageous to use a language different from the default one, as e.g. UML would lead to a documentation far too extensive or not expressive enough for the problem at hand and therefore would not be problem adequate. In such cases it can be advisable to support the information model fragment modeled in a problem adequate language with a corresponding information model fragment documented in the default language, as this simplifies the integration of the patterns.

As a last documentation guideline we would like to advise the *pattern designer* to complement each information model fragment with a textual description promoting understandability.

4.2.3 Integrating Information Model Patterns

Aspects of integrating information model patterns during the creation of an organization-specific information model are not detailed in this article. Nevertheless, two integration approaches are subsequently sketched - relying on experience in conceptual modeling.

The steps for achieving an integration of information model patterns are similar to the ones applied in software engineering. If UML is used for documenting the information model pattern, a simple way to integrate two information model patterns is to identify one or more identical classes within both patterns. These classes can then be used as a point of integration. Another approach involves introducing a new relationship between two classes from different patterns. Potentially this could employ inheritance, e.g. the class *business application* from one pattern could inherit life cycle attributes from a class *LifecycledElement* introduced in another pattern.

⁵In the following called *default language*.

5 Example of the EA Management Pattern Approach

This section exemplifies the pattern-based approach by describing patterns to address concerns related to redundant applications in an application landscape. Thereby, the viewpoint and the methodology are sketched briefly, with the focus lying on the information model pattern, as this constitutes the core subject of the article. Due to readability reasons we introduce the viewpoint first and then complement it with its underlying information model pattern.

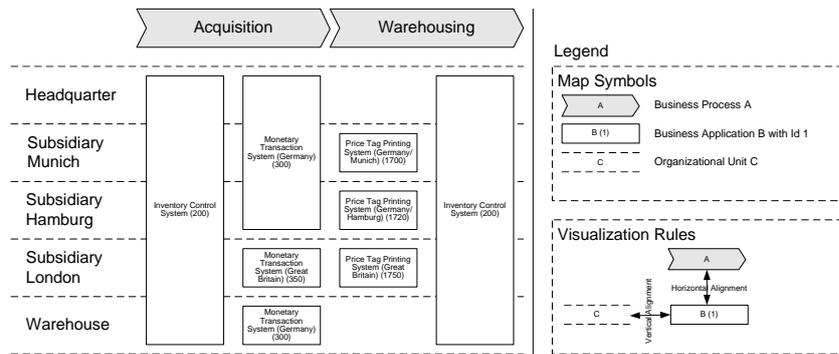


Figure 5: Exemplary process support map

5.1 Viewpoint: Process Support Map (Id V-PSM-1)

Figure 5 shows a *process support map*, a kind of software map we discovered as a visualization used for addressing concerns as mentioned above in practice [LaMW05a].

Basically, the process support map is a software map utilizing positioning of symbols to show, which business processes are supported by which business applications in which organizational units. Thereby, chevrons visualizing a process chain, seen as a sequence of processes, make up the x-axis. The y-axis is made up of labels representing organizational units. The rectangles in the main area of the map symbolize business applications, and their positioning expresses which process is supported by which business application in which organizational unit. Information model pattern *I-BPS-1* (section 5.2) supplies the concepts on which the process support map is built. Methodology *M-ARI-1* (section 5.3) works on visualizations conforming to this viewpoint.

5.2 Information Model Pattern: Business Process Support (I-BPS-1)

According to section 4.2, the information model pattern is described in three parts.

Overview section:

Id: I-BPS-1
 Name: Business Process Support

Alias: IT Building Plan
 Summary: The pattern provides a structure suitable for organizing information about which business applications support which business processes in which organizational units.

Solution section

As this pattern is mainly constructed to serve methodologies based on visualizations that have no strict background in formal methods, it is described in UML, backed up by textual definitions and explanations. The concepts of figure 6 are defined as follows:

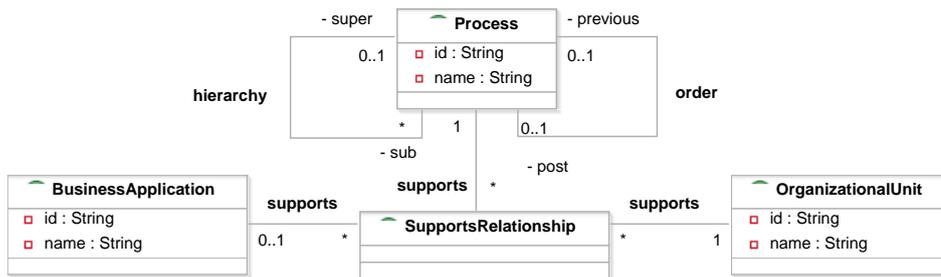


Figure 6: Information model fragment of the business process support pattern

Process: A process can, according to [Krcm05], be defined as a sequence of logical individual functions with connections between them. [DiFH03] states input and output factors and a defined process objective as important characteristics of a process. The process class here should not be identified with single process steps or individual functions, but with high-level processes at a level similar to the one used in value chains. The process class owns the following attributes⁶:

- id** unique alphanumeric identifier
- name** a descriptive name of the process

OrganizationalUnit: An organizational unit represents a subdivision of the organization according to its internal structure. E.g. the entities showing up in an organigram can be used as organizational units.

Business Application: A business application is a software system which is part of an information system of an organization. An information system is according to [Krcm05] understood as a sociotechnical system, which is, besides the software system, made up of the infrastructure the software system is based on, and a social component, namely the employees or stakeholders concerned with it. Thereby, infrastructure

⁶The corresponding attributes of the other classes are subsequently not further detailed.

and social component are not considered as belonging to the business application, while the characterization "business" restricts the term to applications that support at least one process of the respective organization.

SupportsRelationship: The class *SupportsRelationship* represents the support of a process by a business application at a specific organizational unit. Basically, it constitutes, together with its three associations, a ternary relationship between *BusinessProcess*, *OrganizationalUnit*, and *BusinessApplication*. This is necessary in order to be able to tell exactly which organizational unit uses which business application to support a given process. The class is utilized here instead of directly relying on the ternary relationship in UML in order to allow adaptations of the fragment like e.g. introducing attributes to the relationship. This might e.g. be useful if the relationship is time-dependent.

The associations of the information model fragment serve the following purposes:

hierarchy: A process can be part of a larger, encompassing process (super-process) and can include other processes (sub-processes).

order: A process, seen on the level as stated above, can be part of a value chain, which is an (at least partially) ordered sequence of processes. Thus, a process can have a predecessor process and a successor process, expressed by the order relationship.

supports: Instances of the class *SupportsRelationship* use links according to the three associations that are called *supports* to indicate that a business application supports a certain process at a specific organizational unit.

Consequence section:

A possible *appliance* is viewpoint *V-PSM-1*, with its attached methodologies, which is able to visualize information structured according to this information model pattern⁷.

An important *consequence* of this pattern is the challenge of collecting the necessary data. Existing business applications and the execution of processes at different organizational units might be considered well-known in most organizations. Contrary, the support of an business application for a specific business process is more often only implicit knowledge of certain employees, which might considerably add to the burden of information gathering.

⁷Of course other visualizations are possible, but not detailed here.

5.3 Methodology: Business Application Redundancy Identification (Id M-ARI-1)

This methodology briefly describes a procedure to address the concern *business application redundancy identification*. Redundancy in this context means, that the same functionality supporting a process, may be realized by different business applications in different organizational units.

In order to contribute to addressing this concern, the user should employ the viewpoint *V-PSM-1* (see section 5.1) to identify processes, where the different organizational units use different application systems to support the same process. An example for this can be seen in Figure 5, where the process *Acquisition* is supported by two different business applications named *Monetary Transaction System (Germany)* and *Monetary Transaction System (Great Britain)* in the different organizational units.

The next step should be an analysis of these potential redundancies. First of all, they can turn out to be no redundancies at all, as it is possible e.g. in Figure 5, that *Monetary Transaction System (Germany)* supports other subprocesses of the process *Acquisition* than *Monetary Transaction System (Great Britain)*. A software map that does not show the processes on the level on which this is visible, but only a more aggregated view, cannot show this fact explicitly. Moreover, if there are redundancies, they might have been deliberately introduced, e.g. in order to achieve a higher flexibility. In such cases, it may be reasonable to retain the redundancy. In case that no such reasons can be found, the results from the analysis regarding redundancies can be used as input for activities defining visions or plans for the evolution of the application landscape. This can include definitions of project proposals that serve the elimination of the redundancies.

6 Resume and Outlook

In this article, we tried to address the challenges faced in the creation of an information model for *EA Management* and presented our approach to support the designing process with predefined information model patterns.

While *EA Management* frameworks like Zachmann [Zach87], TOGAF [TOG05], etc. offer guidelines for designing, planning, and implementing *EA Management*, they currently do not detail an *EA Management* information model. Enterprises at an initial state of *EA Management* have to implement these frameworks either using existing information models or creating their own information model beginning from scratch, which is a way not without pitfalls.

The approach presented in this paper tries to support enterprises introducing *EA Management* by providing a structured way for laying the basis for an organization-specific information model. Furthermore, the approach presented does not focus on the information itself, but on the concerns of stakeholders typically found in *EA Management*. These concerns are then addressed via

EA Management patterns. Having provided an initial example, how such a pattern could look like, we are currently describing more patterns, as we have encountered them in our project *software cartography* as part of the *state of the art*, e.g. at industry partners or in literature. Once having collected these patterns, we plan to conduct a series of interviews with people concerned with *EA Management* related tasks, in order to gain information about both relevance and usage context of the patterns identified. As a final step of this evaluation, we plan to consolidate these patterns in an *EA Management* pattern catalogue, which is intended to form a knowledge base for the construction of organization-specific information models supporting *EA Management*.

References

- [Bern03] *Bernus, Peter*: Enterprise models for enterprise architecture and ISO 9000:2000. In: Annual Reviews in Control 27 (2003), pp. 211-220.
- [BrWi05] *Braun, Christian; Winter, Robert*: A Comprehensive Enterprise Architecture Metamodel and Its Implementation Using a Metamodeling Platform. EMISA 2005, pp. 64 - 79.
- [DiFH03] *Disterer, Georg; Fels, Friedrich; Hausotter, Andreas (Eds.)*: Taschenbuch der Wirtschaftsinformatik. Carl Hanser Verlag, München, Wien 2003.
- [DoD04] *Department of Defense*: DoD Architecture Framework Version 1.0, Volume I: Definitions and Guidelines. Department of Defense (DoD), USA 2004. http://www.defenselink.mil/nii/doc/DoDAF_v1_Volume_I.pdf.
- [ELSW06] *Ernst, Alexander; Lankes, Josef; Schweda, Christian; Wittenburg, André*: Tool Support for Enterprise Architecture Management - Strengths and Weaknesses. In: The Tenth IEEE International EDOC Conference, Hong Kong 2006, pp. 13-22.
- [Fran02] *Frank, Ulrich*: Multi-Perspective Enterprise Modeling (MEMO) - Conceptual Framework and Modeling Languages. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences 35 (2002), pp. 1258-1267.
- [GHJV94] *Gamma, Erich; Helm, Richard; Johnson, Ralph; Vlissides, John*: Design Patterns - Elements of Reusable Object-Oriented Software. Addison-Wesley Longman, Reading 1994.
- [IEEE00] *IEEE Computer Society*: IEEE Std 1471-2000: IEEE Recommended Practice for Architectural Description of Software-Intensive Systems. IEEE Computer Society 2000.

- [Krcm05] *Krcmar, Helmut*: Informationsmanagement. 4th edition, Springer, Berlin, Heidelberg 2005.
- [KrSH93] *Kronlöf, Klaus; Sheehan, Anne; Hallmann, Matthias*: The Concept of Method Integration. In: *Kronlöf, K. (Ed.): Method Integration*. John Wiley & Sons Ltd., West Sussex, England, 1993, pp. 1 - 18.
- [LaWe04] *Langenberg, Kerstin; Wegmann, Alain*: Enterprise Architecture: What Aspects is Current Research Targeting? EPFL Technical Report IC/2004/77.
- [LaMW05a] *Lankes, Josef; Matthes, Florian, Wittenburg André*: Softwarekartographie: Systematische Darstellung von Anwendungslandschaften. In: *Wirtschaftsinformatik 2005*, Bamberg 2005, pp. 1443-1462.
- [LaMW05b] *Lankes, Josef; Matthes, Florian; Wittenburg, André*: Architekturbeschreibung von Anwendungslandschaften: Softwarekartographie und IEEE Std 1471-2000. In: *Software-Engineering 2005*, Essen 2005, pp. 43-54.
- [Lank05] *Lankhorst, Marc*: Enterprise Architecture at Work: Modelling, Communication and Analysis. Springer, Berlin, Heidelberg, New York 2005.
- [META02] *META Group*: Enterprise Architecture Desk Reference. META Group, Inc. 2002.
- [OMG06] *OMG*: IT Portfolio Management Facility (ITPMF) Available Specification, dtc/06-05-01. Object Management Group 2006.
- [PMCC01] *Purchase, Helen; McGill, Matthew; Colpoys, Linda; Carrington, David*: Graph drawing aesthetics and the comprehension of UML class diagrams: an empirical study. In: *Australian symposium on information visualisation - Vol. 9.*, Australian Computer Society, Inc., Sydney 2001, pp. 129 - 137.
- [Riih05] *Riihinen, Jaakko*: Enterprise Architecture Tool Survey TU Munich. Email correspondence between J. Riihinen (Director, Chief Enterprise Architect, Nokia) and R. Schlossar (BOC Information Systems), 2005-03-21.
- [Sebis05] *sebis*: Enterprise Architecture Management Tool Survey 2005. Technische Universität München, Chair for Informatics 19 (sebis), Munich 2005.
- [TOG05] *The Open Group*: TOGAF 'Enterprise Edition' Version 8.1. The Open Group 2005.
- [Zach87] *Zachman, John*: A framework for information systems architecture. In: *IBM Systems Journal* 26 (1987) 3, pp. 276-292.

Ein hierarchischer, architekturbasierter Ansatz zur Unterstützung des IT/Business Alignment

Ronny Fischer, Robert Winter
Institut für Wirtschaftsinformatik
Universität St. Gallen
CH-9000 St. Gallen
{ronny.fischer, robert.winter}@unisg.ch

Abstract

In diesem Beitrag wird ein hierarchischer, mehrstufiger Ansatz für IT/Business Alignment vorgestellt, der auf der Unternehmensarchitektur als zentralem Koordinationsinstrument basiert.

Den Ausgangspunkt bilden die Diskussion mehrstufiger, hierarchischer Systeme sowie die Beurteilung von Instrumenten für das IT/Business Alignment aus Sicht der Notwendigkeit einer konsistenten Gestaltung komplexer Strukturen und Prozesse.

Es wird ein Ansatz für die Gestaltung und Weiterentwicklung der Unternehmensarchitektur vorgestellt, der IT/Business Alignment unterstützt. In diesem Ansatz werden Business- und IT-Artefakte gleich gewichtet, alle relevanten Artefakte auf einem hohen Abstraktionsniveau repräsentiert sowie die Zusammenhänge zwischen den Artefakten unterschiedlicher Modellierungsebenen explizit abgebildet. Zur Bewirtschaftung dieser Artefakte werden Entwicklungs- und Kommunikationsprozesse skizziert.

1 Einleitung

Aus Sicht von IT-Managern ist IT/Business Alignment ihr wichtigstes Thema überhaupt [Luft05]. Zwar beschränkt sich die empirische Basis von LUFTMAN's Untersuchung auf 182 Unternehmen in den USA und die Jahre 2003-2004. Da aber IT/Business Alignment über alle Managementebenen, Branchen und Erfahrungsklassen hinweg konsistent Priorität 1 hat [Luft05, 274ff.], darf davon ausgegangen werden, dass dies auch für europäische Unternehmen

nicht völlig anders ist. Die Wirtschaftsinformatik sollte deshalb geeignete Konzepte bereitstellen, um IT/Business Alignment zu unterstützen.

„Alignment means that applying IT in an appropriate and timely way, in harmony with business strategies, goals, and needs.“ [Luft05, 271] Die Tatsache, dass es auch nach vielen Jahren Forschung und Entwicklung im Bereich IT/Business Alignment offensichtlich noch Probleme bei der Abstimmung des IT-Einsatzes mit fachlichen Zielen und Anforderungen gibt, wird von Luftman darauf zurückgeführt, dass nicht ein einziges Maßnahmenbündel zum Ziel führt, sondern eine Kombination verschiedenster Maßnahmen [Luft05, 281]. Die Vielzahl von Komponenten, die ein Erfolg versprechender Ansatz für IT/Business Alignment umfassen muss, wird u. a. in LUFTMAN's Reifestufenansatz deutlich: In Analogie zu den fünf Reifegraden anderer Reifestufenmodelle (z.B. CMMI) wird in [Luft00] vorgeschlagen, „Strategic Alignment Maturity“ in Abhängigkeit der Erfüllung bestimmter Bedingungen als (1) spontanen, (2) definierten, (3) etablierten, (4) aktiv gesteuerten oder (5) optimierten Prozess einzuordnen.

IT/Business Alignment wird als wichtiges Instrument zur Erreichung einer effektiven Unternehmensorganisation betrachtet. „Alignment“ stellt sich damit als organisatorische Gestaltungsaufgabe dar, welche die Abstimmung zwischen allen Gestaltungsebenen einer Unternehmung/einer Organisation (siehe Kapitel 4) berücksichtigen muss [PeSo05]. Unternehmungen sind jedoch äußerst komplexe sozioökonomische Gebilde, deren konsistente Gestaltung und Weiterentwicklung über alle relevanten Artefakte hinweg kaum möglich erscheint. Werden derart komplexe Systeme in Teilsysteme zerlegt und mittels eines hierarchischen, mehrstufigen Verfahrens gestaltet bzw. verändert, ist die Erhaltung der Konsistenz jedoch möglich [Wint91]. Neben der Konsistenzerhaltung können Zielpräferenzen in die Gestaltung des mehrstufigen Gestaltungs- bzw. Veränderungsverfahrens eingearbeitet werden.

In hierarchischen, mehrstufigen Systemen wird nach [MeMT70, 37] unterschieden zwischen (1) Abstraktionsebenen (Modellierungsaspekt: *Strata*), (2) Problemlösungsschichten (Problemlösungsaspekt: *Layers*) und (3) Koordinations- / Entscheidungseinheiten (Arbeitsteilungsaspekt: *Echelons*). *Strata* werden gebildet, um Problembeschreibungen zu „vergrößern“ und Lösungen vergrößerter Probleme schrittweise zu verfeinern. *Layers* werden gebildet, um Teil-Problembeschreibungen zu sequenzialisieren und Problemlösungen schrittweise zu ergänzen. *Echelons* werden gebildet, um Problembeschreibungen zu dekomponieren und Problemlösungen verteilt zu erarbeiten.

Für das IT/Business Alignment sind alle diese Abstraktionen notwendig:

- *Strata* werden in Form von Abstraktionshierarchien für Artefakte / Artefaktbeziehungen sowie von (ebenesspezifischen wie auch ebenenübergreifenden) Sichten gebildet. Geeignete Strata sind z.B. Produktgruppen (statt Produktvarianten), Prozesse (statt Aktivitäten), Organisationsziele (statt detaillierter Messgrößen) und Applikationen (statt Softwarekomponenten).
- *Layers* werden gebildet, um zunehmende Implementierungsnähe / abnehmende Realisierungsunabhängigkeit und den Grundsatz „IT follows Business“ umzusetzen. Geeignete Layers sind Strategie, Organisation, Integration, Software und schließlich Technologie (siehe Kapitel 4).
- *Echelons* werden gebildet, um die groben Zusammenhänge (in Form einer Unternehmensarchitektur) ganzheitlich zu gestalten und auf dieser Grundlage Detailarchitekturen (z.B. Produktdatenmanagementsystem, Performancemanagementsystem, Sicherheitsarchitektur, Softwarearchitektur, detaillierte Ablaufplanung) jeweils separat, aber dennoch konsistent gestalten zu können.

Dieser Beitrag beschreibt Unternehmensarchitektur als wichtige Komponente eines mehrstufigen, hierarchischen Gestaltungsansatzes zur Realisierung von IT/Business Alignment. Als konzeptionelle Grundlage wird im folgenden Kapitel 2 IT/Business Alignment definiert, und es werden traditionelle Ansätze zu seiner Umsetzung diskutiert. Kapitel 3 definiert Unternehmensarchitektur und fasst den aktuellen Stand der Forschung zu diesem Thema zusammen. In Kapitel 4 wird ein Ansatz zur Beschreibung, Gestaltung und Bewirtschaftung der Unternehmensarchitektur vorgestellt, der als Grundlage für hierarchisches IT/Business Alignment dienen kann. Das abschließende Kapitel 5 fasst die Ergebnisse zusammen und diskutiert mögliche Weiterentwicklungen des Ansatzes.

2 IT/Business Alignment

Im Allgemeinen wird unter dem Begriff IT/Business Alignment der Grad der Übereinstimmung von IT-Strategie und IT-Infrastruktur auf der einen Seite mit den strategischen Geschäftszielen und der Geschäftsarchitektur (siehe Kapitel 3) auf der anderen Seite verstanden. In der Literatur finden sich zahlreiche Definitionen, die dieses unscharfe Begriffsverständnis stützen (z.B. [LuLO93; MRTG00; Nick04; ReBe00, 82]).

Je spezifischer Definitionen für IT/Business Alignment sind, desto deutlicher werden unterschiedliche, oftmals sogar gegensätzliche Auffassungen: Während die Mehrheit der Autoren Alignment als einen anzustrebenden Zustand betrachtet (z.B. [BrWe93; CHBC97]), verstehen andere Autoren darunter den Prozess, der diesen Zustand herbeiführt (z.B. [Burn97, 85; MRTG00, 19]).

Ebenso werden unterschiedliche Auffassungen hinsichtlich der Reichweite von Alignment vertreten. Während z.B. [MRTG00, 19] und [PeSo05] explizit darauf hinweisen, dass Alignment alle Ebenen der Unternehmensarchitektur (siehe Kapitel 3) betrifft, fokussieren viele andere Autoren ausschließlich auf die Strategieebene (z.B. [CHBC97]). Auch [GoGo00] merken an, dass die strategische Perspektive, d.h. Alignment auf der Ebene von IT- und Geschäftsstrategie, in der Literatur deutlich im Vordergrund steht, die operative Abstimmung zwischen IT- und Geschäftsstruktur hingegen wenig Beachtung erfährt.

Forschungsarbeiten, die sich mit den Auswirkungen des IT/Business Alignment auf Organisationen auseinandersetzen, adressieren vorwiegend drei Themenbereiche [Nick04, 80f.]:

- Identifikation von Symptomen, die auf ein mangelhaftes IT/Business Alignment in Organisationen hinweisen.
- Darlegung des Wertbeitrags, der sich durch eine optimale Abstimmung des IT-Einsatzes mit fachlichen Zielen und Anforderungen ergibt.
- Identifikation organisatorischer Prozesse, die für die Erreichung eines bestimmten Reifegrades der Übereinstimmung zwischen informationstechnologischer Realisierung und fachlichen Anforderungen notwendig sind.

Ad 1: In diversen Untersuchungen wurde eine Vielzahl an Symptomen festgestellt, die auf ein unzureichendes IT/Business Alignment in Organisationen hinweisen. Aus Sicht der IT führt eine mangelnde Abstimmung mit der Geschäftsstrategie typischerweise zu einem eingeschränkten Handlungsspielraum für die IT. Die IT wird in diesem Fall lediglich als Kostenfaktor und nicht als strategischer Wettbewerbsfaktor betrachtet. Wählt man die fachliche Perspektive, resultieren aus einem ungenügenden Alignment von IT- und Geschäftsstrategie zu geringe Wertbeiträge getätigter IT-Investitionen und damit letztlich eine verminderte Wettbewerbsfähigkeit der Organisation als Ganzes [TaKG00].

Ad 2: Viele Forschungsarbeiten widmen sich der Analyse der Nutzenpotenziale des IT- Business-Alignments. XIA und KING stellen heraus, dass sich IT/Business Alignment nicht nur sig-

nifikant positiv auf die Ausnutzung des Leistungsvermögens der IT-Infrastruktur auswirkt, sondern auch einen direkten Einfluss auf die Gesamtperformance der Organisation hat und damit alle Bereiche derselben betrifft [XiKi02]. Laut CHAN führt die Abstimmung von IT- und Geschäftsstrategie zu einem größeren Nutzen für die IT als lediglich der Abgleich formaler Organisationsstrukturen auf IT- und Fachseite [Chan02]. Im Hinblick auf den wahrgenommenen Wertbeitrag der IT resümieren TALLON, KRAEMER und GURBAXANI, dass dieser in Organisationen umso größer ist, wenn es ihnen gelingt, ihre Geschäfts- und IT-Strategie bestmöglich aufeinander abzustimmen [TaKG00].

Ad 3: Organisatorische Prozesse, die zur Erreichung eines bestimmten Alignment-Reifegrades in einer Organisation beitragen, konnten im Rahmen diverser Studien identifiziert werden. REICH und BENBASAT haben z.B. gezeigt, dass die Verbreitung von fachlichem Domänenwissen innerhalb des IT-Bereichs bzw. von IT-Domänenwissen in den Fachbereichen das kurz- als auch das langfristige Alignment beeinflussen [ReBe00]. Eine Untersuchung potenzieller Treiber und Hemmnisse im Hinblick auf die Erreichung von IT/Business Alignment durch LUFTMAN, PAPP und BRIER förderte als entscheidende, bezüglich der Zielerreichung positiv wirkende Faktoren die Unterstützung der IT durch die Unternehmensleitung, eine integrierte Entwicklung von Geschäfts- und IT-Strategie sowie das Verständnis von IT-Aspekten durch die Fachbereiche zu Tage [LuPB99]. Demgegenüber standen als schwerwiegendste Hemmnisse die mangelhafte Zusammenarbeit des IT-Bereichs mit den Fachbereichen, eine schlechte Priorisierung von Vorhaben seitens des IT-Bereichs sowie die Nichteinhaltung bestimmter Festlegungen durch den IT-Bereich.

Hinsichtlich der praktischen Umsetzung von IT/Business Alignment bleiben viele Publikationen nur vage. Oft werden lediglich grobe Gestaltungsrichtlinien vorgeschlagen. Konkrete Beispiele für Alignment-Instrumente, die tatsächlich in der Praxis genutzt werden können, finden sich nur wenige [MRTG00, 8]. Beispielsweise identifiziert [Chan02] im Rahmen einer Fallstudienuntersuchung in acht Unternehmen aus unterschiedlichen Branchen und von unterschiedlicher Größe u. a. die Verzahnung von Geschäfts- und IT-Planung, eine Beteiligung von IT-Experten an der Geschäftsplanung, die Existenz von Projektsponsoren auf der Fachseite für IT-Vorhaben sowie eine enge persönliche Beziehung zwischen CEO und CIO als erfolgskritische Instrumente für das IT/Business Alignment. Deutlich konkreter sind hingegen die Ausführungen von [GrSa01], die eine IT Balanced Scorecard als Alignment-Instrument vorgeschlagen.

Nur selten wird in der Literatur die Auffassung vertreten, dass Alignment als eine kombinierte Management- und Gestaltungsaufgabe verstanden und umgesetzt werden sollte. Der in diesem Beitrag vorgestellte Ansatz eines methoden- und modellgestützten IT/Business Alignment folgt hingegen dem Verständnis von [MRTG00, 4, 19f.], wonach Alignment einen kontinuierlichen Prozess darstellt, welcher management- und gestaltungsbezogene Subprozesse umfasst, die bewusst und nachvollziehbar zusammengehörige Business- und IT-Artefakte miteinander verknüpfen. Das Ziel dieser Verknüpfung ist dabei, einen signifikanten, dauerhaften Wertbeitrag für die Organisation zu liefern. Dem Design Research-Paradigma folgend, ist das Ziel der Gestaltungsprozesse die Konstruktion und Evaluation neuer, innovativer Artefakte, welche die Grenzen organisationaler Fähigkeiten erweitern und identifizierte Problemstellungen lösen [HMPR04, 75ff.]. Vor dem Hintergrund dieses Begriffsverständnisses ergeben sich mehrere Abweichungen gegenüber „traditionellen“ Auffassungen: IT/Business Alignment

- wird als kontinuierlicher Anpassungsprozess und nicht als statischer Zustand betrachtet,
- umfasst und erfordert den Abgleich auf und zwischen allen Ebenen der Unternehmensarchitektur (vgl. Kapitel 3) und nicht nur eine Abstimmung auf strategischer Ebene,
- schließt neben Managementaktivitäten auch Konstruktionshandlungen ein, d.h. besteht aus Management- und Gestaltungsprozessen,
- wirft als konsistente Gesamtgestaltung des Gebildes „Unternehmung“ im Sinne von Punkt 2 ein Komplexitätsproblem auf: Der Abgleich auf und zwischen allen Ebenen der Unternehmensarchitektur kann aufgrund der Mannigfaltigkeit der Zusammenhänge zwischen den einzelnen Elementen nicht simultan erfolgen. Dieser Problematik muss durch einen mehrstufigen, hierarchischen Gestaltungsansatz Rechnung getragen werden. Ein solcher Gestaltungsansatz wird in Kapitel 4 beschrieben.

3 Unternehmensarchitektur

Gemäß ANSI/IEEE Std 1471-2000 [IEEE00] ist Architektur definiert als “The fundamental organization of a system, embodied in its components, their relationships to each other and the

environment, and the principles governing its design and evolution". Unter Unternehmensarchitektur (Enterprise Architecture) wird deshalb die fundamentale Strukturierung eines Unternehmens oder einer Behörde verstanden, sei es als Ganzes, zusammen mit Lieferanten und/oder Kunden oder als Teil, aber in jedem Fall ergänzt durch die entsprechenden Gestaltungs- und Weiterentwicklungsregeln [TOG03].

Die Grundlage für die Modellierung einer Unternehmensarchitektur bilden Metamodelle und Konstruktionsregeln. Während Metamodelle die zulässigen Modellelementtypen und deren Beziehungen zueinander spezifizieren, soll die Anwendung der Konstruktionsregeln die systematische Konstruktion sinnvoller Unternehmensarchitekturen unterstützen.

3.1 Unternehmensarchitektur als mehrstufiges, hierarchisches System

Die Unternehmensarchitektur stellt immer ein Gesamtbild des zu beschreibenden Teils des Unternehmens bzw. der Behörde dar [Sche04, 14]. Entsprechend der Ausführungen in Kapitel 1 stellen damit die obersten Beschreibungs-Strata eines Unternehmens den Betrachtungsgegenstand der Unternehmensarchitektur dar.

Da auch bei aggregierter Betrachtung die Vielzahl der in der Unternehmensarchitektur abgebildeten Artefakte eine simultane Gestaltung verhindert, ist ein sukzessives Konstruktionsverfahren erforderlich. Dazu werden mehrere Layers (im Sinne von nacheinander zu durchlaufenden Architekturebenen) gebildet. Unternehmensarchitektur umfasst damit nicht nur einige Aggregationsebenen, sondern muss zusätzlich als System hierarchisch aufeinander aufbauender (Teil-)Architekturen verstanden werden, die durch zunehmende Implementierungsnähe charakterisiert sind [Hafn05, 28]. Zur Gestaltung und Weiterentwicklung der Unternehmensarchitektur wird üblicherweise ein sukzessives Vorgehen beschrieben, das zunächst auf die strategische Positionierung aus betriebswirtschaftlicher Sicht fokussiert, auf dieser Grundlage eine geeignete Ablauf- und Aufbauorganisation spezifiziert, dann das Informationssystem, d.h. das Zusammenspiel menschlicher und maschineller Aufgabenträger zur Unterstützung der fachlichen Anforderungen betrachtet und schließlich die zur Umsetzung benutzten Technologieartefakte beschreibt [Wint05, 2].

Wenn die Unternehmensarchitektur im Sinne einer hierarchischen Gestaltung bzw. Weiterentwicklung des Gesamtsystems „Unternehmen“ zur Ausrichtung anderer, spezifischer bzw. detaillierter Teilmodelle (z.B. Sicherheitsarchitektur, Datenarchitektur, detaillierte Ablaufmodelle, Softwarearchitektur) aufeinander benutzt wird, hat sie daneben auch den Charakter eines (übergeordneten) Echelons. Die Teilmodelle werden zunächst nebenläufig gestaltet und weiterentwi-

ckelt; Aus der Unternehmensarchitektur resultieren aber Anpassungen, die die Konsistenz des Gesamtsystems bewahren.

Da fachliche Aspekte und IT-Aspekte nicht in sinnvollem Detail umfassend und simultan aufeinander ausgerichtet werden können, bedarf es eines hierarchischen, mehrstufigen Gestaltungs- und Weiterentwicklungsprozesses. Wir verstehen Unternehmensarchitektur als den Kern eines solchen Prozesses. Unternehmensarchitektur bildet die obersten Strata, alle wichtigen Gestaltung-Layer und das Koordinations-Echelon des hierarchischen, mehrstufigen Systems „konsistente Gestaltung des Gesamtunternehmens“ [WiFi06] (Abbildung 1).

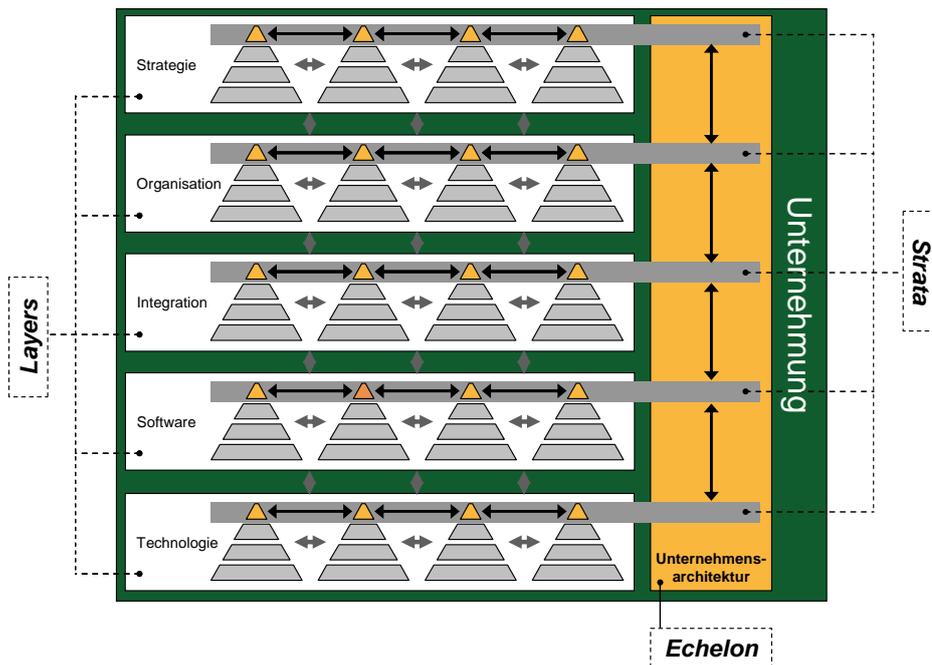


Abbildung 1: Unternehmensarchitektur als mehrstufiges, hierarchisches System

3.2 Inhalte, Bedeutung und Ziele der Unternehmensarchitektur

Das weit verbreitete Unternehmensarchitektur-Framework der Open Group (TOGAF – The Open Group Architecture Framework), unterscheidet in der Version 8.1 [TOG03] die Layers (1) Business Architecture, (2) Data/information Architecture, (3) Application Architecture sowie (4) Technology (IT) Architecture. TOGAF ist hinsichtlich dieser Aufteilung nahezu deckungsgleich mit dem Federal Enterprise Architecture Framework (FEAF, [CIOC99]). FEAF wird von einer Vielzahl der Bundesbehörden in den USA genutzt und differenziert die Layers (1) Business Architecture und (2) IS Architecture, wobei diese entsprechend der drei Aspekte Data Architecture, Application Architecture und Technology Architecture strukturiert ist. Auch ZACHMAN stellt im Zusammenhang mit seinem bekannten Framework for Information Systems Ar-

chitecture (Zachman-Framework [SoZa92; Zach87]) als Kernaussage fest, dass es nicht eine einzige Architektur für das Informationssystem des Unternehmens gibt, sondern eine Menge architektonischer Repräsentationen. Einzelne Repräsentationen sind dabei auf verschiedenen Zwecke und Zielgruppen ausgerichtet. ZACHMAN unterscheidet die fünf Layers (1) Scope, (2) Enterprise Model, (3) System Model, (4) Technology Model und schließlich (5) Components, für die jeweils verschiedene Aspekte differenziert werden.

Sämtlichen Ansätzen ist gemeinsam, dass – ganz im Sinne eines hierarchischen, mehrstufigen Gestaltungs- und Weiterentwicklungsansatzes – die Informationssystem-Gestaltung fachlichen Anforderungen folgt und damit die Ergebnisse jeder Architekturgestaltung die Freiheitsgrade der nachfolgenden Gestaltungsebenen reduzieren [Wint03c, 92ff.].

Zusammenfassend kann festgehalten werden, dass als wichtigste Gesamtzusammenhänge in Organisationen von den meisten Autoren

- die Geschäftsarchitektur (Gesamtzusammenhang der strategischen Positionierung in Leistungsnetzwerken),
- die Prozessarchitektur (Gesamtzusammenhang der Leistungsentwicklung, Leistungserstellung und des Leistungsvertriebs in einer Organisation),
- die Integrationsarchitektur (Gesamtzusammenhang der Zuordnung von Informationssystem-Funktionalitäten bzw. -Diensten zu fachlichen Aktivitäten bzw. Informationsbedarfen),
- die Softwarearchitektur (Gesamtzusammenhang der Softwarekomponenten und Datenstrukturen) sowie
- die IT-Architektur (Gesamtzusammenhang der technischen Komponenten des Informationssystems)

betrachtet werden [Wint03c, 93ff.] (Abbildung 1). Zusätzlich können innerhalb der einzelnen Layers verschiedene Architektursichten eingeführt werden, wenn die Komplexität der abzubildenden Sachverhalte dies erfordert [Krcm90; Sinz99, 1036]. Während z. B. das ISA-Modell die fachlichen Layers noch ganzheitlich abbildet und erst auf Informationssystem-Ebene verschiedene Sichten unterscheidet [Krcm90], differenziert das Zachman-Framework auf allen Layers sechs Sichten [Zach87]. Innerhalb der Layers und Sichten wird häufig die Bildung zusätzlicher Strata im Sinne von Abstraktionshierarchien vorgeschlagen, z. B. für die mehrstufige Abbil-

dung von Produkt- und Prozessmodellen sowie von Kennzahlenmodellen oder Softwarekomponentenmodellen.

Die Bedeutung der Unternehmensarchitektur resultiert aus der strategischen Relevanz der Informationsverarbeitung, der Langfristigkeit der Aktivitäten bei der Nutzung von Informations- und Kommunikationstechnologien, der Notwendigkeit der Gesamtschau des Unternehmens und der Erfordernis der unternehmerischen Führung der Ressource Information und Kommunikation. Treffend konstatiert daher SINZ, dass sich „das Management umfassender Unternehmensarchitekturen [...] mittlerweile zu einem wichtigen Aufgabenbereich auf strategischer und operativer Ebene entwickelt [hat]“ [Sinz04].

Kapitel 4 beschreibt im Folgenden einen hierarchischen Gestaltungsansatz für das IT/Business Alignment, bei dem die Unternehmensarchitektur als wichtigste Komponente im Vordergrund steht.

4 Hierarchischer Gestaltungsansatz für das IT/Business Alignment

In Kapitel 1 wurde ein mehrstufiger, hierarchischer Gestaltungsansatz als geeignete Herangehensweise identifiziert, um das organisatorische Gestaltungsproblem des IT/Business Alignment effizient lösen zu können. In Kapitel 2 wurde gefordert, dass IT/Business Alignment als kontinuierlicher Prozess gestaltet und etabliert werden sollte, der Management- und Gestaltungsprozesse umfasst, um bewusst und nachvollziehbar zusammengehörige Business- und IT-Artefakte auf allen Betrachtungsebenen miteinander zu verknüpfen. In Kapitel 3 wurden als relevante Betrachtungsebenen (Layers) die Strategiebene (Geschäftsarchitektur), die Organisationsebene (Prozessarchitektur), die Integrationsebene (Integrations- bzw. Applikationsarchitektur) sowie die Systemebenen (Softwarearchitektur, IT-Architektur) identifiziert.

Im Folgenden werden zunächst die wesentlichen Artefakttypen innerhalb der einzelnen Betrachtungsebenen vorgestellt, die den o. g. Anforderungen entsprechen. Im zweiten Abschnitt werden Prozesse skizziert, welche diese Artefakte bewirtschaften und damit IT/Business Alignment im hier geforderten Verständnis unterstützen.

4.1 Zentrale Artefakte der Unternehmensarchitektur

Die hier vorgestellten Unternehmensarchitektur-Artefakttypen resultieren aus einer Untersuchung in [WiFi06], deren Ergebnisse auf dem Vergleich weit verbreiteter Unternehmensarchi-

tektur-Frameworks und der darin enthaltenen Artefakttypen basieren. In [BrWi05] wurde ein Metamodell vorgestellt, das eine Vielzahl der vorgeschlagenen Artefakttypen und deren Beziehungen umfasst.

Im Gegensatz zu vielen anderen Unternehmensarchitektur-Ansätzen fordern [WiFi06] und [BrWi05] explizit, dass keines der behandelten Artefakte in vollem Detaillierungsgrad repräsentiert werden sollte. Vielmehr ist es nach diesem Unternehmensarchitektur-Verständnis das Ziel, alle relevanten Artefakte auf hohem Aggregationsgrad zu repräsentieren, um sich dadurch auf die konsistente Gestaltung und korrekte Repräsentation der Zusammenhänge zwischen diesen Artefakten konzentrieren zu können. Für jeden inhaltlichen Cluster von Artefakten sind dann Schnittstellen zu den jeweiligen operativen Bestandsführungssystemen bereitzustellen.

In den folgenden Abschnitten werden die zentralen Artefakttypen der drei obersten Gestaltungsbzw. Beschreibungs-Layer Strategieebene, Organisationsebene und Integrationsebene vorgestellt. Die Darstellung basiert auf der Detaildarstellung in [WiFi06] und dem Metamodell in [BrWi05]. Auf die Aufzählung der jeweiligen Beziehungen zwischen den Artefakttypen eines Beschreibungs-Layers wird hier aus Platzgründen verzichtet.

4.1.1 Strategieebene

Auf Strategieebene steht die Spezifikation der strategischen Ausrichtung der Unternehmung im Vordergrund. Im Einzelnen sind die folgenden Artefakttypen Gegenstand der Betrachtung:

- Zielsystem der Unternehmung als Hierarchie von Organisationszielen, Erfolgsfaktoren, Kennzahlen und strategischen Maßnahmen/Projekten
- Produkt-/Servicemodell der Unternehmung einschließlich des Leistungsaustausches mit anderen Geschäftseinheiten, Kunden und Lieferanten innerhalb von Geschäftsnetzwerken
- Kernkompetenzen, Marktsegmente, Vertriebs- bzw. Distributionskanälen
- Geschäftsgrundsätze

4.1.2 Organisationsebene

Auf Organisationsebene werden – ausgehend vom zuvor spezifizierten Leistungs- und Zielsystem – Abläufe, Verantwortlichkeiten und aufbauorganisatorische Strukturen festgelegt. Das generelle Gestaltungsziel der Organisationsebene sind Effektivität und Effizienz von Leistungs-

entwicklung, Leistungserbringung und Leistungsvertrieb. Wichtige Ergebnisse der Organisationsgestaltung sind:

- Spezifikation der Aufbauorganisation: Organigramme, Rollenbeschreibungen inklusive Qualifikationsprofilen sowie hierarchische Übersichtsdarstellungen der Unternehmensstandorte
- Spezifikation der Ablauforganisation: Funktionsmodelle, Prozessmodelle einschließlich Eingangs- und Ausgangsgrößen sowie Service-Level-Vereinbarungen, Workflowspezifikationen, Performanceindikatoren
- Spezifikation der Informationslogistik: Informationslandkarten und Informationsflüsse auf aggregiertem Niveau

4.1.3 Integrationsebene

Auf Integrationsebene werden – ausgehend von den zuvor spezifizierten Organisationsstrukturen, Informationsbedarfen und Prozessabläufen – Integrationsbereiche innerhalb des Informationssystems bestimmt. Im Normalfall wird ein eng gekoppelter Integrationsbereich durch eine Applikation realisiert, während lose Kopplungen zwischen Integrationsbereichen durch Schnittstellen zwischen Applikationen realisiert werden. Das generelle Gestaltungsziel der Applikationsebene ist die optimale Integration (bzw. Entkopplung) von Applikationen aus fachlicher Sicht. Wichtige Ergebnisse der Architekturgestaltung auf Integrationsebene sind:

- Spezifikation von Applikationen (inkl. Informations-, Leistungs-, Daten- und Kontrollflüssen), Applikationskomponenten und Applikationsdomänen
- Spezifikation fachlicher Services (Enterprise Services) einschließlich Servicehierarchisierung.

4.1.4 Abhängigkeiten zwischen den Architekturebenen

Ein hierarchischer Ansatz zur Unterstützung von IT/Business Alignment basiert auf der konsistenten Verknüpfung fachlicher Artefakte auf der einen Seite mit IT-bezogenen Artefakten auf der anderen Seite. Im Hinblick auf die in den vorhergehenden Abschnitten vorgestellten Unternehmensarchitektur-Artefakttypen finden sich fachliche Artefakte hauptsächlich auf Strategie- und Organisationsebene, während sich IT-bezogene Artefakte hauptsächlich auf der Integrati-

onsebene sowie den – hier aus Platzgründen nicht ausgeführten – Systemebenen Softwareebene und IT-Ebene. Der größte Nutzen eines ganzheitlichen wie in [BrWi05] vorgestellten Unternehmensarchitektur-Metamodells für das IT/Business Alignment entfaltet sich somit durch ebenenübergreifende Verknüpfungen.

Im Folgenden werden die wichtigsten Verknüpfungen zwischen den Artefakttypen der verschiedenen in diesem Beitrag behandelten Architekturebenen in Anlehnung an das Metamodell in [BrWi05, 74] beschrieben:

(a) Abhängigkeiten zwischen Strategie- und Organisationsebene:

- Eine Geschäftseinheit nimmt eine oder mehrere Rollen in einem oder mehreren Wertschöpfungsnetzwerken wahr. Die auf Organisationsebene definierten Prozesse müssen sich jeweils auf eine oder mehrere dieser Rollen beziehen.
- Prozessleistungen (als Ergebnis von Teilprozessen bzw. Aktivitäten) in Prozessablaufmodellen beziehen sich direkt oder indirekt auf entsprechende Spezifikationen in den Leistungsmodellen der jeweiligen Geschäftseinheiten.
- Führungsgrößen in Prozessführungsmodellen beziehen sich direkt oder indirekt auf Kennzahlen in den Zielsystemmodellen der jeweiligen Geschäftseinheiten.
- Serviceaktivitäten (als Teilprozesse bzw. Aktivitäten) in Prozessablaufmodellen beziehen sich direkt oder indirekt auf (selbsterstellte) Leistungen in Kundenprozessmodellen.

(b) Abhängigkeiten zwischen Organisations- und Integrationsebene:

- Geschäftsfunktionen bzw. Informationsobjekte aus den entsprechenden Teilmodellen der Fachkonzeptspezifikation beziehen sich direkt oder indirekt auf Funktionen bzw. Informationsobjekte in Prozessablaufmodellen.
- Applikationen in Applikationslandschaftsmodellen referenzieren direkt oder indirekt auf Applikationen in Prozessablaufmodellen.
- Auf die unter (a) charakterisierten Geschäfts- bzw. Organisationseinheiten beziehen sich wiederum die Spezifikationen von Ownership und Informationsnutzung auf Applikationsebene.

Direkte Beziehungen liegen vor, wenn referenzierendes und referenziertes Artefakt auf der gleichen Aggregationsebene spezifiziert sind. Indirekte Beziehungen liegen dagegen dann vor,

wenn das referenzierende Artefakt auf einer anderen Aggregationsstufe spezifiziert ist als das referenzierte Artefakt.

4.2 Bewirtschaftungs- und Kommunikationsprozesse

Die Prozesse zur Bewirtschaftung und Kommunikation der Architekturartefakte sind mit der Zielstellung verbunden, die auf Basis der Artefakttypen aus Kapitel 4.1 entwickelten Teilmodelle der Strategie- und Organisationsebene auf der einen Seite sowie die Teilmodelle der Integrationsebene und den Systemebenen auf der anderen Seite dauerhaft effizient und effektiv miteinander zu koordinieren, um auf diese Weise IT/Business Alignment im hier geforderten Verständnis zu unterstützen.

Für die Bewirtschaftung von Unternehmensarchitekturen existieren verschiedene Ansätze, die aber in den meisten Fällen nicht die notwendige Reife und den erforderlichen Detaillierungsgrad aufweisen [Hafn05]. Die folgenden Ausführungen basieren auf der Architectural Development Method (ADM, Bestandteil von TOGAF [TOG03] - siehe Kapitel 3) sowie den von HAFNER und WINTER [HaWi05] vorgeschlagenen Bewirtschaftungs- und Kommunikationsprozessen. Diese beiden Ansätze wurden ausgewählt, weil sie methodisch ausreichend detailliert dokumentiert sind, gegenüber vorhandenen Architektur-Frameworks neutral sind. Damit ist ihre (potenzielle) Anwendbarkeit im Kontext unterschiedlichster Rahmenbedingungen nicht eingeschränkt.

ADM rückt den Top-Down-Entwurf von Unternehmensarchitekturen in den Mittelpunkt. Es wird ein zyklischer, in acht Phasen gegliederter Prozess für die Entwicklung und Pflege der Unternehmensarchitektur einer Organisation vorgeschlagen. Hingegen strukturieren [HaWi05] das Architekturmanagement in die vier miteinander verbunden Hauptprozesse Architektur-Führung, Architektur-Entwicklung, Architektur-Kommunikation und Architektur-Vertretung.

Die ersten sieben Phasen der ADM setzen sich ausführlich mit der Umsetzung einer zu Beginn des Architekturzyklus aufgestellten Zielarchitektur auseinander. Das Vorgehen hierbei kann als Sukzessivgestaltung im Sinne des hierarchischen Ansatzes charakterisiert werden. Bei [HaWi05] korrespondieren die Hauptprozesse Architektur-Führung und Architektur-Entwicklung mit den zuvor genannten Phasen der ADM. Innerhalb des Prozesses Architektur-Führung werden strategische Anforderungen identifiziert und zusätzlich dazu innerhalb des Prozesses der Architektur-Entwicklung kontinuierlich operative Anforderungen in die Betrachtung

tung einbezogen. Die Anforderungen werden konsolidiert und priorisiert, so dass den Bedürfnissen entsprechende, neue Architekturartefakte entwickelt werden können.

Durch den Prozess Architektur-Kommunikation bei [HaWi05] werden die Anspruchsgruppen für Schulungen, Informationsmaterial, usw. identifiziert und diesen bedarfsgerecht Informationen über die Architektur zur Verfügung gestellt. Informationen aus der Kommunikation der Architektur und ihrer konkreten Durchsetzung bilden die Grundlage für die Bewertung der Verbreitung und Wirksamkeit der Architektur. Letztlich können daraus neue Anforderungen an das Architekturmanagement resultieren und einen erneuten Durchlauf der Prozesse Architekturführung bzw. -Entwicklung anstoßen. ADM adressiert den Kommunikationsaspekt nicht explizit durch eine eigene Phase. Jedoch widmet sich die ADM-Phase acht der Pflege und dem Unterhalt der Unternehmensarchitektur. In dieser Phase zeichnet das Architekturmanagement für die kontinuierliche Überwachung neuer Entwicklungen verantwortlich, die zu geringfügigen Anpassungen der Architektur oder gar einem Zyklusneustart führen können.

Gemeinsam ist beiden Ansätzen die Berücksichtigung der Architektursteuerung. Bei ADM übt das Architekturmanagement im Rahmen der Implementierung der technischen Lösung eine Governance-Funktion aus, um die Architekturkonformität des realisierten Systems zu wahren. Nach [HaWi05] bietet der Hauptprozess Architektur-Vertretung Unterstützung in Form von reinen Beratungsleistungen bis hin zu direkter Projektmitarbeit für Projektvorhaben, welche architekturbezogene Fragestellungen berühren.

Kritisch muss in Bezug auf ADM angemerkt werden, dass der eindeutige Schwerpunkt der Methode auf der Ausgestaltung der Technologie-/IT-Architektur liegt, was sich anhand des Detaillierungsgrades der Dokumentation dieser Phase offenbart.

5 Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurde Unternehmensarchitektur als wichtige Komponente eines mehrstufigen, hierarchischen Gestaltungsansatzes zur konsequenten Umsetzung von IT/Business Alignment vorgestellt. Den Ausgangspunkt bildete die Auseinandersetzung mit der Theorie hierarchischer, mehrstufiger Systeme nach [MeMT70] als konzeptionelle Grundlage. Im Rahmen dieser Betrachtung wurden die Vorteilhaftigkeit einer hierarchischen Organisationsform für das IT/Business Alignment gezeigt und ein zentrales Unternehmensmodell als Kernkomponente eines entsprechenden Ansatzes vorgeschlagen.

Als Weiterentwicklung bestehender Ansätze der Unternehmensmodellierung sollte ein solches Unternehmensmodell geschäftsbezogene und IT-Artefakte gleich gewichten, alle relevanten Artefakte auf einem hohen Abstraktionsniveau repräsentieren und die Zusammenhänge zwischen den Artefakten unterschiedlicher Modellierungsebenen explizit abbilden.

Zur Bewirtschaftung dieser Artefakte wurden Bewirtschaftungs- und Kommunikationsprozesse skizziert, die das in diesem Beitrag geforderte Verständnis von IT/Business Alignment unterstützen.

Die Validität sowohl der vorgestellten Unternehmensarchitektur-Artefakte wie auch der entsprechenden Bewirtschaftungs- und Kommunikationsprozesse wird zurzeit in mehreren Pilotprojekten untersucht. In einem dieser Pilotprojekte werden auch die IT-Governance-Prozesse (einschl. IT/Business Alignment als wichtige Komponente) neu gestaltet, wobei die bestehenden Reifegradmodelle eine große Rolle spielen. Erfahrungen aus diesen Umsetzungen werden in die Verfeinerung des hier vorgestellten Konzepts einfließen. Insbesondere ist geplant, den Referenzcharakter von Bewirtschaftungs- und Kommunikationsprozessen der Unternehmensarchitektur zu analysieren und damit eine Grundlage zu schaffen, um die Hypothese einer besseren Eignung des hierarchischen Ansatzes für IT/Business Alignment auch auf breiterer empirischer Basis zu untersuchen.

Literatur

- [BrWi05] *Braun, Christian; Winter, Robert: A Comprehensive Enterprise Architecture Metamodel and Its Implementation Using a Metamodeling Platform. In: Proceedings, GI Workshop on Enterprise Modelling and Information Systems Architectures, Klagenfurt 2005.*
- [BrWe93] *Broadbent, Marianne; Weill, Peter: Improving Business and Information Strategy Alignment: Learning from the Banking Industry. In: IBM Systems Journal 32 (1993) 1, S. 162-179.*
- [Burn97] *Burn, Janice M.: A Professional Balancing Act: Walking the Tightrope of Strategic Alignment. In: Sauer, Christopher; Yetton, Philip W. (Hrsg.): Steps to the Future: Fresh Thinking on the Management of IT-Based Organizational Transformation. Jossey-Bass, San Francisco, California 1997, S. 55-88.*
- [Chan02] *Chan, Yolande E.: Why haven't we mastered alignment? The importance of the informal organization structure. In: MIS Quarterly Executive 1 (2002) 2, S. 97-112.*

- [CHBC97] *Chan, Yolande E.; Huff, Sid L.; Barclay, Donald W.; Copeland, Duncan G.*: Business Strategic Orientation, Information Systems Strategic Orientation, and Strategic Alignment. In: *Information Systems Research* 8 (1997) 2, S. 125-150.
- [CIOC99] *CIO-Council*: Federal Enterprise Architecture Framework Version 1.1. September 1999.
- [GoGo00] *Gordon, Judith R.; Gordon, Steven R.*: Structuring the Interaction between IT and Business Units: Prototypes for Service Delivery. In: *Information Systems Management* 17 (2000) 1, S. 7-16.
- [GrSa01] *Grembergen, Wim van; Saull, Ronald*: Aligning Business and Information Technology through the Balanced Scorecard at a Major Canadian Financial Group: its Status Measured with an IT BSC Maturity Model. In: *Proceedings, 34th Hawaii International Conference on Systems Sciences, Maui, Hawaii 2001*.
- [Hafn05] *Hafner, Martin*: Entwicklung einer Methode für das Management der Informationssystemarchitektur im Unternehmen. Dissertation, Universität St. Gallen, St. Gallen 2005.
- [HaWi05] *Hafner, Martin; Winter, Robert*: Vorgehensmodell für das Management der unternehmensweiten Applikationsarchitektur. In: *Proceedings, 7. Internationale Tagung Wirtschaftsinformatik 2005, Bamberg 2005*, S. 627-646.
- [HMPR04] *Hevner, Alan R.; March, Salvatore T.; Park, Jinsoo; Ram, Sudha*: Design Science in Information Systems Research. In: *MIS Quarterly* 28 (2004) 1, S. 75-105.
- [IEEE00] *IEEE*: IEEE Recommended Practice for Architectural Description of Software Intensive Systems (IEEE Std 1471-2000). IEEE Computer Society, New York, NY 2000.
- [Krcm90] *Krcmar, Helmut*: Bedeutung und Ziele von Informationssystemarchitekturen. In: *Wirtschaftsinformatik* 32 (1990) 5, S. 395-402.
- [Luft00] *Luftman, Jerry N.*: Assessing Business-IT Alignment Maturity. In: *Communications of the Association for Information Systems* 4 (2000) December 2000.
- [Luft05] *Luftman, Jerry N.*: Key Issues for IT Executives 2004. In: *MIS Quarterly Executive* 4 (2005) 2, S. 269-285.
- [LuLO93] *Luftman, Jerry N.; Lewis, Paul R.; Oldach, Scott H.*: Transforming the enterprise: The alignment of business and information technology strategies. In: *IBM Systems Journal* 32 (1993) 1, S. 198-221.
- [LuPB99] *Luftman, Jerry N.; Papp, Raymond; Brier, Tom*: Enablers and inhibitors of business-IT alignment. In: *Communications of the AIS* 1 (1999) 3, S. 2-32.
- [MRTG00] *Maes, Rik; Rijsenbrij, Daan; Truijens, Onno; Goedvolk, Hans*: Redefining business - IT alignment through a unified framework. *PrimaVera Working Paper 2000-19*, Universiteit van Amsterdam, Amsterdam 2000, S. 25.
- [MeMT70] *Mesarovic, M.D.; Macko, D.; Takahara, Y.*: Theory of Hierarchical, Multilevel, Systems. Academic Press, New York, London 1970.
- [Nick04] *Nickels, David W.*: IT-Business Alignment: What We Know That We Still Don't Know. In: *Proceedings, The 7th Annual Conference of the Southern Association for Information Systems, Savannah Marriott Riverfront 2004*, S. 79-84.

- [PeSo05] *Pereira, Carla Marques; Sousa, Pedro*: Enterprise Architecture: Business and IT Alignment. In: Proceedings, 2005 ACM symposium on Applied computing (SAC 2005), Santa Fe, New Mexico 2005, S. 1344-1345.
- [ReBe00] *Reich, Blaize Horner; Benbasat, Izak*: Factors That Influence the Social Dimension of Alignment Between Business and Information Technology Objectives. In: MIS Quarterly 24 (2000) 1, S. 81-113.
- [Sche04] *Schekkerman, Jaap*: How to Survive in the Jungle of Enterprise Architecture Frameworks: Creating or Choosing an Enterprise Architecture Framework. 2. Aufl., Trafford Publishing, Victoria, British Columbia 2004.
- [Sinz99] *Sinz, Elmar J.*: Architektur von Informationssystemen. In: *Rechenberg, Peter; Pomberger, Gustav (Hrsg.)*: Informatik-Handbuch. Hanser, München 1999, S. 1035-1046.
- [Sinz04] *Sinz, Elmar J.*: Unternehmensarchitekturen in der Praxis - Architekturdesign am Reissbrett vs. situationsbedingte Realisierung von Informationssystemen. In: Wirtschaftsinformatik 46 (2004) 4, S. 315-316.
- [SoZa92] *Sowa, John F.; Zachman, John A.*: Extending and formalizing the framework for information systems architecture. In: IBM Systems Journal 31 (1992) 3, S. 590-616.
- [TaKG00] *Tallon, Paul P.; Kraemer, Kenneth L.; Gurbaxani, Vijay*: Executives' Perceptions of the Business Value of Information Technology: A Process-Oriented Approach. In: Journal of Management Information Systems 16 (2000) 4, S. 145-173.
- [TOG03] *The Open Group*: TOGAF (The Open Group Architecture Framework) Version 8.1 "Enterprise Edition". <http://www.opengroup.org/architecture/togaf8/>, Abruf am 2006-03-13.
- [Wint91] *Winter, Robert*: Mehrstufige Produktionsplanung in Abstraktionshierarchien auf der Basis relationaler Informationsstrukturen. Dissertation, Springer, Berlin et al. 1991.
- [Wint03] *Winter, Robert*: Modelle, Techniken und Werkzeuge im Business Engineering. In: *Österle, Hubert; Winter, Robert (Hrsg.)*: Business Engineering. Springer, Berlin et al. 2003, S. 87-118.
- [Wint05] *Winter, Robert*: Unternehmensarchitektur und Integrationsmanagement. In: *Sokolovsky, Zbynek; Löschenkohl, Sven (Hrsg.)*: Handbuch Industrialisierung der Finanzwirtschaft. Gabler, Wiesbaden 2005, S. 575-599.
- [WiFi06] *Winter, Robert; Fischer, Ronny*: Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. In: Proceedings, EDOC Workshop on Trends in Enterprise Architecture Research (TEAR 2006) within The Tenth IEEE International EDOC Conference (EDOC 2006), Hong Kong 2006.
- [XiKi02] *Xia, Weidong; King, William R.*: Determinants of Organizational IT Infrastructure Capabilities: An Empirical Study. http://misrc.umn.edu/workingpapers/fullpapers/2002/0210_030102.pdf, Abruf am 2005-12-16.
- [Zach87] *Zachman, John A.*: A Framework for Information Systems Architecture. In: IBM Systems Journal 26 (1987) 3, S. 276-292.

Fostering the Evaluation of Reference Models: Application and Extension of the Concept of IS Design Theories

Michael Schermann, Tilo Böhmman, Helmut Krcmar

Chair for Information Systems
Technische Universität München
Boltzmannstraße 3
85748 Garching b. München
{michael.schermann|boehmann|krcmar}@in.tum.de

Abstract

Despite the imperative to substantiate innovative research results expressed in reference models, little methodical guidance exists for evaluating reference models yet. We propose that IS design theories [WaWE92] can provide theoretical guidance for reference model evaluation since reference models can be formulated as a set of design principles that consist of testable propositions, kernel theories, and intended applications. We show how to facilitate the reconstruction process by applying the idea of pattern languages. Such decomposed reference models allow evaluating each design principle separately and thus formulating a more concise and elementary evaluation objective. We demonstrate the benefits of reconstructing reference models as design theories on the Service Data Management reference model that has been developed by the authors.

1 Introduction, Problem Statement, and Challenges

Since the beginning of the Information systems discipline in German-speaking countries many reference models have been constructed and published [FeLo04b], among them most notably Scheer's Y-CIM or Becker's Retail-H [BeSc04; Sche98]. The term reference model has been adopted by companies in many industries (e.g. Software, Health, Banking) to denote best practices in process design and software design [BeKn02; FeLo04a]. Based on the importance of reference models in the German IS community, reference models and the process of

reference modeling have become research objects themselves, e.g. by supporting the adaptation of reference models or facilitating the management of reference models [BeDK04; Thom06].

Choosing reference models and substantiating their claims require sound evaluation. With the number of reference models rising, potential users, e.g. companies, are faced with a problem of choosing references models and hence evaluating the quality and appropriateness of potential usefulness [FeLo04b]. Furthermore, researchers want to evaluate the utility of their reference models and thus substantiate their proposed claims of reference [BöSK06]. The fundamental claim of reference models is that they accelerate model-based development phases, e.g. requirements engineering and system design, by adapting the reference model instead of pursuing individual modeling [BeSc04; FeLo04a]. Thus, reference models usually have a prescriptive notion as they propose how information systems or processes *should* be designed. However, most of the available reference models lack of evaluation results regarding their utility, suitability, and quality [BöSK06; FeLo04b].

Despite the importance of evaluating reference models, little methodical guidance exists for evaluating reference models yet. In their effort to facilitate evaluation, researchers have found that evaluating information models and particularly reference models is especially difficult due to methodological, philosophical, and practical reasons [BöSK06; FeLo03b; Fran00; Fran98a]. First, evaluating the utility of reference models in a positivist understanding would require gaining access to a large number of users that actually have applied a reference model to reduce the impact of confounding factors in the reference model evaluation. Second, reference models are supposed to be adapted to specific needs of the reference model user. Thus, evaluating reference models has to cope with a large number of confounding factors [BöSK06]. Third, constructors of reference models often do not reveal underlying assumptions, theoretical foundations, as well as the immutable core of their reference models [BePf06]. Overall, reference models do not yet provide necessary elements to evaluate the utility and their claims.

Reconstructing reference models as design theories provides the missing link. Design theories have been proposed as scientific method to capture design experience and provide prescriptive information on how to design information systems in specific domains [MaMG02; WaWE92]. We argue that design theories can be seen as a counter piece to reference models. An important aspect of design theories is to reveal underlying theoretical assumptions as well as provide hypotheses that can be refuted or substantiated in empirical analyses. Thus, reconstructing reference models by the structure of design theories facilitates the identification of design

propositions in reference models. Furthermore, the methodology of developing design theories requires linking these design propositions to so-called kernel theories that provide the theoretical base for design theories. Reconstructing reference models as design theories also requires the proposition of testable hypotheses and thus facilitate the evaluation of reference models and hence the theoretical and practical advancement of reference models.

Overall, we propose a way of reconstructing reference models as testable theories as it has been demanded, e.g. by Becker and Pfeiffer [BePf06]. Hence, we set the following research questions:

- What are the benefits of applying the concept of design theory to reference models?
- How can we facilitate the process of reconstruction?
- What are the benefits of reconstructing reference models as design theories?

The remainder of this paper is organized as follows. In section 2 we introduce the concept of design theory. As reference models are complex design proposals, it is necessary to decompose them into design principles [MaMG02]. To facilitate the decomposition process, we introduce the idea of patterns in section 3 [Alex79]. We further show that the structure of patterns helps to identify required elements of design theories. *Reconstructing a reference model will result in a pattern language that consists of the design principles proposed by the reference models. Each pattern reflects a design principle that can be evaluated individually.* In section 4 we demonstrate the utility of our approach on the example of the SDM reference model that has been developed by the authors [BWFK04]. The paper finishes with a conclusion of the results and provides an outlook on further research. Figure 1 shows the main arguments of this paper.

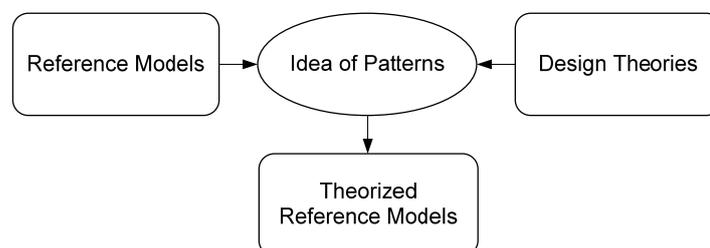


Figure 1: The line of argumentation of this paper

This paper is of exploratory and conceptual nature. Hence, we provide argumentative support when answering our research questions. However, we base our argument upon available empirical and conceptual research results.

2 Design Theories

In this section, we introduce design theories as a methodology of capturing domain knowledge and design experiences in an empirical refutable way. Furthermore, we analyze similarities and differences of design theories and reference models. We will conclude that applying the structure of design theories allows reference model constructors to explicate underlying theoretical assumptions and provide testable hypotheses. This reveals the benefit proposals of the reference model, and thus fosters the academic and practical evaluation of reference models.

2.1 Characteristics of Design Theories

Designing and developing new information systems, e.g. to improve business processes by automation or to enable new ways of doing business is an integral part of work for both IS researchers and IS practitioners [HMPR04; Mert95; Wiss94]. Therefore, one of the pivotal research objectives of IS researchers is to provide theories and practical guidance on facilitating efficient and effective design of information systems. Grounding on the seminal paper by Walls et al. [WaWE92] various authors have used the construct of design theories as a vehicle for capturing and formulating design principles that describe how information systems should be build [MaMG02]. In the following, we will highlight only the main aspects of design theories that are required for our line of argumentation¹.

In the context of design research, designing artifacts means to develop and enhance theories. Generally, the process of design is understood as planning, specifying, and subsequently implementing artificial artifacts [Simo69]. As design research aims at solving problems [HMPR04], the central focus of design science is to support the specification of future artifacts, e.g. new kinds of information systems [Fran97; Fran98a; WaWE92]. Thus, design "...is a set of hypotheses, and ultimately can be proven only by construction of the artifact it describes. The feasibility of a design can, however, be supported by scientific theory to the extent that the design embodies principles of the theory" [WaWE92, p. 38]. Hence, formulating design specifications can be seen as the same process of formulating theories.

Design theories are prescriptive and thus goal-oriented. In contrast to the explanatory and predictive nature of theories in natural science, theories in design science are of prescriptive nature. As design theories aim at providing guidance on how to solve a specific problem: "if acted upon, [they] will cause an artifact of a certain type to come into being" [Greg06, p. 619]

¹ Detailed discussion of design theories is provided in [MaMG02; WaWE92] and the referenced literature there.

Design theories prescribe certain design principles that will lead to applications, which are more effective.

Design theories build upon kernel theories. As the designed artifacts are going to be deployed in a certain environment, the ability of attaining the goals is determined by the governing natural and social laws of that environment. Thus, Walls et al. argue that developing design theories requires considering existing theories, e.g. explanatory, predictive and normative theories from natural or social sciences: “The prescriptive plane provides the common ground for integrating these different types of theories” [WaWE92, p. 41]. Hence, design theories are composite theories, as they rely on theories, e.g. predictive theories [Greg06; WaWE92]. The constraints and intended applications of these underlying theories influence the properties of the resulting artifact and provide the base for evaluating the quality of the artifacts and thus the design theory itself.

Design theories prescribe both the artifact and the process of creating that artifact. Besides defining the properties of the intended artifacts, Walls et al. state that design theories should incorporate the process of designing the artifact. They argue that natural and social laws of the environment also determine the process of designing the artifact. Thus, the design process heavily influences the design result – the artifact [WaWE92].

2.2 Structure of Design Theories

Design theories consist of two components: the *design product* component specifies the properties the artifact has to possess to meet certain requirements, as well as propositions on how to test the quality of this relationship. The *design process* component describes the process that is required to design an artifact in the way that it meets the stated requirements [WaWE92]. The first component *design product* consists of four elements [WaWE92]: In the element *class requirements* the design theory developer specifies the problem and subsequently the goals the design theory is supposed to solve and attain. In the element *class design*, the theory developer specifies the structural and functional properties and characteristics of the intended artifact². The element *kernel theories* specifies existing theories, e.g. from social science or mathematics, that constraint or support statements made in the class-design section. The final element of the

² Walls et al. use the prefix *meta* for requirements and design to denote that both aspects refer to a class of artifacts instead of a specific artifact (e.g. retail information systems versus the retail information system for company ABC) [WaWE92]. We argue, that using the prefix *meta* is misleading as both elements refer to an instantiation relationship [Stra96]. Thus, in the remainder of this paper we will refer to both sections as class requirements and class design.

design product component is a collection of *testable hypotheses* that allow evaluating the capability of the class design to meet the class requirements. Overall, the component design product specifies the class of artifacts the design theory proposes to facilitate.

The second component *design process* consists of three elements [WaWE92]: The element *design method* is specifying the process of designing the intended artifact from the class design in a way that the artifact meets the specified requirements. The element *kernel theories* again refers to existing theories that determine or influence the design process. The element *testable design process hypotheses* refer to propositions that can be derived from the design process and their underlying kernel theories and allow evaluating whether applying the design method results in the intended artifacts.

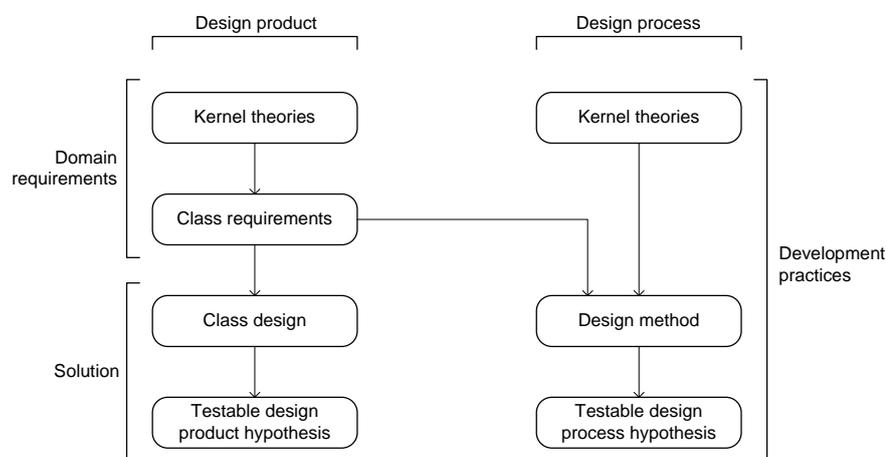


Figure 2: Structure of design theories (according to [MaMG02; WaWE92])

Overall, as Figure 2 shows, design theories capture design knowledge and experiences on both the artifact and the construction process. Design theories are the composition of “...user requirements, a type of system solution (with distinctive features), and a set of effective development practices” [MaMG02, p. 180].

2.3 Benefits of the Structure of Design Theories for Evaluating Reference Models

Design theories provide a framework for design solution proposals in a theoretical and testable way. We argue that by adopting the framework of design theories for reference models we can realize the following benefits:

- *Design theories require the specification of kernel theories when constructing reference models.* Current reference models often do not reveal their underlying theoretical assumption [BePf06]. Thus, adopting the design theory framework

requires reconstructing or disclosing underlying assumptions of the reference models.

- *Design theories provide a framework to reformulate reference models as testable hypotheses.* Design theories result in testable hypotheses that provide the foundation of empirical research on reference models, their utility, and their usage [FeLo04a; WaWE92]. Thus, applying the idea of design theories to reference models facilitate the reconstruction of reference models as genuine theories of IS research.
- *Design theories provide the concept of design principles that guide choosing and adapting reference models.* Design principles can be used to group connected requirements to coherent units that propose a certain utility. By enabling references, e.g. dependencies between design principles [WaWE92] one can identify immutable design principles of reference models. Here, reference models may restrict the adaptations.

In sum, the answer to our first research question is that design theories provide the framework for (re)constructing testable reference models.

3 (Re)constructing Reference Models as Design Theories

To facilitate the reconstruction process we apply Alexander's pattern approach [Alex73].

3.1 Patterns in the Context of Reference Models

Alexander's foundational conceptualization of design is that good design solution resolves perceived misfits in a context [Alex73]³. To facilitate good design, design requirements are deconstructed in a hierarchical way. A certain aspect of design solution will meet each requirement. The general solution is the combination of all solutions. Overall, the main argument is that design issues can be solved by combining coherent and rather independent solutions to specific problems [Alex73]. These coherent solutions are called *patterns* [Alex73; Alex79]. A pattern generally comprises the following elements [Schu03]: the *context* comprises

³ Alexander's ideas refer to design issues in the field of architecture. However, the notion of patterns has been applied to many areas in various disciplines, especially information systems development [GHJV94; Schu03].

causes which lead to the problem described in a pattern and the conditions under which the problem occurs. The context should support assessing the relevance of a pattern [BMRS98]. The *problem* describes contradictions causing the perceived misfits in the context of the pattern. These aspects of the pattern problem section are often called forces [BMRS98]. The next section of a pattern explains the proposed *solution* by dissolving the forces described before. An illustration of consequences of applying the pattern is given as well [BMRS98]. The closing section of a pattern is composed of *references* to related patterns [Schu03]. In sum, a pattern represents a complex structure of knowledge from an application-oriented perspective. The goal of patterns is to explicate experiences and established expert knowledge [Schu03]. As patterns are rarely used independently, Alexander broadens the pattern idea to a system of interrelated patterns that he called *pattern language* [Alex79]. The semantic power of such pattern languages is determined by the references between patterns, which consequently allow capturing solutions for more complex problems [Schu03].

What is the benefit of applying the idea of patterns for reconstructing reference models? Reference models tend to be very complex [BDKK02; BeSc04; Sche98]. Furthermore, reference models generally focus on providing complete design proposals that have to be adapted. Patterns are coherent entities that describe a solution to a specific problem in the sense of design principles as proposed by [MaMG02]. Decomposing reference models into patterns enables identifying the design principles formulated by the reference model. Hence, reconstructing the reference model as pattern language allows reformulating the reference model as a set of design principles and thus forming a design theory [MaMG02]. Such theorized reference models allow evaluating each design principle (i.e. construct of the pattern language) separately and thus derive more concise and elementary evaluation objectives. Furthermore, the idea of pattern languages facilitates reconstructing dependencies between different elements of a reference model and thus supports the identification of core elements. By stating consequences when applying the pattern's solution, the pattern concept facilitates explaining the impact of applying a pattern as well as formulating hypotheses on the benefit of the pattern.

3.2 (Re)construction Framework

Overall, the pattern approach facilitates reconstructing reference models as design theories. Based on the concept of patterns we can now develop a framework for reconstructing reference models. This framework is depicting the elements that are required for the reconstruction process. Thus, the framework ensures that a reconstruction process leads to design principles as

required by [MaMG02]. As depicted in Figure 3 the basic structure of theorized reference models is derived from the structure of design theories as proposed by Walls et al. [WaWE92]. A *theorized reference models* consists of *patterns*. These patterns have references to each other and thus form a pattern language. We introduce the reference types *prerequisite* and *specialization*. Please note that these references can point to external design principles as well. A pattern consists of a *context*, a *problem*, and a *solution*. The context refers to kernel theories that apply to the specific pattern.

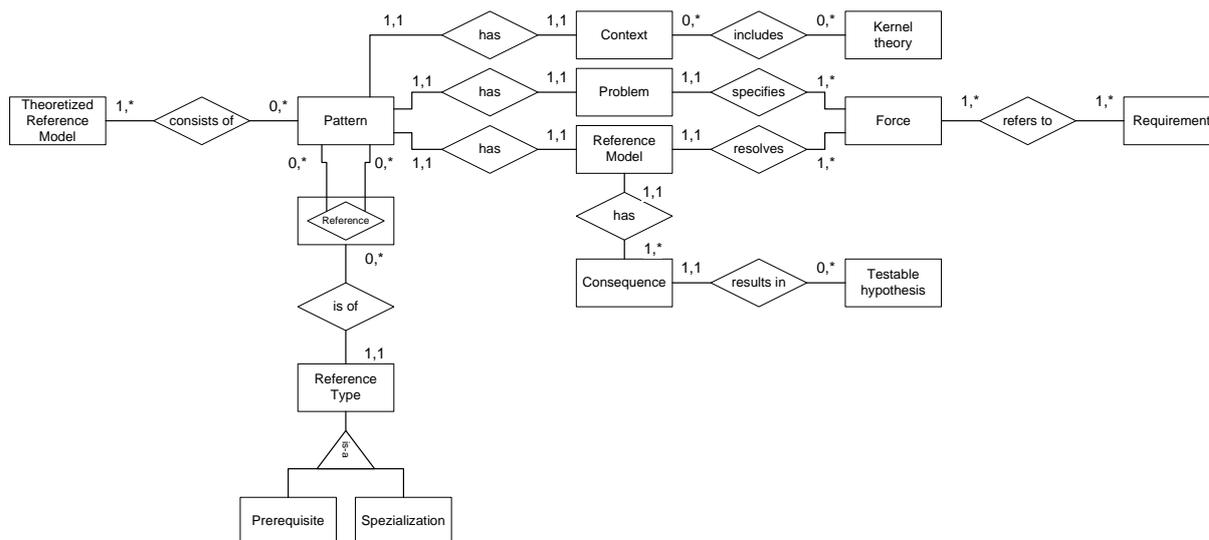


Figure 3: Structure of theorized reference models (according to [Alex79; WaWE92])

Furthermore, figure 3 reveals that it is not required to assign kernel theories. Walls et al. argue that in information systems it might not be possible to identify appropriate kernel theories [WaWE92]. Hence, Markus et al. broaden the definition of kernel theory to include practitioner theories-in-use, e.g. [SaLe02] and theory candidates. The problem analyzes *forces* that are the result of user *requirements*. The *reference model* (or a specific part of it) resolves these forces and has certain *consequences* when applying it. These consequences, either good or bad, are the basis for *testable hypotheses*.

The process of constructing patterns can be found e.g. in [Köhn05; Schu03]. Please note that the concept of patterns can also be used to describe common analysis and design processes, as explained in e.g. [Köhn05]. Thus, patterns can also be used to describe the design process section of design theories.

In sum, this framework for theorized reference models combines the proposed structure of design theories and patterns and guides the reconstruction process. Thus, we have answered research question two on how to facilitate the process of reconstruction.

4 Demonstration: Reconstruction of the SDM Reference Model

In this section we demonstrate our approach on the Service Data Management (SDM) reference model that has been developed by the authors [BWFK04]. We use the SDM reference model, since we are fully aware of the underlying objective and the intended applications and do not rely on interpretations. Hence, we hope to formulate a more accurate reconstructed reference model.

4.1 Introduction to the SDM Reference Model

The IT services industry will likely have a worldwide market volume of about US\$ 760bn. by 2009 [HDLA05]. As IT services (i.e. services that rely on information technology) become more complex, systematic development and efficient delivery of IT services is an important challenge [BuSG03]. IT service providers face challenges similar to that of industrial enterprises: establishing an integrated management of services throughout their lifecycle across different stages of the service value chain [DaJY05]. Hence, an integrated view on all aspects of service engineering and delivery is needed. We call this view service data management [Böhm04; BWFK04]. The objective of the SDM reference model is to capture data structures for service data management.

4.2 The SDM Reference Model as Design Theory

Figure 4 summarizes three fundamental aspects⁴ of the SDM reference model and depicts them as patterns: the *Service Architecture*, the *Service Module*, and the *Service Level Agreement*. These patterns form the design principles of our design theory. The following tables show these design principles in detail.

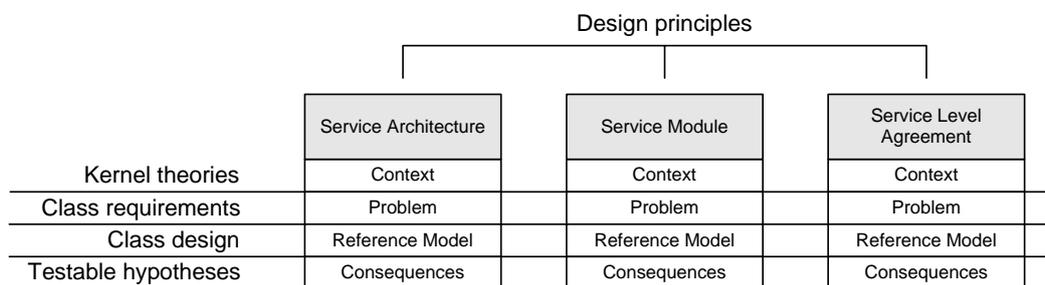


Figure 4: The SDM patterns in the light of the categories of design theories

⁴ A more detailed description of the patterns, especially of the section reference model, would go beyond the scope of this paper.

Service Architecture	
Context (with theoretical references)	Similar to industrial products, IT services are complex systems of various functionalities that are provided by many internal organization units and sub-providers. In industrial production industry, product architectures are used to componentize product elements [Sche98] and thus reduce coordination costs [Coas37]. This idea also has been transferred to software engineering [PoBL05].
Problem (with forces)	<ul style="list-style-type: none"> • Mass-customization for IT services [Böhm04] requires standardized service elements that can be combined. • Many stakeholders, e.g. marketing, sales, and engineering, have different views on IT services. • Especially managing long-term IT services requires considering existing service contracts and their impact on the service infrastructure.
Reference Model	<pre> classDiagram class ServiceArchitecture[Service Architecture] class Service class Catalogue class Product class ConfigurationBase[Configuration Base] class Configuration ServiceArchitecture "1,1" -- "0,*" Service : consists of ServiceArchitecture "1,1" -- "0,*" Catalogue : has Catalogue "1,1" -- "0,*" Product : consists of Catalogue "1,1" -- "0,1" ConfigurationBase : has Service "1,*" -- "0,*" Product : is in Product "0,*" -- "1,1" Configuration : instantiated by ConfigurationBase "1,1" -- "0,*" Configuration : consists of </pre>
Consequences (testable hypotheses)	<ul style="list-style-type: none"> • The differentiation in architecture, catalogue, and configurations reduces coordination costs between stakeholders in IT service engineering and delivery. • Service architectures allow mass customization of IT services. • Service architectures enable tracking of impacts of possible changes in the service capabilities.
References	<ul style="list-style-type: none"> • Product Architectures, e.g. [Sche98] (external prerequisite) • Service Module (prerequisite)

Table 1: The pattern Service Architecture

Service Module	
Context (with theoretical references)	Modern IT services are complex sets of functionalities and rely on technical, organizational, and human resources. Thus, services can be characterized as complex systems [Bung77; Ropo79]. Efficient management of such complex systems requires mechanisms to reduce complexity [BaCl00].
Problem (with forces)	<ul style="list-style-type: none"> • Decomposing service functionality requires describing visible and accessible characteristics. • The dependencies between service functionalities have to be identified and documented.

Reference Model	
Consequences (testable hypotheses)	<ul style="list-style-type: none"> • IT service can be decomposed in service modules [Böhm04] • It is possible to develop standardized definitions of IT services by specifying an interface. • It is possible to develop service products from standardized service module interfaces.
References	<ul style="list-style-type: none"> • Meta-model of BWB-constructs [RoGr02] (external prerequisite)

Table 2: The pattern Service Module

Service Level Agreement	
Context (with theoretical references)	Efficient service delivery has to provide the contracted service functionality at the agreed quality [BuSG03; StMJ00]. However, services generally do not exhibit characteristics that customers can inspect prior to acquiring a service [Böhm04]. Furthermore services rely on the integration of external factors, e.g. input of the service customer [Burr04].
Problem (with forces)	<ul style="list-style-type: none"> • Integration of external factors requires definition of responsibilities of service provider and service client. • Contracting services require defining the outcome of the service contract. • As services change over time, the quality definitions have to change as well. • Services have various states that result in different quality requirements.
Reference Model	
Consequences (testable hypotheses)	<ul style="list-style-type: none"> • Service quality can be described as a set of objectives that are measured and assigned to specific parties. • It is possible to measure each service quality criterion. • All types of IT services have distinct states, e.g. maintenance, operating, etc.
References	<ul style="list-style-type: none"> • Web Service Level Agreements [LKDK03] (external prerequisites)

Table 3: The pattern Service Level Agreement

4.3 Implications for evaluating reference models

We have proposed design theories as a suitable framework for reconstructing reference models to facilitate the evaluation and thus to enhance the benefit and utility of reference models. So, what benefits can be derived from the example for evaluating reference models?

- Decomposing the reference model into coherent patterns reduces the complexity. Patterns can be evaluated individually by testing the provided hypotheses. As said in the example, the modularization of IT service has already been applied successfully in [Böhm04]. Thus, this hypothesis has been substantiated.
- Referring to existing theoretical foundations, i.e. kernel theories, in the context section allows reference model constructors to reveal underlying assumptions. Furthermore, the context describes intended applications of the specific pattern.
- The references between patterns help to analyze the immutable core (prerequisite patterns) and guide reference model adaptation and configuration. The patterns show existing links to other reference models and hence help to avoid double work [FeLo04b].
- Patterns can be applied individually and reduce the overhead of learning and adaptation. Thus, the individual utility can be determined more easily.
- Results from evaluating design principles will lead to local changes in the patterns. Thus, our approach facilitates the incremental enhancement of reference models.

Overall, reconstructing reference models as design theories based on the pattern idea provide a beneficiary framework for constructing and evaluating reference models.

5 Conclusion, Limitations, and Outlook

In this paper we have proposed the reconstruction of references models as IS design theories [WaWE92] to substantiate reference models as innovative research outcomes by providing a theoretical and practical foundation for evaluating reference models. To support the reconstruction, we have applied the pattern approach to facilitate formulating the reference models as set of design principles. By the example of the SDM reference model, we have demonstrated the feasibility and utility of reconstructing reference models as design theories.

However, our approach has some limitations:

- We could not identify any patterns describing the design process yet, which is necessary to formulate a complete design theory. Existing approaches on how to use reference models could be analyzed and adapted for the specific requirements of the IT service industry.
- The reconstruction process either has to be done by the authors or relies on the capability of interpreting information models and associated documentation. Here recent research on collaborative reference modeling and “open models” could be applied [Broc04; KoSF06].
- The pattern approach does not support multiple perspectives on reference models [BDKK02; BeDK04]. However, this shortcoming has already been identified in the pattern community and various solutions have been proposed, e.g. as discussed by [Köhn05].
- Managing theorized reference models requires efficient management of their patterns. Approaches for version management of reference models could be combined with approaches for managing pattern languages [Cunn05; Thom06].

Despite these limitations, we conclude that reconstructing reference models as design theories is a promising approach that can foster both the academic and practical utility of reference models. Thus, our future work will include providing tool support for reconstructing reference models as well as addressing the above-mentioned limitations.

References

- [Alex73] *Alexander, C.:* Notes on the Synthesis of Form. 7 ed., Harvard Univ. Press, Cambridge, Mass. 1973.
- [Alex79] *Alexander, C.:* The timeless way of building. Oxford University Press, New York 1979.
- [BaCl00] *Baldwin, C.Y.; Clark, K.B.:* The power of modularity. MIT Press, Cambridge, Mass. 2000.
- [BDKK02] *Becker, J.; Delfmann, P.; Knackstedt, R.; Kuropka, D.:* Konfigurative Referenzmodellierung. In: *Becker, J.; Knackstedt, R. (Eds.): Wissensmanagement mit Referenzmodellen: Konzepte für die*

Anwendungssystem- und Organisationsgestaltung. Physica-Verlag, Heidelberg 2002, pp. 25-144.

- [BeDK04] *Becker, J.; Delfmann, P.; Knackstedt, R.*: Konstruktion von Referenzmodellierungssprachen: Ein Ordnungsrahmen zur Spezifikation von Adaptionsmechanismen für Informationsmodelle. In: *Wirtschaftsinformatik* 46 (2004) 4, pp. 251-264.
- [BeKn02] *Becker, J.; Knackstedt, R.* (2002). *Referenzmodellierung 2002: Methoden - Modelle - Erfahrungen* (90). Münster: Westfälische Wilhelms-Universität Münster, Institut für Wirtschaftsinformatik.
- [BePf06] *Becker, J.; Pfeiffer, D.*: Konzeptionelle Modellierung: Ein wissenschaftstheoretischer Forschungsleitfaden. In: *Lehner, F.; Nösekabel, H.; Kleinschmidt, P. (Eds.): Multikonferenz Wirtschaftsinformatik 2006*. GITO, Berlin 2006, pp. 3-19.
- [BeSc04] *Becker, J.; Schütte, R.*: *Handelsinformationssysteme*. 2 ed., Redline Wirtschaft, Frankfurt am Main 2004.
- [BMRS98] *Buschmann, F.; Meunier, R.; Rohnert, H.; Sommerlad, P.; Stal, M.*: *Pattern-orientierte Software-Architektur: Ein Pattern-System*. Addison-Wesley-Longman Verlag, Bonn 1998.
- [Böhm04] *Böhm, T.*: *Modularisierung von IT-Dienstleistungen - Eine Methode für das Service Engineering*. Deutscher Universitäts-Verlag, Wiesbaden 2004.
- [BöSK06] *Böhm, T.; Schermann, M.; Krcmar, H.*: Application-Oriented Evaluation of the SDM Reference Model: Framework, Instantiation and Initial Findings. In: (Eds.): *Referenzmodellierung 2006 in conjunction with Multikonferenz Wirtschaftsinformatik 2006 (MKWI '06)* 2006. Passau, Germany.
- [Broc04] *vom Brocke, J.*: Internetbasierte Referenzmodellierung - State-of-the-Art und Entwicklungsperspektiven. In: *Wirtschaftsinformatik* 46 (2004) 5, pp. 390-404.
- [Bung77] *Bunge, M.*: *Ontology I: The furniture of the world*. (Bd. 3), Reidel, Dordrecht 1977.
- [Burr04] *Burr, W.*: *Chancen und Risiken der Modularisierung von Dienstleistungen aus betriebswirtschaftlicher Sicht*. 2004.

- [BuSG03] *Bullinger, H.-J.; Scheer, A.-W.; Grieble, O.:* Service Engineering: Entwicklung und Gestaltung innovativer Dienstleistungen. Springer, Berlin 2003.
- [BWFK04] *Böhmman, T.; Winkler, T.; Fogl, F.; Krcmar, H.:* Servicedatenmanagement für IT-Dienstleistungen: Ansatzpunkte für ein fachkonzeptionelles Referenzmodell. In: *Becker, J.; Delfmann, P. (Eds.):* Referenzmodellierung: Grundlagen, Techniken und domänenbezogene Anwendung. Physica, Heidelberg 2004, pp. 99-124.
- [Coas37] *Coase, R.H.:* The Nature of The Firm. In: *Economica New Series 4 (1937)*, pp. 386-405.
- [Cunn05] Cunningham, W. (2005). Portland Pattern Repository. <http://c2.com/ppr/>, Accessed: 2005-Oct-27.
- [DaJY05] Da Rold, C.; Jester, R.; Young, A. (2005). *The Future of Outsourcing* (Gartner Research Report). Stamford, CT, USA: Gartner Inc.
- [FeLo03b] *Fettke, P.; Loos, P.:* Multiperspective Evaluation of Reference Models - Towards a Framework. In: *(Eds.):* International Workshop on Conceptual Modeling Quality (IWCMQ'03) 2003. Chicago, Illinois, pp. 80-91.
- [FeLo04a] *Fettke, P.; Loos, P.:* Referenzmodellierungsforschung. In: *WIRTSCHAFTSINFORMATIK 46 (2004) 5*, pp. 331-340.
- [FeLo04b] *Fettke, P.; Loos, P. (2004). Systematische Erhebung von Referenzmodellen - Ergebnisse einer Voruntersuchung* (Working Paper 19). Mainz: Lehrstuhl für Wirtschaftsinformatik und BWL, ISYM - Information Systems & Management, Johannes Gutenberg-Universität.
- [Fran00] *Frank, U.:* Modelle als Evaluationsobjekt: Einführung und Grundlegung. In: *Häntschel, I.; Heinrich, L.J. (Eds.):* Evaluation und Evaluationsforschung in der Wirtschaftsinformatik. Oldenbourg, München 2000, pp. 339-352.
- [Fran97] *Frank, U.:* Erfahrung, Erkenntnis und Wirklichkeitsgestaltung: Anmerkungen zur Rolle der Empirie in der Wirtschaftsinformatik. In: *Grün, O.; Heinrich, L.J. (Eds.):* Wirtschaftsinformatik: Ergebnisse empirischer Forschung. Springer, Berlin 1997, pp. 21-35.

- [Fran98a] *Frank, U.:* Die Evaluation von Artefakten: Eine zentrale Herausforderung der Wirtschaftsinformatik. In: (Eds.): Workshop Evaluation und Evaluationsforschung in der Wirtschaftsinformatik 1998. Linz, Austria.
- [GHJV94] *Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J.:* Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Publishing Company, Reading, Massachusetts 1994.
- [Greg06] *Gregor, S.:* The Nature of Theory in Information Systems. In: MIS Quarterly 30 (2006) 3, pp. 611-642.
- [HDLA05] *Hale, K.; De Souza, R.; Lo, T.; Adachi, Y. (2005).* *Forecast: IT Services, Worldwide, 2005-2009* (Gartner Research Report): Gartner Inc.
- [HMPR04] *Hevner, A.R.; March, S.T.; Park, J.; Ram, S.:* Design Science in Information Systems Research. In: MIS Quarterly 28 (2004) 1, pp. 77-105.
- [Köhn05] *Köhne, S. (2005).* *Didaktischer Ansatz für das Blended Learning: Konzeption und Anwendung von Educational Patterns.* Dissertation, Universität Hohenheim.
- [KoSF06] *Koch, S.; Strecker, S.; Frank, U.:* Conceptual Modelling as a New Entry in the Bazaar: The Open Model Approach. In: *Damiani, E.; Fitzgerald, B.; Scacchi, W.; Scotto, M.; Succi, G. (Eds.):* Second International Conference on Open Source Systems (OSS) 2006. Como, Italy.
- [LKDK03] *Ludwig, H.; Keller, A.; Dan, A.; King, R.P.; Franck, R. (2003).* *Web Service Level Agreement (WSLA) Language Specification 1.0.*
- [MaMG02] *Markus, M.L.; Majchrzak, A.; Gasser, L.:* A Design Theory for Systems that support emergent Knowledge Processes. In: MIS Quarterly 26 (2002) 3, pp. 179-212.
- [Mert95] *Mertens, P.:* Wirtschaftsinformatik: von den Moden zum Trends. In: *König, W. (Eds.):* Wirtschaftsinformatik '95: Wettbewerbsfähigkeit, Innovation, Wirtschaftlichkeit. Physica-Verlag, Heidelberg 1995, pp. 25-64.
- [PoBL05] *Pohl, K.; Böckle, G.; Linden, F.v.d.:* Software product line engineering: foundations, principles, and techniques Springer, Berlin 2005.

- [RoGr02] *Rosemann, M.; Green, P.:* Developing a meta model for the Bunge-Wand-Weber ontological constructs. In: *Information Systems* 27 (2002), pp. 75-91.
- [Ropo79] *Ropohl, G.:* Eine Systemtheorie der Technik: zur Grundlegung der allgemeinen Technologie. Carl Hanser Verlag, München 1979.
- [SaLe02] *Sarker, S.; Lee, A.S.:* Using a Positivist Case Research Methodology to Test Three Competing Theories-In-Use of Business Process Reengineering. In: *Journal of the Association for Information Systems* 2 (2002) 7, pp. 1-74.
- [Sche98] *Scheer, A.-W.:* Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse. 2 ed., Springer, Berlin 1998.
- [Schu03] *Schumacher, M.:* Security engineering with patterns: origins, theoretical models, and new applications. Springer, Berlin 2003.
- [Simo69] *Simon, H.A.:* The Sciences of the Artificial. The M.I.T. Press, Cambridge, Massachusetts, USA 1969.
- [StMJ00] *Sturm, R.; Morris, W.; Jander, M.:* Foundations of Service Level Management. Indianapolis, SAMS 2000.
- [Stra96] *Strahringer, S.:* Metamodellierung als Instrument des Methodenvergleichs: Eine Evaluierung am Beispiel objektorientierter Analysemethoden. Shaker, Aachen 1996.
- [Thom06] *Thomas, O.:* Version Management for Reference Models: Design and Implementation. In: (*Eds.*): Referenzmodellierung 2006 in conjunction with Multikonferenz Wirtschaftsinformatik 2006 (MKWI '06) 2006. Passau, Germany.
- [WaWE92] *Walls, J.G.; Widmeyer, G.R.; El Sawy, O.A.:* Building an Information System Design Theory for Vigilant EIS. In: *Information Systems Research* 3 (1992) 1, pp. 36-59.
- [Wiss94] *Wissenschaftliche Kommission Wirtschaftsinformatik:* Profil der Wirtschaftsinformatik. In: *WIRTSCHAFTSINFORMATIK* 36 (1994) 1, pp. 80-81.

Modellbasierter Entwurf strukturanaloger Architekturen auf Basis der Partitionierung von Graphen

Stephan Aier

Institut für Wirtschaftsinformatik
Universität St.Gallen
stephan.aier@unisg.ch

Marten Schönherr

Institut für Wirtschaftsinformatik und quantitative Methoden
Technische Universität Berlin
mschoenherr@sysedv.tu-berlin.de

Abstract

In großen Unternehmen treffen hohe Komplexität auf die Forderung nach hoher Flexibilität, um Produkte und Prozesse den sich schneller ändernden Anforderungen anpassen zu können. Informationssysteme stellen dabei einen unterstützenden, gleichzeitig aber auch einen hemmenden Faktor dar, da sie oft weniger schnell verändert werden können, als die durch sie unterstützten Prozesse. Da die Abhängigkeiten zwischen Prozessen und IT, die gemeinsam die Unternehmensarchitektur bilden, in der Summe zunehmen, liegt die Aufgabe des Architekturmanagements darin, nachhaltige Strukturen zu schaffen. Ein Ansatz zur Erreichung dieses Ziels kann die Modularisierung der Architektur unter Berücksichtigung der Abhängigkeiten zwischen Prozessen und IT sein. Dieser Beitrag zeigt, wie die Partitionierung von Graphen durch Clusteringalgorithmen hilft, solche Module in bestehenden Architekturen zu identifizieren.

1 Problemstellung

Die IT-Systemlandschaften der meisten großen Unternehmen haben heute eine Komplexität erreicht, die nur noch schwer zu beherrschen ist. Gleichzeitig soll die IT sich immer schneller ändernde Geschäftsprozesse unterstützen. Um solche komplexen Unternehmensarchitekturen als Gesamtheit von Organisationsstrukturen, Geschäftsprozessen und IT-Systemen effizient

managen zu können, werden unter anderem verschiedene Technologien prozessorientierter IT-Systemintegration wie beispielsweise Enterprise Application Integration (EAI) oder Serviceorientierte Architekturen (SOA) diskutiert. Neben diesen Technologien sind vor allem Methoden für die Gestaltung komplexer Unternehmensarchitekturen von Bedeutung.

Ein mögliches Gestaltungsziel für Unternehmensarchitekturen, welches durch die Diskussion über SOA wieder stärker in den Mittelpunkt gerückt ist, ist die Modularisierung. Es bleibt jedoch oft unklar, nach welchen möglichst generischen Kriterien Module zu bilden sind. Dies gilt für technische wie für fachliche Aspekte. Die hier entwickelte Methode wendet Clusteringalgorithmen an, um in einer bestehenden, komplexen Unternehmensarchitektur latent vorhandene Module zu identifizieren. Die Algorithmen arbeiten auf Basis eines Enterprise Architecture Modells, welches insbesondere Aspekte der prozessorientierten Systemintegration abbildet. Durch die auf formalen Modellen basierende ganzheitliche Sichtweise auf Geschäftsprozesse und IT-Systeme sollen Architekturmodule identifiziert werden, die als Basis für eine nachhaltige Unternehmensarchitektur dienen. Dazu wird im folgenden Abschnitt kurz unser Verständnis von Unternehmensarchitekturen dargestellt, da es die Grundlage für das folgende Architekturmodell und die Architekturgestaltung bildet. Daran anschließend werden Gestaltungsansätze für nachhaltige Architekturen beschrieben, um dann unseren Architektur- und Clusteringsansatz sowie dessen technische Implementierung darzustellen.

2 Verständnis der Unternehmensarchitektur

Der Begriff der *Unternehmensarchitektur* (Enterprise Architecture, EA) ist zentraler Bestandteil dieses Beitrags. Sowohl unsere Gestaltungsvorschläge für Unternehmensarchitekturen als auch die technische Implementierung für die Modellierung und Analyse dieser Architekturen bauen darauf auf. Darum wird im Folgenden ein Überblick über das zugrunde liegende Architekturverständnis gegeben.

Verkürzt kann eine Architektur als eine abstrakte, ganzheitliche Betrachtung von Strukturen und Mustern mit Planungscharakter aufgefasst werden [Bas⁺03, S. 19 ff.]. Architekturen sind in der Regel das Ergebnis eines Planungsprozesses und stellen nach ihrer Definition selbst einen Masterplan für die ganzheitliche Realisierung zukünftiger Maßnahmen dar.

Diese allgemeinen Charakteristika lassen sich auf die Gestaltung von Unternehmen anwenden und werden so zur Unternehmensarchitektur. Unter einer Unternehmensarchitektur wird das

Zusammenwirken organisatorischer, technischer und psychosozialer Aspekte bei der Planung und Entwicklung betrieblicher soziotechnischer Informationssysteme verstanden [Gron03, S. 45]. Im Folgenden sollen vor allem die organisatorische und die technische Dimension der Unternehmensarchitektur erläutert werden. Dazu werden die Begriffe *Organisationsarchitektur* und *IT-Architektur* verwendet (Abbildung 1).

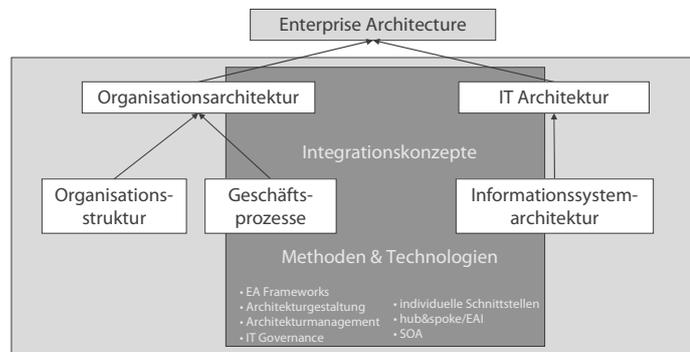


Abbildung 1: Bestandteile der Unternehmensarchitektur

Die *Organisationsarchitektur* enthält alle nichttechnischen Bestandteile der Unternehmensarchitektur und kann mit dem instrumentalen Organisationsbegriff verglichen werden, der die Gesamtheit aller generellen expliziten Regelungen zur Gestaltung der Aufbau- und Ablauforganisation umfasst [Blei91, S. 35]. Entsprechend wird die Organisationsarchitektur hier in die Organisationsstruktur und die Geschäftsprozesse unterschieden.

Komplementär zur Organisationsarchitektur steht die *IT-Architektur*. Sie umfasst alle technischen Bestandteile der Unternehmensarchitektur. Insbesondere beinhaltet sie die technischen Informationssysteme, welche ihrerseits eine eigene Architektur, die Informationssystemarchitektur, aufweisen. In der Literatur sind die Begriffe Organisationsarchitektur und IT-Architektur häufig verwendet, jedoch meist anders definiert. Abhängig von der fachlichen Herkunft des Autors umfasst die Organisationsarchitektur auch technische Komponenten [Nad⁺94] bzw. die IT-Architektur auch organisatorische Fragestellungen [Krcm90]. Im Folgenden sollen jedoch beide Bereiche separat und somit bezogen auf ihre Bedeutung gleichwertig betrachtet werden.

Die Aufgabe von *Integrationskonzepten* besteht zum einen darin, die technischen Komponenten der Architektur technisch zu verbinden, zum anderen dienen Integrationskonzepte zur konzeptuellen Zusammenführung von technischen Elementen mit den Elementen der Organisation – was oft unter dem Begriff Business-IT-Alignment diskutiert wird [WiLa06]. Zur Bewältigung dieser Integrationsaufgabe bedarf es *Technologien* – wie EAI oder SOA – und *Methoden*. Letztere stehen im Mittelpunkt dieses Beitrags.

3 Gestaltungsansätze nachhaltiger Unternehmensarchitekturen

Im folgenden Abschnitt sollen im ersten Teil das Konzept der intern orientierten Nachhaltigkeit dargelegt werden, um dann die Strategie der strukturellen Analogie und die Modularisierung als Gestaltungsansätze nachhaltiger Unternehmensarchitekturen abzuleiten. Dabei handelt es sich um die Fortführung der in [AiDo04; AiDo05] dargestellten Ansätze. Diese bilden in der Folge die Grundlage für die Modellierung, Analyse und Optimierung von Unternehmensarchitekturen.

3.1 Konzept der intern orientierten Nachhaltigkeit

In der Literatur existieren unzählige Definitionen und Konzepte von Nachhaltigkeit. Im Kern steht jedoch das Verständnis, das ein System langfristig nicht mehr Ressourcen verbrauchen darf als neu verfügbar werden. Aus einer Analyse dieser Definitionen lassen sich vier Strategien extrahieren, die im Ergebnis ihrer Anwendung zu mehr Nachhaltigkeit führen sollen. Dies sind die Strategie der Effizienz, die Strategie der Suffizienz, die Strategie der Konsistenz und die Strategie der Partizipation.

Der *Suffizienzstrategie* erhebt die Forderung, durch Verzicht ein ausreichendes Maß an Genügsamkeit zu realisieren und basiert auf der Einsicht, dass es keine Unbeschränktheit gibt, da jedes System Grenzen in Raum und Zeit aufweist [Hube95, S. 40]. Die *Effizienzstrategie* zielt auf die Steigerung der Produktivität, um dadurch Leistungen wirtschaftlich, d.h. mit dem kleinsten möglichen Ressourcenverbrauch zu erstellen. Zentrale Konzepte sind dabei Wiederverwendung und Langlebigkeit. Die *Konsistenzstrategie* hat entweder die vollständige Abschirmung von Systemen von deren Umwelt oder die Sicherstellung deren Verträglichkeit, Stimmigkeit mit dem sie umgebenden (Super-)System zum Ziel [Hube95, S. 41 f.]. Wenn Systeme keine Verbindungen zur umgebenden Umwelt haben, müssen diese Systeme auch nicht mit der Umwelt verträglich sein, da es keine Wirkungsbeziehungen zwischen ihnen gibt. Bestehen jedoch Beziehungen, so wird Konsistenz der Systeme mit ihrer Umwelt gefordert. Die *Partizipationsstrategie* schließlich fordert die Teilhabe, d.h. die Einbindung der von einer Systemgestaltung Betroffenen. Dies ist nötig, um das System zum einen bestmöglich für die „Betroffenen“ zu gestalten und zum anderen, um deren Akzeptanz des Systems sicherzustellen [Gron03, S. 222].

Wird Nachhaltigkeit als ein für die Unternehmung relevantes Thema betrachtet, so geht es meist um die Reduktion negativer externer Effekte auf die Umwelt oder die Gesellschaft (physische und soziale Umwelt) welche durch das Handeln des Unternehmens entstehen [LeSt03, S. 259 f.]. Dies soll im Folgenden als *extern orientierte Nachhaltigkeit* bezeichnet werden. Das hier

zugrunde gelegte Verständnis von Nachhaltigkeit soll vollkommen losgelöst von Themen der Ökologie sein. Vielmehr soll das Ziel der Betrachtung eine unternehmensinterne Sicht der Nachhaltigkeit mit dem primären Ziel der langfristigen, effizienten Unternehmensführung sein [Gron03; Klam00; Scha02]. HAHN/HUNGENBERG definieren das Oberziel einer jeden Unternehmung als Erhaltung und erfolgreiche Weiterentwicklung, als Erfüllung der Individualziele aller an der Unternehmung interessierten Gruppen [HaHu01, S. 13]. Somit kann der hier entwickelte interne Nachhaltigkeitsbegriff als ein Unterziel dieses Oberziels betrachtet werden. Dies soll im Folgenden als *intern orientierte Nachhaltigkeit* bezeichnet werden [AiDo04; AiDo05].

3.2 Strategie der strukturellen Analogie von Organisations- und IT-Architektur

Der Zusammenhang zwischen der Organisation einerseits und der IT andererseits hat in der wissenschaftlichen Literatur große Beachtung erfahren. Trotz der langjährigen Beschäftigung der Wissenschaft mit diesem Thema und den zahlreichen Veröffentlichungen lassen sich jedoch keine absoluten, verallgemeinernden Aussagen zu den Beziehungen zwischen IT und Organisation finden [MaRo88; LeHu98, S. 271 f.]. Das Bestehen monokausaler Zusammenhänge zwischen Organisation und IT wird heute weitgehend verneint. Vielmehr wird angenommen, dass Informationstechnologie eine den Gestaltungsspielraum des Organisators erweiternde Option darstellt – „IT als Enabler“ [Fres00, S. 142 f.]. Gleichzeitig wird aber auch deutlich, dass IT ein „Disabler“ für die Veränderung von Organisationsstrukturen und Geschäftsprozessen sein kann, wenn die IT nicht in ausreichendem Maße an die sich stetig verändernden Anforderungen anpassbar wird [Hage03, S. 66].

Um eine Nachhaltigkeit der Unternehmensarchitektur zu erreichen, ist die Berücksichtigung dieser Abhängigkeiten in einer integrierten Gestaltung unabdingbar. Die Strategie der strukturellen Analogie soll die geforderte Stimmigkeit zwischen IT- und Organisationsarchitektur durch ihre strukturähnliche Gestaltung unterstützen [Gron03, S. 216; Wall96, S. 63; AiDo04]. Der Nachhaltigkeitsaspekt der strukturellen Analogie folgt mittelbar aus der Konsistenzstrategie, nach welcher IT- und Organisationsarchitektur in übereinstimmend zu gestalten sind.

IT- und Organisationsarchitekturen sind strukturanalog zueinander, wenn sie in ihren Merkmalsausprägungen weitgehend übereinstimmen. Besonders geeignet für den Strukturvergleich zwischen Organisations- und IT-Struktur erscheinen die Organisationsdimensionen Spezialisierung (Arbeitsteilung), Entscheidungsdelegation (Kompetenzverteilung) und Formalisierung. Eine Möglichkeit zur Realisierung der Strategie der strukturellen Analogie ist die Modularisierung der Unternehmensarchitektur.

3.3 Strategie der Modularisierung

Flexibilität ist eine Voraussetzung für nachhaltige Architekturen, da sich nur flexible Architekturen dauerhaft an die Anforderungen ihrer Umwelt anpassen können. Die Forderung der Komplexitätsreduktion entstammt der entscheidungstheoretischen Sicht der Nachhaltigkeit. Danach stellen sich negativ irreversible Entscheidungen vor allem als Problem komplexer Systeme dar, in welchen nicht alle Wirkungen von Entscheidungen zur irreversiblen Ressourcenverwendung erkannt werden können [Krca03, S. 23 f.]. Nachfolgend wird gezeigt, warum modulare Architekturen prinzipiell nachhaltig sein können. In Anlehnung an die im vorhergehenden Abschnitt beschriebene strukturelle Analogie zwischen Organisation und IT ist der Gedanke der Modularisierung auf beide Domänen anzuwenden (Abbildung 2).

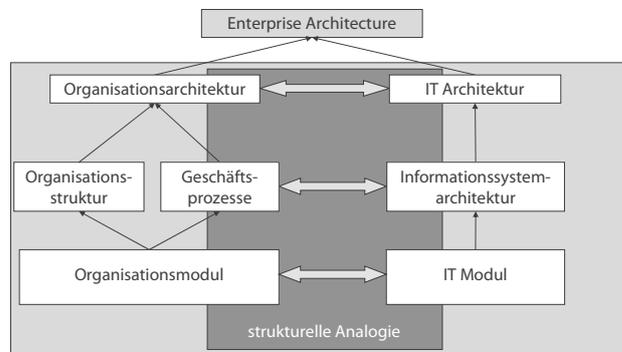


Abbildung 2: Strukturelle Analogie in einer modularisierten Architektur

Modularisierung bedeutet allgemein die Strukturierung eines Systems in relativ kleine, teilautonome und überschaubare Subsysteme (Module, Fraktale, Cluster). Die Komplexitätsreduktion ergibt sich dabei aus der Subsystembildung innerhalb des Systems Unternehmung. Die Subsystembildung wirkt komplexitätsreduzierend, da sie zum einen die subsysteminterne Komplexität im Sinne der Kapselung vor der Subsystemumwelt verbirgt, zum anderen durch die Reduktion auf wenige bekannte Schnittstellen eine Entkopplung der Subsysteme im Sinne einer Reduktion von Abhängigkeiten bewirkt. Mit anderen Worten: Komplexität ist systemtheoretisch ein Selektionszwang und Systemdifferenzierung durch Subsystembildung eine Möglichkeit der Selektion zur Verringerung der Komplexität [Stue96, S. 73]. Im Zusammenhang damit steht die Fähigkeit der einzelnen Subsysteme, selbständig zu agieren und sich selbst organisieren zu können. Die Flexibilisierung der Strukturen und damit der Abläufe ergibt sich durch die nun leichtere Rekonfigurationsmöglichkeit der entkoppelten Module über ihre definierten Schnittstellen.

Die wesentliche Frage ist nun, wie in der Praxis Module gebildet werden sollen, die zu einer strukturanalogen Architektur führen. Um die o.g. Ziele einer Modularisierung zu erreichen wird

vorgeschlagen, die Abhängigkeiten zwischen verschiedenen Modulen zu minimieren und die Abhängigkeiten zwischen Elementen innerhalb eines Moduls zu maximieren indem solche Elemente in ein Modul verschoben werden [AiSc03, S. 27 ff.]. Diese allgemein gehaltene Anweisung lässt sich praktisch nicht ohne weitere Unterstützung durch Methoden und Werkzeuge anwenden. Darum wird im folgenden Abschnitt ein Ansatz entwickelt, der mit Clusteringalgorithmen, unter Berücksichtigung bestehender Geschäftsprozesse, IT-Systeme und deren Interaktionen, Module identifiziert.

4 Modellierung als Grundlage eines Architektur-Clusterings

Modelle bilden die Basis des Clusterings der Unternehmensarchitektur. Darum wird zuerst der verfolgte Modellierungsansatz vorgestellt. Darauf folgen die Einführung entsprechender Clusteringalgorithmen sowie die Darstellung der prototypischen Implementierung.

4.1 Ableitung der Modellierungsnotation

Um Clusteringalgorithmen auf Unternehmensarchitekturen anwenden zu können, müssen die hier relevanten Aspekte der Architekturen in einem Modell abgebildet werden. In unserem Ansatz liegt der Schwerpunkt darauf, prozessorientierte Systemintegration modellieren zu können. Dafür sind mindestens die folgenden Elemente notwendig:

- Geschäftsprozesse, d.h. eine Abfolgen von Aktivitäten (Aufgaben),
- IT-Systeme,
- Relationen, die die Nutzung bestimmter IT-Systeme entlang eines Geschäftsprozesses widerspiegeln,
- und die Beziehungen und Schnittstellen zwischen IT-Systemen entlang eines Geschäftsprozesses.

Wir nennen die Sicht, die durch die Abbildung der IT-System-Beziehungen entlang von Geschäftsprozessen entsteht, *dynamische Sicht*. Sie stellt eine Erweiterung statischer Sichten dar, die oft nur IT-Systeme und deren Beziehungen unabhängig von Geschäftsprozessen abbilden. In statischen Sichten ist nicht zu erkennen, welche IT-System-Schnittstellen entlang der Aus-

fürungen eines bestimmten Prozesses genutzt werden. Die Abbildbarkeit dynamischer Sichten ist somit eine Voraussetzung für die Modellierung prozessorientierter Systemintegration.

Bei dem Entwurf unseres Modellierungsansatzes war es nicht das Ziel, eine vollständig neue Notation zu entwickeln. Vielmehr bestand das Ziel darin, Unternehmensarchitekturen als Basis für die Anwendung von Clusteringalgorithmen modellieren zu können. Der Schwerpunkt lag dabei auf der Darstellung der Verknüpfung zwischen Geschäftsprozessen, IT-Systemen und der Integration dieser IT-Systeme entlang der Geschäftsprozesse.

Als Basisnotation für die Modellierung kamen nach der Analyse verschiedener Notationen die Unified Modeling Language (UML) sowie ereignisgesteuerte Prozessketten (EPK) in die engere Auswahl. Beide Notationen unterscheiden sich im Bereich der reinen Prozessabbildung faktisch nur unwesentlich. Die Vorteile der EPK liegen in der Möglichkeit IT-Systeme verschiedenen Prozessschritten zuordnen zu können, sowie in der weiteren Verbreitung bei Fachanwendern. Darüber hinaus lässt sich eine einfache EPK auch automatisiert in ein UML Activity Diagramm transformieren. Die entgegengesetzte Transformation ist zwar auch möglich, jedoch komplexer. Darum wurde als Ausgangspunkt für die Visualisierung die EPK gewählt. Darüber hinaus wurde die Integration verschiedener IT-Systeme in der dynamischen Prozesssicht, sowie in einer statischen IT-Sicht visualisiert. Um in beliebigen Detaillierungsebenen konsistent modellieren zu können, wurden Eingangs- und Ausgangselemente hinzugefügt (Abbildung 3). Ein solches Modell kann auch als Graph bzw. Netzwerk betrachtet werden. Im folgenden Abschnitt wird eine kurze Einführung in die Graphentheorie sowie in die Algorithmen zur Partitionierung von Graphen gegeben, um danach unsere Implementierung dieser Algorithmen für das Clustering von Unternehmensarchitekturen vorzustellen.

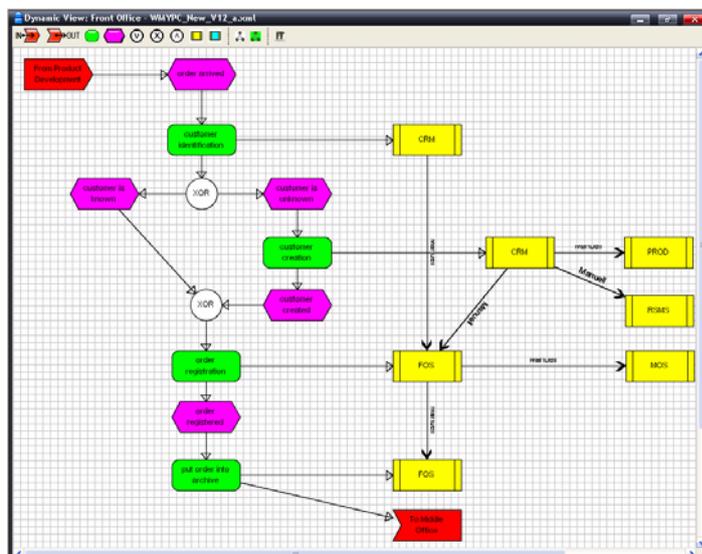


Abbildung 3: Modellierungsnotation auf EPK-Basis

4.2 Graphentheorie und Clusteringansätze

Ein *Graph/Netzwerk* besteht aus einer Menge von *Knoten* (vertices) V und einer Menge von *Kanten* (edges) E . Im hier vorliegenden Kontext kann das Architekturmodell als Graph aufgefasst werden. Alle strukturell wesentlichen Elemente (Module, Funktionen, IT-Systeme) lassen sich als Knoten und die Verbindungen zwischen diesen Elementen als Kanten abstrahieren. Ereignisse und Operatoren werden nicht in den Graphen übernommen, da sie für eine strukturelle Analyse nur von untergeordneter Bedeutung sind.

Graphen deren Kanten durch Zahlenwerte attribuiert werden, werden *gewichtete Graphen* genannt. Kanten, die einen definierten Anfang und ein definiertes Ende haben, heißen *gerichtete Kanten*. Kanten bei denen das nicht der Fall ist heißen *ungerichtete Kanten*. Die *Entfernung* zwischen zwei Knoten ist definiert als der kürzeste Pfad zwischen diesen. Bei gewichteten Graphen wird oft auch von der *gewichteten Entfernung* gesprochen [OMa⁺05, S. 4 f.]. Ein solches Netzwerk mit n Knoten lässt sich nun als Nachbarschaftsmatrix A der Form $n \times n$ beschreiben.

$$A_{ij} = \begin{cases} 1 & \text{wenn die Knoten } i \text{ und } j \text{ verbunden sind} \\ 0 & \text{sonst} \end{cases} \quad (1)$$

In einem gewichteten Netz stellt A_{ij} das Gewicht der Kante, welche die Knoten i und j verbindet, dar. Weiterhin lässt sich der *Grad* (degree) k eines Knotens i definieren als die Anzahl der Kanten, die mit diesem Knoten verbunden sind:

$$k_i = \sum_j A_{ij} \quad (2)$$

Die vorliegenden Modelle können in einen solchen Graphen transformiert werden. Dazu ist es zunächst unerheblich, welchem Typ die Elemente entsprechen. Sie werden gleichrangig als Knoten abgebildet. Die Kanten werden als ungerichtete Kanten in den Graphen übernommen, da die Richtungsinformation hier nicht wesentlich für die strukturellen Zusammenhänge ist.

Ein Gestaltungsziel für Architekturen bestand in der Nachhaltigkeit. Zur Umsetzung dieses Ziels wurde eine Modularisierung der Architektur vorgeschlagen. Als die wichtigsten Kriterien für die Modulbildung werden die Minimierung der Abhängigkeiten zwischen verschiedenen Modulen und die Maximierung der Abhängigkeiten zwischen Elementen innerhalb eines Moduls, im Sinne einer Verschiebung solcher Elemente in ein Modul, vorgeschlagen.

Bereits bei kleineren Ausschnitten einer Architektur, wird die Anwendung einer solchen Vorschrift ohne Werkzeugunterstützung schnell schwer überschaubar. Darum wurde versucht, Algorithmen zu finden und anzuwenden, die aus der Struktur eines Graphen die Zusammengehö-

rigkeit verschiedener Elemente bewerten. Dabei wird von der inhaltlichen Dimension des Modells abstrahiert und ausschließlich die Struktur ausgewertet. Entsprechend des Optimierungsansatzes sollten alle Elemente mit einer hohen Zusammengehörigkeit ein Modul bilden.

Solche Fragestellungen werden unter den Schlagworten *Partitionierung* bzw. *Clustering* von Graphen diskutiert. Ein Cluster ist dabei eine Menge von Elementen, die in der einen oder anderen Form ähnlich zueinander sind [OMa⁺05, S. 18.]. Im hier vorliegenden Fall ergibt sich Ähnlichkeit aus der Tatsache, dass verschiedene Elemente eine gemeinsame Untermenge von Nachbarn haben. Insbesondere in soziologischen Forschungsrichtungen finden Clusteringansätze und netzwerkbezogene Analysen Anwendung wenn es darum geht, Gruppen und Organisationen also soziale Systeme zu analysieren [WaFa99; Scot05].

Girvan/Newman entwickelten ein Clusteringalgorithmus, um Communities in sozialen Netzwerken zu bestimmen. Ein Netzwerk besteht dabei aus einer Menge von Personen (Knoten) die Beziehungen zueinander haben (Kanten). Die Beziehung kann beispielsweise durch das gegenseitige Kennen bestehen. Die Frage ist nun, welche Personen kennen sich in einem Netzwerk direkt und lassen sich diese Personen und die Communities (Cluster) zu denen sie gehören identifizieren [GiNe02, S. 7821]? Abbildung 4 stellt ein solches Netzwerk dar in welchem die Cluster (Communities) hervorgehoben worden sind. Girvan/Newman analysierten bestehende Clusteringalgorithmen nach ihrer Leistungsfähigkeit Cluster in Netzen zu erkennen, deren Strukturen bekannt sind. Dabei handelt es sich meist um hierarchisch arbeitende Algorithmen in welchen Cluster durch Baumstrukturen abgebildet werden (Abbildung 5).

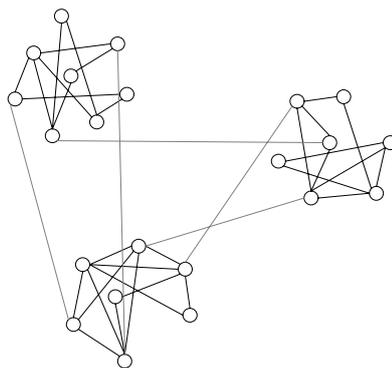


Abbildung 4: Netzwerk mit einer Community-Struktur [GiNe02, S. 7822]

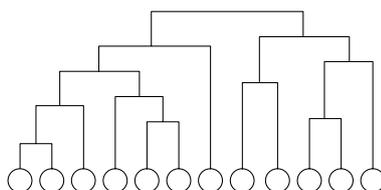


Abbildung 5: Hierarchisches Clustering in Baumstrukturen [GiNe02, S. 7822]

In bestimmten Konstellationen, insbesondere bei Knoten die am Rand eines Netzwerks liegen und nur durch eine einzige Kante mit dem Netz verbunden sind, liefern diese Algorithmen schlechte Ergebnisse. Elemente werden aus einem Cluster ausgeschlossen obwohl sie richtiger Weise dazugehören sollten. Die Grundidee des „*Betweenness*“-Algorithmus von Girvan/Newman ist es nicht, die zentralen Kanten eines Netzes zu finden, sondern die Kanten die am wenigsten zentral sind und somit am meisten *zwischen* (*between*) den Communities liegen.¹ In der Vergangenheit wurde die Knoten-Betweenness als ein Maß des Einflusses eines Knotens auf ein Netzwerk untersucht. Zuerst von Freeman vorgeschlagen, wird die *Betweenness-Centrality* eines Knotens i definiert als die Anzahl der kürzesten Pfade zwischen Paaren anderer Knoten die durch i verlaufen [Free77]. Girvan/Newman verallgemeinern nun Freemans Betweenness-Centrality und definieren die *Kanten-Betweenness* einer Kante als die Anzahl der kürzesten Pfade zwischen Paaren von Knoten die über diese Kante laufen. Wenn in einem Netzwerk nun Communities bestehen, dann sind diese nur über wenige Kanten miteinander verbunden. Das heißt, dass die kürzesten Pfade zwischen diesen Communities über diese wenigen Kanten laufen. Darum werden diese Kanten eine hohe Kanten-Betweenness haben. Durch das Entfernen dieser Kanten, können die Communities separiert werden und es kann die zugrunde liegende Community-Struktur aufgedeckt werden. Der Algorithmus wird darum auch *Edge-Remover-Clusterer* genannt.

Der Edge-Remover-Algorithmus für gewichtete Graphen ist wie folgt anzuwenden [Newm04]:

1. Berechnung der Betweenness für alle Kanten in einem Netzwerk.
2. Division der Betweenness durch das Gewicht der jeweiligen Kante
3. Entfernung der Kante mit der höchsten resultierenden Betweenness.
4. Neuberechnung der resultierenden Betweenness für alle übrigen Kanten.
5. Wiederholung der Schritte ab Schritt 3 bis keine Kanten übrig bleiben.

Das dargestellte Problem und der Lösungsalgorithmus wurde von uns auf Modularisierung von Architekturen übertragen. Bei Bedarf können die Kanten – insbesondere die der Nutzung und Verbindung von und zwischen IT-Systemen – mit Gewichten versehen werden.

¹ Im Folgenden wird angelehnt an den Originalbeitrag von [GiNe02] weiter von *Betweenness* gesprochen als einem Grad, wie sehr ein Element zwischen den anderen liegt. Es gibt, unseres Erachtens nach, keine entsprechende deutsche Übersetzung.

Durch die wiederholte Anwendung des Algorithmus im Schritt fünf, lässt sich auch eine mehrstufige Modularisierung erreichen. Eine weitere Frage ist, wann eine gute Modularisierung durch das Entfernen von Kanten erreicht ist bzw. wann andernfalls noch weitere Kanten entfernt werden sollten. Girvan/Newman definieren dazu die *Modularität* Q [NeGi04, S. 7 f]. Sie berechnen den Anteil der Kanten eines Graphen welche sich innerhalb der Communities befinden mit [Newm04, S. 7 f]:

$$\frac{\sum_{ij} A_{ij} \delta(c_i, c_j)}{\sum_{ij} A_{ij}} = \frac{1}{2m} \sum_{ij} A_{ij} \delta(c_i, c_j) \quad (3)$$

Dabei ist c_i die Community der der Knoten i angehört. Die Funktion $\delta(u, v)$ ist dann 1 wenn $u = v$, sonst ist sie 0. Weiterhin ist $m = \frac{1}{2} \sum_{ij} A_{ij}$ die Anzahl der Kanten des Graphen. Wenn nun die Grad k_i für alle Knoten i beibehalten werden, jedoch die Kanten zufällig im Netzwerk verteilt werden, dann ist die Wahrscheinlichkeit, dass zwischen den Knoten i und j eine Kante existiert $k_i k_j / 2m$. Daraus ergibt sich die Modularität Q zu

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j). \quad (4)$$

Daraus ergeben sich für Q Werte zwischen 0 und 1. Ein Wert von 0 bedeutet, dass es nach einem Clustering keine Kanten mehr in einer Community gibt, die man dort gemäß dem Zufall erwarten würde. Reale Werte für Q , welche ein gutes Clustering vermuten lassen, liegen zwischen 0,3 und 0,7. Prinzipiell können bei der sukzessiven Entfernung von Kanten nach dem Girvan/Newman Algorithmus mehrere lokale Maxima auftreten. Diese Maxima beschreiben gute (hierarchische) Modularisierungen.

4.3 EA Builder Software System

Der Modellierungs- und Clusteringansatz wurde prototypisch im Kontext von Unternehmensarchitekturen in einem Softwaresystem namens EA Builder implementiert (Abbildung 3) [<http://www.ea-builder.com>].

Im Gegensatz zu vielen der am Markt erhältlichen EA-Tools [Mat⁺05] unterstützt das Metamodell des EA Builders die Abbildung prozessorientierter Systemintegration über die so genannte dynamische Sicht. Dies ist unabhängig davon, welcher Integrationsansatz gewählt wurde.

Die Leistung des eingesetzten Clusteringalgorithmus wurde an einer Reihe spezieller Testszenerien untersucht [Aier06]. Im Folgenden wird die Funktionsweise des EA Builders anhand des

fiktiven Unternehmens WMYPC (We Make Your PC) demonstriert. Das Unternehmen baut und vertreibt individuell konfigurierte Computersysteme – angefangen beim kleinen Multimediale System bis hin zu größeren Serversystemen. Die Geschäftsprozesse werden durch sechs individuell implementierte Informationssysteme sowie ein Standardsoftwaresystem unterstützt.

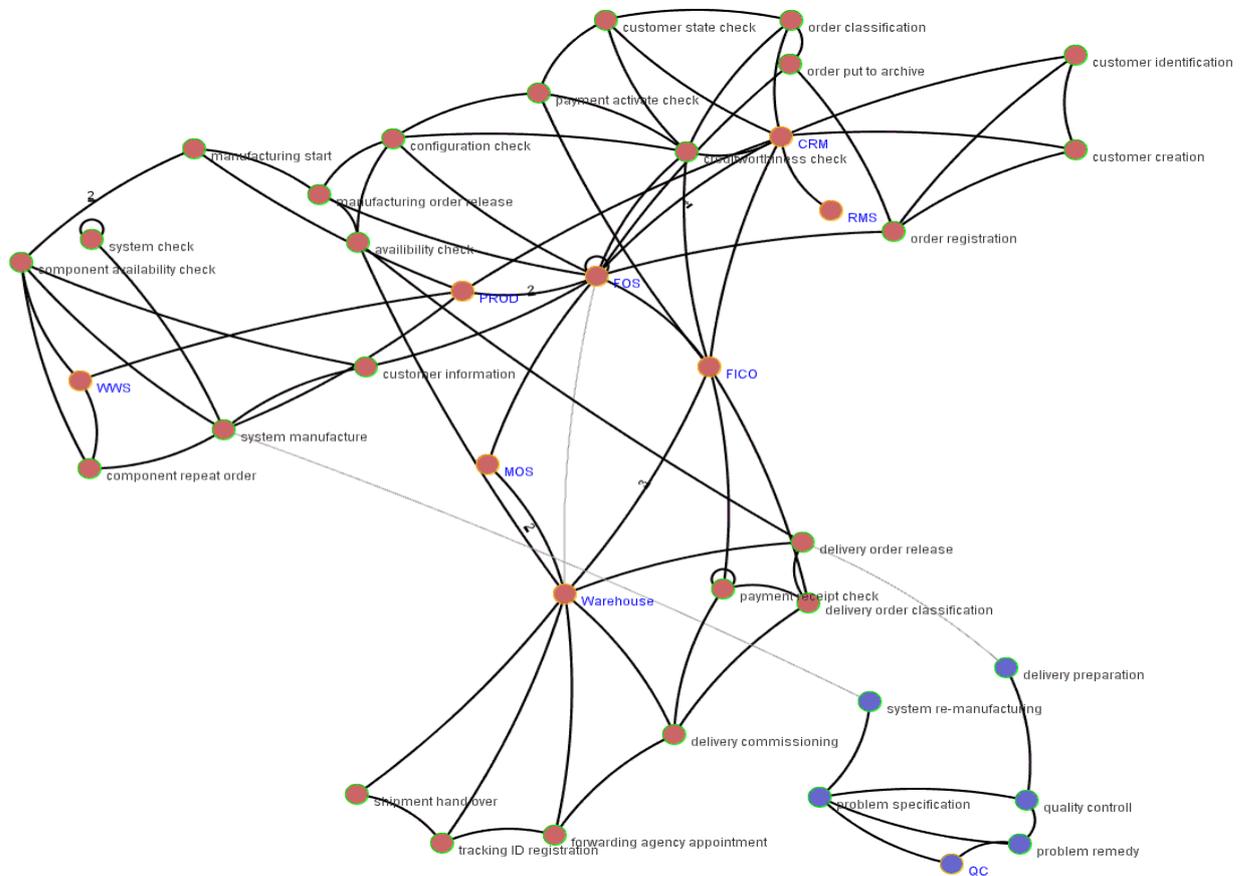


Abbildung 6: Modell transformiert in einen Graphen, drei Kanten entfernt

Abbildung 6 zeigt das zugehörige EA-Modell, welches in einen Graphen transformiert wurde. Bei 20 entfernten Kanten erreicht die Modularitätsfunktion ihr Maximum. Nach der Anwendung des Clusteringalgorithmus und dem Entfernen von 20 Kanten (Abbildung 7) ist der Graph in fünf Cluster aufgeteilt.

Das klassische Szenario für die Anwendung unseres Ansatzes ist die Transformation komplexer IT-Landschaften beispielsweise in eine serviceorientierte Architektur. Der Ansatz kann als Bottom-up-Ansatz bezeichnet werden, da die Cluster aus einer bestehenden Ist-Architektur abgeleitet werden. Verglichen mit einem Top-down-Ansatz führt dies zu einer potenziell evolutionären Architekturtransformation.

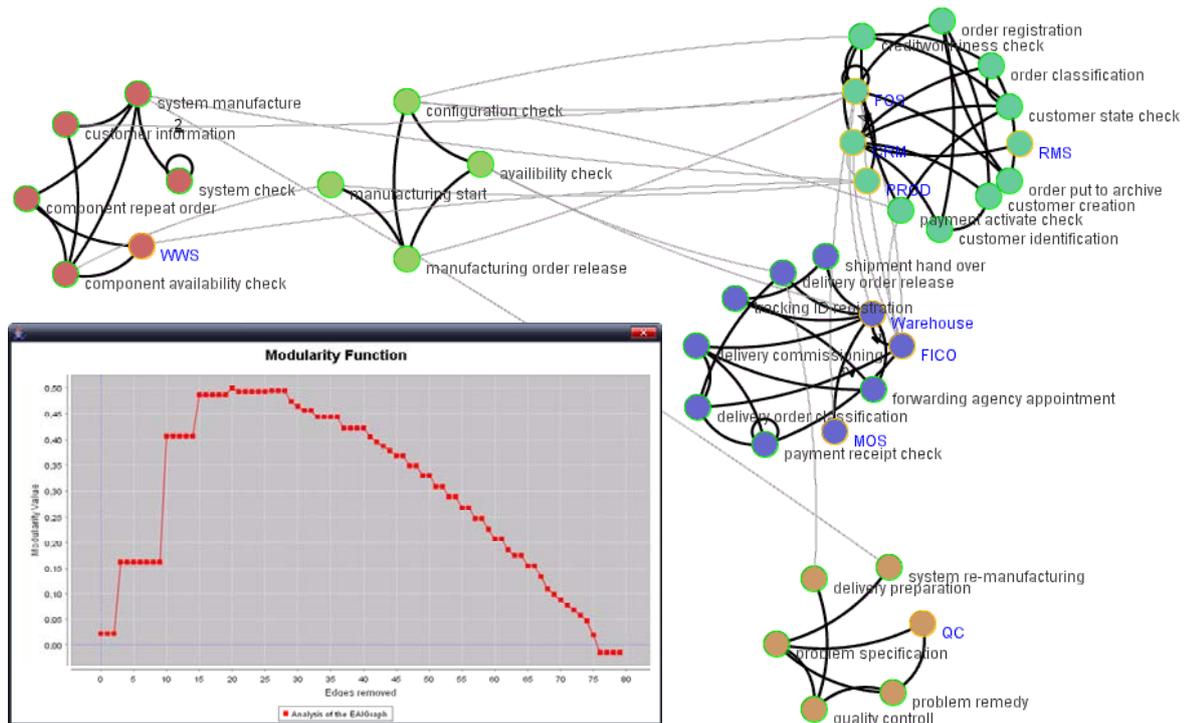


Abbildung 7: Screenshot des EA Builder Modells transformiert in einen Graphen, nach entfernen von 20 Kanten entstehen fünf Cluster, gruppierte Ansicht, Verlauf der Modularity-Funktion

Die zu Beginn genannte Forderung eines Business-IT-Alignment – hier strukturelle Analogie – wird durch diesen Ansatz unterstützt. Nach der Transformation des EA-Modells in einen Graphen, wird für das Clustering nicht mehr zwischen Prozessen und IT-Systemen differenziert. Das bedeutet, dass auch keine isolierte Optimierung von Prozess- bzw. IT-System-Strukturen erfolgt – obwohl auch diese möglich wäre. Die gebildeten Cluster reflektieren das tatsächliche Zusammenspiel von Prozessen und IT-Systemen und führen damit potenziell zu struktureller Analogie zwischen beiden.

5 Fazit

Ausgehend von der Nachhaltigkeit als Gestaltungsziel für Unternehmensarchitekturen, haben wir die Modularisierung der Architekturen als mögliche Umsetzungsstrategie vorgeschlagen. Betont wurde dabei die Notwendigkeit, Geschäftsprozesse und sie unterstützende IT-Systeme gemeinsam zu betrachten. Dafür wurde das Konzept der strukturellen Analogie genutzt. Basierend auf einem Architekturmodell, bestehend aus den hier als relevant erachteten Elementen, wurde der Einsatz von Clusteringalgorithmen zur Identifikation von Modulen in der Architektur vorgeschlagen. Dieser Ansatz wurde prototypisch im EA Builder-Werkzeug implementiert.

Die so entstandenen Module (Cluster) beschreiben jedoch „nur“ die latent vorhandene Ist-Struktur. Diese Ist-Struktur ist jedoch nicht notwendiger Weise mit dem Organigramm oder einer Karte der Anwendungslandschaft identisch, da es die tatsächliche, nicht die geplante oder vermutete Struktur beschreibt. Abhängig von einer Architektur- oder Integrationsstrategie muss nun entschieden werden, wie mit den Informationen aus dem Clustering weiter verfahren wird. Wird beispielsweise eine serviceorientierte Architektur (SOA) als Paradigma zugrunde gelegt, sind die identifizierten Module Kandidaten für so genannte Service Domains [BaHe04] für welche dann individuelle Integrations- und Managementkonzepte entwickelt werden können.

Der vorgeschlagene Ansatz liefert kein abgeschlossenes Vorgehensmodell für die Gestaltung nachhaltiger Architekturen. Der Nutzen liegt vielmehr darin, latent vorhandene Informationen über komplexe Architekturen zu erhalten, die durch eine einfache Visualisierung nicht sichtbar werden. Diese Informationen bilden dann wiederum die Grundlage für die gezielte Beeinflussung der Architekturen.

Weiterer Forschungsbedarf wird insbesondere in der Analyse des Einflusses weiterer Eigenschaften der modellierten Elemente auf das Clustering-Ergebnis gesehen. So können Elementen und Verbindungen zwischen diesen Attribute gegeben werden, die dann über Gewichte im Clustering berücksichtigt werden. In diesem Zusammenhang erscheint eine Erweiterung des Metamodells um weitere Elementtypen ebenfalls sinnvoll.

Literatur

[Aier06] Aier, Stephan: How Clustering Enterprise Architectures helps to Design Service Oriented Architectures. In: Proceedings of the IEEE International Conference on Services Computing (SCC'06), (Chicago, USA, 2006), IEEE Computer Society, Los Alamitos, CA, USA, S. 269–272.

[AiDo04] Aier, Stephan; Dogan, Turgut: Nachhaltigkeit als Gestaltungsziel von Unternehmensarchitekturen. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Serviceorientierung und nachhaltige Architekturen. Gito, Berlin 2004, S. 75–122.

[AiDo05] Aier, Stephan; Dogan, Turgut: Indikatoren zur Bewertung der Nachhaltigkeit von Unternehmensarchitekturen. In: Ferstl, O.K.; Sinz, E.J. (Hrsg.): Wirtschaftsin-

formatik 2005: eEconomy, eGovernment, eSociety. Physica, Heidelberg 2005, S. 607–626.

[AiSc03] Aier, Stephan; Schönherr, Marten: Flexibilisierung von Organisations- und IT-Architekturen durch EAI. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Management komplexer Architekturen. Gito, Berlin 2003, S. 3–56.

[Bas+03] Bass, Len; Clements, Paul; Kazman, Rick: Software Architecture in Practice. 2. Auflage, Pearson Education Inc., Boston 2003.

[BaHe04] Bath, Uwe; Herr, Michael: Implementation of a service oriented architecture at Deutsche Post MAIL. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Serviceorientierung und nachhaltige Architekturen. Gito, Berlin 2004, S. 279–297.

[Blei91] Bleicher, Knut: Organisation: Strategien, Strukturen, Kulturen. 2. Auflage, Gabler, Wiesbaden 1991.

[Free77] Freeman, L. C.: A Set of Measures of Centrality based upon Betweenness. In: Sociometry 40 (1977) S. 35–41.

[Fres00] Frese, Erich: Grundlagen der Organisation: Konzept – Prinzipien – Strukturen. 8. Auflage, Gabler, Wiesbaden 2000.

[GiNe02] Girvan, Michelle; Newman, Mark E. J.: Community Structure in Social and Biological Networks. In: Proceedings of the National Academy of Science 99 (2002) 12, S. 7821–7826.

[Gron03] Gronau, Norbert: Wandlungsfähige Informationssystemarchitekturen – Nachhaltigkeit bei organisatorischem Wandel. Gito, Berlin 2003.

[Hage03] Hagen, Claus: Integrationsarchitektur der Credit Suisse. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Management komplexer Architekturen. Gito, Berlin 2003, S. 61–83.

- [HaHu01] Hahn, Dietger; Hungenberg, Harald: PuK – Wertorientierte Controllingkonzepte. 6. Auflage, Gabler, Wiesbaden 2001.
- [Hube95] Huber, Joseph: Nachhaltige Entwicklung durch Suffizienz, Effizienz und Konsistenz. In: Fritz, P.; Huber, Joseph; Levi, Hans Wolfgang (Hrsg.): Nachhaltigkeit in naturwissenschaftlicher und sozialwissenschaftlicher Perspektive. Hirzel, Stuttgart 1995, S. 31–46.
- [Klam00] Klamma, Ralf: Vernetztes Verbesserungsmanagement mit einem Unternehmensgedächtnis-Repository Fakultät für Mathematik, Informatik und Naturwissenschaften, Rheinisch-Westfälischen Technischen Hochschule Aachen, Aachen, 2000.
- [Krc03] Krcal, Hans-Christian: Systemtheoretischer Metaansatz für den Umgang mit Komplexität und Nachhaltigkeit. In: Leisten, Rainer; Krcal, Hans-Cristian (Hrsg.): Nachhaltige Unternehmensführung – Systemperspektiven. Gabler, Wiesbaden 2003, S. 3–30.
- [Krcm90] Krcmar, Helmut: Bedeutung und Ziele von Informationssystem-Architekturen. In: Wirtschaftsinformatik 32 (1990) 5, S. 395–402.
- [LeSt03] Leitschuh-Fecht, Heike; Steger, Ulrich: Wie wird Nachhaltigkeit für Unternehmen attraktiv? – Business Case für nachhaltige Unternehmensentwicklung. In: Linne, Gudrun; Schwarz, Michael (Hrsg.): Handbuch Nachhaltige Entwicklung. Leske + Budrich, Opladen 2003, S. 257–266.
- [LeHu98] Lewin, Arie Y.; Hunter, Starling D.: Information Technology & Organizational Design: A Longitudinal Study of Information Technology Implementations in the U.S. Retailing Industrie, 1980–1996. In: Glaser, Horst; Schröder, Ernst F.; Werder, Axel v. (Hrsg.): Organisation im Wandel der Märkte. Gabler, Wiesbaden 1998, S. 251–286.
- [MaRo88] Markus, M. Lynne; Robey, Daniel: Information Technology and Organizational Change: Causal Structure in Theory and Research. In: Management Science 34 (1988) 5, S. 583–589.

- [Mat+05] Matthes, Florian; Wittenburg, André; Ernst, Alexander M.; Lankes, Josef; Schweda, Christian M.: Enterprise Architecture Management Tool Survey, Software Engineering for Business Information Systems (sebis), Ernst Denert-Stiftungslehrstuhl, Chair for Informatics 19, TU München, München, 2005.
- [Nad+94] Nadler, David A.; Gerstein, Marc S.; Shaw, Robert B.: Organisationsarchitektur. Campus, Frankfurt/Main, New York 1994.
- [Newm04] Newman, Mark E. J.: Analysis of Weighted Networks. In: Phys. Rev. E 70 (2004) 056131,
- [NeGi04] Newman, Mark E. J.; Girvan, Michelle: Finding and Evaluating Community Structure in Networks. In: Phys. Rev. E 69 (2004) 026113,
- [OMa+05] O'Madadhain, Joshua; Fisher, Danyel; Smyth, Padhraic ; White, Scott ; Boey, Yan-Biao: Analysis and Visualization of Network Data using JUNG. In: Journal of Statistical Software (2005) noch nicht erschienen.
- [Scha02] Schöffner, Lothar: Der Beitrag der Veränderungsforschung zur Nachhaltigkeit von Organisationsentwicklung. Hampp, München et al. 2002.
- [Scot05] Scott, John: Social Network Analysis: A Handbook. 2. Auflage, SAGE, London u.a. 2005.
- [Stue96] Stünzer, Lilia: Systemtheorie und betriebswirtschaftliche Organisationsforschung. Duncker & Humblot, Berlin 1996.
- [Wall96] Wall, Friederike: Organisation und betriebliche Informationssysteme – Elemente einer Konstruktionslehre. Gabler, Wiesbaden 1996.
- [WaFa99] Wasserman, Stanley; Faust, Katherine: Social Network Analysis: Methods and Applications. Cambridge Univ. Press, Cambridge u.a. 1999.
- [WiLa06] Winter, Robert; Landert, Karl: IT/Business Alignment als Managementherausforderung - Editorial zum Schwerpunktthema. In: Wirtschaftsinformatik 48 (2006) 5, S. 309.

„Open Model“ – ein Vorschlag für ein Forschungsprogramm der Wirtschaftsinformatik

Ulrich Frank¹, Stefan Strecker¹, Stefan Koch²

¹Lehrstuhl für Wirtschaftsinformatik und Unternehmensmodellierung
Institut für Informatik und Wirtschaftsinformatik
Universität Duisburg-Essen, Campus Essen

²Wirtschaftsuniversität Wien
Department für Informationsverarbeitung und Prozessmanagement
Institut für Informationswirtschaft

Abstract

Referenzmodelle sind ein zentraler Untersuchungsgegenstand der Wirtschaftsinformatik. Ihre Entwicklung ist nicht nur wissenschaftlich reizvoll, sie stellen gleichzeitig eine wirksame Unterstützung der Praxis in Aussicht. Diesen Verheißungen zum Trotz verlief die Entwicklung und – vor allem – Verbreitung wissenschaftlich fundierter Referenzmodelle bisher eher zurückhaltend. Im vorliegenden Beitrag werden zunächst Gründe für den bisher unzureichenden Erfolg von Referenzmodellen analysiert. Anschließend wird – inspiriert durch den Erfolg von Open-Source-Software – untersucht, ob die gemeinschaftliche Entwicklung offener Referenzmodelle geeignet ist, die Erstellung und Verbreitung solcher Modelle nachhaltig zu fördern. Dazu werden verschiedene Optionen zur Entwicklung offener Referenzmodelle unterschieden und bewertet. Vor diesem Hintergrund wird ein Vorschlag zur Durchführung einschlägiger Projekte entworfen.

1 Einleitung

Referenzmodelle in der Wirtschaftsinformatik sind i. d. R. konzeptuelle Modelle. Ein konzeptuelles Modell ist vor allem auf eine Rekonstruktion zentraler Begriffe einer betrachteten Domäne mittels einer geeigneten Modellierungssprache gerichtet. Referenzmodelle sind idealisierte konzeptuelle Modelle ausgewählter Domänenklassen. Sie sind mit dem Ziel verbunden, angemess-

sene Orientierungen für eine größere Zahl von Unternehmen zu liefern. Die Erstellung von Referenzmodellen erfolgt sowohl in deskriptiver, als auch in präskriptiver Absicht: Einerseits sollen sie wesentliche Merkmale der betrachteten Domänen angemessen beschreiben, andererseits sollen sie auch Vorschläge für innovative Formen der Gestaltung und Nutzung betrieblicher Informationssysteme enthalten. Es ist seit langem Konsens, dass Referenzmodelle eine überaus attraktive Orientierung für die Wirtschaftsinformatik darstellen. So ist ihre Entwicklung eine anspruchsvolle Forschungsaufgabe. Gleichzeitig versprechen gute Referenzmodelle einen hohen wirtschaftlichen Nutzen in der Praxis. So liefern sie einerseits eine Grundlage für die Entwicklung und den Einsatz von Software, die durch vielfache Wiederverwendung ein scheinbares ökonomisches Paradoxon aufzulösen verspricht: höhere Qualität zu geringeren Kosten. Andererseits bieten Referenzmodelle eine hervorragende Basis für die – ggfs. unternehmensübergreifende – Integration betrieblicher Informationssysteme, da sie den zu integrierenden Systemteilen als gemeinsames semantisches Referenzsystem dienen. Hier ist im Hinblick auf die statische Integration etwa an Daten- oder Objektmodelle, im Hinblick auf die dynamische Integration an Prozessmodelle zu denken. Angesichts solcher Vorteile ist es verwunderlich, dass die Zahl von Referenzmodellen, die in der Praxis wie in der Forschung entwickelt wurden, bisher sehr überschaubar ist [FeLo04].

Vor diesem Hintergrund werden zunächst die besondere Bedeutung von Referenzmodellen für die Forschung in der Wirtschaftsinformatik und die daran anknüpfenden Herausforderungen betrachtet. Anschließend wird untersucht, wie diesen Herausforderungen mit der Entwicklung offener Referenzmodelle begegnet werden kann. Der Beitrag knüpft an [KoSF06] an, in dem die Idee einer „Open Model“-Initiative zuerst vorgestellt wurde. Da das dargestellte Vorhaben eine breite Beteiligung voraussetzt, ist er nicht zuletzt als Diskussionsbeitrag gedacht. Dieser Umstand wird u. A. dadurch betont, dass die zentralen Hypothesen, auf die sich die Argumentation stützt, als solche kenntlich gemacht werden.

2 Referenzmodelle als Gegenstand und Ziel der Forschung

Die besondere Bedeutung, die der Entwicklung von Referenzmodellen in der Wirtschaftsinformatik zukommt, lässt sich durch einen Vergleich mit anderen Disziplinen unterstreichen. Dabei ist zu berücksichtigen, dass die Wirtschaftsinformatik eine anwendungsorientierte Diszi-

plin ist: Die Ergebnisse der Forschung sollen geeignet sein, die Wettbewerbsfähigkeit von Unternehmen zu fördern.

2.1 Die besondere Bedeutung von Referenzmodellen für die Wirtschaftsinformatik

Referenzmodelle in der Wirtschaftsinformatik weisen eine deutliche Ähnlichkeit zu den Artefakten auf, die in den Ingenieurwissenschaften und insbesondere in der Informatik entwickelt werden. Daneben gibt es aber auch einige Besonderheiten. Im Unterschied zu gängigen Modellen oder Referenzartefakten der Ingenieurwissenschaften sind Referenzmodelle auf Handlungskomplexe gerichtet, für die es i. d. R. keine eindeutige Anforderungsspezifikation gibt. Vielmehr reflektieren Referenzmodelle vermutete tatsächliche oder für sinnvoll erachtete Anforderungen, basieren also auf Hypothesen über eine Klasse von Handlungssystemen. Zudem sind sie nicht auf die Erstellung physischer Systeme gerichtet. Vielmehr sind sie vor allem eine sprachliche Rekonstruktion, die eine Grundlage für weitere Analysen oder für die Transformation in andere Repräsentationen wie z. B. Programmcode darstellt. Der Aufwand, der mit ihrer Anwendung verbunden ist, kann deshalb deutlich geringer als in den Ingenieurwissenschaften sein. Im Unterschied zu Modellen bzw. Referenzarchitekturen der Informatik sind Referenzmodelle zumeist auf einer fachlichen Ebene angesiedelt, beschreiben also auch den Handlungskontext, in dem Informationssysteme eingesetzt werden. Dazu fokussieren sie auf eine Rekonstruktion der Fachbegriffe der jeweiligen Domäne. Das ermöglicht die Beschreibung von Systemen auf einem deutlich höheren semantischen Niveau als dies in der Informatik üblich ist. Dadurch wird gegenüber den Modellen der Informatik zweierlei in Aussicht gestellt: ein höherer Wiederverwendungskomfort und ein höheres Integrationsniveau von Informationssystemen.

Gehaltvolle Theorien, die eine technologische Transformation in erfolgreiche Systeme und die sie umgebenden Handlungskontexte ermöglichen, sind in der Wirtschaftsinformatik bisher nicht entstanden – und sie sind auch nicht in Sicht. Die Suche nach Mustern erfolgreichen Handelns, wie sie in *Information Systems* betrieben wird, ist aber auch deshalb unbefriedigend, weil sie zwangsweise rückwärts gerichtet ist und damit innovative, heute noch nicht existierende Ansätze ausklammert. Im Unterschied dazu stellen Referenzmodelle Vorschläge für die Gestaltung effizienterer Handlungssysteme dar – in die allerdings auch Annahmen bzw. – falls verfügbar – geeignete Theorien über die jeweils abgebildete Domäne einfließen. Das mit der konzeptuellen Modellierung verbundene Ziel, Systeme anschaulich darzustellen – und das heißt nicht zuletzt: mit Mitteln einer vertrauten Fachsprache – fördert zudem die Verbreitung von Referenzmodellen und macht sie gleichzeitig geeignet, als Medium für den Austausch mit der Praxis zu dienen.

Referenzmodelle können also als Objekt und Objektivierung der Wirtschaftsinformatik verstanden werden.

2.2 Das Paradoxon der Referenzmodellforschung

Wenn man der im vorhergehenden Abschnitt dargestellten Argumentation folgt, stellen Referenzmodelle einen reizvollen Forschungsgegenstand dar und versprechen der Praxis gleichzeitig einen hohen Nutzen. Für eine anwendungsorientierte Disziplin erscheint dies als eine gleichsam ideale Konstellation. Gleichzeitig stoßen die wissenschaftliche Erstellung von Referenzmodellen sowie ihre Verbreitung auf beachtliche Widrigkeiten, die in der Vergangenheit den Erfolg der Referenzmodellforschung erheblich behindert haben. Sie betreffen sowohl originär wissenschaftliche Aspekte, als auch die Randbedingungen des praktischen Einsatzes von Referenzmodellen. Von besonderer Bedeutung sind dabei die Anreize für Forscher wie auch für ggfs. zu beteiligende Unternehmen.

Ein wesentliches Merkmal der Referenzmodellierung ist Komplexität. Das betrifft den Umfang, den Referenzmodelle gewöhnlich erreichen, vor allem aber die Abstraktion, die erforderlich ist, um einer Vielzahl potentieller Anwendungsfälle gerecht zu werden. In der Regel wird dazu eine ausführliche Analyse ausgewählter Unternehmen nötig sein.

Hypothese 1: Der Aufwand, der mit der Erstellung von Referenzmodellen verbunden ist, übersteigt häufig die Möglichkeiten, über die einzelne Forscher – etwa im Rahmen eines Dissertations- oder Habilitationsprojekts – verfügen. Selbst Forschungsgruppen an Universitäten sehen sich hier vor kaum zu bewältigenden Herausforderungen.

Hypothese 2: Trotz des hohen Anspruchs, der mit der Entwicklung von Referenzmodellen verbunden ist, ist ihre Anerkennung als wissenschaftliche Leistung keineswegs garantiert.

Die Anreize, die ein Forschungsthema für einzelne Wissenschaftler mit sich bringt, hängen auch von den damit verbundenen Publikationsmöglichkeiten ab. Auch in dieser Hinsicht sieht sich die Referenzmodellforschung einigen sperrigen Hindernissen gegenüber. So sind die Publikationsgelegenheiten für sog. „Design Science“ international noch bescheiden. Vor allem aber sind Referenzmodelle i. d. R. viel zu umfangreich, um ihre Publikation in einer Zeitschrift oder in einem Konferenzband zu ermöglichen. Die Beschränkung auf einzelne Aspekte eines Referenzmodells ist u. U. eine Option, bedeutet aber, dass ein großer Teil der Gesamtleistung nicht dargestellt werden kann. Dieser Nachteil von Referenzmodellen erhält durch die Umstellung auf kumulative Habilitationsverfahren zusätzliches Gewicht.

Hypothese 3: Die Chancen, über die Beteiligung an der Entwicklung von Referenzmodellen karrierewirksame Publikationen zu erstellen, sind in den gegenwärtigen Wissenschaftsstrukturen relativ gering.

Ein gleichsam inhärenter Aspekt des Erfolgs von Referenzmodellen ist ihre Verbreitung. Sie setzt u. a. voraus, dass potentielle Nutzer von den Vorteilen eines Referenzmodells überzeugt sind. Neben flankierenden vertrauensbildenden Marketing-Maßnahmen wird das Vertrauen in die Qualität und den Nutzen von Referenzmodellen wohl auch dadurch gefördert, dass Nutzer an der Entwicklung beteiligt waren.

Hypothese 4: Der Anreiz für ein einzelnes Unternehmen, mit einem Wissenschaftler oder einer wissenschaftlichen Institution zusammenzuarbeiten, ist wegen zweifelhafter Erfolgsaussichten oftmals nicht hinreichend groß.

Unsere kurze Analyse zeigt das scheinbare Paradoxon auf, das die Referenzmodellforschung kennzeichnet: Auf der einen Seite stellen Referenzmodelle für eine anwendungsorientierte Disziplin wie die Wirtschaftsinformatik wie auch für prospektive Anwender in der Praxis eine überaus attraktive Anreize dar. Auf der anderen Seite wird ihre Entwicklung und Verbreitung dadurch behindert, dass für Wissenschaftler und Anwender in der Praxis die Anreize dafür, sich selbst zu engagieren, in vielen Fällen nicht überzeugend sind.

Hypothese 5: Nur wenn es gelingt, dieses scheinbare Paradoxon aufzulösen, können Referenzmodelle den Stellenwert in der Wirtschaftsinformatik erlangen, den sie verdienen.

Die Analyse gegenwärtiger Hindernisse der Referenzmodellforschung führt uns zu einer Reihe von Anforderungen, die in den folgenden Annahmen sowie damit verbundenen Fragen formuliert sind:

Hypothese 6: Die Entwicklung von Referenzmodellen erfordert konzertierte Projekte von mehreren Forschungsgruppen und prospektiven Nutzern. Welche Größe der Entwicklergemeinschaft ist sinnvoll?

Hypothese 7: Die Entwicklung und Verbreitung von Referenzmodellen erfordert die Beteiligung prospektiver Nutzer. Wie soll diese Beteiligung aussehen?

Hypothese 8: Die Beteiligung von Nutzern erfordert wirksame Anreize. An welche Anreize ist hier zu denken und wie können sie geschaffen werden?

Hypothese 9: Die engagierte und nachhaltige Mitwirkung von Forschern erfordert Anreize, die auch den Randbedingungen wissenschaftlicher Karrieren Rechnung tragen. An welche Anreize ist hier zu denken und wie können sie geschaffen werden?

3 Lehren aus der Open-Source-Software-Entwicklung

Um zu untersuchen, ob eine Initiative zur Entwicklung offener Referenzmodelle geeignet ist, die skizzierten Anforderungen zu erfüllen, betrachten wir zunächst Open-Source-Software, insbesondere die Gründe ihres Erfolgs. Vor diesem Hintergrund wird dann untersucht, welche Aspekte von Open Source Software sich auf die Entwicklung offener Referenzmodelle übertragen lassen.

3.1 Free/Libre/Open-Source-Software: Wesentliche Merkmale und Erfolgsfaktoren

Mit Wurzeln in akademischen Gemeinschaften U.S.-amerikanischer Ingenieurs- und Informatikfakultäten („Hackerkulturen“ [LeTi02]) hat sich aus der Idee freier bzw. quelloffener Software (Free/Libre/Open-Source-Software, abgekürzt FLOSS, im Folgenden auch verkürzt OSS) [Pere99; Stal02] ein vielschichtiges (software-) technisches, wirtschaftliches und gesellschaftliches Phänomen entwickelt [BHPC04]: In Gemeinschaften geografisch verteilter Akteure entstehen in kollektiven Prozessen komplexe Softwaresysteme unter Nutzung elektronischer Medien. Sowohl System- (z. B. GNU/Linux, Apache, MySQL) als auch Anwendungssoftware (z. B. Mozilla Firefox, OpenOffice) sind als Beispiele für die erfolgreiche Durchführung von OSS-Projekten weltweit bekannt geworden. Die kennzeichnenden Merkmale der Entwicklung von OSS sind in Erfahrungsberichten beteiligter Akteure [Raym01; MoFH02], Beobachtungen und empirischen Studien Außenstehender [StGo06; BHPC04] dokumentiert. Die folgende Charakterisierung basiert auf diesen Quellen und differenziert folgende Aspekte: Organisation, Ablauf, Anreize und Erfolgsfaktoren.

Konstituierend ist eine Lizenzierung von Software, mit der ein Urheber bzw. Rechteinhaber Lizenznehmern Zugang zum Quellcode der Software sowie nichtexklusive Nutzungs- und Verwertungsrechte zur Bearbeitung, Vervielfältigung und Verbreitung von originärem und modifiziertem Quelltext einräumt, ohne dafür Lizenzgebühren zu verlangen. Mit der Entwicklung von OSS ist eine spezifische Form der Projektorganisation verbunden, die auf einem offenen und öffentlichen Zugang basiert. Ein OSS-Projekt wird von einer Gemeinschaft („community“) von, zumindest teilweise, freiwillig mitwirkenden Softwareentwicklern unter Einbindung von Endanwendern betrieben. Diese Entwickler- und Anwendergemeinschaft teilt gemeinsame Werte, Überzeugungen und Normen [StGo06] und entwickelt darauf aufbauend eine spezifische Projektkultur zur Stiftung einer gemeinsamen Identität [BHPC04]. Die räumliche Verteilung

der Mitwirkenden erzwingt eine vernetzte und asynchrone Zusammenarbeit über elektronische Medien (E-Mail, Newsgroups, World Wide Web, ICQ usw.).

OSS-Projekte zeichnen sich durch ein hohes Maß an Arbeitsteilung aus, wobei die Aufgabenverteilung und -zuordnung auf Selbstselektion basiert. Der Grad an Arbeitsteilung bedingt einen hohen Koordinationsbedarf. Die Koordination erfolgt durch Signalisierung eines Einzelnen oder eines Teams von Mitwirkenden an die Gemeinschaft. Zur Koordination in OSS-Projekten haben sich typische Rollenkonzepte, Macht- und Entscheidungsstrukturen entwickelt (weitere Ausführungen zur Organisation von OSS-Projekten finden sich in einem begleitenden Arbeitsbericht [FrSK06]).

Das primäre Anreizsystem in OSS-Projekten basiert auf der Reputation des Einzelnen, der sich durch wesentliche Beiträge zum Projekt die Anerkennung und das Ansehen seiner Mitstreiter erwirbt. Reputation wird in einer Projekthierarchie durch die Zuweisung hervorgehobener Rollen honoriert und dient damit der Erreichung eines angesehenen sozialen Status innerhalb einer Gemeinschaft. Die Erarbeitung von Reputation dient als Anreiz neue Mitstreiter zu gewinnen; die Erhaltung der Reputation motiviert bestehende Mitwirkende ihre Beiträge in angemessener Zeit und Qualität einzureichen, welche dann eine gegenseitige Begutachtung („peer review“) durchlaufen. Die möglichen negativen Folgen aus einem Verlust an Reputation wirken dabei als zusätzlicher Anreiz. Über die Gemeinschaft hinaus wirkt die Erarbeitung einer hohen Reputation als Investition in das eigene Humankapital im Sinne einer Steigerung des eigenen Wertes auf dem Arbeitsmarkt. Neben, auf die extrinsische Motivation gerichteten, Anreizen weisen empirische Studien diverse intrinsische Motivationsgründe von Individuen nach, darunter die Freude am Umgang mit dem Endprodukt und die gleichsam künstlerische Freude an der Erstellung von qualitativ hochwertiger Software [u. a. BoRo03, HeNH03].

Es liegen noch keine gesicherten Erkenntnisse darüber vor, welche Kriterien für den Erfolg eines OSS-Projekts ausschlaggebend sind [CrHA06]. Die vorliegenden Berichte und Studien lassen allerdings erste Schlüsse zu, die im Folgenden dargestellt sind:

- Der Erfolg von OSS-Projekten gründet u. a. auf einer zwangsfreien, virtualisierten und dezentralisierten Organisation. Die Projektkultur und die ihr zugrunde liegende Weltanschauung wirken positiv auf gegenseitiges Vertrauen und Kommunikationsqualität und damit auf den Projekterfolg [StGo06; Gall01].

- Eine hohe intrinsische Motivation der Entwickler, ausgelöst u. a. durch die Identität von Entwickler und Nutzer, führt zu einer ausgeprägten Fokussierung auf die Projektziele und damit auf den Projekterfolg.
- Die Zugangsmodalitäten fördern die Entwickler- und Nutzerbeteiligung, womit positive Skaleneffekte durch weitgehende Arbeitsteilung und Parallelisierung verbunden sind (z.B. in Bezug auf Softwaretests und Fehlerbehebung) [Raym01].
- Die Arbeitsteilung durch Selbstselektion fördert die Bearbeitung von erfolgswirksamen Aufgaben, die anderer Kompetenzen als die der Programmierung bedürfen (z. B. Dokumentationserstellung, Anwenderunterstützung, Öffentlichkeitsarbeit), da Mitwirkende ihre individuellen Kompetenzen in die Projektarbeit einbringen können.
- Die Selbstkontrolle der Beteiligten bzw. die soziale Kontrolle durch die Gemeinschaft senken den Koordinationsbedarf [Gall01]. Reputation fungiert als wirksames Instrument der Selbstkontrolle. „Peer review“ ist ein wirksames Instrument der sozialen Kontrolle [MaMA00].
- Etablierte Regeln betonen u. a. den Einbezug aller Projektmitwirkenden von Beginn an sowie frühe und schnelle Freigabezyklen („Release early, release often. And listen to your customers“ [Raym01]) und erhöhen die Effektivität und Effizienz des Entwicklungsprozesses.
- Die Modularisierung von Software erhöht die potentielle Arbeitsteiligkeit und reduziert Koordinationsbedarfe [BoRo03].

3.2 Zum Verhältnis von „Open Source“ und „Open Model“

Unter einem offenen Referenzmodell („Open Model“, kurz OM) verstehen wir in Analogie zu OSS ein Referenzmodell, welches durch den Urheber bzw. Rechteinhaber mit einer Lizenz versehen wird, die Lizenznehmern Zugang zum Referenzmodell sowie nichtexklusive Nutzungs- und Verwertungsrechte zur Bearbeitung, Vervielfältigung und Verbreitung von originärem und modifiziertem Referenzmodell einräumt, ohne dafür Lizenzgebühren zu verlangen. Dieselben Rechte werden auch für die Modelldokumentation gefordert. Neben diesen grundlegenden rechtlichen Merkmalen gibt es eine Reihe weiterer Gemeinsamkeiten. Referenzmodelle wie Programme sind

- digital repräsentierte Artefakte, die einer Modularisierung und damit einer arbeitsteiligen Entwicklung zugänglich sind. Sie eignen sich damit prinzipiell für die gemeinsame Bearbeitung durch verteilte Gemeinschaften.
- zielgerichtete, anspruchsvolle Konstruktionen, deren Konkretisierung von Entwurfsentscheidungen abhängen, die eine erhebliche fachliche Kompetenz erfordern.
- Konstruktionen, die einer formalen Syntax genügen und damit automatisierten Analysen und Transformationen zugänglich sind.
- Konstruktionen, die einer begleitenden Dokumentation bedürfen, um Wartung und Weiterentwicklung zu gewährleisten.
- Konstruktionen, deren Qualität nicht durch Rückgriff auf etablierter Metriken gemessen werden kann. Vielmehr ist in beiden Fällen ein geeigneter Prozess der Qualitätssicherung zu etablieren.
- Konstruktionen, die potentiell eine erhebliche wirtschaftliche Bedeutung haben. Dies gilt zum einen für die kostenlose Nutzung, zum anderen für die Schaffung neuer Geschäftsmodelle, die diese Nutzung begleiten.

Daneben sind einige mehr oder weniger deutliche Unterschiede zu beachten:

- Referenzmodellen sind im Vergleich zu Programmen durch eine höhere Abstraktion von technischen Randbedingungen und eine ausgeprägte Berücksichtigung domänenspezifischer Fachsprachen gekennzeichnet. Die dazu erforderliche Kompetenz unterscheidet sich deutlich von der, die für die Codierung klar spezifizierter Programmteile nötig ist.
- Programme belohnen ihre Entwickler durch das befriedigende Erlebnis, ein ausführbares Artefakt geschaffen zu haben, das unmittelbar für die Bewältigung von Aufgaben eingesetzt werden kann. Referenzmodelle sind i. d. R. nicht ausführbar. Ihr Nutzen entfaltet sich erst durch weitere Anpassungen, Transformationen und Interpretationen.
- Quellcode zeichnet sich durch vergleichsweise geringe Interpretationsspielräume hinsichtlich der Intention des Programmierers aus, während Referenzmodelle inhärent durch einen hohen Interpretationsspielraum gekennzeichnet sind. Letztlich sind

Referenzmodelle nur unter Bezug auf den intendierten Modellierungszweck und die Zuhilfenahme der Modelldokumentation sinnvoll zu interpretieren. Aus diesem Grund ist die Beurteilung der Qualität von Modellen tendenziell mit einer größeren Herausforderung verbunden als dies bei der Evaluation von Programmen der Fall ist.

- OSS profitiert davon, dass es weltweit eine breite Basis von Software-Entwicklern gibt. Demgegenüber ist die Entwicklung konzeptueller Modelle, mehr noch die Entwicklung von Referenzmodellen auf einen vergleichsweise kleinen Kreis von Fachleuten beschränkt.
- OSS profitiert von dem Mythos, der die Softwareentwicklung immer noch umgibt. Die Erstellung von Referenzmodellen wird im Unterschied dazu nicht durch Legenden und Mythen aufgewertet.
- OSS profitiert von der Anerkennung, die die „freie“ Entwicklung von Software als Akt der Emanzipation von der Allmacht internationaler Konzerne in weiten Kreisen der Gesellschaft genießt. Für die Entwicklung von Referenzmodellen ist dieser Effekt prinzipiell auch nutzbar. Allerdings ist die Herausforderung etablierter Produkte nicht so offenkundig.

Betrachten wir nun vor dem Hintergrund dieses Vergleichs die Frage, inwieweit die in 3.1 ermittelten Gründe für den Erfolg von OSS auf OM übertragbar sind. In Tab. 1 sind entsprechende Einschätzungen dargestellt, die im Folgenden erläutert werden.

- Normen, Überzeugungen und Werte: Die OSS-Gemeinschaften zugrunde liegende Weltanschauung ist weitgehend nicht Quellcode-spezifisch und daher gut auf die offene Referenzmodellierung übertragbar.
- Reputation als Anreizmechanismus: Referenzmodellierung ist eine intellektuell anspruchsvolle Tätigkeit. Es ist daher zu erwarten, dass ein bedeutender Beitrag die Anerkennung der Gemeinschaft findet und das Ansehen des Modellierers erhöht. Angesichts des hohen Aufwands, der mit der Referenzmodellierung verbunden ist, scheint es fraglich, ob ein hervorgehobener sozialer Status innerhalb der Gemeinschaft ausreichend Anreiz bietet. Es ist zu vermuten, dass flankierende Maßnahmen

zur Förderung von Reputation über die Grenzen eines OM-Projekts hinaus notwendig sind (z. B. in der wissenschaftlichen Modellierungscommunity).

- „peer review“: Mit der fundierten Begutachtung eines Modells ist ein hoher zeitlicher Aufwand verbunden. Zudem erfordert die Begutachtung eine hohe fachliche Kompetenz. Die Anfertigung eines Gutachtens bedarf daher zusätzlicher Anreize, insbesondere mit Blick auf die Einbindung von Domänenexperten.
- Etabliertes Regelwerk: Viele der von Raymond und anderen formulierten Regeln lassen sich gut auf die Entwicklung von Modellen übertragen [Raym01].

Erfolgsfaktor	Übertragbarkeit auf OM
Normen, Überzeugungen und Werte	Gut
Reputation als Anreizmechanismus	Offen
„peer review“	Offen
Etabliertes Regelwerk	Gut
Einbindung individueller Kompetenzen	Gut
Entwickler- und Nutzerbeteiligung	Offen
Modularität	Offen

Tabelle 1: Übertragbarkeit von Gründen für den Erfolg von OSS auf OM

- Einbindung individueller Kompetenzen: Auch ein OM-Projekt bedarf neben Modellierungskompetenz weiterer Kompetenzen, vorrangig domänenspezifische Fachkompetenz, aber auch Kompetenzen im Hinblick auf Modelldokumentation, Anwenderunterstützung und Öffentlichkeitsarbeit. Die Einbindung von Mitwirkenden mit entsprechenden Kompetenzen ist prinzipiell gut möglich. Allerdings bedarf die Beteiligung spezieller und dringend benötigter Kompetenzen (etwa zur Einbindung von Fachexperten aus der Praxis) weitergehender Anreize.
- Entwickler- und Nutzerbeteiligung: Die Beteiligung von Entwicklern wie Anwendern ist – entsprechende Zugangsmodalitäten vorausgesetzt – möglich. Die Übertragung der Skaleneffekte wird vor allem von der Modularisierbarkeit von Modellen abhängen.
- Modularität: Prinzipiell ist eine Modularisierung von Referenzmodellen z. B. durch integrierende Metamodelle möglich. Die Zerlegung *eines* Referenzmodells in meh-

rere bearbeitbare Teile, die zu einem späteren Zeitpunkt wieder in konsistentes Ganzes integriert werden können, erscheint allerdings als Herausforderung für OM-Projekte.

Der Vergleich zwischen OSS und OM liefert gute Gründe für die Annahme, dass die Etablierung von Gemeinschaften zur Entwicklung offener Referenzmodelle erfolgreich sein kann. Dies liegt nicht zuletzt auch daran, dass einige der in 2 im Hinblick auf die Entwicklung von Referenzmodellen formulierten Hypothesen für ein OM-Vorhaben nach dem Vorbild von OSS-Projekten offenkundig erfüllt wären, wie auch sich die daran anknüpfenden Fragen zumindest teilweise beantworten lassen (s. 4). Gleichzeitig wird deutlich, dass der Nutzen von OM schwerer zu vermitteln ist als der von OSS. Zur weiteren Evaluation eines solchen Vorhabens ist es deshalb erforderlich, die spezifischen Vorteile, die mit OM verbunden sind, deutlich herauszuarbeiten und die daraus resultierenden Anforderungen zu betrachten.

4 Kriterien zur Konfiguration von OM-Vorhaben

Es liegt auf der Hand, dass ein missionarischer Aufruf zur Erstellung offener Referenzmodelle allein kaum hinreicht. Es ist vielmehr nötig, eine Strategie zu entwickeln, die geeignet erscheint, die für die Etablierung offener Referenzmodelle notwendige kritische Masse an Entwicklern und Anwendern zu erzeugen. Eine solche Strategie umfasst neben der Berücksichtigung der bereits diskutierten Erfolgsfaktoren die Auswahl geeigneter Domänen, Modellierungszwecke, Abstraktionsebenen sowie eine Projektorganisation einschließlich einer Vorgehensweise (s. 5). Eine ausführliche Analyse dieser Aspekte findet sich in [FrSK06]. Die folgenden Hypothesen beschränken sich auf Aspekte, die für OM besonders gut geeignet scheinen, da es für die nachhaltige Etablierung einer solchen Initiative besonders wichtig ist, dass die ersten Projekte erfolgreich verlaufen. Die Hypothesen sind jeweils unter einer „ceteris paribus“-Annahme formuliert. Hinsichtlich der beteiligten Akteure unterscheiden wir Wissenschaftler und Unternehmen, die jeweils sowohl in der Rolle des Entwicklers als auch des Nutzers eines Referenzmodells auftreten können.

Domäne/Gegenstand: Um die Skaleneffekte von Referenzmodellen nutzen zu können, sollte es in der ausgewählten Domäne eine größere Zahl potentieller Nutzer geben, die hinreichende Gemeinsamkeiten aufweisen.

Zweck:

- Grundlage für die Entwicklung von Code: immer dann reizvoll, wenn die Softwareentwicklung in überzeugender Weise von einem Referenzmodell profitiert – z. B. durch die Förderung von Integration, Wiederverwendung und Wettbewerbsfähigkeit (s. u.).
- Grundlage für die Kommunikation zwischen (existierenden) Systemen, z. B. zur Spezifikation von Schnittstellen (lose Kopplung): immer dann reizvoll, wenn unternehmensübergreifende Wertschöpfungsprozesse modelliert werden und/oder die betrachtete Domäne durch „Legacy“-Systeme geprägt ist, die auf absehbare Zeit nicht ersetzt werden können.
- Orientierung für die Strukturierung komplexer, arbeitsteiliger Handlungsmuster, z. B. durch Referenzmodelle von Geschäftsprozessen: immer dann reizvoll, wenn die Domäne durch einen hohen Wettbewerbsdruck gekennzeichnet ist, insbesondere bei absehbarem Strukturwandel.

Akteure: Grundsätzlich gehen wir davon aus, dass die Entwicklung offener Referenzmodelle dadurch begünstigt wird, dass die Entwickler auch zu den prospektiven Nutzern zählen.

Unternehmen: Die Beteiligung an der Entwicklung eines offenen Referenzmodells ist reizvoll, wenn

- das Referenzmodell eine Verbesserung der Wettbewerbsfähigkeit gegenüber übermächtigen Konkurrenten in Aussicht stellt.
- das Referenzmodell zur Überwindung von Markteintrittsbarrieren beiträgt.
- das Referenzmodell ein Medium für ergänzende Dienstleistungen darstellt.
- die Mitwirkung an der Entwicklung eines Referenzmodells eine wertvolle Qualifizierung der beteiligten Mitarbeiter verspricht – nicht zuletzt durch den Austausch mit wissenschaftlichen Einrichtungen.
- die Preisgabe individueller Differenzierungsvorteile durch den Einsatz von Referenzmodellen durch den erwarteten Nutzen überkompensiert wird.

Wissenschaftler: Die Beteiligung an der Entwicklung eines offenen Referenzmodells ist reizvoll, wenn

- die Entwicklung eines Referenzmodells mit einem hohen Anspruch verbunden ist, der in deutlichem Zusammenhang zu disziplinspezifischen Forschungsaufgaben steht.
- die Mitwirkung an der Gestaltung von Referenzmodellen einen differenzierteren Zugang zur Anwendungspraxis verspricht.
- das Referenzmodell in Aussicht stellt, selbst als Forschungsgegenstand zu dienen, z. B. um Analyse- und Transformationsverfahren zu entwickeln.
- ein Referenzmodell eine Bereicherung der Lehre verspricht.

Abstraktionsniveau: Referenzmodelle können für unterschiedliche Abstraktionen genutzt werden. Das gilt zum einen für die Abstraktion von spezifischen Besonderheiten in konkreten Einsatzfällen, also dafür, wie generisch ein Modell ist. Zum anderen kann zwischen Metamodellen zur Spezifikation von Modellierungssprachen und Modellen als Anwendung von Modellierungssprachen unterschieden werden. Die Entscheidungssituation ist hier ambivalent, so dass letztlich im Einzelfall anhand der genannten Entscheidungskriterien geprüft werden muss, welche Option gewählt wird. Die jeweilige Empfehlung beruht auf der Annahme, dass frühzeitige Erfolge wichtig sind.

Generizität: Einerseits: Je weiter ein Referenzmodell von spezifischen Merkmalen einzelner Anwendungsfälle abstrahiert – also je anwendungsferner es ist, desto anspruchsvoller und aufwändiger ist seine Erstellung, was aus wissenschaftlicher Sicht reizvoll ist. Gleichzeitig werden größere Wiederverwendungs- und Integrationspotentiale geschaffen.

Andererseits: Je anwendungsnäher ein Referenzmodell ist, desto größer ist sein unmittelbarer Nutzen für diejenigen, deren Anforderungen berücksichtigt wurden. Gleichzeitig sind auch die Realisierungschancen günstiger zu bewerten.

Entscheidungskriterien: Machbarkeit, Varianz, verfügbare Ressourcen

Empfehlung: Beginn mit Bottom-Up, um frühzeitig erste Ergebnisse zu produzieren; später ergänzt um einen Top-Down-Ansatz.

Modell vs. Metamodell: Einerseits: Die Entwicklung von Modellierungssprachen wirkt auf Unternehmen tendenziell abschreckend. Zum einen, weil sie Investitionen in existierende Modellierungssprachen gefährdet, zum anderen, weil sie die Entwicklung anwendungsnaher Referenzmodelle auf längere Zeit zu verzögern droht.

Andererseits: Die Entwicklung von Modellierungssprachen ist besonders anspruchsvoll und gehört zu anerkannten Forschungsthemen der Wirtschaftsinformatik und der Informatik.

Entscheidungskriterien: Qualität existierender Sprachen und Werkzeuge, verfügbare Ressourcen, Bedeutung von Standards

Empfehlung: Entwicklung zunächst mit existierenden Modellierungssprachen und Werkzeugen; ggfs. parallel dazu Entwicklung besser geeigneter Sprachen und Werkzeuge.

Zur Prüfung eines möglichen Vorhabens ist neben der Frage, ob ein geeigneter Zweck vorliegt, zu klären, welche Kriterien aus der Sicht der beteiligten Gruppen erfüllt sind.

5 Hinweise zur Organisation und Durchführung

Anknüpfend an den in 4 entwickelten Bezugsrahmen bedarf die Initialisierung und nachhaltige Entwicklung von OM-Projekten einer, auf die Analyse in 3, abgestimmten Vorgehensweise. Hinweise zu einem Vorgehen zeigt die folgende Übersicht (in [FrSK06] stellen wir alternative Projektvorschläge zur Diskussion):

- Zur Initialisierung eines OM-Vorhabens bietet sich der Zusammenschluss weniger Forschergruppen an, die zusammen mit ausgewählten Praktikern eine Projektgemeinschaft begründen und durch wesentliche Beiträge einen überzeugenden Anfangserfolg des Projekts in Aussicht stellen (s. Hypothese 6).
- Es bietet sich gegenwärtig an, solche Anwendungsdomänen auszuwählen, die für eine offene Referenzmodellierung besonders Erfolg versprechend sind (s. 5). Überzeugende Anfangserfolge wirken als Katalysator für die Entwickler- und Nutzerbeteiligung.
- Die Einbindung etablierter Gremien aus Wissenschaft und Praxis verspricht, die Relevanz und Nachhaltigkeit des Vorhabens zu unterstreichen und dessen Sichtbarkeit zu erhöhen.
- Nach einer initialen Projektphase stellt die Öffnung des Vorhabens für die Beteiligung weiterer Entwickler und Anwender eine nachhaltige Entwicklung sowohl der Gemeinschaft, als auch der entstehenden Artefakte in Aussicht.

Für die intendierten Akteure sind Anreize zu gestalten und zu implementieren. Dazu sind Rollen zu definieren [KoSF06] und rollenspezifische Anreize zu bestimmen.

- Die Rollen Modellverwalter, Modellkonstrukteur, Modellgutachter, Modellnutzer und Begleitforscher bieten sich für Wissenschaftler an, wobei sich je nach wissenschaftlicher Disziplin eine spezifische Rolle ergeben mag (so ist für die Rolle Begleitforscher z. B. an Organisationsforscher und Soziologen zu denken). Für Unternehmen kommen vornehmlich die Rollen Modellgutachter und Modellnutzer in Frage. Für Domänenexperten ist hierbei zuerst an die Rolle des Modellgutachters zu denken und für Mitarbeiter in den IT-Abteilungen an die Rolle der Modellnutzer (s. Hypothese 7).
- Ein spezifischer Anreiz für Fachexperten könnte durch dedizierte Workshops zum Austausch mit beteiligten Wissenschaftlern erreicht werden (s. Hypothese 8).
- Die Beteiligung von Forschern erfordert eine weithin akzeptierte Anerkennung von Modellierungsleistungen als Publikations-äquivalent. Dazu sind spezifische Begutachtungsprozesse zu etablieren, die einen Beitrag zu einem OM-Projekt einer dem akademischen „peer review“ vergleichbaren Untersuchung unterziehen. Ein Ansatz wäre die Einrichtung eines anerkannten Gremiums, das eingereichte Modelle und Modelldokumentationen auf Anfrage z. B. doppel-blind begutachtet (s. Hypothese 9).

6 Abschließende Anmerkungen

Motiviert durch die hohe Attraktivität, die Referenzmodelle für Wissenschaft und Praxis aufweisen, sowie die ernüchternde Bilanz der bisherigen Referenzmodellforschung untersucht der vorliegende Beitrag die Potentiale einer Initiative zur gemeinschaftlichen Entwicklung offener Referenzmodelle in Anlehnung an die OSS-Bewegung. Da die Erforschung von OSS-Projekten noch in einem frühen Stadium ist und zudem die Übertragbarkeit vorliegender Indizien teilweise mit Vorsicht zu betrachten ist, muss eine solche Untersuchung zu einem erheblichen Teil spekulativ bleiben. Der Beitrag ist denn auch darauf gerichtet, die vorgestellten Hypothesen und die daraus resultierenden Empfehlungen zur Diskussion zu stellen – mit der Intention, die Ziele wie auch die organisatorischen Rahmenbedingungen entsprechender Projekte zu konkretisieren.

Gleichzeitig wollen wir nicht verschweigen, dass wir mit dem Beitrag auch für eine OM-Initiative werben möchten. Jenseits noch offener Fragen und damit verbundener Risiken sind wir der Ansicht, dass eine solche Initiative für die Modellierungsforschung in der Wirtschaftsinformatik wie auch für beteiligte Unternehmen attraktive Perspektiven eröffnet - eine Einschätzung, in der wir uns nach der Vorstellung erster Ideen im Rahmen der GI-Fachgruppe MobIS und des GI-Querschnittsfachausschuss Modellierung bestätigt sehen. Um die Diskussionen zu fokussieren und einen organisatorischen Rahmen für die Gründung einer ersten OM-Initiative zu schaffen, wird z. Z. unter <http://openmodels.org> eine Online-Community aufgebaut.

Literaturverzeichnis

- [BHPC04] Brügge, B., Harhoff, D., Picot, A., Creighton, O., Fiedler, M., Henkel, J.: Open-Source-Software. Eine ökonomische und technische Analyse. Springer, Berlin 2004.
- [BoRo03] Bonaccorsi, A.; Rossi, C.: Why open source software can succeed. In: Research Policy 32 (2003) 7, S. 1243–1258.
- [CrHA06] Crowston, K.; Howieson, J.; Annabi, H.: Information systems success in free and open source software development: Theory and measures. In: Software Process: Improvement and Practice (Special Issue on Free/Open Source Software Processes) 11 (2006) 2, S. 123–148.
- [FeLo04] Fettke, P.; Loos, P.: Systematische Erhebung von Referenzmodellen – Ergebnisse einer Voruntersuchung. Working Papers of the Research Group Information Systems & Management, Paper 19, Mainz 2004. http://www.staff.uni-mainz.de/fettke/free/fettke_2004_isym_paper_019.pdf, Abruf am 24.07.2006.
- [FrSK06] Frank, U.; Strecker, S.; Koch, S.: „Open Model“ – ein Vorschlag für ein Forschungsprogramm der Wirtschaftsinformatik (Langfassung). Institut für Informatik und Wirtschaftsinformatik, Universität Duisburg-Essen, ICB-Research Report No. 8, erscheint voraussichtlich im Februar 2007.

- [Gall01] Gallivan, M. J.: Striking a balance between trust and control in virtual organizations: a content analysis of open source software case studies. In: *Information Systems Journal* 11 (2001), S. 277–304.
- [KoSF06] Koch, S.; Strecker, S.; Frank, U.: Conceptual Modelling as a New Entry in the Bazaar: The Open Model Approach. In: *Damiani, E.; Fitzgerald, B.; Scacchi, W.; Scotto, M.; Succi, G.* (Hrsg.): *Open Source Systems*. Springer, Berlin 2006, S. 9–20.
- [LeTi02] Lerner, J.; Tirole, J.: The simple economics of open source. In: *Journal of Industrial Economics*, 50 (2002) 2, S. 197–234.
- [MaMA00] Markus, M. L.; Manville, B.; Agres, C. E.: What makes virtual organizations work? In: *Sloan Management Review*, 42 (2000) 1, S. 13–26.
- [MoFH02] Mockus, A.; Fielding, R. T.; Herbsleb, J. D.: Two case studies of open source software development: Apache and Mozilla. In: *ACM Transactions on Software Engineering and Methodology*, 11 (2002) 3, S. 309–346.
- [HeNH03] Hertel, G.; Niedner, S.; Hermann, S.: Motivation of software developers in open source projects: An internet-based survey of contributors to the Linux kernel. In: *Research Policy*, 32 (2003) 7, S. 1159–1177.
- [Pere99] Perens, B.: The Open Source Definition. *DiBona, C.; Ockman, S.; Stone, M.* (Hrsg.): *Open Sources: Voices from the Open Source Revolution*, O’Reilly, 1999.
- [Raym01] Raymond, E. S.: *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. 2. Auflage. O’Reilly, Sebastopol 2001.
- [Stal02] Stallman, R. M.: *Free Software, Free Society: Selected Essays of Richard M. Stallman*. GNU Press, Boston, Massachusetts 2002.
- [StGo06] Stewart, K. J.; Gosain, S.: The Impact of Ideology on Effectiveness in Open Source Software Development Teams. In: *MIS Quarterly*, 30 (2006) 2, S. 291–314.

Einführung in den Track

Softwareprozessverbesserung

Prof. Dr. Georg Herzwurm

Universität Stuttgart

Prof. Dr. Marcus A. Rothenberger

University of Nevada Las Vegas

Dr. Ralf Kneuper

Beratung für Softwarequalitätsmanagement
und Prozessverbesserung, Darmstadt

Effizienz, Qualität, und Zuverlässigkeit in der Softwareentwicklung sind kritische Faktoren in vielen Organisationen. Viele Softwareentwickler und Unternehmen streben danach, ihr Vorgehen zu verbessern, indem sie geeignete Methoden und Technologien anwenden, wie z.B. SEIs Capability Maturity Model Integration, wieder verwendbare Softwarekomponenten, Web-Services, Middleware und Extreme Programming. Keine dieser Technologien kann die alleinige Lösung zur Effizienz- und Qualitätssteigerung in der Softwareentwicklung sein; jedoch haben viele dieser Technologien das Potential, Teilaspekte der Softwareentwicklung zu verbessern.

Dieser Track lädt sowohl technische als auch organisatorische Beiträge ein. Manuskripte sollten sich dem Einsatz von Technologien und Methoden widmen, die zu einer Verbesserung von Produktivität und Qualität in der Softwareentwicklung führen können, oder Erfahrungen mit derartigen Technologien und Methoden berichten.

Programmkomitee:

Prof. Dr. Thorsten Spitta, Universität Bielefeld

Harry M. Sneed, AneCon GmbH

Prof. Dr. Stefan Eicker, Universität Essen

Prof. Dr. Wolfram Pietsch, FH Aachen

Prof. Dr. Werner Mellis, Universität Köln

Anton Dillinger, SAP

Stephan Salmann, Corporate Quality

Prof. Dr. Gerhard Chroust, Universität Linz

Dr. Ernest Wallmüller, Qualität & Informatik, Schweiz

Gerhard Fessler, Stuttgart

Prof. Dr. Marc Haines, University of Wisconsin-Milwaukee

Prof. Dr. Mark Srite, University of Wisconsin-Milwaukee

Ralf Remy, remy mc concepts

From Communication to Implementation

A framework for understanding and reducing unintentional complexity in software development processes

Matthias Biggeleben

Chair of Information Systems Engineering

Johann Wolfgang Goethe University

60325 Frankfurt / Main

biggeleben@wiwi.uni-frankfurt.de

Abstract

Since the invention of “software engineering” in 1968, software development has been suffering from efficiency problems. Software development is bridging the gap between verbally formulated requirements and programming languages. This work equates development with communication. Communication efforts refer to either essential or accidental complexity. This work hypothesizes that accidental complexity is inherent in implementation processes. However, it can be mitigated. This paper discusses an implementation framework that attacks accidental complexity. The framework is tested in an experiment in order to study the hypothesized efficiency gains in a daily programming task. Finally, this work discusses potential reasons for the existence of accidental complexity in software development.

1 Introduction

Since the invention of the term “software engineering” in 1968 [NaRa68] and in spite of all invented methods to improve software engineering [Berr04; Somm01], software development (SD) still suffers from efficiency problems. 23% of SD projects are cancelled due to failure and 49% exceed project resources [Stan01], i.e. projects do not meet deadlines or projects are simply too expensive. Similarly, [KeMR00] note, that between 30% and 40% of all software projects exhibit some degree of escalation.

Generally, a SD process starts with an initial customer’s need, which is structured, documented and discussed by techniques of requirements engineering [Somm01]. The process results – if successful – in an implemented and running system. Simplified, SD is bridging the *gap* between requirements (the “left side”) and a perfected implementation (the “right side”; see Figure 1).

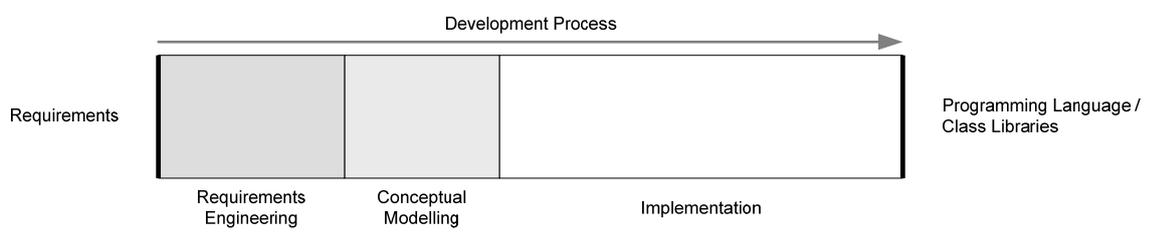


Figure 1: Simplified SD process

SD can be viewed from different perspectives, e.g. programming languages, compilers, operating systems or system architectures. Most perspectives have a technical nature. However, SD can also be understood as a form of communication. Usually, SD is performed within a team. Customers, requirements engineers, modelers, software developer and other relevant stakeholders have to communicate with each other during the course of an SD project. In addition, developers have to communicate with machines through programming languages. Even in the case of a one-man development, there is still a communication process between the developer and the machine. In the following, this work uses the term "*communication*" to unify both social and man-machine communication.

This work concentrates on *shortening* the gap by shifting the very “right side” of the overall development process, leaving other methods of software engineering unaffected. If this shift is realizable, the procedures developers usually follow will be truncated. Consequently, the process will demand less effort and less communication. This work proposes an *implementation framework* to realize that shift (see Figure 2).

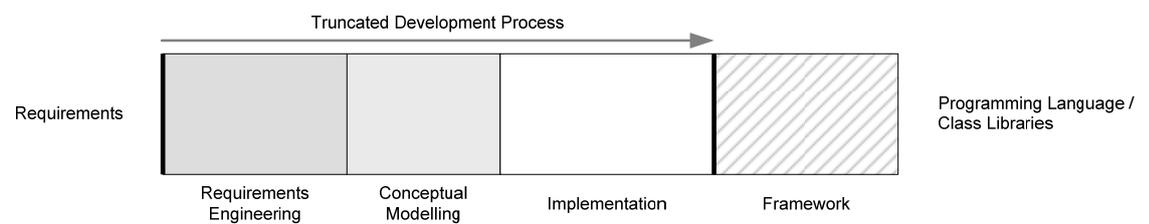


Figure 2: Truncated SD process

The crucial motivation for the framework is to free programmers from anything, which is not in straight line with problem solution or unnecessarily error-prone, i.e. the identification and elimination of *accidental complexity* [Broo87; Broo95]. In addition, the framework provides and integrates established concepts developers might need to know and implement. By doing so, the framework attempts to prevent communication hazards, which are supposed to exist in development processes.

The paper is structured as follows. Section 2 discusses related work. This section also elaborates three types of communication hazards. Section 3.1 proposes an abstract implementation framework that addresses the presented communication hazards. Accidental complexity is the most promising candidate to attack. Yet, accidental complexity strongly depends on technical aspects, i.e. how concepts can be implemented. Thus, section 3.2 summarizes analyzed aspects of SD. This analysis results from the development of various framework prototypes. These developments can be regarded as a collection of experiments. Details of the research methodology and the hypothesis of this paper are presented in section 4.

Afterwards, section 5 will illustrate a *test* of the hypothesis by analyzing a daily programming problem. The test is based on the current framework prototype. Finally, section 6 discusses the test results and attempts to derive a model of *general* existence of accidental complexity in SD and explain how it can be mitigated.

2 Related Work

Requirements are neither clear at the start nor stable during the development process. Following [Jack94] only two things are known about requirements: they will change and they will be misunderstood. This insight is not due to defects of requirements engineering. In fact, changes are natural. [Lehm80] points out, that real world software – once installed and becoming part of the application domain (termed E-type systems) – alters its own requirements. The software co-evolves with its domain and vice versa [LeRa02]. As a consequence, requirements are adapted, some are removed and new ones are added, which is organizationally managed by change requests.

During development, developers continuously gain knowledge about requirements through communication processes. Analogously and according to [BuMo79, 1] who describe the origin of knowledge, “one might begin to understand the world and communicate this as knowledge to

fellow human beings”. [GaNe01, 611] describe knowledge as “justified true belief” and state that “belief refers to an individual’s or group’s idea about what is truth”. As a result, truth is socially constructed. Analogously, software requirements are social constructs as well.

However, it is impossible to verbalize requirements in a form that a machine automatically generates an implementation [Holt03]. For this reason, a gap emerges that cannot be bridged instantly. A machine simply does not understand us. In consequence, there has to be at least one person who is capable of mapping *concepts* formulated (or thought) in spoken language to concepts of a machine represented in a programming language. This transformation is inevitable. Here, the term *concept* is used with respect to [KaLo84]. Kamlah and Lorenzen define *concept* as "the meaning of a term [...]: the meaning of a term is that which the term makes understood on the basis of its explicit agreement, which, however, can also be made understood by other signs" [KaLo84, 73].

This work identifies two sets of concepts. The first set contains concepts, termed *R-concepts* (requirements), which are used to define requirements and which refer to the “left” side of development processes. The second set contains concepts, termed *I-concepts* (implementation), a machine understands and provides and which refer to the “right” side of development processes. Both sets refer to a corresponding language community [KaLo84]. With respect to their members, both language communities might be disjoint. This emphasizes the importance of a person who is able to collate concepts of both communities.

The transformation of R-concepts into I-concepts is time-consuming and expensive, since software is complex. “Software entities are more complex for their size than perhaps any other human construct, because no two parts are alike” [Broo95, 182]. Thus, developers have to invest a remarkable amount of time in communication, because it is necessary to communicate each software entity. Conversely, there are communication processes which can be identified as unnecessary with respect to problem solution.

Despite extensive use of the word, there is no generally accepted definition of complexity. Complexity is a multi-faceted term which has many possible meanings [FlCa93; Klir85]. Since complexity is something perceived by an observer, the complexity of the system being observed can be described as a measure of the perceived effort that is required to understand and cope with the system [Back02]. Here, complexity can be understood in terms of number of parts and number of relationships [FlCa93], as well as requisite variety [Ashb65].

Following [Broo95, 182] complexity can be divided into the *essence* and the *accidents*. *Essential complexity* is inherent and unavoidable by definition, since any *simpler* solution would not solve the problem. In contrast, *accidental complexity* is avoidable, since it refers to complexity that we create on our own, i.e. it originates from the process or the techniques of solving a problem. It is not part of the solution. Accordingly, communication processes that focus accidental complexity are unnecessary. However, there has to be an alternative method that avoids accidental complexity. If there is one, accidental complexity can be removed and communication efforts will be reduced.

Additionally, there are communication processes that are unnecessary because developers do not know a particular concept. Instead, they have to invest time to develop and express concepts on their own, e.g. concepts like the model-view-controller or object-relational mapping. The linguist Whorf [Whor56] hypothesizes a relationship between the expressive power of a language and the ability to think a particular thought. If required words are unknown, a person will not be able to express the thought and might not even be able to formulate it in other words. This hypothesis can be transferred to SD [McCo04]. For example, if simple concepts like hashes or dictionaries are not known, they will probably not be used within development, although their convenience might even be known from keys and indexes within databases.

As a result, this work identifies three types of communication hazards:

- to communicate accidental complexity,
- to communicate unknown concepts and
- *not* to communicate relevant concepts.

These communication hazards still refer to communication as a unification of both social and man-machine communication. Each communication hazard is regarded as a source of inefficiency.

With respect to the reduction of inefficiency, a major focus of software engineering has been the organization and *acceleration* of the process of bridging the gap. Popular representatives of this idea are the build-and-fix model [Scha02], the waterfall model [Royc70; Somm01], structured programming [DaDH72], extreme programming [BeAn04], domain specific approaches etc. However, past decades have shown that the efficiency problem has not changed and that Brooks' "silver bullet" has not been found [Berr04; Broo87; Broo95]. Consequently, this work

does not directly focus on the process, but the basis of SD, i.e. the quantity and usability of provided *I-concepts* on the implementation level.

3 Implementation Framework

3.1 Abstract Implementation Framework

The implementation framework is based on a 3rd generation programming language. Accordingly, accidental complexity of mapping high-level language constructs to machine or assembler code does not exist. Additionally, the framework will address an object-oriented programming language. However, object-orientation has not removed accidental complexity [Broo87]. The choice for an object-oriented language just results from the popularity and widespread use of these languages, e.g. C++, C# and Java.

A programming language simply consists of essential language constructs, e.g. assignments, arithmetic expressions, loops, conditions etc. Frequent data structures and functions (which do not extend the syntax of a language) are usually provided by a collection of class-libraries. Commonly, class-libraries offer reusable functionality [KiLa92]. Moreover, generality and extensibility are their design objectives [KiLa92]. Accordingly, class-libraries reduce accidental complexity, because they offer already implemented functions.

Though generality and extensibility are design objectives of class-libraries, these features may also emerge as obstacles. For example, an application has to load an XML document from a web server which requires authorization, validate the document against an XML schema definition and extract some data via XPath [W3C99]. This should be three lines of code. Using class libraries straightforwardly developers need a multiple of lines, because class-libraries contain small but highly reusable structures that have to be composed for problem solution.

In consequence, the implementation framework has to identify the most required functions, which normally have to be composed by developers manually. Thus, the framework has to *pre-compose* frequent functionality, so that a particular functionality is utilizable within one line of code. If a pre-composed function requires customization, this will be management by descriptors that are separated from the code level.

Another major source of accidental complexity is bad design [Broo87]. Consequently, the framework supports and even forces a primary segmentation of the implementation.

Accordingly, the framework demands less effort and ability from programmers. Following [RaTo92] *ability* which includes familiarity with relevant concepts and design still has the strongest effect on software developers' performance. Basically, the framework is divided into *four* segments: *data*, *code*, *graphical user interface* (GUI) and *communication*, which is consistent with *three-tier* architectures like the client/server model (the communication segment connects the tiers). However, the segmentation is transparent for developers.

At this point, the framework appears to be equivalent to architectures like *Java 5 EE* (successor of J2EE; [Sun06]) or the *.NET Framework* [Micr06d]. For distinction purposes, a programming language, its class libraries and corresponding architectures are subsumed under the term "*implementation base*". Nevertheless, the proposed implementation framework does not intend to replace or compete with such implementation bases. It is rather intended to put the implementation framework on such an implementation base. This implicates, that – in spite of a compound implementation base – there is much accidental complexity left in SD processes, which legitimates a further *layer* between developers and the implementation base. However, the framework does not encapsulate the implementation base. Thus developers still have direct access (see Figure 3).

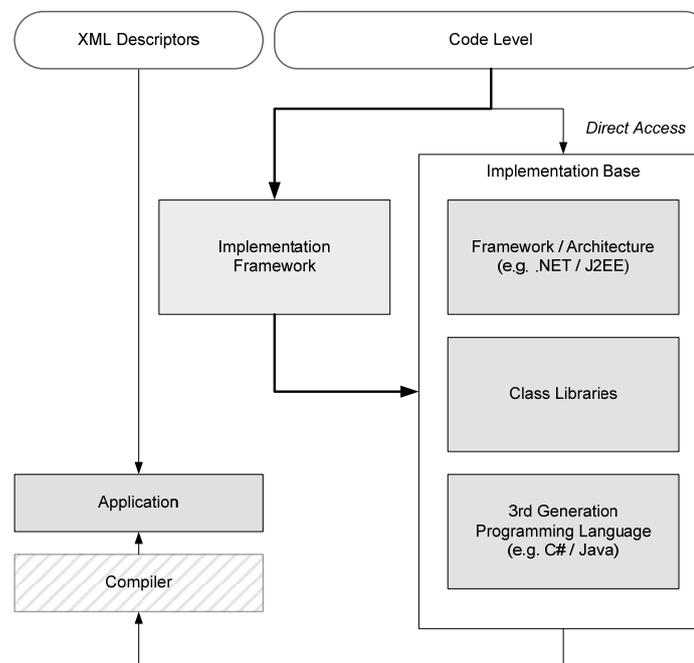


Figure 3: Integration of the implementation framework and an implementation base

For the legitimization of a further layer, the *abstract* implementation framework has to be fulfilled by a *general* explanation for the existence of accidental complexity. In particular, the existence of accidental complexity and ways of mitigation strongly depend on technical aspects.

3.2 Technical Aspects

For abbreviation purposes, this section does not present technical aspects in detail, but gives an impression which aspects of SD have been technically analyzed. This analysis examined differences between *typed* and *untyped* programming languages with respect to productivity [McCo04; Oust98; Prec05], the coherency between a language and productivity in general [BISW96; MaWD96], the lines-of-code paradox [Jone94], arguments for typed languages [Oust98; RySB05], arguments for untyped languages [Oust98], future features of programming languages ("*Anonymous Classes*", "*implicitly typed variables*", etc. [Micr06b]), data processing based on XML, GUI description based on XML (*XML application markup language* [Micr03]), GUI element access via the *document object model* [W3C04], late-binding, code reflection, promising approaches to integrate script-languages [Hugu04; Micr06c], object-relational mapping [Micr06a; Sun05] and finally standardizing communication via web services [W3C02].

Summarizing, the technical analysis encouraged the assumption that accidental complexity still exists in daily programming scenarios. Moreover, the analysis provided ideas to mitigate this complexity.

4 Research Methodology

This work *hypothesizes*, that accidental complexity denotes a communication hazard that negatively impacts the efficiency of SD processes. Using the proposed framework in SD processes reduces accidental complexity and positively impacts the efficiency.

With respect to software engineering, [DBOB03, 53] state, that "it is certainly arguable whether a positivist approach can ever be appropriate for a discipline so dependent on people and the environment, where carefully controlled and repeatable experiments, which change only one variable at a time, are often difficult or impossible to design and implement". For that reason, this work predominantly addresses a *subjective* and *interpretive understanding* [Lee91].

A collection of simple controlled experiments have been made to provide first data for the *subjective understanding*. These experiments were derived from a superordinate exemplary SD project. With respect to comprehensibility, the project goal was to develop a simplified replica of the software "*EndNote*" [Thom06], which is a tool for publishing and managing bibliographies. The choice for replicating this tool was made due to a lot of experienced crashes, the lack of database support and the missing support for a lot of citation styles (actually *EndNote* is not able to handle the citation style of this paper). Besides, this software type is regarded as popular within the research community.

The SD project was analyzed for frequent programming issues. The dominant issues are database queries (database support), XML (for importing existing *EndNote* libraries), XSLT (for describing and generating output styles), GUI handling and client-server communication (between the replica and e.g. *Microsoft Word*). Relating to the identification and mitigation of accidental complexity, each issue was subject to a controlled experiment.

All experiments have been conducted by the author. Other persons were not involved. First, a programming issue was isolated from its context. Afterwards the issue was implemented. With respect to the experimental context [KPPJ02], the author is especially familiar with C, C#, Java, JavaScript, Perl, PHP, SQL, XML and XSLT with more than two decades of programming experience. All issues were implemented in C# using Microsoft's .NET Framework (version 1.1) and Visual Studio .NET 2003. Run-time benchmarks have been performed under Microsoft Windows XP Professional (SP2) on a Pentium 4 machine (2.8 GHz; 512 MB RAM).

After the implementation, the resulting code was analyzed with respect to accidental complexity. For example, the essence of a database query is the SQL statement, the corresponding database and the query results. Any further line of code is regarded as an unnecessary communication effort and a possible source of error and therefore accidental.

The results of each analysis were *interpreted* and thus they delivered insights and synergies which helped constructing and adapting the implementation framework. Thus, the experiments were repeated using the first prototypes of the framework. Again, the resulting code was tested for accidental complexity and possible negative side-effects on other issues. With respect to a comparison, the communication efforts of both series, i.e. implementing a solution with or without the framework, have been counted. Here, a communication effort refers to function calls, assignments, loops, object instantiations, type definitions etc. All communication efforts have been weighted equally.

5 Testing the Hypothesis

To test the presented hypothesis, accidental complexity as a communication hazard has to be revealed and mitigated. Accordingly, a common task of SD is analyzed in the following: *database queries*. This refers to one of the experiments derived from the *EndNote* replica development. The analysis addresses the question if this task contains accidental complexity. This section is based on the current prototype of the implementation framework.

The requirements are as follows. The program has to respond to an AJAX request (asynchronous JavaScript and XML) [Garr05; Paul05] or a web-service. The corresponding method has to connect to a database (using a native driver), execute an SQL query, collect its results, serialize them into an XML document, transform the XML document into an AJAX/web-service compatible schema and finally serialize the XML document to a plain string. The program has to support two different database management systems (MySQL 5.0 & Microsoft SQL Server 2005).

The required communication efforts have been classified and counted. A developer has to spend a total of 78 communication efforts (see Figure 4).

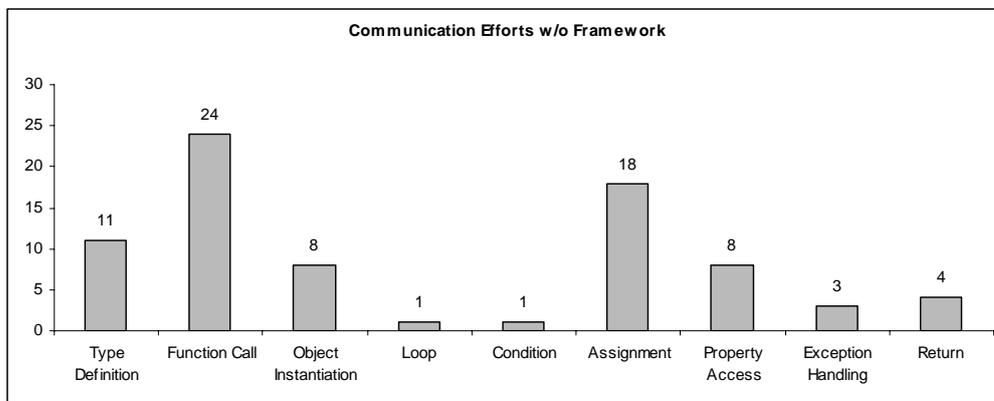


Figure 4: Communication efforts (without framework)

These efforts refer to the composition of basic language constructs and the recomposition of "atomic" elements provided by class-libraries. The latter implies that he or she knows how to implement database access, XML document creation and XSLT transformation. Using the implementation framework, the total effort and thus accidental complexity can be minimized (see Figure 5).

```

public string DemoRequest()
{
    // load global configuration
    Global.LoadConfiguration("configuration.xml");
    // prepare & execute the statement (uses default connection),
    // SQL query is given by name, transform result set to xml, process xslt,
    // return xml as string (implicit conversion)
    return Pool.PrepareByName("Query1").XmlResultSet.Xslt("ajax.xsl");
}

```

Figure 5: The exemplary development task using the implementation framework

The given example demonstrates that the task can be solved within two lines code. The first line loads the configuration (see Figure 6), i.e. all required XML descriptors. It is possible and planned to create and maintain this descriptor by an external application, so that its maintenance demands marginal efforts.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<pif>
  <configuration>
    <system name="Development">
      <database>
        <connections default="testdb">
          <connection name="testdb">
            <type>mysql</type>
            <host>localhost</host>
            <database>foobar</database>
            <user></user>
            <password></password>
          </connection>
        </connections>
        <queries>
          <query name="Query1">
            <sql type="mysql">
              SELECT proprietary_function(foo)
              FROM bar
            </sql>
            <sql type="mssql">
              SELECT another_proprietary_function(foo)
              FROM bar
            </sql>
          </query>
        </queries>
      </database>
    </system>
  </configuration>
</pif>

```

Figure 6: Exemplary configuration file

At this time, a default database connection is defined, which also defines the database system type, e.g. MySQL. If the connection fails, a debug dialog will open. The user can immediately kill the process within the debug dialog. Moreover, the debug dialog does some checks to help the developer (send ping to host, connect without a particular database) and it shows the

corresponding configuration lines and the location of the configuration file. If the configuration file is changed, the developer has the possibility to reload the configuration during run-time. The second line (see Figure 5) prepares and executes the SQL query, defined by its name. All corresponding result sets are serialized to XML. This XML data is immediately transformed by an XSL transformation. The results are implicitly converted to a string and returned. Using the implementation framework, the communication efforts decrease to a total of 5. The results of the exemplary development task are compared in Tab. 1. The comparison considers communication efforts (CE), lines of code (LOC) and effective lines of code (ELOC), i.e. without braces, empty lines and comments, as well as a run-time benchmark (10.000 iterations). In addition, the comparison presents rounded ratios.

	With & Without Comparison	
	w/	w/o
CE	5	78
LOC	9	121
ELOC	2	48
CE ratio	1	: 16
LOC ratio	1	: 13
ELOC ratio	1	: 24
Benchmark (Total)	21.565 ms	120.482 ms
Benchmark	464 iterations/s	83 iterations/s
Performance ratio	6	: 1

Tab. 1: Comparison of the exemplary development task both with ("w/") and without ("w/o") the implementation framework

The comparison clearly revealed that the implementation framework reduces communication efforts within the experimental setting. All measures concerning programming effort, i.e. CE, LOC and ELOC, decreased. Especially, the effective lines of code present a ratio of 1:24. In contrast, the run-time benchmark shows that the implementation framework provides better performance (6:1) based on code optimizations and suitable caching mechanisms.

Finally, the ratio of CE and ELOC reveal accidental complexity. While both cases (i.e. with and without the implementation framework) solve the same problem, the framework is able to eliminate approx. 95% of the initial complexity on the implementation level.

6 Discussion

The presented experiment demonstrated the existence of accidental complexity in SD and that this complexity can be mitigated. Thus, the existence of communication hazards could be confirmed, which corroborates the given hypothesis. The development of the implementation framework prototype has already revealed lots of further accidental complexity. At this point, it is interesting to ask if there is a *general explanation* for accidental complexity.

As stated, a 3rd generation programming language is usually provided with class libraries. Analogous to relational database design, class libraries give the impression of *normalization*. In general, the goal of normalization is reduction of redundancy [SiKS06]. This is consistent with class libraries' generality and extensibility.

Given a particular collection of class libraries, there is almost no redundancy, since each construction is *decomposed* into reusable parts, so that the degree of reusability is maximized. In consequence, a developer has to *recompose* "atomic" elements to a sequence of instructions. Thinking in relational database design, there is no functional dependency within any potential sequence. But there might be sequences, which we now call "frequent functions". Accordingly, a *recomposition* of a frequent function is equal to accidental complexity. It can be avoided by "*pre-composed*" functions, i.e. adding *redundancy*, on a framework level. This redundancy is also able to attack conceptual essence [Broo95, 196], since it provides mechanisms developers no longer have to communicate, implement and test and therefore they no longer have to conceptualize these mechanisms either.

Reducing accidental complexity ahead of a project to reduce costs might be criticized as a self-fulfilling prophecy. However, the costs of the development and customization of the framework have to be added to the total project costs. In general, the framework approach pays if its development costs are less than the project costs savings. Still, if this is not the case, it can be an advantageous investment in the long run.

7 Future Research

Future research will concentrate on completing the implementation framework. With respect to the presented hypothesis, a series of controlled experiments with students will follow. These experiments will be based on a test group, which uses the implementation framework, and a control group, which does not use the framework. Both groups will implement identical requirements, which allows a comparison of the development productivity of each group. The results will be helpful to elaborate a hypothesis about the degree to which the framework approach compresses the project schedule. Furthermore, it is decided to perform a case study with an enterprise that is engaged in SD.

Finally, the author would like to thank his reviewers for their recommendations on this work. Furthermore, the author would like to thank the German Federal Ministry of Education and Research, which funded this work under record no. 01AK706.

References

- [Ashb65] *Ashby, W. R.:* An introduction to cybernetics. London 1965.
- [Back02] *Backlund, A.:* The concept of complexity in organisations and information systems. In: *Kybernetes* 31 (2002) 1, pp. 30-43.
- [BeAn04] *Beck, Kent; Andres, Cynthia:* Extreme Programming Explained: Embrace Change. Addison-Wesley Professional, 2004.
- [Berr04] *Berry, Daniel M.:* The Inevitable Pain of Software Development: Why There Is No Silver Bullet. In: *Proceedings of the Radical Innovations of Software and Systems Engineering in the Future*, Venice, Italy, 2004, pp. 50-74.
- [BISW96] *Blackburn, Joseph D.; Scudder, Gary D.; Van Wassenhove, Luk N.:* Improving Speed and Productivity of Software Development: A Global Survey of Software Developers. In: *IEEE Transactions on Software Engineering* 22 (1996) 12, pp. 875-895.
- [Broo87] *Brooks, Fred P.:* No Silver Bullet: Essence and Accident in Software Engineering. In: *IEEE Computer* 20 (1987) 4, pp. 10-19.

- [Broo95] *Brooks, Fred P.:* The Mythical Man-Month: Essays on Software Engineering. Addison-Wesley, Reading, MA 1995.
- [BuMo79] *Burrell, G.; Morgan, G.:* Sociological Paradigms and Organisational Analysis. Aldeshorst et al. 1979.
- [DaDH72] *Dahl, O.-J.; Dijkstra, E. W.; Hoare, C. A. R.:* Structured Programming. Academic Press, London 1972.
- [DBOB03] *Dawson, Ray; Bones, Phil; Oates, Briony J.; Brereton, Pearl; Azuma, Motoei; Jackson, Mary Lou:* Empirical methodologies in software engineering. In: Proceedings of the Eleventh Annual International Workshop on Software Technology and Engineering Practice, 2003., Amsterdam, The Netherlands, 2003, pp. 52-58.
- [FlCa93] *Flood, R. L.; Carson, E. R.:* Dealing with complexity. Plenum Press New York, 1993.
- [GaNe01] *Galliers, Robert D.; Newell, Sue:* Back to the future: from knowledge management to the management of information and data. In: Proceedings of the European Conference on Information Systems, Bled, Slovenia, 2001, pp. 609-615.
- [Garr05] *Garret, Jesse James:* Ajax: A New Approach to Web Applications. <http://www.adaptivepath.com/publications/essays/archives/000385.php>, 2005, Last access 2005-03-29.
- [Holt03] *Holten, Roland:* Integration von Informationssystemen. Theorie und Anwendung im Supply Chain Management. 2003.
- [Hugu04] *Huginin, Jim:* IronPython - A fast Python implementation for.NET and Mono. <http://www.ironpython.com/>, 2004, Last access 2006-07-05.
- [Jack94] *Jackson, M.A.:* The Role of Architecture in Requirements Engineering. In: Proceedings of the Proceedings of the IEEE International Conference on Requirements Engineering, Colorado Springs, 1994, p. 241.

- [Jone94] *Jones, C.: Assessment and Control of Software Risks.* Yourdon Press, Englewood Cliffs, N.J. 1994.
- [KaLo84] *Kamlah, Wilhelm; Lorenzen, Paul: Logical Propaedeutic. Pre-School of Reasonable Discourse.* University Press of America, Lanham 1984.
- [KeMR00] *Keil, Mark; Mann, Joan; Rai, Arun: Why Software Projects Escalate: An Empirical Analysis and Test of Four Theoretical Models.* In: *MIS Quarterly* 24 (2000) 4, pp. 631-664.
- [KiLa92] *Kiczales, Gregor; Lamping, John: Issues in the design and specification of class libraries.* In: *ACM SIGPLAN Notices* 27 (1992) 10, pp. 435 - 451.
- [Klir85] *Klir, G. J.: Complexity: some general observations.* In: *Systems Research* 2 (1985) 2, p. 131-140.
- [KPPJ02] *Kitchenham, Barbara A.; Pfleeger, Shari Lawrence; Pickard, Lesley M.; Jones, Peter W.; Hoaglin, David C.; Emam, Khaled El; Rosenberg, Jarrett: Preliminary Guidelines for Empirical Research in Software Engineering.* In: *IEEE Transactions on Software Engineering* 28 (2002) 8, pp. 721-734.
- [Lee91] *Lee, Allen S.: Integrating positivist and interpretive approaches to organizational research.* In: *Organization Science* 2 (1991) 4, pp. 342-365.
- [Lehm80] *Lehman, M. M.: Programs, Life Cycles and Laws of Software Evolution.* In: *Proceedings of the Proceedings of the IEEE, 1980*, pp. 1060-1076.
- [LeRa02] *Lehman, M. M.; Ramil, Juan F.: Software Evolution and Software Evolution Processes.* In: *Annals of Software Engineering* 14 (2002) 1-4, pp. 275 - 309.
- [MaWD96] *Maxwell, Katrina D.; Van Wassenhove, Luk; Dutta, Soumitra: Software Development Productivity of European Space, Military and Industrial Applications.* In: *IEEE Transactions on Software Engineering* 22 (1996) 10, pp. 706-718.
- [McCo04] *McConnel, Steve: Code Complete.* Microsoft Press, Washington 2004.

- [Micr03] *Microsoft*: Chapter 1: The "Longhorn" Application Model. <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnintlong/html/longhornch03.asp>, 2003, Last access 2006-07-03.
- [Micr06a] *Microsoft*: ADO.NET 2.0. <http://msdn.microsoft.com/data/ref/adonet/>, 2006, Last access 2006-07-05.
- [Micr06b] *Microsoft*: C# Version 3.0 Specification. http://download.microsoft.com/download/5/8/6/5868081c-68aa-40de-9a45-a3803d8134b8/CSharp_3.0_Specification.doc, 2006, Last access 2006-07-03.
- [Micr06c] *Microsoft*: IronPython. <http://www.codeplex.com/Wiki/View.aspx?ProjectName=IronPython>, 2006, Last access 2006-07-05.
- [Micr06d] *Microsoft*: Microsoft.NET Framework 3.0 (formerly WinFX). <http://msdn.microsoft.com/winfx/>, 2006, Last access 2006-07-05.
- [NaRa68] *Naur, P.; Randell, B.*: Software Engineering: Report an a Conference Sponsored by the NATO Science Commission. In: Proceedings of the Garmisch, Germany, 1968.
- [Oust98] *Ousterhout, John K.*: Scripting: Higher-Level Programming for the 21st Century. In: IEEE Computer 31 (1998) 3, pp. 23-30.
- [Paul05] *Paulson, Linda Dailey*: Building Rich Web Applications with Ajax. In: IEEE Computer 38 (2005) 10, pp. 14-17.
- [Prec05] *Prechelt, Lutz*: An Empirical Comparison of Seven Programming Languages. In: IEEE Computer 33 (2000) 10, pp. 23-29.
- [RaTo92] *Rasch, Ronald H.; Tosi, Henry L.*: Factors Affecting Software Developers' Performance: An Integrated Approach. In: MIS Quarterly 16 (1992) 3, pp. 395-413.
- [Royc70] *Royce, W. W.*: Managing the Development of Large Software Systems: Concepts and Techniques. In: Proceedings of the Proceedings of WesCon, 1970.

- [RySB05] *Ryder, Barbara G.; Soffa, Mary Lou; Burnett, Margaret: The Impact of Software Engineering Research on Modern Programming Languages.* In: ACM Transactions on Software Engineering and Methodology 14 (2005) 4, p. 431–477.
- [Scha02] *Schach, S. R.: Object-oriented and Classical Software Engineering.* McGraw-Hill, New York 2002.
- [SiKS06] *Silberschatz, Avi; Korth, Hank; Sudarshan, S.: Database System Concepts.* McGraw-Hill, New York 2006.
- [Somm01] *Sommerville, Ian: Software Engineering.* Pearson Education Limited, Essex 2001.
- [Stan01] *Standish Group International, Inc.: Extreme CHAOS.* In: Research report, ordering information available at www.standishgroup.com (2001).
- [Sun05] *Sun Microsystems Inc.: J2EE - Enterprise JavaBeans Technology.* <http://java.sun.com/products/ejb/>, 2005, Last access 2005-03-29.
- [Sun06] *Sun Microsystems Inc.: Java EE At a Glance.* <http://java.sun.com/javaaee/>, 2006, Last access 2006-07-07.
- [Thom06] *Thomson Corporation: EndNote.* <http://www.endnote.com/>, 2006, Last access 2006-07-14.
- [W3C02] *W3C: SOAP Version 1.2 Part 0: Primer.* <http://www.w3.org/TR/2002/WD-soap12-part0-20020626/>, 2002, Last access 2005-03-29.
- [W3C04] *W3C: Document Object Model.* <http://www.w3.org/DOM/>, 2004, Last access 2004-12-15.
- [W3C99] *W3C: XML Path Language (XPath).* <http://www.w3.org/TR/xpath>, 1999, Last access 2005-03-29.
- [Whor56] *Whorf, Benjamin: Language, Thought and Reality.* MIT Press, Cambridge, MA 1956.

Interaktive Entscheidungsunterstützung für die Auswahl von Software-Komponenten bei mehrfachen Zielsetzungen

Thomas Neubauer

Secure Business Austria - Security Research
1040 Wien
neubauer@securityresearch.at

Christian Stummer

Institut für Betriebswirtschaftslehre
Universität Wien
1210 Wien
christian.stummer@univie.ac.at

Abstract

In der betrieblichen Praxis kommt der komponentenbasierten Software-Entwicklung hoher Stellenwert zu. Angesichts mehrfacher Zielsetzungen und vielfältiger Nebenbedingungen ist dabei insbesondere die Auswahl der „besten“ Kombination von Komponenten ein nicht-triviales Entscheidungsproblem. Bislang wurden hierfür vor allem die Nutzwertanalyse bzw. der Analytic Hierarchy Process zur Entscheidungsunterstützung vorgeschlagen, wobei aber beide eine Reihe von Unzulänglichkeiten aufweisen. Diese Arbeit will dazu nunmehr eine Alternative anbieten. Darin werden in einem ersten Schritt zunächst (zulässige) Pareto-effiziente Kombinationen von Software-Komponenten berechnet und die Entscheidungsträger dann im zweiten Schritt interaktiv bei der Suche nach jener Variante unterstützt, die einen Ziele-Mix in Aussicht stellt, der den jeweiligen individuellen Präferenzen am besten entspricht. Das neue Verfahren zeichnet sich im Vergleich zu herkömmlichen Ansätzen insbesondere durch den Verzicht auf umfangreiche a priori Präferenzinformationen (wie z.B. Zielgewichtungen) aus. Darüber hinaus kann es ohne großen Anpassungsaufwand in bestehende Vorgehensmodelle zur Auswahl von Software-Komponenten integriert werden.

1 Einleitung

Die komponentenbasierte Software-Entwicklung unterscheidet sich vom traditionellen Vorgehen insbesondere dadurch, dass (bestehende) Komponenten als Grundlage für die Entwicklung komplexer Softwarelösungen genutzt werden. Tatsächlich kommt ihr heutzutage hoher Stellenwert zu [Ruhe2002], da in immer kürzeren Abständen qualitativ hochwertige und zuverlässige Software auf den Markt gebracht werden muss. Zudem steigen die funktionalen Anforderungen an Software, so dass insbesondere kleinere Unternehmen bei der Erfüllung ihrer Aufträge zunehmend davon abhängig sind, auf vorhandene Komponenten zurückgreifen zu können und das Produkt nicht in allen Details von Grund auf neu entwickeln zu müssen [Alves2003]. Empirische Studien [Alves2003, Basili1996] zeigen ferner, dass durch den Rückgriff auf bewährte Komponenten Fehler im Gesamtsystem wesentlich reduziert werden. Und schließlich wird die komponentenbasierte Software-Entwicklung auch durch Technologien wie CORBA, JavaBeans/EJB, DCOM/ActiveX oder .Net sowie die Verfügbarkeit von verschiedenen Tools für die Konfiguration und den Einsatz solcher Lösungen vorangetrieben [Andrews2005].

Allerdings müssen Software-Produkte in der Regel an spezifische Anforderungen angepasst werden. Dementsprechend spielt bei der komponentenbasierten Software-Entwicklung die Auswahl der „richtigen“ Komponenten eine wesentliche Rolle mit Auswirkungen auf alle weiteren Phasen des Entwicklungszyklus. Ineffiziente Entscheidungen haben daher nicht nur Einfluss auf Korrektheit und Zuverlässigkeit der komponentenbasierten Anwendung, sondern können auch zu massiven Kostensteigerungen in der Entwicklung und/oder der nachfolgenden Wartung führen [Maiden1998,Ruhe2002,Ruhe2003].

In der Literatur finden sich zahlreiche Vorgehensmodelle für die Evaluierung und Auswahl von Software-Komponenten (z.B. OTSO von [Kontio1995]). Nahezu alle diese Ansätze berücksichtigen mehrfache Zielsetzungen (wie Kosten, Kompatibilität, Einfachheit der Installation, usw.), wobei die meisten Autoren entweder eine Nutzwertanalyse (Weighted Scoring Method; WSM) oder den Analytic Hierarchy Process (AHP) zur Entscheidungsfindung empfehlen (vgl. u.a. [Alves2003,Maiden1998,Navarrete2005,Ncube2002,Wanyama2005]). Tatsächlich weisen aber beide Verfahren wesentliche Schwachstellen auf, wie insbesondere den Bedarf an umfangreichen Informationen über die a priori Präferenzen der Entscheidungsträger bei der Nutzwertanalyse, die kombinatorische Explosion der paarweisen Vergleiche bei der Verwendung des

AHP oder problematische Annahmen über die Form der Nutzenfunktion (für die regelmäßig lineare Nutzenverläufe unterstellt werden). Des Weiteren bieten solche Ansätze dem Entscheidungsträger lediglich eine einzelne Lösung, während es ein interaktives Verfahren erlauben würde, unterschiedliche Szenarien zu erkunden und zu analysieren bzw. aktiv am Entscheidungsprozess teilzunehmen und ihn zu kontrollieren.

In dieser Arbeit stellen wir einen entsprechenden, zweiphasigen Ansatz vor. Die erste Phase widmet sich der Ermittlung von Lösungen (d.h. Kombinationen von Software-Komponenten), die einerseits gegebene Nebenbedingungen (wie Ressourcenbeschränkungen oder Abhängigkeiten zwischen zwei oder mehreren Komponenten) erfüllen und andererseits Pareto-effizient¹ hinsichtlich der gegebenen Zielsetzungen sind. In der zweiten Phase werden Entscheidungsträger bei der interaktiven Erkundung des solcherart bestimmten Lösungsraums unterstützt, bis sie die für sie individuell „beste“ Zusammenstellung von Komponenten mit dem für sie attraktivsten Mix an Zielwerten gefunden haben. Der Ansatz kann im Übrigen problemlos in bestehende Vorgehensmodelle für die komponentenbasierte Software-Entwicklung integriert werden. Diese Arbeit bietet nun im Anschluss zunächst einen Überblick zum Forschungsstand hinsichtlich der Auswahl von Software-Komponenten bei mehrfachen Zielsetzungen. Danach beschreiben wir schrittweise unseren Vorschlag zur interaktiven Entscheidungsunterstützung. Die Arbeit schließt letztlich mit einem Resümee und einem Ausblick auf in diesem Zusammenhang interessante weiterführende Forschungsfragen.

2 Die Auswahl von Software-Komponenten im Überblick

Im Folgenden liegt das Augenmerk auf Software-Komponenten, die im Rahmen einer Wiederverwendung Teil eines neuen Softwareprodukts werden können und dazu in der Regel schon von vornherein nicht spezifisch für ein bestimmtes Projekt erstellt worden sind (für eine Diskussion vgl. [Torchiano2004]). Die komponentenbasierte Software-Entwicklung zielt nun darauf ab, mit Hilfe solcher mehrfach nutzbarer Komponenten komplexe Software-Projekte in kürzerer Zeit bzw. mit geringerem Budget durchführen zu können. In der Praxis werden hierbei allerdings oftmals die Risiken unterschätzt, die mit der Evaluierung, Auswahl und Integration der Komponenten verbunden sind, so dass mitunter beträchtliche Verzögerungen bzw. Budget-

¹ Eine Lösungsalternative gilt als „Pareto-effizient“, wenn keine andere (zulässige) Lösung existiert, die in allen betrachteten Zielsetzungen zumindest gleich gut und in mindestens einer Zielsetzung besser abschneidet.

überschreibungen für die Entwicklung oder Wartung der Systeme zu beobachten sind [Tran1997].

In der Vergangenheit wurden mehrere Vorgehensmodelle zur Strukturierung der komponentenbasierten Software-Entwicklung vorgeschlagen. Obgleich sie sich über weite Strecken stark ähneln, gibt es bislang noch kein allgemein als Standard akzeptiertes Vorgehen [Ruhe2002]. Andrews unterscheidet dazu zwischen „architektur-“ und „anforderungsgetriebenen“ Vertretern, wobei die erste Gruppe etwa OTSO [Kontio1995], BAREMO [LozanoTello2002], oder STACE [Kunda2003] umfasst und SCARLET [Maiden2002] (als Nachfolger von PORE [Maiden1998]) sowie CRE [Alves2001] zur zweiten Gruppe zählen [Andrews2005].

Gemeinsam ist den Verfahren, dass an irgendeiner Stelle verschiedene Lösungsalternativen bewertet werden müssen. Eindimensionale Kennzahlen aus der Finanzwirtschaft wie etwa der Kapitalwert greifen aber für die komplexe Bewertung von IT-Investitionen regelmäßig zu kurz, so dass hierfür bislang die Nutzwertanalyse oder der AHP eingesetzt worden sind. Beide Verfahren erfordern jedoch die Bekanntgabe der Form der Nutzenfunktion (wobei meist der Einfachheit halber Linearität unterstellt wird) sowie von Gewichtungen für jedes zu berücksichtigende Zielkriterium. Die Nutzwertanalyse setzt dabei voraus, dass (i) sich die Entscheidungsträger von vornherein vollständig über alle Gewichtungen im Klaren sind, ohne jemals zuvor eine Lösungsalternative gesehen zu haben, und (ii) auch bereit sind, diese Präferenzen vollständig und vorbehaltlos preiszugeben, damit mit einer auf diese Weise definierten Zielfunktion Nutzenwerte für jede Kombination von Software-Komponenten ermittelt werden können. Demgegenüber stützt sich der AHP auf eine Hierarchie von Zielen und errechnet die benötigten Gewichtungen durch paarweise Vergleiche aller Kriterien je Hierarchieebene (mit Hilfe der Eigenwertmethode). Damit können nun bei korrekter Anwendung – trotz vielfacher methodischer Kritik – tendenziell konsistentere Ergebnisse erzielt werden als mit der Nutzwertanalyse; dies muss allerdings mit enormem Aufwand für die in großer Zahl notwendigen paarweisen Vergleiche erkauft werden. Maiden illustriert das anhand einer Fallstudie zu einem Projekt mit 130 Anforderungen, bei dem die Verwendung des AHP daran scheiterte, dass die Entscheidungsträger mehr als 42.000 Vergleiche durchführen hätten müssen [Maiden1998].

Generell scheint jeder Versuch, die vielfältigen Aspekte der Komponentenauswahl bei der Software-Entwicklung auf einen einzigen „Nutzenwert“ zu reduzieren, problematisch, da durch die Aggregation verschiedener Ziele zu einem einzigen Indikator nicht zuletzt individuelle Attribute und somit die Information über besondere Schwächen oder Stärken verloren gehen.

Daher kann eine hohe Bewertung in einem Ziel eine schlechte Leistung in einem anderen Kriterium (gleichsam automatisch) ausgleichen und solcherart ein potentiell Risiko unentdeckt bleiben [Ncube2002]. Forscher wie Praktiker vertreten deshalb vielfach die Auffassung, dass traditionelle Methoden bei der Bewertung von Software-Komponenten ungeeignet sind [Martinsons1998,Ryan2004] und Entscheidungsträger bessere Unterstützung bei der Bewertung von Kombinationen von Software-Komponenten und der entsprechenden Verteilung von Ressourcen hinsichtlich der damit verbundenen Nutzen und Risiken benötigen.

3 Ein interaktives Entscheidungsunterstützungssystem

Der nachfolgend vorgestellte neue Ansatz möchte eine geeignete Alternative bieten. Er zeichnet sich insbesondere dadurch aus, dass Entscheidungsträger weder umfangreiche a priori Präferenzinformationen bereitstellen noch zahlreiche paarweise Vergleiche anstellen müssen. Stattdessen wird zuerst – ohne aktives Zutun der Entscheidungsträger – die Menge der Paretoeffizienten Kombinationen von Software-Komponenten bestimmt und im Anschluss die Möglichkeit geboten, darin interaktiv nach jener Lösung zu suchen, die den individuellen Präferenzen am besten entspricht. Vor einer detaillierten Diskussion der beiden Phasen des interaktiven Entscheidungsunterstützungssystems soll zunächst noch ein typisches Vorgehensmodell für die komponentenbasierte Software-Entwicklung skizziert werden, um zu zeigen, in welcher Phase unser Ansatz zum Einsatz kommen würde.

3.1 Das Vorgehensmodell OTSO

Das von Kontio vorgeschlagene Modell OTSO (Off-The-Shelf Option; [Kontio1995]) gliedert sich in die nachfolgend beschriebenen fünf Phasen.

1. Definition der Kriterien: Im ersten Schritt werden die relevanten Zielkriterien bestimmt mit besonderem Augenmerk auf die Infrastruktur der Organisation, die Applikationsarchitektur, das Applikationsdesign, Anforderungen an die Applikation, Projektziele und -einschränkungen und die grundsätzliche Verfügbarkeit von Software-Komponenten bzw. -bibliotheken. Die Auswahl der Kriterien muss für jedes Vorhaben separat (z.B. unter Verwendung von GQM [Basili1987]) durchgeführt werden, da etwa funktionale Anforderungen regelmäßig für jede Applikation unterschiedlich sind. Typische Kriterien gliedern sich grob in vier Kategorien, nämlich: (i) Funktionale Kriterien, die auf Basis der Geschäftsprozesse definiert werden und deren

Erfüllung üblicherweise ein besonders hoher Stellenwert zukommt; (ii) Qualitätskriterien, wie beispielsweise die Fehlerrate, Leistungsmaßzahlen, Benutzerfreundlichkeit oder Sicherheit; (iii) Strategische Kriterien mit einem Fokus auf Kosten, Zeitaspekten oder verfügbarem Personal und dessen Qualifikationen; und (iv) Einsatzspezifische und architekturelevante Kriterien, die eine Rolle spielen, sofern dadurch Vorgaben oder Einschränkungen für das zu realisierende Softwareprodukt bedingt sind.

2. Identifikation der Komponenten: Dieser Schritt umfasst die Identifikation von potentiellen Komponenten, die beispielsweise untergliedert werden können in (i) Betriebssysteme (etwa AIX, FreeBSD, Linux, Microsoft Windows Server, QNX, Solaris), (ii) Middleware (BEA Weblogic, IBM Websphere, OpenORB, Oracle Application Server, Visibroker und weitere Komponenten wie z.B. CRM, DMS, ERP, SCM, SRM, die von unterschiedlichen Firmen angeboten werden), (iii) Datenbanken (Hypersonic SQL, Microsoft Access, Microsoft SQL Server, MySQL, Oracle DB, PostgreSQL) und (iv) Support Software (Agent++, Apache Tomcat, Hypertext Preprocessor, IBM Java Runtime Environment, Intel COPS Client, Log4J, Telia BER Coder, WebMacro, Xerces, XMLDB) [Morisio2002].

3. Screening: In dieser Phase sollen – zunächst einmal isoliert von den anderen – vielversprechende Komponenten für eine anschließende, umfassendere Evaluierung gefiltert werden.

4. Evaluierung: In der Evaluierungsphase erfolgt die Konsolidierung und gründliche Evaluierung der zuvor identifizierten Komponenten in Bezug auf die im ersten Schritt festgelegten Kriterien sowie die Dokumentation der Ergebnisse.

5. Analyse: In der Analysephase bestimmen die Entscheidungsträger die für die Umsetzung der Anforderungen „beste“ Kombination an Komponenten. Dazu wird in OTSO der AHP eingesetzt.

3.2 Die Bestimmung effizienter Kombinationen von Software-Komponenten

Unser Ansatz ist zum Einsatz in der letzten Phase des eben umrissenen Vorgehensmodells konzipiert und würde dort den althergebrachten AHP ersetzen. Der erste Schritt zielt dabei auf die Ermittlung der hinsichtlich der betrachteten Ziele Pareto-effizienten Kombinationen von Software-Komponenten. Dazu ist das zugrunde liegende multikriterielle kombinatorische Optimierungsproblem (zur multiobjective combinatorial optimization (MOCO) vgl. [Ehrgott2000]) zu lösen. Binäre Entscheidungsvariablen $x_i \in \{0,1\}$ geben darin an, ob eine Komponente i in einer Lösung verwendet wird oder nicht ($x_i = 1$ falls ja, ansonsten $x_i = 0$). Eine Kombination

von Komponenten wird durch einen Vektor $x = (x_1, \dots, x_N)$ mit Einträgen für alle N zur Wahl stehenden Komponenten abgebildet. Im MOCO-Problem sind nun K Zielfunktionen $u_k(x)$ ($k = 1, \dots, K$) (z.B. für die Funktionalität, Verwendbarkeit oder die Zuverlässigkeit) zu maximieren bzw. gegebenenfalls (etwa bei den Kosten) zu minimieren. Die Funktionen $u_k(x)$ können dabei beliebige Formen annehmen, solange sie für alle zulässigen Kombinationen von Software-Komponenten definiert sind. Dabei können auch vielfältige Abhängigkeiten zwischen zwei oder mehreren Komponenten (bedingt z.B. durch Synergie- oder Kannibalismuseffekte) berücksichtigt werden; vgl. [Stummer2003] für eine detaillierte Diskussion am Beispiel der Auswahl von F&E-Portfolios.

Jedweder Berechnungsalgorithmus, der in dieser ersten Phase zum Einsatz kommt, muss alle (oder zumindest nahezu alle) effizienten MOCO-Lösungen ermitteln und dabei insbesondere Nebenbedingungen aus zwei Gruppen berücksichtigen. Solche aus der ersten Gruppe betreffen typischerweise Ressourcenbeschränkungen (etwa bei den Entwicklungs- oder Wartungskosten). Sie haben daher in der Regel die Form $\sum_i r_{iq} x_i \leq R_q$ ($q = 1, \dots, Q$), wobei r_{iq} für die Menge an

von Komponente i nachgefragten Ressourcen vom Typ q steht und Parameter R_q das Gesamtbudget für diesen Ressourcentyp angibt. Im Falle von Synergie- oder Kannibalismuseffekten müssen wiederum entsprechende Korrekturterme ergänzt werden. Nebenbedingungen aus der zweiten Gruppe gewährleistet, dass jedenfalls eine Mindestzahl bzw. nicht mehr als eine Maximalanzahl von Komponenten aus einer vorab festgelegten Teilmenge ausgewählt werden. Zum Beispiel könnte vorgegeben werden, dass in jeder gültigen Lösung aus jeder Architekturschicht mindestens eine Komponente enthalten sein muss, falls der Benutzer etwa eine mehrschichtige Architektur abbilden möchte. Wenn das sechs Komponenten mit den Indizes 1

bis 6 betrifft, lautet die entsprechende Nebenbedingung: $\sum_{i=1}^6 x_i \geq 1$.

Das MOCO-Problem ist NP-schwer, da sich der Suchraum mit jeder hinzukommenden Komponente verdoppelt. Abhängig vom Umfang – bestimmt insbesondere durch die Anzahl der zur Wahl stehenden Software-Komponenten, die Zahl an Zielen sowie Zahl und Typ der Abhängigkeiten zwischen den Komponenten – kann es entweder exakt durch vollständige Enumeration aller Kombinationen von Komponenten (bei 30 Komponenten wären das mehr als eine Milliarde ($2^{30} > 1,07 \cdot 10^9$) zu evaluierende Alternativen) oder mit Hilfe von (Meta-) Heuristiken gelöst werden. Als Daumenregel gilt, dass letztere ab etwa vierzig Komponenten

zum Einsatz kommen. Sie können zwar keine Garantie für das Auffinden aller Pareto-effizienten MOCO-Lösungen geben, ermitteln aber regelmäßig in einem Bruchteil der für die vollständige Enumeration benötigten Zeit bereits einen Großteil der gesuchten Lösungen; einen Überblick zu Metaheuristiken für MOCO-Probleme bietet [Ehrgott2004], Anwendungsbeispiele mit Leistungsvergleichen finden sich etwa bei [Doerner2006,Stummer2005]. Nichtsdestotrotz kann die vollständige Enumeration auch bei einer höheren Anzahl an Komponenten noch in praktikabler Zeit durchlaufen werden, sofern den Suchraum stark einschränkende Abhängigkeiten existieren (z.B. wenn aus jeder Architekturschicht exakt eine Alternative zu wählen ist).

3.3 Zwei Varianten für die interaktive Suche im Lösungsraums

In der zweiten Phase unterstützt das System den Entscheidungsträger bei der endgültigen Auswahl jener Kombination von Software-Komponenten, die am besten seinen Präferenzen entspricht. Angesichts eines Lösungsraums, der unter Umständen mehrere tausend Elemente umfasst, kommt eine simple Auflistung sicherlich nicht in Frage. Stattdessen bietet das System zwei alternative Varianten für die interaktive „Erforschung“ des Lösungsraums.

Die erste erlaubt es, sich durch Verschieben von unteren bzw. oberen Schranken für beliebige Ziele rasch im Lösungsraum zu bewegen. Das Entscheidungsunterstützende System (EUS) beginnt dazu mit der Anzeige von K „fliegenden“ Säulen (vgl. Abbildung 1).

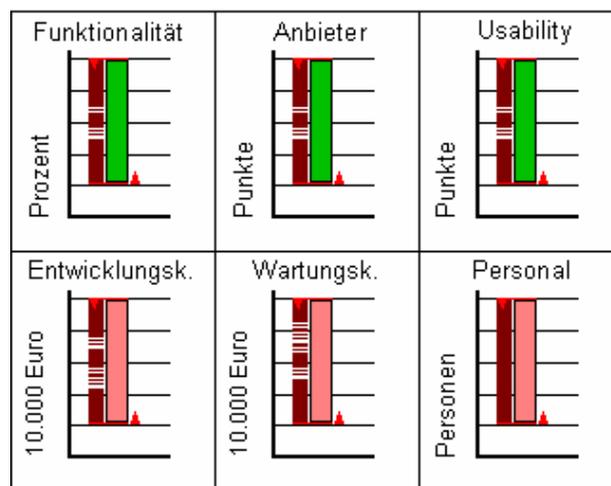


Abbildung 1: Status des EUS zu Beginn der interaktiven Auswahl

Für jedes Ziel (vgl. Abbildung 2) bietet das System Informationen über (i) die Zielwerte, die mit zumindest einer der effizienten Lösungen machbar wären (die kurzen Markierungen auf der linken Seite können dazu optisch zu schmalen Säulen zusammenwachsen, wenn sie knapp bei-

einander liegen) sowie (ii) den Zielgrößenbereich, der mit den – unter Bedachtnahme auf die bereits gesetzten Ober- bzw. Untergrenzen – noch verbliebenen Lösungen erreicht werden kann.

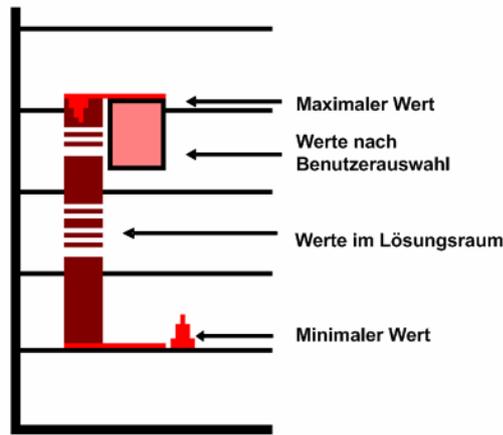


Abbildung 2: Detaillierte Darstellung

Die beweglichen, horizontalen Linien mit den kleinen Pfeilen auf einer Seite symbolisieren die oberen und unteren Schranken und erlauben die schrittweise Einschränkung (z.B. durch Anheben der unteren Schranke in einem Ziel) oder Ausweitung (z.B. durch Hinaufsetzen einer oberen Schranke) der Menge jener Pareto-effizienten MOCO-Lösungen, die gegen keine der bislang gesetzten Schranken verstoßen. Damit kann der Entscheidungsträger spielerisch Vorgaben machen und erhält vom System unmittelbare Rückmeldung über deren Auswirkungen. So illustriert etwa Abbildung 3 die Rückmeldung, nachdem die Obergrenze für die Entwicklungskosten gesenkt worden ist.

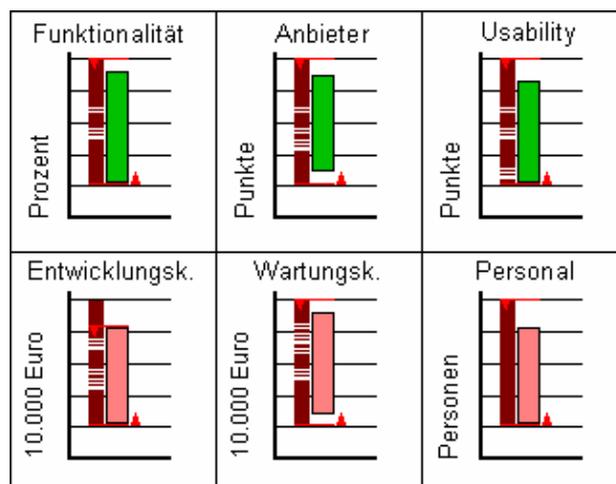


Abbildung 3: Status des DSS nach einer ersten Einschränkung

Dadurch scheiden nämlich jene Lösungsalternativen mit besonders hohen Entwicklungskosten (häufig auch korreliert mit höherem Personalbedarf), jedoch oftmals auch niedrigeren Wartungskosten bzw. höherer Funktionalität aus. Dementsprechend schrumpfen die, durch die fliegenden Balken repräsentierten, Intervalle der nunmehr noch in den anderen Zielen erreichbaren Werte. Eine anschließende Erhöhung der unteren Schranke in der Kategorie „Funktionalität“ führt zu einer weiteren Reduktion der Alternativen auf jene, die beide Vorgaben (d.h. reduzierte Entwicklungskosten und höhere Mindest-Funktionalität) erfüllen, wobei hierbei insbesondere kostengünstigere Alternativen wegfallen (vgl. Abbildung 4).

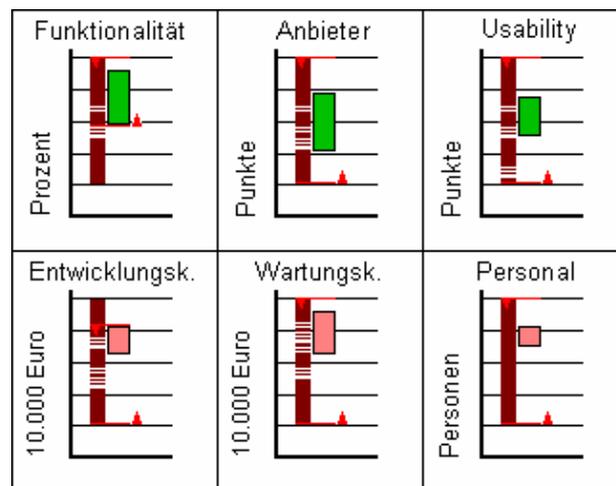


Abbildung 4: Status des DSS nach einer zweiten Einschränkung

In weiteren Iterationen erlaubt das System dem Entscheidungsträger die spielerische Modifikation beliebig vieler oberer und unterer Schranken und ermöglicht ihm damit, unmittelbar die Konsequenzen seiner Vorgaben zu erfahren. Er gewinnt auf diese Weise ein besseres „Gefühl“ für das Entscheidungsproblem und sollte schließlich auf jene Lösung stoßen, die für ihn den „besten“ Mix an Zielwerten bietet und bei der er nicht mehr bereit ist, weitere Kompromisse zwischen den Zielen einzugehen. Dabei hat er – im Gegensatz zu herkömmlichen Verfahren – weder explizit Gewichtungen für einzelne Ziele anzugeben noch muss er tausende paarweise Vergleiche über sich ergehen lassen. Stattdessen erhält er umfassende Informationen über das Auswahlproblem und kann sich darauf verlassen, dass jede angebotene Lösung effizient ist und somit keine Alternative existiert, die objektiv „besser“ wäre.

Im zweiten von uns vorgeschlagene Ansatz für die interaktive Suche im Lösungsraum präsentiert das EUS dem Entscheidungsträger in jeder von mehreren Runden die Zielwerte für bis zu sieben Kombinationen von Software-Komponenten, aus denen die jeweils attraktivste zu wählen ist (vgl. Abbildung 5 für ein Beispiel mit vier Lösungen). Die Beschränkung auf maxi-

mal sieben Alternativen ist psychologisch begründet, weil Menschen sieben plus/minus zwei Elemente gerade noch gut vergleichen können.

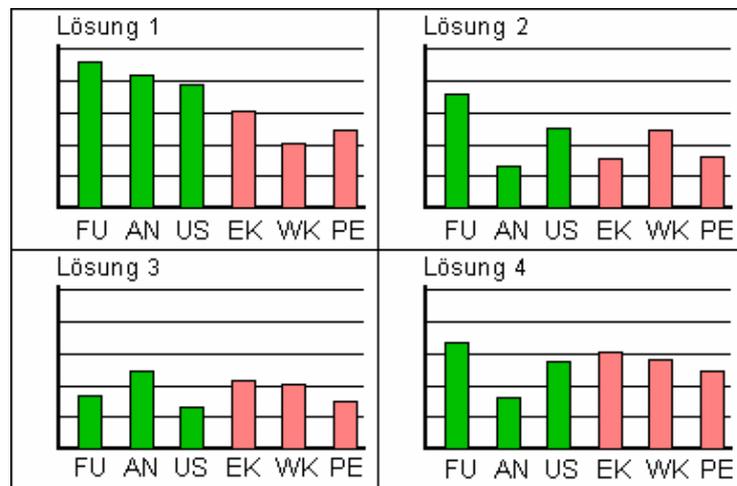


Abbildung 5: Auswahl mit Hilfe von Cluster-Repräsentanten

Anfangs unterscheiden sich die Zielwerte der Lösungsvorschläge in hohem Maße, so dass hier vor allem Richtungsentscheidungen (etwa hohe Funktionalität vs. geringe Kosten) zu treffen sind. Später wird die Suche verfeinert und die Unterschiede zwischen den angebotenen Lösungen werden zusehends geringer. Methodisch kommt das k-means-Clustering zum Einsatz. Aus den damit aus der Gesamtmenge aller effizienten Lösungen bestimmten sieben Clustern werden dem Entscheidungsträger Repräsentanten angeboten. Jenes Cluster, aus dem der dann gewählte Vertreter stammt, wird in der nächsten Iteration weiter in Sub-Cluster unterteilt usw., bis sich der Entscheidungsträger letztlich für ein Cluster entscheidet, das nur mehr eine einzige Lösung enthält. Bildlich gesprochen „taucht“ er dabei gleichsam in den Lösungsraum ein und bewegt sich zuerst in großen, dann später immer kleineren Kreisen an die für ihn individuell attraktivste Lösung heran. Dies funktioniert in aller Regel recht rasch, weil im Durchschnitt jedes Cluster nur ein Siebentel der noch verbliebenen Lösungen enthält. Selbst bei etwa 20.000 effizienten Lösungen sind daher im Mittel lediglich um die fünf, mit hoher Wahrscheinlichkeit aber jedenfalls nicht mehr als zehn Iterationen nötig.

4 Resümee und Ausblick

Entscheidungsträger sehen sich bei der komponentenbasierten Software-Entwicklung regelmäßig mit der Herausforderung konfrontiert, aus einem großen Pool von Komponenten die

„beste“ Kombination zusammen zu stellen. Diese Aufgabe wird nicht nur durch eine oftmals große Zahl an alternativen Komponenten (auf verschiedenen architektonischen Ebenen), sondern auch durch Abhängigkeiten zwischen einzelnen Komponenten bzw. nicht zuletzt durch mehrere divergierende Zielsetzungen verkompliziert. In diesem Aufsatz haben wir einen zweiphasigen Ansatz zur Entscheidungsunterstützung bei der Auswahl von Software-Komponenten entwickelt, der als Erweiterung in bestehende Vorgehensmodelle für die komponentenbasierte Softwareentwicklung integriert werden kann und insbesondere Probleme und Unzulänglichkeiten, die bei herkömmlichen Verfahren wie der Nutzwertanalyse oder dem AHP auftreten, überwindet.

Wie alle anderen Verfahren hängt auch unseres von der Verfügbarkeit geeigneter Zielfunktionen bzw. der Qualität der in der Evaluierungsphase erhobenen Daten ab. So ist es sicherlich nicht trivial, eine geeignete Funktion für die Zuverlässigkeit einer Kombination von Software-Komponenten aufzustellen. Damit wollen wir uns im Zuge weiterer Forschungsaktivitäten beschäftigen, wobei wir hierfür etwa das Einbeziehen der unmittelbar betroffenen Praktiker im Rahmen von Workshops ins Auge gefasst haben (für ein Beispiel vgl. [Neubauer2006]).

Darüber hinaus gibt es eine Reihe weiterer interessanter Fragestellungen. So sollten Aspekte der Ungewissheit berücksichtigt werden, um bei Bedarf bewusst robuste Zusammenstellungen von Software-Komponenten erzwingen zu können, die dann weniger anfällig für sich ändernde Systemanforderungen wären. Im MOCO-Modell ließe sich das in einem ersten Schritt mit stochastischen Zufallsvariablen und dann über Quantilswerte als separate Ziele umsetzen (für ein Beispiel vgl. [Medaglia2006]). Des Weiteren könnten vermehrt Ideen aus der Literatur zu Gruppenentscheidungen bzw. der Verhandlungsanalyse übernommen werden. Und schließlich sollen Feldstudien Aufschluss über die Eignung unterschiedlicher (grafischer) Benutzerschnittstellen geben.

Literaturverzeichnis

[Alves2001] C. Alves, J. Castro (2001) CRE: A systematic method for COTS components selection. Proceedings of the XV Brazilian Symposium on Software Engineering.

- [Alves2003] C. Alves, A. Finkelstein (2003) Investigating conflicts in COTS decisionmaking. *International Journal of Software Engineering and Knowledge Engineering*, 13. Jg., H. 5, S. 473-495.
- [Andrews2005] A. A. Andrews, A. Stefik, N. Picone, S. Ghosh (2005) A COTS component comprehension process. *Proceedings of the 13th International Workshop on Program Comprehension (IWPC'05)*, IEEE, S. 135-144.
- [Basili1987] V. R. Basili, H. Rombach (1987) Tailoring the software process to project goals and environments. *Proceedings of the 9th International Conference on Software Engineering*. IEEE, S. 345-357.
- [Basili1996] V. R. Basili, L. C. Briand, W. L. Melo (1996) How reuse influences productivity in object-oriented systems. *Communications of the ACM*, 39. Jg., H. 10, S. 104-116.
- [Doerner2006] K. F. Doerner, W. J. Gutjahr, R. F. Hartl, C. Strauss, C. Stummer (2006) Nature-inspired metaheuristics for multiobjective activity crashing. *Omega*, im Druck.
- [Ehrgott2000] M. Ehrgott, X. Gandibleux (2000) A survey and annotated bibliography of multiobjective combinatorial optimization. *OR Spectrum*, 22. Jg., H. 4, S. 425-460.
- [Ehrgott2004] M. Ehrgott, X. Gandibleux (2004) Approximative solution methods for multi-objective combinatorial optimization. *Top*, 12. Jg., H. 1, S. 1-63.
- [Kontio1995] J. Kontio (1995) OTSO: A systematic process for reusable software component selection. Institute for Advanced Computer Studies and Department of Computer Science, University of Maryland, Arbeitspapier.
- [Kunda2003] D. Kunda (2003) STACE: Social technical approach to COTS software evaluation. *Component-Based Software Quality: Methods and Techniques*, Springer LNCS 2693, S. 64-84.
- [LozanoTello2002] A. Lozano-Tello, A. Gomez-Perez (2002) BAREMO: How to choose the appropriate software component using the Analytic Hierarchy Process. *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering*, ACM Proceedings 27, S. 781-788.
- [Maiden1998] N. Maiden, C. Ncube (1998) Acquiring COTS software selection requirements. *IEEE Software*, 15. Jg., H. 2, S. 46-56.

- [Maiden2002] N. Maiden, H. Kim, C. Ncube (2002) Rethinking process guidance for selecting software components. Proceedings of the First International Conference on COTS-Based Software Systems (ICCBSS 2002), Springer LNCS 2255, S. 151-164.
- [Martinsons1998] M. Martinsons, R. Davidson, D. Tse (1998) The balanced scorecard: A foundation for the strategic management of information systems. Decision Support Systems Journal, 25. Jg., H. 1, S. 71-78.
- [Medaglia2006] A. Medaglia, S. Graves, J. Ringuest (2006) A multiobjective evolutionary approach for linearly constrained project selection under uncertainty. European Journal of Operational Research, im Druck.
- [Morisio2002] M. Morisio, M. Torchiano (2002) Definition and classification of COTS: A proposal. Proceedings of the First International Conference on COTS-Based Software Systems (ICCBSS 2002). Springer LNCS 2255, S. 165-175.
- [Navarrete2005] F. Navarrete, P. Botella, X. Franch (2005) How agile COTS selection methods are (and can be)? Proceedings of the 31st EUROMICROConference on Software Engineering and Advanced Applications (EUROMICRO-SEAA'05), IEEE, S. 160-167.
- [Ncube2002] C. Ncube, J. C. Dean (2002) The limitations of current decision-making techniques in the procurement of COTS software components. Proceedings of the First International Conference on COTS-Based Software Systems (ICCBSS 2002), Springer LNCS 2255, S. 176-187.
- [Neubauer2006] T. Neubauer, C. Stummer, E. Weippl (2006) Workshop-based multiobjective security safeguard selection. Proceedings of the First International Conference on Availability, Reliability and Security (ARES'06). IEEE, S. 366-373.
- [Ruhe2002] G. Ruhe (2002) Intelligent support for selection of COTS products. Revised Papers from the Workshop on Web, Web-Services, and Database Systems (NODE 2002), Springer LNCS 2593, S. 34-45.
- [Ruhe2003] G. Ruhe (2003) Software engineering decision support: A new paradigm for learning software organizations. Proceedings of the 4th International Workshop on Advances in Learning Software Organizations (LSO 2002), Springer LNCS 2640, S. 104-113.

- [Ryan2004] S. D. Ryan, M. S. Gates (2004) Inclusion of social sub-system issues in IT-investment decisions: An empirical assessment. *Information Resources Management Journal*, 17. Jg., H. 1, S. 1-18.
- [Torchiano2004] M. Torchiano, M. Morisio (2004) Overlooked aspects of COTS-based development. *IEEE Software*, 21. Jg., H. 2, S. 88-93.
- [Tran1997] V. Tran, D.-B. Liu, B. Hummel (1997) Component-based systems development: challenges and lessons learned. *Software Technology and Engineering Practice*. IEEE, S. 452-462.
- [Stummer2003] C. Stummer, K. Heidenberger (2003) Interactive R&D portfolio analysis with project interdependencies and time profiles of multiple objectives. *IEEE Transactions on Engineering Management*, 50. Jg., H. 2, S. 175-183.
- [Stummer2005] C. Stummer, M. Sun (2005) New multiobjective metaheuristic solution procedures for capital investment planning. *Journal of Heuristics*, 11. Jg., H. 3, S. 183-199.
- [Wanyama2005] T. Wanyama, B. Homayoun (2005) Towards providing decision support for COTS selection. *Proceedings of the 18th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2005)*, IEEE, S. 908-911.

Ein Kennzahlensystem zur Optimierung von SAP Systemen

Eine Fallstudie am Beispiel der deutschen Automobilindustrie

Peggy Sekatzek

BMW Group
80788 München
Peggy.Sekatzek@bmw.de

Helmut Krcmar

Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
85748 Garching
Krcmar@in.tum.de

Abstract

In vielen Fällen hat der Einsatz betrieblicher Standardsoftware (SSW) die gesetzten Erwartungen nicht erfüllt ([BCG00]; [MuÖs99]). Hohe Kosten für Einführung und Betrieb vermindern den erwarteten monetären Nutzen teilweise erheblich ([MaAP02]). Häufiger Kostentreiber ist in diesem Zusammenhang ein hoher Modifikationsgrad der Software, der neben hohen Einführungskosten auch eine aufwändigere Wartung nach sich zieht. Eine Lösung des Problems kann in einer konsequenten Überwachung der Systeme hinsichtlich ungenutzter Eigenentwicklungen gesehen werden. Dieser Beitrag beschreibt, wie ausgehend von einer Schwachstellenanalyse bei Einführung und Betrieb von betrieblicher Standardsoftware der SAP AG ein Kennzahlensystem entwickelt wurde, welches SAP Systeme hinsichtlich Kosten und potentiellen Kostentreibern wie z.B. ungenutzten Eigenentwicklungen transparent macht. In einer Fallstudie bei der BMW Group wurde das Kennzahlensystem hinsichtlich Praxistauglichkeit validiert. Das Vorgehen bei der Systemerstellung sowie Messergebnisse aus der Fallstudie werden in dem Beitrag vorgestellt.

1 Einleitung

Unternehmen haben hohe Erwartungen an den Einsatz ihrer betrieblichen Standardsoftware (SSW) wie SAP oder Oracle gesetzt. Ziele waren die Optimierung der Prozesse, unternehmensweite Standardisierung, Integration, Flexibilität, aber auch der Einsatz einer zukunftsfähigeren und kostengünstigeren Softwarelösung, als sie eigenentwickelte Software darstellt ([OIRo02]).

Diesen Erwartungen stehen in der Praxis aber vermeintlich hohe Kostenaufwände bei Einführung und Betrieb von SSW gegenüber: Studien belegen, dass dies zu einem nicht unwesentlichen Anteil von Softwaremodifikationen und -erweiterungen herrührt ([ThIO]; [MaAP02]; [BuKö97]). Der Fachbereich definiert zu einem sehr frühen Zeitpunkt der Softwareeinführung die von ihm benötigten Funktionalitäten, die, sofern sie nicht im Standardumfang der Software enthalten sind, mit Hilfe von Modifikationen und Erweiterungen hinzugefügt werden. Diese werden von ihm später möglicherweise nicht mehr gebraucht. Die IT-Abteilung hingegen hat keine Transparenz über das Nutzungsverhalten der Anwender, so dass ungenutzte Eigenentwicklungen weiter gewartet werden, wodurch unnötige Kosten entstehen. Diese Einsparungspotentiale, die die Unternehmen bei konsequenter Evaluierung ihrer SSW Systeme erreichen könnten, bleiben häufig ungenutzt ([CIO06]). Die Herausforderung besteht für die Unternehmen also in der ständigen Überwachung und Anpassung der SAP Landschaft.

Dieser Beitrag beschreibt ein andauerndes Forschungsvorhaben: Ausgehend von einer Literaturanalyse wurden kritische Faktoren des SAP Einsatzes identifiziert. In einer anschließenden Fallstudie, die in der BMW Group durchgeführt wurde, wurden diese Faktoren hinsichtlich ihrer Praxisrelevanz überprüft. Auf dieser Basis wurde ein Kennzahlensystem zur Überwachung von SAP Systemen entworfen. Das System erlaubt einen Vergleich zwischen verschiedenen SAP Systemen hinsichtlich Kosten, Planung, Standardisierung und -nutzung, Lizenznutzung sowie Hardware-spezifischen Faktoren.

2 Forschungsdesign

Das Forschungsvorgehen besteht aus zwei Teilen. Zuerst wurden anhand der Literatur Kostentreiber beim Einsatz von betrieblicher Standardsoftware analysiert und

zusammengefasst. Für den zweiten Teil wurde als Forschungsmethodik die Fallstudie ausgewählt, die im Zeitraum von Juni 2005 bis Juni 2006 in der BMW Group, wie bei Yin ([Yin94]) und Eisenhardt ([Eise98]) beschrieben, durchgeführt wurde. Hier wurde zuerst die Kostensituation in Hinblick auf SAP untersucht. Mit Hilfe von Dokumentenanalysen und semi-strukturierten Experteninterviews wurden daraus die für das Unternehmen relevanten Kostentreiber abgeleitet.

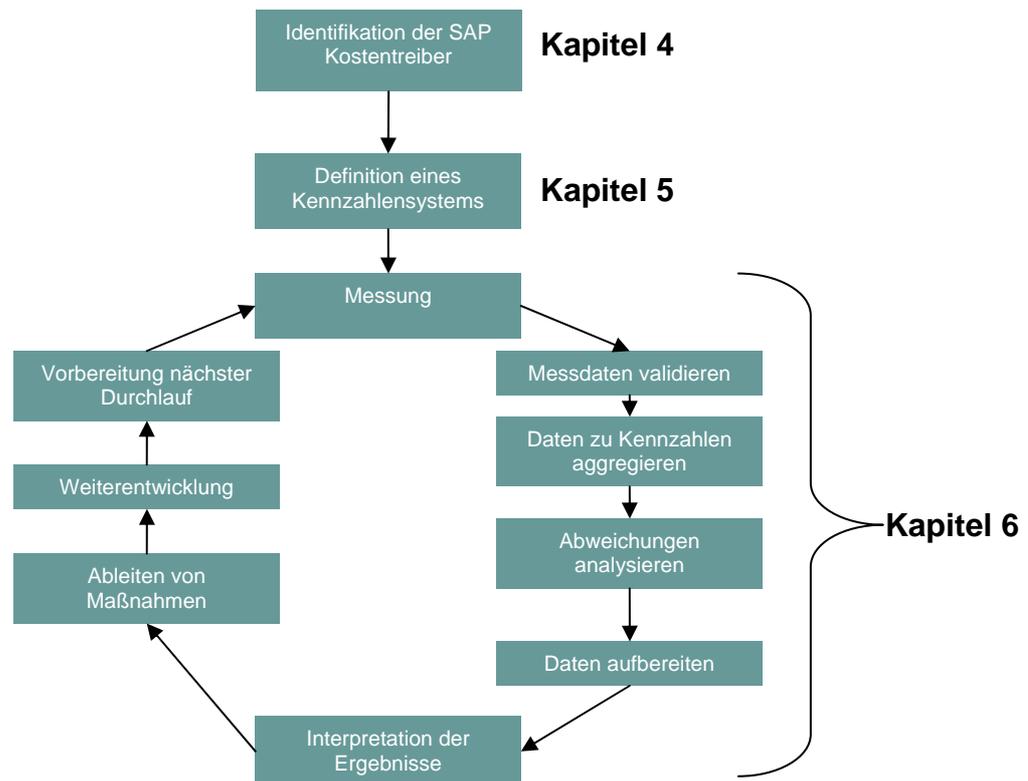


Abbildung 1: Vorgehensmodell
Quelle: Eigene Darstellung

Im nächsten Schritt wurde das Kennzahlensystem entworfen, welches die SAP Systeme hinsichtlich Kosten, Standardisierung, Lizenznutzung und Hardware bewertbar macht. Das Kennzahlensystem wurde mit Experten diskutiert und über einen Zeitraum von fünf Monaten iterativ verfeinert. Nach seiner Fertigstellung wurde das Modell konzernweit ausgerollt. Alle in- und ausländischen SAP Produktivsysteme wurden vermessen. Die Ergebnisse des ersten Vermessungsdurchganges werden in diesem Beitrag vorgestellt.

Abbildung 1 zeigt das Vorgehen bei der Erstellung des Kennzahlensystems sowie bei der Kennzahlenerhebung und die Zuordnung der einzelnen Schritte zu den jeweiligen Kapiteln dieses Beitrags.

3 Angrenzende Forschungsvorhaben

Bei der Evaluierung von SSW findet sich häufig eine Einteilung in monetäre und nicht-monetäre Bewertungsansätze. Im Rahmen dieser Arbeit wurde ein Kennzahlensystem entwickelt, welches SSW Systeme auf beide Arten vom Standpunkt der IT aus betrachtet. Im Gegensatz hierzu wäre eine Betrachtungsweise aus Business-Sicht denkbar. In der Business-Sicht werden der Nutzen bzw. die Auswirkungen einer SSW auf das Business betrachtet. Die IT-Sicht dagegen untersucht die SSW hinsichtlich IT-Kriterien, wie z.B. Systemstandardisierung- bzw. Nutzung.

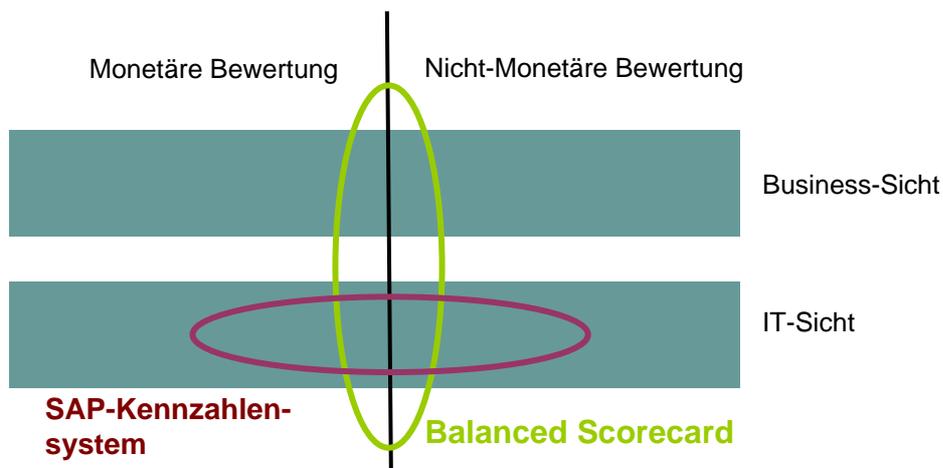


Abbildung 2: Klassifizierung von Kennzahlensystemen
Quelle: Eigene Darstellung

Abbildung 2 zeigt eine mögliche Klassifizierung von Kennzahlensystemen in diese zwei Dimensionen sowie die Einordnung des vorliegenden Kennzahlensystems und der Balanced Scorecard. Weiter Klassifizierungsmöglichkeiten finden sich bei Krcmar ([Krcm05]) sowie bei Irani/Love ([IrLo02]).

3.1 Monetäre Ansätze zur SSW Evaluierung

Traditionell basiert die Evaluierung von IS/IT-Evaluierung auf monetären Kriterien. Zur Erfolgsmessung bedient man sich in den meisten Fällen des Return on Investment (ROI) oder den dynamischen Investitionsmethoden wie dem Net Present Value (NPV) ([Stef01]; [SeRG01]). Dieser stellt den meist genutzten Weg dar, den monetären Erfolg einer Investition zu berechnen. Im Folgenden wird ferner das Verfahren zur Berechnung der Total Cost of

Ownership vorgestellt, da es die Ausgangsbasis für die Erstellung des Kennzahlensystems darstellt.

Das im Jahre 1987 durch die Gartner Group erstellte Modell zur Erfassung der Total Cost of Ownership (TCO) fasst alle Kosten zusammen, die für Anschaffung und Betrieb eines IT-Systems während dessen gesamter Lebensdauer anfallen ([Krcm05]; [TrTS03]; [WiHe]). Angeregt durch den steigenden Wert der IT in Unternehmen war es das Ziel, ein Modell zur Erfassung der Gesamtkosten von IT-Systemen zu entwerfen, das über die reine Betrachtung der Anschaffungskosten von Hard- und Software hinausgeht. Der Hauptvorteil des Modells liegt in der Darstellung der tatsächlichen Kosten einer Investition sowie in seiner Einfachheit, durch welche es eine Vergleichsbasis für ähnliche Alternativen schafft. Als Grundlage für interne und externe Benchmarks dient es der Identifikation von Kostentreibern. Nach dem Erfolg von TCO entstanden neben dem Modell von Gartner zahlreiche Variationen u.A. von Forrester Research, der Meta Group sowie der SAP AG ([SAP05a]; [TrTS03]; [WiHe]). Obwohl dieser Ansatz in der Vergangenheit zum Teil kontrovers diskutiert wurde, wird er als Grundlage des Kennzahlensystems genutzt, um einen späteren Kostenvergleich zwischen verschiedenen Unternehmen bzw. mit Benchmarkdaten zu ermöglichen ([MuÖs99]).

Im Folgenden werden als Bestandteil eines TCO-Modells die Kategorien Einmalkosten, Betriebskosten, Wartungskosten, Infrastrukturkosten und Lizenzkosten gesehen.

3.2 Nicht-Monetäre Ansätze zur SSW Evaluierung

Monetäre Messungen können nicht das komplette Bild einer IT-Investition wiedergeben. Viele Versuche, qualitative Elemente in die Evaluierung der IT-Projekte aufzunehmen wurden untersucht. In der Literatur finden sich mehrere Ansätze zur Erfolgsmessung von betrieblicher Standardsoftware wie das Process Performance Messmodell, Workflow basierte Messmodelle, Statische Messmethoden und die Balanced Scorecard ([SeRG01]; [Stef01]; [KenNe02]). Jede dieser Methoden hat kontextabhängige Vor- und Nachteile. Dennoch liegt die Abgrenzung der einzelnen Ansätze nicht im Scope dieses Beitrags, lediglich die Balanced Scorecard (BSC), die einen der meist verbreiteten Ansätze darstellt, soll im Folgenden vorgestellt werden.

Die Konzeption dieses Instruments basiert auf einer Studie von Kaplan und Norton ([KaNo92]). Die originäre BSC wird aus vier Perspektiven zusammengesetzt, die ein in sich ausgewogenes Kennzahlensystem bilden sollen, welches vorlaufende Indikatoren und Leistungstreiber integriert. Diese Perspektiven sind um die aus der Unternehmensvision oder -mission abgeleitete Unternehmensstrategie angeordnet. Sie decken in der Regel die monetäre Perspektive, die Kundenperspektive, die interne Prozessperspektive und die Lern- und Wachstumsperspektive ab. ([KaNo92]; [JLMM04]; [Wefe00])

Das im Rahmen dieses Beitrags diskutierte Kennzahlensystem nutzt die grundlegenden Ansätze und Vorteile der Balanced Scorecard, zum Beispiel indem es nicht-monetäre Kennzahlen gleichwertig zu monetären Kennzahlen einbezieht und Kausalketten zwischen den Kennzahlen herstellt. Der Unterschied zur herkömmlichen BSC ist, dass das Kennzahlensystem die Evaluierung eines SAP-Systems aus IT-Sicht durchführt. Auf Basis einer Schwachstellenanalyse wurden spezifische Teilaspekte eines SAP-Systems wie z.B. der Standardisierungsgrad direkt adressiert. Schwerpunkt war bei der Entwicklung also nicht die Ganzheitlichkeit, wie sie eine BSC gewährleistet, sondern die spezifische Betrachtung von kritischen Faktoren eines SAP Systems. Indem die für die Kennzahlen benötigten Daten direkt aus jedem beliebigen SAP System gewonnen werden können, ist eine Einsetzbarkeit in jedem Unternehmen gewährleistet. Eine BSC dagegen muss individuell auf die Gegebenheiten eines Unternehmens angepasst werden.

4 Kostentreiber-Analyse

Auf Basis einer Literaturanalyse sowie von im Rahmen der Fallstudie gesammeltem Informationsmaterial wurden Kostentreiber, die bei IT-Projekten im Allgemeinen sowie bei Einführung und Einsatz betrieblicher Standard Software anfallen, herausgearbeitet ([ShSS04]; [HoLi99]; [HoLG98]; [KyYo02]). In den folgenden Tabellen werden diese getrennt voneinander vorgestellt, da das Kennzahlensystem speziell auf die Kostentreiber der Standardsoftware zugeschnitten ist.

Kostentreiber bei IT-Projekten (allgemein)
Fehlendes Top-Management Commitment
Fehlende Anforderungsdefinitionen
Mangelnde Planungssicherheit
Nicht gedeckter Eigenleistungsanteil
Nicht ausreichende Freistellung von Projektmitarbeitern
Unzureichende Projektsteuerung und –koordination
Zu lange Projektlaufzeiten
Zu ambitionierte Kosten- und Terminziele
Wesentliche Änderung grundlegender Prämissen während der System-Einführungsphase
Projektkomplexität

Tabelle 1: Kostentreiber in IT-Projekten

Bei der Einführung sowie dem Betrieb von betrieblicher Standardsoftware sind vor allem die folgenden Kostentreiber zu beachten:

Standardsoftware-spezifische Kostentreiber
Nutzung unreifer Produkte / Technologien
Unzureichendes internes Produkt-Knowhow
Lokale Anforderungen der Business Units
Hohe Parallelisierung Phasen / Rollouts
Zusätzlicher Schulungs- und Betreuungsaufwand
Anbindung von Legacy
Erstellung von Modifikationen/ Erweiterungen
Wartung von Modifikationen/ Erweiterungen
Mangelnde Planungssicherheit bei der Lizenzplanung
Unwirtschaftliche Nutzung von SAP Lizenzen

Tabelle 2: Kostentreiber bei Einführung und Betrieb von betrieblicher Standardsoftware

Die identifizierten Kostentreiber wurden den Kostenkategorien aus dem TCO-Modell zugeordnet (siehe Abbildung 3). Ebenfalls wurde auf Basis von unterschiedlichen Analystenstudien der prozentuale Anteil der Kostenkategorien an den gesamten TCO vorgenommen. Die Abbildung zeigt nur einen Teilausschnitt des im Rahmen dieses Beitrags erstellten TCO-Kostentreiber-Wirkmodells. Zu beachten ist, dass nur die Kostentreiber, die im nachfolgenden Kennzahlensystem adressiert werden, in dieser Abbildung dargestellt wurden.

In einem weiteren Schritt wurden anhand der identifizierten Schwachstellen Kennzahlen definiert, um die wichtigsten Kostentreiber messbar machen zu können. Als Kennzahlenkategorien für die erste Version des Kennzahlensystems wurden Standardisierung, Lizenzen und Hardware-spezifische Faktoren identifiziert.

Mit dem entwickelten Kennzahlensystem können nicht alle identifizierten Kostentreiber erfasst werden. Kostentreiber wie eine überhöhte Fremdleistungsquote sowie Fehler beim Projektmanagement liegen in der Verantwortung des Unternehmens. Trotzdem können mit dem System diejenigen Kostentreiber, deren Erhebung direkt im System liegt, schnell erfasst werden. Das Modell ist auf andere Unternehmen übertragbar, da die benötigten Daten direkt aus den SAP Systemen bezogen werden können.

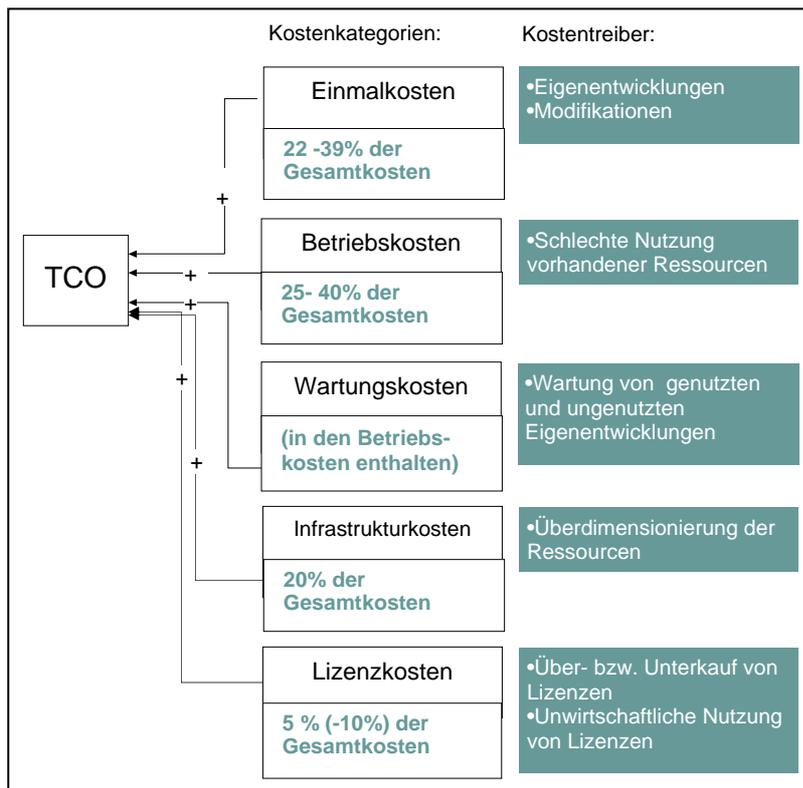


Abbildung 3: Kostentreiber (Teilausschnitt)¹

5 Design des Kennzahlensystems

Kennzahlen können als Verdichtungen quantitativer Daten gesehen werden, die über zahlenmäßig erfassbare Ausschnitte der Realität informieren ([Krcm05]). Während manche Kennzahlensysteme das Unternehmen als Gesamtheit abzubilden versuchen, sind andere, wie auch das vorliegende Kennzahlenmodell, dafür konzipiert, lediglich einen Teilbereich eines Unternehmens zu betrachten ([Kütz03]). Mit dem vorliegenden Kennzahlensystem sollen die

¹ Prozentwerte auf Basis verschiedener Analystenstudien (Gartner etc.)

Überwachung der SAP Landschaft eines Unternehmens sowie die folgenden Ziele verfolgt werden:

- Unternehmensweite Transparenz über den SAP Einsatz
- Internes und externes Benchmarking
- Ableitung von Optimierungsmaßnahmen
- Wirksamkeitsnachweis angewandter Maßnahmen

Das SAP Kennzahlenmodell besteht aus zwei Teilen, zum einen den Key Goal Indikatoren (KGIs), d.h. Zielen wie Optimierung der Kosten oder optimale Nutzung vorhandener Ressourcen, die mit dem Kennzahlensystem erreicht werden sollen. Zum anderen definiert es Key Performance Indikatoren (KPIs), die zur Messung der Zielerreichung dienen ([Kütz03]). Die KPIs sind in Kennzahlenkategorien eingeteilt:

1. Kennzahlen zur Kostenüberwachung (Kostenkennzahlen)
2. Kennzahlen zur Lizenznutzung
3. Kennzahlen zur Standardisierung
4. Kennzahlen zur Hardwarenutzung

Die Kategorie *Kostenkennzahlen* enthält Kennzahlen zur Überwachung von Kostengrößen, die über den gesamten Lifecycle eines SAP Systems anfallen. Basis für diese Kennzahlen bildet das TCO-Modell.

Beispielkennzahl: U.A. findet sich in dieser Kategorie die Kennzahl *Jährliche Wartungskosten pro User*, die die Darstellung des Kostenverlaufs eines Systems über die Jahresscheiben ermöglicht und somit Schwankungen erkennbar macht. Eine erste Normierung wurde durch die Division durch die Userzahlen erreicht. Somit können zum einen unternehmensinterne Systeme unabhängig von ihrer Systemgröße verglichen werden, zum anderen können die Kosten mit am Markt verfügbaren Best-practice-Werten verglichen werden. Berechnet wird die Kennzahl wie folgt:

$$K1 = \frac{\sum WK}{n}, \text{ mit}$$

$\sum WK$ = Summe der jährlich angefallenen Wartungskosten und
 n = Anzahl aktiver User im SAP System

Die Kategorie *Kennzahlen zur Lizenznutzung* dient der Steuerung des effizienten Kaufes und Einsatzes von Userlizenzen.

Beispielkennzahl: In dieser Kategorie findet sich u.A. die Kennzahl *Lizenznutzungsgrad*. Diese stellt das Verhältnis von angemeldeten Usern zu gekauften Usern dar und dient als Indikator für die Nutzungsintensität der User-Lizenzen eines Systems. Ziel der Kennzahl ist die Reduktion im System befindlicher ungenutzter SAP-Lizenzen, die so in einen Pool zurückfließen können, aus dem sie an anderer Stelle wieder verwendet werden können. Kostenintensive Nachkäufe können so vermieden werden. Der Lizenznutzungsgrad eines Systems berechnet sich wie folgt:

$$K2 = \frac{m}{n}, \text{ mit}$$

m = Anzahl der in 90 Tagen angemeldeten SAP User und
 n = Anzahl aller aktiven User im System

In der Kategorie *Kennzahlen zur Standardisierung* sind Kennzahlen enthalten, die Auskunft über den Standardisierungsgrad einer eingeführten SAP Lösung geben. U.A. sind in dieser Kategorie die Kennzahlen Standardabdeckungs- und -nutzungsgrad enthalten.

Beispielkennzahl: Die Kennzahl *Standardabdeckungsgrad* gibt an, welcher Anteil einer SAP Lösung im SAP Standard realisiert ist und welcher Anteil aus Erweiterungen bzw. Modifikationen besteht. Ziel ist eine Implementierung, die sich möglichst nahe am SAP Standard orientiert, womit eine Vermeidung der Kosten, die durch Modifikationen und Erweiterungen des Standards entstehen, angestrebt wird.

Neben dem Standardabdeckungsgrad, welcher statisch ist, misst der *Standardnutzungsgrad* die Intensität der Nutzung von SAP Standardfunktionalität gegenüber der Nutzung von Erweiterungen und Modifikationen. Ziel ist die Bereinigung eines Systems von nicht genutzten Modifikationen und Erweiterungen, wodurch der Wartungsaufwand erheblich

reduziert werden kann. Auf der anderen Seite können häufig genutzte Eigenentwicklungen so ggf. in den Standard überführt werden. Berechnet wird der Standnutzungsgrad wie folgt:

$$K3 = \frac{o}{p}, \text{ mit}$$

o = Genutzte Programme im SAP Standard (Zeitraum: Drei Monate) und

p = Gesamtzahl aller genutzten Programme der verwendeten Module (Zeitraum: Drei Monate)

Die Kategorie *Kennzahlen zur Hardwarenutzung* enthält Kennzahlen zur Steuerung der effizienten Nutzung der vorhandenen Infrastruktur.

Beispielkennzahl: U.A. findet sich in dieser Kategorie die Kennzahl *Serverauslastung*, die die durchschnittliche Auslastung der genutzten Server angibt. Ziel ist eine optimale Nutzung der vorhandenen Serverkapazitäten.

Über einen längeren Zeitraum gemessen, können die Kennzahlenwerte eines Systems im Zeitverlauf analysiert werden. Sollen auch zwischen verschiedenen SAP Systemen eines Unternehmens Vergleiche gezogen werden, ist eine sorgfältige Einteilung der SAP Systeme notwendig (siehe Fallstudie BMW Group).

6 BMW Group – Die Fallstudie

In diesem Abschnitt wird eine in der BMW Group durchgeführte Fallstudie vorgestellt. Das Kapitel beginnt mit der Vorstellung der BMW Group und ihrer SAP Landschaft. Anschließend werden die Ziele der Fallstudie sowie erste Messergebnisse dargestellt.

6.1 SAP Einsatz in der BMW Group

Die Entwicklung des Einsatzes von SAP Software bei der BMW Group lässt sich in drei Phasen einteilen: Phase 1 begann mit der Einführung eines zentralen Buchhaltungssystems. Damals war der Einsatz von SAP geprägt durch lokale, autonome Systeme, ohne dass es eine zentrale Steuerungsinstanz gab.

Mit Phase 2 begann der Übergang von lokalen Lösungen zu in Programmen und Projekten zusammengefassten Lösungen, die über Schnittstellen verbunden wurden. Erste Template-

Ansätze (zentrale Entwicklung und Rollouts in Standorte) wurden entwickelt. Die Phase 3 beinhaltet die Ausweitung auf eine unternehmensweite Integration über die Geschäftsprozesse hinweg.

6.2 Durchführung der Fallstudie und Messergebnisse

Die Fallstudie geht der Frage nach, wie die Systeme der BMW Group in Bezug auf Kosten, Standardisierung, Lizenznutzung und Hardware-spezifischen Daten positioniert sind.

Um einen Überblick über die durch SAP verursachten Kosten zu erhalten, wurde im Unternehmen ein Total Cost of Ownership (TCO) Modell für die gesamte SAP Landschaft ausgerollt. Dieses bildete die Ausgangsbasis zur Berechnung der Kosten-Kennzahlen.

Um eine optimale Vergleichbarkeit auch zwischen den Systemen herzustellen, wurden diese zuerst anhand der folgenden Kriterien in Komplexitätskategorien eingeteilt: Anzahl der fachlichen Schnittstellen, Anzahl der eingesetzten SAP-Module, Anzahl der aktiven User und Verhältnis der produktiven zu nicht-produktiven Systemen.

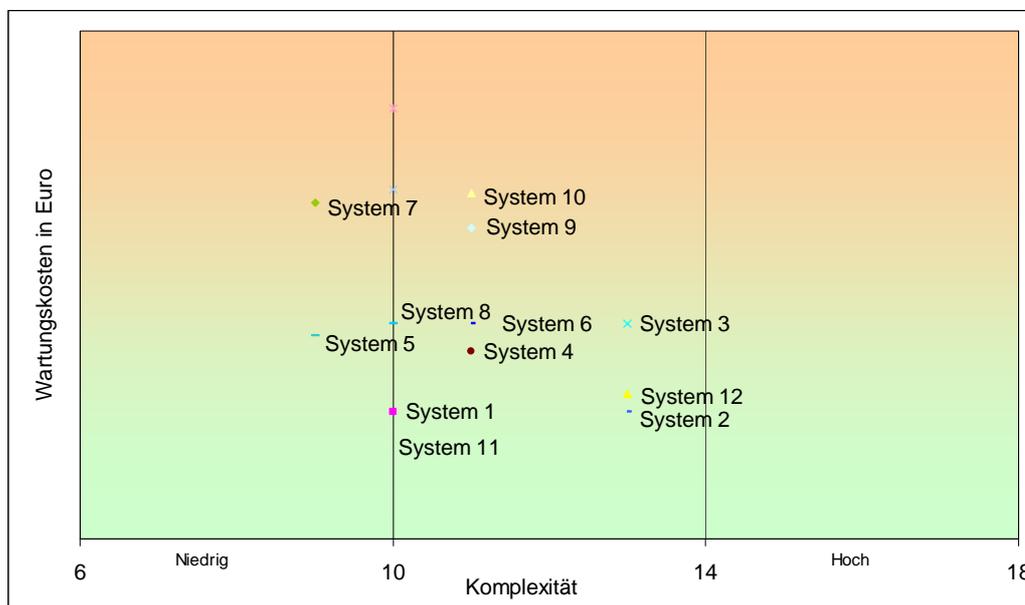


Abbildung 4: Einteilung der Systeme in Komplexitätsklassen
Quelle: Eigene Darstellung

Abbildung 4 zeigt die drei Komplexitätsklassen (hoch, mittel, gering) sowie die Einteilung einiger SAP-Systeme. Der Komplexitätsgrad befindet sich an der horizontalen Achse, die

Wartungskosten pro User werden an der y-Achse eingetragen. Die Wartungskosten pro User eines Systems lassen sich nun mit den übrigen Systemen innerhalb einer Komplexitätsklasse vergleichen. Im Rahmen dieses Beitrags können nicht alle Messergebnisse im Detail besprochen werden. Ein Ausschnitt der ersten Messergebnisse sowie daraus abgeleitete Maßnahmen werden im Folgenden vorgestellt.

Lizenznutzungsgrad: Im Unternehmen wird ein User-Bereinigungsprozess durchgeführt. Im Rahmen der Kennzahlenerhebung wurde der Grad der Lizenznutzung pro SAP System vor bzw. nach der Userbereinigung erhoben. Mit Hilfe der Kennzahlenerhebung konnte gezeigt werden, dass die Userbereinigung eine Steigerung des Lizenznutzungsgrades um 30% ermöglicht hat und somit die Anzahl der ungenutzten User im System gesenkt werden konnte. Hierdurch können ebenfalls Lizenzkosten gespart werden, indem ungenutzte User in den Systemen identifiziert werden konnten.

Standardnutzungsgrad: Die Erhebung des Standardnutzungsgrades identifizierte die Systeme, bei denen größtenteils die Standardfunktionalität genutzt wurde, ebenso wie die Systeme, bei denen hauptsächlich Eigenentwicklungen und Modifikationen zum Einsatz kamen. In einem zweiten Schritt sollen nun die nicht genutzten Eigenentwicklungen identifiziert und aus den Systemen entfernt werden. Genauso müssen nicht genutzte Standardpotentiale näher untersucht werden, um die vorhandene Funktionalität best möglichst auszunutzen.

Serverauslastung: Durch die erste Erhebung dieses KPIs konnten die Systeme, deren Serverauslastung nach Sizing-Richtlinien zu gering ausfiel, identifiziert werden. Weiter wurden zwei Systeme identifiziert, deren Auslastung gemäß Richtlinien als zu hoch eingestuft wurde. Für die erste Kategorie wurden Maßnahmen abgeleitet, die eine bessere Auslastungsmöglichkeit prüfen sollen. Für die zweite Kategorie wurden Maßnahmen zur Durchführung eines Hardware-Upgrades initialisiert.

6.3 Erfahrungen beim Einsatz des Kennzahlensystems im Unternehmen

Durch die ständige Kommunikation mit den SAP-Verantwortlichen der BMW Group konnte das anfangs theoretische Kennzahlensystem sukzessive für die Praxis adaptiert werden.

Hierzu waren mehrere Messdurchgänge erforderlich. Die Hauptkritik wurde an den Erhebungsmethodiken geäußert, die die Besonderheiten der Praxis teilweise nicht abbilden konnten. Zum Beispiel wurde bei der Ermittlung des Lizenznutzungsgrades anfänglich nicht berücksichtigt, dass ein Teil der SAP-User das System lediglich Quartalsweise nutzt. Bei Erhebung der Kennzahl wurden somit für die betroffenen Systeme zu niedrige Werte erzielt. Die größte Schwierigkeit lag in der Erhebung des Standardabdeckungs- und -nutzungsgrades. Bei der ursprünglichen Methodik wurden nur Eigenentwicklungen als Nicht-Standard klassifiziert, Modifikationen am Standard konnten nicht identifiziert werden. Da diese dem Standard zugerechnet wurden, ergab sich hier stets ein erhöhter Wert. Mittels eines Programms, welches im Rahmen dieser Forschungstätigkeit implementiert wurde, sind die SAP-Programme nun eindeutig als Standard, modifizierter Standard oder Eigenentwicklung klassifizierbar.

7 Zusammenfassung

Der vorliegende Beitrag stellt ein Kennzahlensystem vor, welches einem Unternehmen erlaubt, seine installierten SAP Systeme hinsichtlich Kosten, Nutzungsgrad, Standardisierungsgrad sowie Hardware-spezifischen Kriterien zu vergleichen. Auf diese Weise ist es das Ziel, Transparenz zu erzeugen und Optimierungspotentiale freizusetzen.

Ziel der vorliegenden Studie war es, sowohl Probleme der Praxis zu adressieren als auch einen Beitrag für die Wissenschaft zu leisten. Dies wurde erreicht, indem auf Basis von Literaturanalysen Problemfelder des Einsatzes betrieblicher Standardsoftware identifiziert wurden. Auf dieser Basis wurde ein Kennzahlensystem entwickelt, das nicht Unternehmens-spezifisch ist, sondern auf jedes beliebige Unternehmen, welches SAP Standardsoftware im Einsatz hat, anwendbar ist.

Durch Anwendung des Kennzahlensystems auf die Produktivsysteme eines internationalen Großunternehmens konnte das Modell hinsichtlich seiner Praxiseignung bestätigt werden. Dem Wunsch nach Transparenz in der SAP Landschaft wurde durch Anwendung des Modells näher gekommen. Darüber hinaus konnte durch das Kennzahlensystem die interne Diskussion initialisiert werden. Durch die Analyse jedes Ergebniswertes mit den entsprechenden Experten konnten erste Optimierungspotentiale in den SAP Systemen erkannt werden.

In der Diskussion zeigte sich allerdings auch, dass eine Erweiterung des Kennzahlensystems erforderlich ist, da noch nicht alle relevanten Faktoren eines SAP Systems adressiert wurden.

Dennoch weist das Kennzahlensystem auch Schwachstellen auf, z.B. werden Nutzenaspekte, die einem Unternehmen durch die Einführung eines SAP-Systems zukommen, im Kennzahlensystem nicht betrachtet. Da ein Total Cost of Ownership-Modell als Grundlage der Kostenkennzahlen dient, wird ein SAP-System mit hohen Kosten, das dem Unternehmen strategische Vorteile bringt, möglicherweise schlecht bewertet.

Aus diesem Grund wird für die anschließenden Forschungsaktivitäten die Erweiterung des Kennzahlensystems dahingehend adressieren, dass neben den Kosten auch Nutzenaspekte berücksichtigt werden. Ein erster Schritt kann in einer Erweiterung der TCO- auf eine Total Benefit of Ownership- Betrachtung (TBO) gesehen werden.

Literaturverzeichnis

- [ArRS05] Arnold, F., J.Röseler, M.Staade (2005). Enterprise Performance Management mit SAP. Bonn, Galileo Press GmbH.
- [Atte03] Atteslander, P. (2003). Methoden der empirischen Sozialforschung, de Gruyter Studienbuch.
- [BCG00] Boston Consulting Group, (2000). "Getting value from the enterprise initiatives, a survey of executives", Verfügbar unter: <http://www.bcg.com>.
- [BiTB00] Bititci, U. S., T.Turner, C.Begemann (2000). "Dynamics of performance measurement systems." International Journal of Operations & Production Management 20(6): 692-704.
- [BoDö02] Bortz, J., N.Döring (2002). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin, Heidelberg, Springer Verlag.
- [BuKö97] Buxmann, P., W.König (1997). "Empirische Ergebnisse zum Einsatz der betrieblichen Standardsoftware SAP R/3." Wirtschaftsinformatik 39(4): 331-338.

- [CIO06] CIO Online (2006). „Jedes dritte Programm wird kein einziges Mal aufgerufen“, Verfügbar unter: <http://www.cio.de/knowledgecenter/erp/820496/index.html>; abgerufen am 10.06.2006.
- [Eise98] Eisenhardt, K. M. (1998). “Building Theories form Case Study Research.” *Academy of Management Reviews* 14(4): 532-550.
- [GuBl] Gunson, J., J.-P. Blasis “The place and key success factors of enterprise resource planning (ERP) in the new paradigms of business management.”
- [HoLG98] Holland, C., B.Light, N.Gibson (1998). “Global ERP Implementation” in proceedings of the American Conference of Information Systems, Baltimore.
- [HoLi99] Holland, C. P., B. Light (1999). “A Critical Success Factors Model for ERP Implementation.” *IEEE Software*: 30-36.
- [IrLo02] Irani, Z., P.E.D.Love (2002). “Developing a frame of reference for ex-ante IT/IS investment evaluation.” *European Journal of Information Systems* 11: 74-82.
- [JLMM04] Jonen, A., V. Lingnau, J.Müller, P.Müller (2004). “Balanced IT-Decision-Card. Ein Instrument für das Investitionscontrolling von IT-Projekten.” *Wirtschaftsinformatik* 46(3): 196–203.
- [KaNo92] Kaplan, Robert S.; Norton, David P. (1992). The Balanced Scorecard Measures that drive performance. In: *Harvard Business Review* 70.
- [KenNe02] Kennerley, M., A.Neely (2002). “A framework of the factors affecting the evolution of performance measurement systems.” *International Journal of Operations & Production Management* 22(11): 1222-1245.
- [Krcm05] Krcmar, H. (2005). *Informationsmanagement*. Berlin, Heidelberg, New York, Springer Verlag.
- [Kütz03] Kütz, M. (2003). *Kennzahlen der IT*, Dpunkt Verlag.

- [KyYo02] Kyung-Kwon, H., K. Young-Gul (2002). "The critical success factors for ERP implementation: an organizational fit perspective." *Information & Management* 40: 25-40.
- [Lamn89] Lamnek, S. (1989). *Qualitative Sozialforschung*. München, Psychologie Verlags Union.
- [MaAP02] Markus, M. L., S. Axline, D. Pertrie et al. (2000). "Learning form adopters`experiences with ERP: problems encountered and success achieved." *Journal of Information Technology* 15: 245-265.
- [MuÖs99] Muschter, S., H.Österle (1999). *Investitionen in Standardsoftware: Ein geschäftsorientierter Ansatz zur Nutzenmessung und -bewertung*. Heidelberg, Physica-Verlag.
- [OlRo02] Oliver, D., C.Romm (2002). "Justifying enterprise resource planning adoption." *Journal of Information Technology* 17: 199-213.
- [SAP05a] SAP AG (2005). *A Model to analyze Total Cost of Ownership*.
- [SAP05b] SAP AG (2005). *Automotive Symposium 2005*.
- [SeRG01] Sedera, D., M.Rosemann, G.Gable (2001). Using performance measurement models for benefit realization with enterprise systes- the queensland government approach (case study). *Global Co-Operation in the New Millennium. The 9th European Conference on Information Systems, Bled, Slovenia*.
- [ShSS04] Shehab, E. M., M.W. Sharp, L. Supramaniam et al. (2004). "Enterprise resource planning, An integrative review." *Business Process Management Journal* 10(4): 359-386.
- [Stef01] Stefanou, C. J. (2001). "A framework for the ex-ante evaluation of ERP software." *European Journal of Information Systems* 10: 204–215.

- [ThIO] Themistocleous, M., Irani, Z., R.M. O`Keefe (2001). "ERP and application integration, Exploratory survey." *Business Process Management Journal* 7(3): 195-204.
- [TrTS03] Treber, U., P.Teipel, A.C.Schwickert (2003). *Total Cost of Ownership- Stand und Entwicklungstendenzen 2003*. Giessen, Justus-Liebig-Universität.
- [Wefe00] Wefers, M. (2000). "Strategische Unternehmensführung mit der IV-gestützten Balanced Scorecard." *Wirtschaftsinformatik* 42(2): 123-130.
- [WiHe] Wild, M.; S.Herges: *Total Cost of Ownership (TCO)- Ein Überblick*.
- [Yin94] Yin, R. K. (1994). *Case Study Research – Design and Methods*. Thousands Oaks, London, New Delhi, Sage Publications.
- [Zrim03] Zrimsek, B. (2003). *Enterprise Resource Planning TCO: More "How" than "What"*, Gartner Group.

Systematische Auswahl einer Methode zur Aufwandschätzung in der Softwareentwicklung

Ulrike Dowie, Georg Herzwurm

Lehrstuhl für Wirtschaftsinformatik (Unternehmenssoftware)

Universität Stuttgart

70174 Stuttgart

{dowie,herzwurm}@wi.uni-stuttgart.de

Abstract

Während sowohl in wissenschaftlichen Arbeiten als auch in der Praxis des Projektmanagements Einigkeit über die Wichtigkeit der Aufwandschätzung herrscht, ist die Diskrepanz bezüglich der Verbreitung verschiedener Methoden zwischen Wissenschaft und Praxis erheblich.

Dieser Beitrag schildert einen systematischen Ansatz zur Auswahl geeigneter Aufwandschätzmethoden in der Softwareentwicklung. Hierbei liefern wissenschaftliche Arbeiten Informationen über verfügbare Methoden sowie deren Stärken und Schwächen, während aus Sicht der Praxis Anforderungen formuliert werden, die an jegliche Aufwandschätzmethode gestellt werden.

Die Anwendbarkeit des Ansatzes wird anhand eines Beispiels erläutert. Gleichzeitig wird gezeigt, wie verschiedene Schätzmethoden integriert werden können.

1 Einleitung

In der Vielzahl wissenschaftlicher Veröffentlichungen, die sich mit Aufwandschätzmethoden befassen, gibt es kaum Anhaltspunkte, welche Methode in einer gegebenen Situation zu wählen ist [MySS05, 380]. Im Folgenden wird ein Ansatz erläutert, aus der Menge möglicher Schätzmethoden eine Auswahl zu treffen, die sowohl Bedürfnisse als auch Fähigkeiten einer softwareentwickelnden Organisation berücksichtigt [MaSh03, 1].

1.1 Bedeutung der Auswahl von Aufwandschätzmethoden

Aufwandschätzung ist als wichtige Aufgabe des Projektmanagers als Bestandteil der Projektsteuerung anerkannt [BuFa04]. Sie dient bspw. zur Ressourcenallokation, Zeitplanung, Angebotserstellung bei Auftragsentwicklung, sowie als Entscheidungsgrundlage beim Abwägen zwischen Qualitätszielen, Time to Market und Produktumfang [BuFa04].

Hierzu existiert eine Vielzahl von Methoden [vgl. zum Überblick Kitc96; MKLP00]. Häufige Widersprüche zwischen Ergebnissen verschiedener Untersuchungen zeigen, dass keine Methode als generell gut oder schlecht beurteilt werden kann [MaSh03; MySS05]. Die Eignung einer Schätzmethode ist offenbar von situativen Faktoren wie den Zielen und Rahmenbedingungen abhängig, die in einer konkreten Schätzsituation vorliegen.

In der Softwareentwicklungspraxis findet jedoch kaum eine Analyse dieser Faktoren und dementsprechende Auswahl geeigneter Schätzmethoden statt. Meist wird verwendet, wozu in der Organisation Wissen und Erfahrung verfügbar sind [Molø04].

Wissenschaftliche Arbeiten sind meist der Entwicklung und Untersuchung jeweils einer Methode gewidmet (z. B. [Lang04]) oder vergleichen einige ausgewählte Methoden hinsichtlich der Schätzgenauigkeit anhand einer bestimmten Datenbasis (von Autoren ermittelte [WaJe98] oder öffentlich verbreitete empirische Datenbestände [MKLP00; KCCS00]).

Sollen zusätzliche Anforderungen an die Methode, weitere Methoden oder andere bzw. eigene historische Daten berücksichtigt werden, sind derartige wissenschaftliche Erkenntnisse wenig hilfreich. Daher ist sowohl aus Sicht der Praxis als auch der Wissenschaft zu untersuchen, welche Schätzmethode in einer bestimmten Schätzsituation a priori auszuwählen ist [MaSh03].

1.2 Ziel und Vorgehensweise

Die Ziele dieses Beitrags sind folgende:

- Ein Konzept zur systematischen Auswahl geeigneter Aufwandschätzmethoden zu entwickeln
- Die Praktikabilität des Konzepts anhand eines Beispiels zu illustrieren
- Die Integration verschiedener Methoden aufzuzeigen (am Beispiel Case-based Reasoning und Expertenschätzung)

Das Vorgehen zur Methodenauswahl greift auf Elemente der Qualitätssicherungsmethode Quality Function Deployment (QFD)¹ zurück. Ziel von QFD ist es, durch Erfassung lösungsunabhängiger Anforderungen den Raum möglicher Lösungen zu erweitern, und durch die Gewichtung der Anforderungen die Lösungen zu priorisieren. Analog dazu soll die Auswahl einer Schätzmethode nachvollziehbar und auf Anforderungen gegründet sein und der Weg für neue, d. h. bisher ungenutzte Methoden geebnet werden.

Hierzu wird zunächst in Kapitel 2 ein Überblick über existierende Aufwandschätzmethoden, insbesondere deren Einsatzvoraussetzungen, und Verfahren zur Integration verschiedener Methoden gegeben. In Kapitel 3 wird dann der Ansatz zur Methodenauswahl erläutert. Dazu werden Annahmen getroffen, deren Erfüllung für die exemplarisch dargestellte Auswahlmatrix notwendig ist. Anschließend wird die Vorbereitung der Auswahl beschrieben, eine Übersicht über die Anforderungserfüllung einzelner Schätzmethoden gegeben und der Auswahlprozess geschildert. In Kapitel 4 wird das Vorgehen anhand eines Anwendungsbeispiels illustriert. Schließlich erfolgt in Kapitel 5 eine Diskussion der Ergebnisse.

2 Stand der Forschung

Aufwandschätzung von Softwareprojekten ist Untersuchungsgegenstand mehrerer Wissensgebiete: Softwareentwicklung, Forecasting und Projektmanagement. Fokus dieses Beitrags ist die Anwendung in der Softwareentwicklung.

2.1 Aufwandschätzmethoden in der Softwareentwicklung

Aufwandschätzmethoden können grob unterteilt werden in solche basierend auf Statistik (parametrische Modelle), Erfahrung bzw. Erfahrungsaufbau (Expertenschätzung) und Machine learning [BESW98, 2]. Zur weiteren Unterteilung finden sich unterschiedliche Ausführungen (z. B. gemäß der erforderlichen Menge historischer Projektdaten; zum Überblick vgl. [Molø04, 36]). Zur Erläuterung des Auswahlansatzes werden für jede Obergruppe Beispiele genannt² und anhand der Einsatzvoraussetzungen charakterisiert.³

¹ zu QFD siehe z. B. [Akao92; HeSM97]

² Nicht behandelt werden Methoden, die zwar als solche bezeichnet werden, aber streng genommen keine Schätzmethoden sind, z. B. „Price to win“ oder „Parkinson’sches Gesetz“ [Kitc96, 190].

³ Um das Auswahlvorgehen zu erläutern, bedarf es keiner lückenlosen Methodenübersicht. Da allerdings je nach Situation unterschiedliche Methoden zur Schätzung geeignet sein können, müssen derart unterschiedliche Methoden zur Auswahl berücksichtigt werden.

2.1.1 Statistische Methoden (parametrische Modelle)

Beispiele statistischer Methoden, die zur Modellierung im Rahmen der Aufwandschätzung eingesetzt werden, sind einfache, schrittweise oder multiple Regressionsanalyse [BEPW06], Analysis of Variance (ANOVA, [BLW99]), Classification and Regression Trees (CART, [BESW98]) und Optimal Subset Selection [BoFW74]. Das bekannteste parametrische Modell, das auf statistischen Methoden basiert, ist CoCoMo [Boeh00].

Voraussetzung für den Einsatz sind Kenntnisse oder begründete Vermutung über die Kausalzusammenhänge zwischen Einflüssen und Aufwand [BEPW06, 8]. Alle Parameter des Modells müssen in numerischer Form vorliegen oder in eine solche transformiert werden. Außerdem muss der Umfang der Datenbasis ermöglichen, statistisch signifikante Korrelationen zwischen Einflussfaktoren und Aufwand bzw. zwischen Einflussfaktoren untereinander zu bestimmen. Die Daten müssen je nach statistischem Verfahren unterschiedliche Eigenschaften aufweisen⁴ (z. B. standardnormal verteilt sein, keine Kollinearität besitzen, vgl. [BEPW06]).

2.1.2 Expertenschätzung

Die Schätzung durch einen oder eine Gruppe von Experten kann formalisiert (z. B. mittels Analytical Hierarchy Process oder Delphi-Methode) oder ad hoc und informell ablaufen. Ziel der Schätzung durch eine Gruppe ist es, durch Austausch und gegenseitiges Feedback unter den Teilnehmern einen Konsens bzgl. des Schätzwertes zu erreichen, in den die Erfahrung und jeweils berücksichtigte Einflüsse aller Teilnehmer eingehen. Voraussetzung des Einsatzes dieser Methoden ist die Erfahrung der Beteiligten mit den Aufgaben⁵, für die sie den Aufwand schätzen (für weitere Details zu Expertenschätzung s. z. B. [Jørg05]).

2.1.3 Machine learning

Hierzu zählen alle Methoden, die Fähigkeiten des menschlichen Verstands nachbilden [MKLP00]. Sie eignen sich insbesondere zur Lösung sehr komplexer Probleme und vermögen aus Erfahrung zu lernen [AaP194, 4; BEPW06, 14]. Bekanntestes Beispiel sind künstliche Neuronale Netze [Lang04]. Weitere Vertreter dieser Methodengruppe sind Case-based Reasoning [AaP194], Genetic Programming [Dola01] und Rule Induction [MKLP00]. Voraussetzungen für den Einsatz dieser Methoden sind eine numerisch vorliegende Datenbasis und die Konfigura-

⁴ Liegen die Eigenschaften bzw. Voraussetzungen nicht vor, ist die Anwendung des jeweiligen statistischen Verfahrens nicht zulässig (s. z. B. [BEPW06], S. 78 und S. 430)

⁵ Erfahrung wird hier verstanden als Durchführung einer Aktivität (bspw. Test) in mehreren vorangegangenen Projekten.

tion des jeweiligen Machine learning-Systems, d. h. Treffen einer Reihe von Aufbau- bzw. Design-Entscheidungen [BEPW, 763; KCCS00, 4].

2.2 Integration verschiedener Methoden

Zahlreiche Autoren kommen zum Schluss, dass in jedem Schätzvorgang verschiedene Methoden eingesetzt werden sollen: „Almost any combination of forecasts proves more accurate than the single inputs“ [BIHo90, 889; s. auch BuFa04, 33; Devn99; Good05; Kitc96; Jørg05; WaJe98]. Vielfach wird empfohlen, die Ergebnisse verschiedener Methoden zu vergleichen, ohne jedoch anzugeben, wie im Falle abweichender Ergebnisse vorzugehen ist [WaJe98, 25]. Kitchenham nennt dieses Vorgehen „validation by alternative estimates“ [Kitc96, 190]. Um Stärken verschiedener Methoden zu vereinen, müssen sie jedoch integriert werden anstatt die Ergebnisse zur Validierung zu vergleichen.

Integrationsansätze unter Beteiligung von Expertenschätzung lassen sich unterscheiden in mechanische / festgelegte und interaktive / freiwillige Kombination der Methoden [Good00].

Autoren, die *mechanische Kombination von Methoden* untersuchen, wenden mathematische Verknüpfungen der gleichzeitig gewonnenen Schätzergebnisse an. Blattberg und Hoch bspw. gewichteten Expertenschätzung und Ergebnisse statistischer Modelle gleich und addierten jeweils 50 % des Schätzwerts [BIHo90]. Boehm et al. integrierten Expertenschätzung und Regressionsanalyse bei der Entwicklung des Modells CoCoMo [Boeh00], indem eine Gruppe von Experten initiale Skalenwerte für die Einflussfaktoren des Modells festlegte. In [Devn99] wird beschrieben, wie mittels Anwendung des Bayes-Theorems Expertenschätzung und statistische Analyseergebnisse verknüpft und damit CoCoMo kalibriert wurden [Devn99, 37ff].

Diese Ansätze sind entweder pragmatischer Natur [BIHo90], d. h. es wurde nicht untersucht, ob eine andere Integrationsweise zu genaueren Schätzergebnissen führt. Oder sie sind mit hohem Kalibrierungsaufwand (zur Anpassung an Gegebenheiten der schätzenden Organisation) verbunden, damit ausreichend genaue Schätzergebnisse resultieren (s. z. B. [Kitc96]). Vorteile der mechanischen Integration, die vor Verfügbarkeit der Schätzergebnisse festgelegt wird, sind die objektive, nachvollziehbare Vorgehensweise, die Manipulation der Ergebnisse verhindert, und die Verminderung der Expertenschätzung inhärenten Inkonsistenz [Good05].

Interaktive Methodenintegration überlässt dem Schätzenden die Gewichtung des statistisch (oder durch ein Machine learning-System) ermittelten Schätzwerts [Good00, 86]. Bei der Schätzung kann der Experte entscheiden, ob und wie er das vorliegende Schätzergebnis verwendet.

Beispielsweise kann der von einem parametrischen Modell ermittelte Schätzwert von einem Experten genutzt werden, seine auf Analogieschluss basierende Schätzung zu korrigieren, indem er die Grenzwerte seines geschätzten Intervalls reduziert.

Nachteil der interaktiven Integration ist, dass Expertenschätzung durchgängig von übermäßigem Selbstvertrauen geprägt ist [GoWr04, 371]. Auch andere systematische Fehler menschlicher Wahrnehmung und menschlichen Urteilens [z. B. BoDö02, 182ff] können zum Tragen kommen [BIHo90]. Schließlich unterliegen Experten politischen Einflüssen innerhalb der Organisation und kombinieren Beobachtungen nicht konsistent [BIHo90, 889f]. Vorteil dieser Integrationsweise ist, dass sie eher akzeptiert wird, weil der Schätzende die Kontrolle über das endgültige Ergebnis der Schätzung behält [BIHo90]. Er kann zudem Veränderungen des Umfelds und besondere Ereignisse berücksichtigen, die aufgrund ihrer Seltenheit nicht modelliert werden. Außerdem hat er die Möglichkeit, bei Aufteilung der Schätzaufgabe zwischen Teilen zu unterscheiden, die durch ein Modell geschätzt werden können und solchen, die mittels Intuition besser schätzbar sind [Good00, 88]. Schließlich kann Schätzung durch einen Experten fehlende oder nicht verlässliche historische Daten kompensieren [Good05, 9].

3 Prozess zur systematischen Methodenauswahl für die Aufwandschätzung

Für die hier präsentierten Zusammenhänge zwischen Anforderungen und Schätzmethoden werden folgende Annahmen getroffen⁶:

Annahme 1: Daten für Modellierung mit statistischen Methoden oder zum Training von Machine learning-Systemen sind vorhanden.

Das bedeutet, dass für mehrere abgeschlossene Projekte verlässliche, vergleichbare Daten, möglichst verhältnisskaliert oder intervallskaliert, verfügbar sind.⁷

Ist Annahme 1 nicht erfüllt, können parametrische Modelle bzw. Machine learning-Systeme nur mit organisationsfremden Daten erstellt bzw. trainiert werden. Dadurch sind sie angesichts

⁶ Perspektive ist die einer softwareentwickelnden Organisation, die verschiedene Produkte als Standardsoftware weiterentwickelt, gelegentlich technisch überarbeitet und regelmäßig neue Produkte entwickelt, und in der bislang ohne Methode Aufwand geschätzt wird.

⁷ Die erforderliche Anzahl der Projekte richtet sich nach der Anzahl der unabhängigen Variablen, die modelliert werden sollen, d. h. nach der abzubildenden Komplexität. Je mehr unabhängige Variablen zur Erklärung der abhängigen Variablen, also des zu schätzenden Aufwands, heran gezogen werden sollen, umso mehr Projekte müssen analysiert werden.

bestimmter Anforderungen weniger zum Einsatz geeignet (z. B. sind Akzeptanz und Schätzgenauigkeit in geringerem Ausmaß gewährleistet [Kitc96]).

Annahme 2: In der Organisation wird zum Zeitpunkt der Auswahl keine formalisierte Methode eingesetzt.

Diese Annahme besagt, dass z. B. keine Kalibrierung des algorithmischen Schätzmodells CoCoMo stattgefunden hat, und dass kein Machine learning-System konfiguriert ist.

Ist Annahme 2 nicht erfüllt, weisen bereits im Einsatz befindliche formalisierte Methoden andere Werte bezüglich der Anforderungserfüllung auf (insbesondere bezogen auf die Anforderung „Ergebnis aufwandsarm verfügbar“).

Sind Annahme 1 oder 2 nicht erfüllt, müssen die Erfüllungsgrade des Templates (s. Tabelle 1) neu bestimmt werden (für bereits eingesetzte Methoden).

3.1 Vorbereitung der Methodenauswahl

Im ersten Schritt werden die Ziele der Schätzung festgehalten. Unter anderem ist zu ermitteln, für wen und welchen Zweck Aufwand geschätzt werden soll.

Im zweiten Schritt werden verfügbare Zeit, Ressourcen und Budget für die Schätzung festgelegt. Die Ergebnisse dieser zwei Schritte, Ziele und Rahmenbedingungen der Schätzung, repräsentieren wesentliche Anforderungen.

Anschließend müssen die Stakeholder identifiziert werden, deren Interessen bei der Auswahl zu berücksichtigen sind (z. B. Budgetverantwortliche, oberes Management). Schließlich werden die Anforderungen⁸ aller Stakeholder an die Schätzmethode und verfügbare Methoden gesammelt bzw. die im Template vorhandenen Listen überarbeitet (s. Tabelle 1). Die Anforderungen werden in den Zeilen, die Schätzmethoden in den Spalten eingetragen⁹.

3.2 Anforderungserfüllung der Schätzmethoden

Jede Schätzmethode trägt zur Erfüllung jeder Anforderung abhängig von ihren Stärken und Schwächen in unterschiedlichem Ausmaß bei. Die Skala der Anforderungserfüllung

⁸ Die Anforderungsermittlung ist für die Methodenauswahl entscheidend. Zum Vorgehen s. [HeSM97], S. 80ff.

⁹ Dies folgt dem Aufbau des House of Quality, in dem Anforderungen den Lösungsmerkmalen gegenübergestellt werden, um die Erfüllung der Anforderungen je Lösungsmerkmal zu untersuchen, vgl. [HeSM97], S. 43f.

(Erfüllungsgrade) reicht von keinem Beitrag zur Erfüllung (mit Wert 0 versehen) über geringe (1) und mittlere Erfüllung (3) bis zu sehr hoher Erfüllung der Anforderung (Wert 9).¹⁰

Tabelle 1 gibt die analysierten Erfüllungsgrade wieder und kann als Template verwendet werden. Enthält eine Zelle mehrere Erfüllungsgrade, kommt es auf das Vorliegen bestimmter Bedingungen an, welcher Wert für eine konkrete Anwendungssituation zutreffend ist. Die notwendigen Bedingungen sind mit Klein- und Großbuchstaben gekennzeichnet und in der Legende erklärt. Die Begründung der ermittelten Erfüllungsgrade wird im Anhang beispielhaft anhand der Anforderung „Ergebnis aufwandsarm verfügbar“ erläutert.

Anforderungen: Ziele u. Rahmenbedingungen d. Schätzung	Schätzmethoden	Schätzmethoden												
		parametr. Modell (Regr. analyse)	parametr. Modell (optimal subset sel.)	CoCoMo	Classification and Regression Trees	Analogieschluss Informell	indiv. Expertenschätzung: ad hoc	indiv. Expertenschätzung: formalisiert	Gruppe von Experten: ad hoc	Gruppe von Experten: formalisiert	künstl. Neuronale Netze	Rule Induction	Genetische Programmierung	Case-based reasoning
Ergebnis aufwandsarm verfügbar	0%	1	1	3	3	9	9	9	3	3	1	1	3	3
Nachvollziehbar	0%	9 ^{dk}	9	9	9	9	1	3	1	3	0	9	1	9
Hohe Schätzgenauigkeit und Treffsicherheit	0%	0 ^{4,4,4,4} 1 ^{0,ST} 3 ^{0,M,ST}	1 ^{4,4,4} 3 ^{0,ST} 9 ^{0,M,ST}	0 ^{4,4,4} 1 ^{0,ST} 3 ^{0,M,ST}	1 ^{4,4,4} 3 ^{0,ST} 9 ^{0,M,ST}	1 ^{4,4} 3 ⁴ 9 ⁴	1 ^{4,4} 3 ⁴ 9 ⁴	1 ^{4,4} 3 ⁴ 9 ⁴	3 ^R 9 ^{R,SE} 9 ^{Kat,R,SE}	3 ^R 9 ^{R,SE} 9 ^{Kat,SE,VE}	3 ^R 9 ^{R,SE} 9 ^{Kat,SE,VE}	1 ^{0,ST,V} 3 ^{0,ST,V} 9 ^{0,ST,V}	1 ^{0,ST,V} 3 ^{0,ST,V} 9 ^{0,ST,V}	1 ^{0,ST,V} 3 ^{0,ST,V} 9 ^{0,ST,V}
Akzeptanz durch oberes Management	0%	9	9	9	1	3	1	3	3	3	3	1	1	1
Akzeptanz durch Projektmitarbeiter	0%	3	3	3	3	3	3	3	9	9	9	1	1	1
Ohne Verwendung historischer Daten	0%	0	0	1	0	0	0	9	9	9	0	0	0	0
Kein Domänenwissen erforderlich	0%	3	3	9	3	1	0	0	0	0	9	3	9	3
leicht anpassbar (an neue Technologie, Entw.prozess, ...)	0%	1	1	1	1	3	9	9	9	9	1	1	1	1
Ohne Zusatzkosten f. Softwarelizenz	0%	1	0	1	0	9	9	9	9	9	0	0	0	0
Kein statistisches Wissen erforderlich	0%	0	0	1	1	9	9	9	9	9	9	9	9	9
Automatisierbar	0%	9	9	9	9	3	0	1	0	1	9	9	9	9
Früh im Entwicklungsprozess verwendbar	0%	3 ^M	3 ^M	3 ^M	3	9	9	9	9	9	3	3	3	9
Leicht manipulierbar	0%	1	0	0	3	9	9	3	9	3	1	3	3	9
Objektiv (intersubjektiv durchführbar)	0%	9	9	9	9	0	0	0	1	3	9	9	9	3
Jederzeit wiederholbar m. möglicher Verbesserung	0%	1 ^M	1	1	1 ^M	9	9	9	9	9	1 ^M	1 ^M	1 ^M	9
Berücksichtigung einer Vielzahl v. Aufwandseinflüssen	0%	3	3	9	3	9	9	9	9	9	3	1	9	9
Unabhängig v. Test- o. Entwicklungsmethodik, -prozess, ...	0%	0	0	0	9	9	3 ^{EU}	3 ^{EU}	3 ^{EU}	3 ^{EU}	9	9	9	9
Absolute Bedeutung														
Relative Bedeutung														
Rangfolge														

LEGENDE

0/1/3/9 keine / geringe / mittlere / sehr hohe Erfüllung
a / A / ka mind. ein ähnliches / sehr ähnliches abgeschlossenes Projekte existiert / kein ähnliches Projekt existiert
D / d hohe / niedrige Datenqualität (hoch. Prämissen stat. Methoden erfüllt, große Anz. Beobachtungen, keine fehl. Werte, verlässl. Daten)
dok Vorgehen und Ergebnisse verständlich dokumentiert
EU Erfahrung mit Entw. umgebung (Technologie, Entw. prozess, fachl. Anforderungen) liegt vor
Kstat / Kex Kombination mit stat. Methode / mit Expertenschätzung
M / m schätzrelevante Merkmalsausprägungen des zu schätzenden Projekts frühzeitig (Planungsphase) bekannt / nicht bekannt
R Risikoanalyse ist Teil d. Schätzung
SE / kSE Schätzerfahrung liegt vor / nicht vor
ST / st hohe / geringe Stabilität des Entw. Umfelds (Technologie, Entw. prozess, fachl. Anforderungen)
V Vielzahl von Einflussfaktoren erfasst
vE verschiedene Erfahrungshintergründe liegen vor

Tabelle 1: Anforderungen (ungewichtet), Schätzmethoden und Erfüllungsgrade

3.3 Vorgehen der Methodenauswahl

Die folgenden Schritte sind jeweils zur Auswahl geeigneter Methoden zu durchlaufen. Basis ist das Template (s. Tabelle 1), das jeweils ergänzt und überarbeitet wird.

1. Auswahl bzw. Ergänzung und Gewichtung der Anforderungen

¹⁰ Über die vier Bewertungsstufen keine, schwache, mittelmäßige und starke Anforderungserfüllung hinaus sind feinere Unterteilungen nur angemessen, wenn sie gerechtfertigt werden können. Gemäß allgemeiner Konvention in der Anwendung von QFD werden diesen Erfüllungsgraden die Werte 0, 1, 3 und 9 zugeordnet [HeSM97, 115].

Die in der jeweiligen Schätzsituation relevanten Anforderungen (neben Zielen und Rahmenbedingungen) müssen identifiziert und anschließend unter Berücksichtigung aller Stakeholder (s. Kapitel 3.1) gewichtet werden.¹¹ Die Gewichte werden neben den Anforderungen im Template erfasst.

2. Ergänzung und gegebenenfalls Auswahl verfügbarer Schätzmethoden

Es wird überprüft, ob die vorliegende Sammlung um (neue) Schätzverfahren ergänzt werden muss. Sind bestimmte Schätzmethoden nicht verfügbar (z. B. statistische Methoden, wenn keinerlei statistisches Wissen vorliegt oder kurzfristig erworben werden kann), können sie von der Auswahl ausgeschlossen werden.

3. Ermittlung der Bedeutung (absolut und relativ) und Rangfolge der Schätzmethoden

Mittels Gewichtung der Anforderungen und Ausmaß der Anforderungserfüllung lassen sich die absolute Bedeutung, die relative Bedeutung und davon abgeleitet die Rangfolge der Schätzmethoden ermitteln. Dazu werden für jede Schätzmethode (in Spalten in Tab. 1) die Erfüllungsgrade mit dem relativen Gewicht der jeweiligen Anforderung multipliziert und aufaddiert. Das Ergebnis ist die absolute Bedeutung der Schätzmethode. Die relative Bedeutung ergibt sich nach Division jedes einzelnen Absolutwerts durch die Summe der absoluten Bedeutungen. Entsprechend den relativen Bedeutungen wird die Rangfolge der Schätzmethoden ermittelt.¹²

4. Auswahl der Schätzmethoden

Die systematische Auswahl der Schätzmethoden folgt deren Rangfolge. Allerdings wird nicht eine Methode gewählt, sondern die zwei oder drei ranghöchsten Methoden. Die Anzahl hängt vom verfügbaren Wissen zur jeweiligen Methode und von der Basis der Methoden ab (Statistik, Expertenschätzung, Machine learning). Sie sollten sich stark unterscheiden, damit Zusatznutzen von mehrfacher Schätzung entstehen kann. Dies folgt den Ergebnissen der Mehrzahl wissenschaftlicher Arbeiten, nach denen eine Kombination verschiedener Methoden einzelnen Methoden überlegen ist (vgl. Kapitel 2.2).

Ohne Kenntnisse einer Methode erreichen damit gewonnene Schätzergebnisse deutlich geringere Treffsicherheit und Genauigkeit, wie experimentell belegt wurde [WaJe98, 26].

¹¹ Um die Gewichtung bei sehr vielen Anforderungen zu vereinfachen, werden die Anforderungen zuerst gruppiert und die Gruppen mit Gewichten versehen. Im nächsten Schritt werden die Anforderungen jeder einzelnen Gruppe gewichtet. Multipliziert man die Einzelgewichtungen mit dem jeweiligen Gruppengewicht, erhält man die absolute Gewichtung jeder Anforderung. Um relative Gewichte zu erhalten, wird jedes absolute Gewicht durch die Summe aller absoluten Gewichte dividiert.

¹² Ein anschauliches Beispiel zur Ermittlung der Bedeutungen findet sich bei [HeSM97, 41f].

Unterschiedliche Fundamente der ausgewählten Methoden sind wichtig, um die Validierung der Ergebnisse zu ermöglichen [Kitc96]. Außerdem können nur unter dieser Voraussetzung die Stärken der einzelnen Methoden bei gleichzeitiger Vermeidung der Schwächen genutzt werden.

4 Anwendungsbeispiel

Das beschriebene Beispiel ist fiktiv, jedoch so realitätsnah wie möglich gewählt. Die Form der Fallstudie zur Überprüfung des Auswahlansatzes eignet sich, weil das „Wie“ der Schätzung interessiert, ohne dass der Forscher die Vielzahl von Einflussfaktoren messen oder steuern kann, denen der Untersuchungsgegenstand (die Schätzmethodeauswahl) unterliegt [Yin03, 8f].

In einem softwareentwickelnden Unternehmen soll die Aufwandschätzung methodisch untermauert werden, um nachvollziehbare, verlässliche Schätzungen zu ermöglichen. Ein Mitarbeiter der Qualitätssicherungsabteilung wird mit der Vorbereitung der Auswahl betraut.

Als Stakeholder werden die Projektleiter identifiziert, die für die Aufwandschätzung verantwortlich sind, Entwickler und Tester, für die die Schätzergebnisse Charakter von Zielvorgaben haben, die Managementebene, die Projektbudgets genehmigt, sowie die Qualitätssicherungsabteilung, die für den Entwicklungsprozess verantwortlich ist und die Kosten der Auswahl trägt.

Mehrere Vertreter jeder dieser Gruppen werden per E-Mail oder telefonisch nach ihren Anforderungen an eine Schätzmethode gefragt. Dazu wird ihnen die Liste von Anforderungen des Templates gegeben, die sie ergänzen bzw. reduzieren können.

Um die Anforderungen zu konsolidieren und zu gewichten, wird je ein Vertreter zu einer Sitzung eingeladen. Nicht teilnehmende, vorher Befragte erhalten die Ergebnisse zugeschickt und werden gebeten, begründete Einwände mitzuteilen. Aufgrund mehrerer Einsprüche werden die ursprünglich geplanten Teilnehmer zu einer zweiten Sitzung eingeladen. Jeder Teilnehmer (drei Projektleiter, zwei Mitarbeiter der Qualitätssicherung, ein Vertreter des Managements mit Verantwortung für Budgetplanung) erhält gleiches Gewicht der Stimme zur Anforderungsgewichtung. Die resultierende Liste gewichteter Anforderungen ist in Tabelle 2 dargestellt.

Die im Template enthaltenen Erfüllungsgrade werden mit einem externen Aufwandschätzexperten überprüft. Insbesondere die Erfüllungsgrade der Anforderung „Hohe Schätzgenauigkeit und Treffsicherheit“ werden entsprechend der Situation, in der sich die Mehrzahl der Projekte befindet, angepasst. Das Ergebnis ist ebenfalls in Tabelle 2 abgebildet.

Anforderungen: Ziele und Rahmenbedingungen der Schätzung	Schätzmethoden	parametrisches Modell basierend auf schrittweiser R	parametrisches Modell basierend auf optimal subset	CoCoMo	CART (Classification and Regression Trees)	Analogieschluss informell	individuelle Expertenschätzung: ad hoc	individuelle Expertenschätzung: formalisiertes Vorgehen	Gruppe von Experten: ad hoc	Gruppe von Experten: formalisiertes Vorgehen	künstl. Neuronale Netze	Rule Induction	Genetic Programming	Case-based reasoning
Ergebnis aufwandsarm verfügbar	6%	1	1	3	3	9	9	9	9	3	1	1	3	3
Nachvollziehbar	3%	9	9	9	9	9	1	3	1	3	0	9	1	3
Hohe Schätzgenauigkeit und Treffsicherheit	15%	1	3	1	3	3	1	3	9	9	1	1	1	3
Akzeptanz durch oberes Management	15%	9	9	9	1	3	1	3	3	3	3	3	3	3
Akzeptanz durch Projektmitarbeiter (Entwickler, Tester,...) -> akzeptierte	15%	3	3	3	3	3	3	3	9	9	1	1	1	3
Ohne Verwendung historischer Daten	0%	0	0	1	0	0	9	9	9	9	0	0	0	3
Kein Domänenwissen erforderlich	0%	3	3	9	3	1	0	0	0	0	9	3	9	3
leicht anpassbar (an neue Technologie, Entw.umgebung, Entw.prozess, fachl. Hintergrund,...)	8%	1	1	1	1	3	9	9	9	9	1	1	1	1
Ohne Zusatzkosten f. Softwarelizenz	0%	1	0	1	0	9	9	9	9	9	0	0	0	0
Kein statistisches Wissen erforderlich	0%	0	0	1	1	9	9	9	9	9	9	9	9	9
Automatisierbar	6%	9	9	9	9	3	0	1	0	1	9	9	9	9
Früh im Entwicklungsprozess verwendbar	20%	3	3	1	3	9	9	9	9	9	3	3	3	3
Leicht manipulierbar	0%	1	0	0	3	9	9	3	9	3	1	3	3	3
Objektiv (verschiedene Personen müssen zum gleichen Ergebnis kommen)	3%	9	9	9	9	0	0	0	1	3	9	9	9	3
Wiederholbar mit möglicher Verbesserung (jederzeit im	5%	1	1	1	1	9	9	9	9	9	1	1	1	9
Berücksichtigung einer Vielzahl von Aufwandseinflüssen	5%	3	3	9	3	9	9	9	9	9	9	3	3	9
Unabhängig einsetzbar (von Test- oder Entwicklungsmethodik, Test- oder Entwicklungsprozess, Entwicklungssprache, Entwicklungsarchitektur,...)	0%	0	0	0	9	9	3	3	3	3	9	9	9	9
Absolute Bedeutung		3,9	4,2	3,9	3,1	5,2	4,7	5,4	7,2	7,0	2,3	2,2	2,2	7,0
Relative Bedeutung		6,7	7,2	6,7	5,3	9,0	8,1	9,4	12,3	12,0	4,0	3,8	3,7	11,9
Rangfolge		9	7	8	10	5	6	4	1	2	11	12	13	3

Tabelle 2: Anwendungsbeispiel: gewichtete Anforderungen, angepasste Erfüllungsgrade

Da „Schätzung durch eine Expertengruppe ad hoc“ und „Schätzung durch eine Expertengruppe: formalisiert“ das gleiche Fundament aufweisen, ist wenig Zusatznutzen aus dem Einsatz beider Methoden zu erwarten. Eine Differenzanalyse wird durchgeführt, um die Methode zu identifizieren, die wichtigeren Anforderungen möglichst deutlich besser erfüllt.

Die Methoden „Schätzung durch eine Expertengruppe ad hoc“ und „Case-based Reasoning“ werden ausgewählt. Zum einen ist bei Betrachtung der relativen Bedeutung ein deutlicher Abstand zur rangnächsten Methode erkennbar. Zum anderen weisen die ausgewählten Methoden komplementäre Fundamente auf (Expertenschätzung und Machine learning), weshalb sie sich zur Integration eignen (zu dieser speziellen Kombination siehe [Kitc96]).

Die Integration der Methoden erfolgt interaktiv. Dies liegt zum einen an mangelnder Erfahrung mit Case-based Reasoning, zum anderen an den häufigen Änderungen und Unvorhersehbarkeit des Entwicklungsumfelds.

Die Konfiguration des Case-based Reasoning-Tools setzt Erfahrung mit Aufwandschätzung voraus. Der Qualitätssicherungsmitarbeiter wird hierbei von zwei erfahrenen Projektleitern

unterstützt. Gemeinsam treffen sie Entscheidungen zur Auswahl der Projektmerkmale, Skalierung der Merkmale, Bestimmung des Ähnlichkeitsmaßes, Anzahl zu verwendender Analogieobjekte und Anpassung des Aufwands der Analogieobjekte [KCCS00, 4].

Mit dem konfigurierten Case-based Reasoning-Tool wird in zwei Pilotprojekten der Aufwand geschätzt. Für das eine finden sich drei sehr ähnliche Vergleichsprojekte, deren Aufwand nur gering streut. Für das andere Pilotprojekt, dessen Umfeld von wesentlich größerer Unsicherheit geprägt ist, lässt sich nur ein Analogieprojekt finden, das sich in einigen Merkmalen stark vom zu schätzenden Pilotprojekt unterscheidet. Die Ergebnisse – ein relativ schmales und ein breites Aufwandsintervall – werden jeweils einer Gruppe von Experten vorgelegt, bestehend aus Projektleiter, einem Entwickler, Designarchitekten und dem Qualitätsverantwortlichen des Projekts. Die Experten nutzen das Ergebnis als Basis ihrer ad hoc-Schätzung, die in einer kurzen Besprechung diskutiert und gemeinsam bestimmt wird. Im Fall des schmalen Aufwandsintervalls wird die Intervallbreite beibehalten und nach oben korrigiert, im anderen Fall wird ein weniger breites Intervall am unteren Rand des ursprünglichen Werts geschätzt.

Insgesamt erfordert das Vorgehen zur Auswahl 50 Arbeitsstunden für Identifikation der Stakeholder und Auswahl der Teilnehmer, Versenden und Auswertung der Anforderungsbefragung, zwei Sitzungen zur Festlegung und Gewichtung der Anforderungen, Überarbeitung des Templates und Ermittlung der auszuwählenden Methoden.

5 Diskussion und Ausblick

Der Aufwand zur Auswahl geeigneter Schätzmethoden mag vor allem dem Praktiker hoch erscheinen. Außerdem sind verschiedene Experten am Prozess zu beteiligen, deren Verfügbarkeit sichergestellt werden muss. Und schließlich kann trotz aller Sorgfalt bei der Methodenauswahl jeder Schätzvorgang kein verlässlicheres Ergebnis liefern als die Verlässlichkeit jedes seiner Teilprozesse, und wird „so unzuverlässig wie der schwächste Teilprozess“ sein [BuFa04, 31]. Die beschriebene, formalisierte Vorgehensweise weist allerdings der ad hoc-Auswahl gegenüber einige Vorteile auf:

- Beteiligung aller Stakeholder schafft Verantwortungsgefühl und erhöht die Akzeptanz der Ergebnisse.
- Das Bewusstsein für die Bedeutung der Aufwandsschätzung wird geschärft.

- Kosten für ungeeignete Schätzverfahren werden gespart (z. B. Lizenzgebühren für Software zur Entwicklung künstlicher Neuronaler Netze).
- Gemeinsames Verständnis dessen kann erzielt werden, was ein Schätzverfahren leisten kann (z. B. an Genauigkeit) und welche Voraussetzungen dafür vorliegen müssen.

Ist die Auswahl geeigneter Schätzmethoden für ein Projekt getroffen, kann das Ergebnis in der gleichen Organisation mit geringem Überarbeitungsaufwand (bzw. unter gleichen Bedingungen ohne Anpassung) erneut verwendet werden. Unterstellt man außerdem eine Lernkurve den Auswahlprozess betreffend, kann die zweite Auswahl deutlich schneller ablaufen.

Die praktische Überprüfung des Ansatzes steht noch aus, ist jedoch schon in Vorbereitung. Mit den Ergebnissen der Aufwandschätzung, die unter Verwendung der ausgewählten Methoden stattfindet, soll untersucht werden, ob und inwieweit die systematisch und begründet ausgewählten Methoden zur Verbesserung des Schätzprozesses beitragen können.

Fundament realistischer Schätzung sind historische Daten, deren Qualität (Korrektheit und Validität) und Angemessenheit [BESW98, 8]. Weitere Grenzen sind der Anwendbarkeit dadurch gesetzt, dass Kenntnisse der verschiedenen Methoden vorliegen und Experten (bezogen auf Entwicklung und Schätzung der einzelnen Entwicklungsaufgaben) verfügbar sein müssen.

Wenn der vorgestellte Ansatz neben der Auswahl angemessene Datenerfassung und Projektnachbereitung unterstützt, kann die Aufwandschätzung vom Risikofaktor zur Risikomanagementtechnik [Boeh89, 117] werden und zum Projekterfolg beitragen.

Literaturverzeichnis

- [AaPl94] Aamodt, A.; Plaza, E.: Case-based reasoning: foundational issues, methodological variations, system approaches. In: AI Communications, 7 (1994) 1, S. 39-59.
- [Akao92] Akao, Yoji: Quality Function Deployment. Moderne Industrie, Landsberg 1992.
- [BEPW06] Backhaus, Klaus; Erichson, Bernd; Plinke, Wulff; Weiber, Rolf (Hrsg.): Multivariate Analysemethoden. 11. Aufl., Springer, Berlin / Heidelberg 2006.
- [BESW98] Briand, Lionel C.; El Emam, Khaled; Surmann, Dagmar; Wieczorek, Isabella; Maxwell, Katrina D.: An Assessment and Comparison of Common Software Cost Estimation Modeling Techniques. Technical Report ISERN-98-27

- [BlHo90] Blattberg, Robert C.; Hoch, Stephen: Database Models and Managerial intuition: 50% Model + 50% Manager. In: Management Science, 36 (1990) 8, S. 887-899.
- [BoDö02] Bortz, Jürgen; Döring, Nicola: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 3. Aufl., Springer, Heidelberg 2002.
- [Boeh00] Boehm, B. W.; Abts, C.; Horowitz, E.; Madachy, R.; Reifer, D.; Clark, B.K.; Steece, B.; Brown, A.W.; Chulani, S.: Software cost estimation with Cocomo II. Prentice Hall, Upper Saddle River (NJ/USA) 2000.
- [Boeh89] Boehm, Barry W.: Software Risk Management. IEEE Computer Society Press, Washington 1989.
- [BoFW74] Boyce, D.E.; Farhi; Weischedel, R.: Optimal Subset Selection, Multiple Regression, Interdependence and Optimal Network Algorithms. In: Beckmann, M.; Künzi, H.P. (Hrsg.): Lecture Notes in Economics and Mathematical Systems. Operations Research. Vol. 103. Springer-Verlag, Berlin et al. 1974.
- [BuFa04] Bundschuh, Manfred, Fabry, Axel: Aufwandschätzung von IT-Projekten, 2. Aufl., mitp-Verlag, Bonn 2004.
- [Devn99] Devnani-Chulani, Sunita: Bayesian Analysis of Software Cost and Quality Models. Dissertation, University of Southern California, Berkeley 1999.
- [Dola01] Dolado, J.J.: On the problem of the software cost function. In: Information and Software Technology, 43 (2001) 1, S. 61-72.
- [Good00] Goodwin, Paul: Improving the voluntary integration of statistical forecasts and judgment. In: International Journal of Forecasting, 16 (2000), S. 85-99.
- [Good05] Goodwin, Paul: How to integrate Management Judgment with Statistical Forecasts. In: Foresight 1 (2005) 1, S. 8-12.
- [GoWr04] Goodwin, Paul; Wright, George: Decision Analysis for Management Judgment. 3. Aufl., John Wiley & Sons, Chichester 2004.

- [HeSM97] Herzwurm, Georg; Schockert, Sixten; Mellis, Werner: Qualitätssoftware durch Kundenorientierung. Die Methode Quality Function Deployment (QFD): Grundlagen, Praxis und SAP R/3 Fallbeispiel. Vieweg, Braunschweig 1997.
- [Jørg05] Jørgensen, Magne: Practical Guidelines for Expert-Judgment-Based Software Effort Estimation. In: IEEE Software, 22 (2005) 3, S. 57-63
- [KCCS00] Kadoda, Gada; Cartwright, Michelle; Chen, Liguang; Shepperd, Martin: Experiences using case-based reasoning to predict software project effort. ESERG Technical Report No. 00-09, Bournemouth University, 2000.
- [Kite96] Kitchenham, Barbara: Software Metrics: Measurement for Software Process Improvement. Blackwell Publishers, Cambridge (USA) 1996.
- [Lang04] Lange, Carsten: Neuronale Netze in der wirtschaftswissenschaftlichen Prognose und Modellgenerierung. Physika-Verlag, Heidelberg 2004.
- [MaSh03] MacDonell, Stephen G.; Shepperd, Martin J.: Combining techniques to optimize effort predictions in software project management. In: Journal of Systems and Software, 66 (2003) 2, S. 91–98.
- [MKLP00] Mair, Carolyn; Kadoda, Gada; Lefley, Martin; Phalp, Keith; Schofield, Chris; Shepperd, Martin: An investigation of machine learning based prediction systems. In: Journal of Systems and Software, 53 (2000) 1, S. 23-29.
- [Molø04] Moløkken-Østvold, K: Eff.ort and schedule estimation of software development projects. Ph.D. Thesis, Dep. of Informatics, University of Oslo, 2004.
- [MySS05] Myrtveit, Ingunn; Stensrud, Erik; Shepperd, Martin: Reliability and Validity in Comparative Studies of Software Prediction Models. In: IEEE Transactions on Software Engineering, 31 (2005) 5, S. 380-391
- [WaJe98] Walkerden, Fiona; Jeffery, Ross: An Empirical Study of Analogy-based Software Effort Estimation. Technical Report 98/8, University of New South Wales, Sydney 1998.
- [Yin03] Yin, Robert: Case study research. 3. Aufl., Sage, Thousand Oaks (2003)

6 Anhang

Bestimmung der Erfüllungsgrade am Beispiel „Ergebnis aufwandsarm verfügbar“ (vgl. Tabelle 1 und Tabelle 2):

Methoden basierend auf statistischen Verfahren

Die Entwicklung eines parametrischen Modells basierend auf schrittweiser Regressionsanalyse erfordert verlässliche Daten abgeschlossener Projekte und ist trotz erhältlicher Toolunterstützung mit erheblichem Modellierungsaufwand verbunden (z. B. Auswahl der Gleichungsform, vorausgehende Datenanalyse).¹³ Daher erfüllt schrittweise Regressionsanalyse die Anforderung in sehr geringem Maße (Wert 1).

Die Methode „optimal subset selection“ liefert bessere Modellierungsergebnisse als einfache oder schrittweise Regression [BoFW74, 1], ist aber mit vergleichbarem Aufwand zur Modellierung verbunden und seltener von kommerzieller Statistiksoftware unterstützt, weshalb die Erfüllung der Anforderung „Ergebnis aufwandsarm verfügbar“ ebenfalls gering ist (1).

Das Schätzmodell CoCoMo enthält Standardwerte für Koeffizienten und Konstanten der Schätzgleichung, sodass prinzipiell zur Anwendung nur die Werte der modellierten Variablen des zu schätzenden Projekts gemessen und in die Gleichung eingesetzt werden müssen. Allerdings muss das Modell für genaue Schätzergebnisse mit lokalen Projektdaten kalibriert werden, sodass die Anforderung „Ergebnis aufwandsarm verfügbar“ nur mittelmäßig erfüllt wird (3).

Bei Schätzung mittels eines „Classification and Regression Trees“ gibt es unterstützende Softwaretools, doch fällt erheblicher Aufwand zur Datenerfassung an, da wie für alle statistischen Ansätze mehr Daten vorliegen (und eingegeben werden) müssen, je mehr unabhängige Variablen zur Schätzung verwendet werden (bzw. je mehr Verzweigungen der Baum aufweist). Außerdem muss festgelegt werden, welche Projektmerkmale die zu schätzende Größe beeinflussen und demnach zur Bildung des Regressionsbaumes verwendet werden, und unter welcher Bedingung ein Zweigende erreicht ist. Insgesamt wird die Anforderung „Ergebnis aufwandsarm verfügbar“ nur mittelmäßig erfüllt (3).

Expertenschätzmethoden

Bei Verwendung der Methode „Analogieschluss informell“ muss sich der Experte vergleichbare Projekte in Erinnerung rufen. Weder Vergleichsdaten noch Vorgehen sind zu beschreiben,

¹³ Selbst wenn bereits ein solches Modell mit firmeneigenen Daten entwickelt wurde, muss es regelmäßig überprüft und gegebenenfalls aktualisiert werden (z. B. bei Veränderung des Entwicklungsprozesses). Ein Schätzergebnis ist nur in den seltenen Fällen aufwandsarm verfügbar, in denen das Modell kurz zuvor mit Daten neuer Projekte überprüft und aktualisiert wurde.

weshalb der Zeitaufwand zwar abhängig von der Bearbeitungszeit des Schätzers, jedoch absolut betrachtet gering ist. Der Grad der Anforderungserfüllung ist demnach sehr hoch (9).

Wird als Schätzmethode „Individuelle Expertenschätzung: ad hoc“ gewählt, ist ebenfalls keine Dokumentation erforderlich. Der Schätzaufwand hängt von der Arbeitsgeschwindigkeit des Experten bei dieser Aufgabe ab. Das Ergebnis ist relativ schnell und aufwandsarm verfügbar, sodass die Anforderungserfüllung dieser Methode als sehr hoch beurteilt wird (9).

Läuft die „individuelle Expertenschätzung“ formalisiert ab, muss das Vorgehen dokumentiert und die herangezogenen Informationen angegeben werden. Hierfür werden häufig Fragebögen bzw. Guidelines oder Templates verwendet, sodass der Zeitaufwand zwar höher als bei dem Verfahren ad hoc, aber absolut betrachtet gering ist. Die Methode erfüllt daher die Anforderung „Ergebnis aufwandsarm verfügbar“ in hohem Maß (9).

Bei einer „Expertenschätzung durch eine Gruppe von Experten“ findet auf der Suche nach einem Konsens Diskussion zwischen den Beteiligten statt. Läuft dieser Austausch formalisiert und über mehrere Feedbackrunden ab (wie bei der Delphimethode), ist mit deutlich mehr Aufwand als bei individueller Expertenschätzung oder ad hoc-Schätzung durch eine Expertengruppe zu rechnen. Der Gesamtaufwand nimmt mit dem Grad der Formalisierung und der Anzahl beteiligter Experten zu. Es fällt weder Einarbeitungs- noch Modellierungsaufwand an, sodass der Erfüllungsgrad der Anforderung „Ergebnis aufwandsarm verfügbar“ bei der ad hoc-Gruppenschätzung hoch (9), bei formalisierter Gruppenschätzung mittelmäßig ist (3).

Schätzmethoden basierend auf Machine learning

„Künstliche Neuronale Netze“ müssen für den Einsatz konfiguriert werden, was mit erheblichem Aufwand zur Bildung des Netzes verbunden ist [KCCS00, 2]. Der Erfüllungsgrad der Anforderung „Ergebnis aufwandsarm verfügbar“ ist daher als niedrig einzustufen (1).

Zur Schätzung mittels „Rule Induction“ müssen lediglich die vorhandenen Daten abgeschlossener Projekte eingegeben werden, damit Regeln zur Schätzung weiterer Projekte durch die Rule Induction-Applikation generiert werden. Allerdings erfordern Regeln, die zu hinreichend genauen Schätzergebnissen führen, Sorgfalt bei der Auswahl der Merkmale, anhand derer die Regeln aufgestellt werden [MKLP00, 26]. Da diese Auswahl je nach Anzahl verfügbarer Merkmale sehr zeitaufwändig werden kann [MKLP00, 26], und bei Softwareentwicklungsaufwand sehr viele mögliche beeinflussende Merkmale existieren, wird der Erfüllungsgrad der Anforderung „Ergebnis aufwandsarm verfügbar“ als niedrig eingestuft (1).

Wird mittels „Genetic Programming“ geschätzt, ist das System anhand verschiedener Parameter zu konfigurieren (z. B. Anzahl der gleichungserzeugenden Durchläufe [Dola01, 64]). Um die vorhandenen Daten ausreichend genau abzubilden, muss der Algorithmus mehrmals mit verschiedenen Parameterwerten durchlaufen werden. Aufgrund des Konfigurationsaufwands wird die Anforderung „Ergebnis aufwandsarm verfügbar“ nur mittelmäßig erfüllt (3).

Die Methode „Case-based Reasoning“ erfordert wie alle Machine learning-Systeme die Eingabe zu analysierender Datensätze und Konfiguration des Systems (z. B. Anzahl auszuwählender Vergleichsprojekte, Auswahl von Vergleichskriterien). Hierfür existiert keine etablierte Vorgehensweise, sodass die Konfiguration meist einem Trial and Error-Prozess mit entsprechendem Aufwand folgt [KCCS00, 20]. Da eine kleine Anzahl Cases für den effektiven Einsatz der Methode ausreicht [KCCS00, 2], wird die Anforderung „Ergebnis aufwandsarm verfügbar“ mittelmäßig erfüllt (3).

Practical Application of SPICE in the Finance Sector

Dr. Christian Völcker

SYNSPACE SA, Rue de Lyon 114, CH – 1203 Genève

cv@synspace.com

Johann Leibig

HVB Systems GmbH, Apianstraße 14, D – 85774 Unterföhring johann.leibig@hvbsystems.com

Josef Maier

HVB Systems GmbH, Apianstraße 14, D – 85774 Unterföhring

TU Berlin, Institut für Wirtschaftsinformatik und Quantitative Methoden, FR6-7, Franklinstr.

28/29, 10587 Berlin

josef.maier@tu-berlin.de

Holger Kühnl

SYNSPACE GmbH, Kartäuserstr. 49, D-79102 Freiburg

hk@synspace.com

Abstract:

In this paper, performance, outcomes and benefits are reported of nearly 140 SPICE-Assessments conducted over a time span of two years in the IT-organization of one of the largest German banking groups. The scope of the assessments includes projects and services covering development, maintenance, migration and production assurance of finance applications. The numerical results of the assessments are used as input for the organization's Balanced Scorecard, while the identified strengths and weaknesses serve as input to refine the organization's standard process and for further improvement of process effectiveness and efficiency. The paper reflects the trend in the finance sector of the increasing awareness and importance of the quality and effectiveness of IT-Processes. It demonstrates that ISO/IEC TR 15504 (SPICE) spans across the landscape of IT Processes in a financial environment and how it can be used for performance evaluation at various levels in an organization. It is described how the assessment process has been optimized to perform the high number of assessments and how the assessment model has been adapted to the particular needs and culture of the organization. In addition, the overall benefits from the assessments are discussed from the organization's point of view.

1 Introduction

Following a trial phase of several years, the international standard ISO/IEC TR 15504 has been published in 1998 [ISO98]. This standard integrates the concepts and experiences made with various assessment methods, including CMM [Pa93] and BOOTSTRAP [Ku93] and provides a reference model as a basis for comparing results from the different assessment models. This reference model contains process definitions that are related to the Software Life Cycle Processes of ISO/IEC 12207:1995 [ISO95] and covers all processes that are necessary for the acquisition, development, maintenance, and operation of software. In addition, the reference model defines a universal measurement framework that can be applied to any process in any organization. With this measurement framework, the capability of any process can be determined on an integer scale from 0 (incomplete) to 5 (optimizing).

Since its publication as Technical Report, the standard – most commonly referred to as SPICE (Software Process Improvement and Capability dEtermination) - has been used in many

companies of various IT sectors as a starting point for process improvement. In particular in Aerospace [Ca04] and Automotive [HIS04, Vö03], ISO/IEC TR 15504 has also been intensively applied as a means for evaluating the software processes of IT suppliers.

It is the flexibility and the universality of SPICE as well as the transparency of the results that ensures its rapidly increasing popularity throughout all IT sectors. The structure of the standard allows defining additional domain specific processes beyond the scope of ISO/IEC TR 15504 when necessary. In order to deploy the concept of process assessment beyond the IT industry, the standard has been revised in the past years with the following key modifications:

- Alignment of the Capability Level attributes with ISO 9001:2000
- Introduction of the concept of process reference models
- Transfer of the process dimension of TR 15504 to a process reference model embedded in ISO/IEC 12207 AMD1:2002 (and AMD2:2004),

The result is now being published as ISO/IEC 15504:2004 [ISO03, ISO04a, ISO04b, ISO04c]. An exemplar assessment model based on the process reference model of ISO/IEC 12207 AMD1 and AMD2 was published in early 2006 as an international Standard (IS).

2 Process Assessment in the Finance Sector – Current Situation

While IT development in Aerospace, Telecommunication and Automotive is traditionally deployed in a complex structure of contractors, subcontractors and sub-subcontractors, in the finance sector (mostly) just one department or subsidiary company provides the development, operation and maintenance of software. Nevertheless, outsourcing IT solutions and services to third parties becomes more and more popular as companies - in particular from Asia or Eastern Europe - offer more and more competitive conditions as an alternative to an internal IT organization. These companies offer cheaper rates and very often they also advertise their very high process capability or maturity as additional asset. Therefore, the internal IT organizations face pressure from principals and competitors on both cost and process quality.

These trends have made cost reduction as well as adherence to the organizations standard process the primary objectives for the executive board of HVB Systems, the IT service organization of the HVB Group, one of the largest financial institutions in Europe. To enforce the achievement of these objectives throughout the organization, a Balanced Scorecard is

defined for each business year, which covers key data related to cost, internal processes, customer satisfaction and employee development.

In order to be able to benchmark internally and externally, HVB Systems has selected SPICE as the method for evaluating the process capability. To emphasize the company's commitment to SPICE, HVB Systems top management selected the average SPICE capability level to be one of the key metrics for the process perspective of the company's Balanced Scorecard. For each year, a minimum average Capability Level is defined as measurable objective in the Balanced Scorecard. If the objective is met, the respective department's bonus is affected positively.

Other IT departments and subsidiaries in Germany and Switzerland apply similar measures or programs in order to improve process adherence, predictability and to reduce cost caused by ineffective processes. Some of them have chosen CMMI, while others decided in favour of SPICE, because they considered it to be more cost effective and easily adaptable to the particular needs of the organization.

3 The Initial Situation

In order to deliver the measures for the HVB Systems Balanced Scorecard of 2004 and 2005, 10 departments had to be assessed in three business areas: application development projects, production assurance and application maintenance. The latter two are continuous activities that include problem prevention and bug fixing during operation as well as the implementation of small functional changes. Since these two business areas generally involve the same people for each application, evaluation of production assurance and application maintenance have been combined into one assessment per application. In 2004 nearly 70 SPICE assessments had to be performed within a period of 7 months in 2004. For 2005, the same scope has been used which resulted in a total of almost 140 assessments in two years.

For the performance of the assessment series, four internal staff members and initially one external Lead Assessor were assigned. In order to meet the timeframe, the need for one additional external Lead Assessor was identified in the planning phase. Due to additional tasks of the internal staff members to be performed in other areas of the department, the number of assessors for 2005 had to be increased to five internal and three external assessors.

During the previous years, various programs for process establishment and process improvement have been performed at HVB Systems that were expected to have significant

positive impact on the process capability. The purpose of the assessments was therefore not only to provide the relevant data for the Balanced Scorecard, but also to gain deeper insight on the performance and suitability of the organization's standard process and to identify areas for further improvement of the standard process towards Capability Level 3 for all processes within the scope of the assessments. As a result, the time in-between the two assessment series of 2004 and 2005 was scheduled to be used for the prioritization and implementation of the identified improvement areas of the standard process.

4 Defining the Assessment method

During the preparation phase, the main challenge was to define an assessment method that would

- minimize the duration of each assessment,
- minimize the interview time,
- ensure consistency of the ratings across the assessor teams,
- take into account the key processes of the organization's standard process,
- aggregate assessment results into one number per assessment, and
- effectively aggregate individual assessment feedback to identify areas for improvement for process performers and the standard processes.

The definition of the assessment method included

- tailoring the assessment model,
- defining the assessment process and the assessment tools to be used.

These tasks are described in detail in the following subsections.

4.1 Tailoring the Assessment Model

In a first step, the organization's standard process activities and work products were mapped to the base practices and management practices of the ISO/IEC TR 15504 assessment model. It turned out that for some activities, the granularity of the process dimension of 15504 was not detailed enough or did not even cover the activities at all. For example, the CUS.2 Supply process consisted of two very distinct processes at HVB Systems, the tender preparation and

assignment process at one end of the software life cycle and the process for preparation and delivery for operations at the other end. This problem has been identified previously in other IT sectors, notably in the space sector [Ca01] and therefore the same solution was taken here: CUS.2 has been split into two processes, CUS.2.1 Supply Preparation and CUS.2.2 Delivery.

For the assessment of application maintenance, a distinct process for recording and tracking the change requests from the customer was required in the assessment model. This process is not adequately covered by ISO/IEC TR 15504, but this problem has been addressed during the migration of the process dimension to ISO/IEC 12207. Amendment 2 defines a ‘Change Request Management’ process that ideally meets the needs of the assessments at HVB Systems [ISO04d]. The base practices and work products for this process have been taken from the draft assessment model of the revised ISO/IEC 15504.

It should be noted that HVB Systems business areas focus on the engineering processes, and exclude software operation, which is under the responsibility of HVB Info, another member of the HVB Group. While HVB Info has structured its processes according to ITIL [BS02, BS03], this model was not found to be suitable for the application maintenance processes at HVB Systems and therefore has not been considered when tailoring the assessment model.

The full set of processes selected for the assessments at HVB Systems is shown in Table 1.

Application Development:	Production Assurance:
CUS.2.1 Supply preparation	CUS.4.1 Operational Use
CUS.2.2 Delivery	SUP.8 Problem Resolution
ENG.1.1 System Requirements Analysis and Design	ENG.2 System and Software Maintenance
ENG.1.2 Software Requirements Analysis	
ENG.1.3 Software Design	
ENG.1.4 Software Construction	Application Maintenance
ENG.1.5 Software Integration	SUP.9 Change Request Management
ENG.1.6 Software Test	ENG.2 System and Software Maintenance
ENG.1.7 System Integration and Test	
MAN.2 Project Management	
SUP.3 Quality Assurance	

Table 1: Processes assessed in three business areas of HVB Systems.

It is remarkable that although HVB Systems does not operate the software, all base practices of the process CUS.4.1 Operational Use except one (BP 3: Operate the software) were found to be applicable.

For each process, a list of questions has been developed around the base practices and a list of relevant work products and activities of the HVB Systems standard process has been established. These questions used wherever possible the company specific terminology with which the interviewees were familiar. In addition, a list of generic as well as process specific questions for each management practice has been developed. These standard questions for base practices and management practices have reduced the time and effort needed for the interviews significantly. Furthermore, comprehensive guidelines and rules for rating have been developed for internal use by the assessors to accelerate the data analysis phase and the rating phase of each assessment. Both questions and rating guidelines helped ensure consistency and comparability of the assessment results across the different two-person assessor teams.

4.2 Defining the Assessment Process and Selecting the Assessment Tools

The assessment process consisted of the typical activities required by ISO/IEC TR 15504 (planning, data collection, data validation, rating, reporting). In order to optimize the efficiency of the process, a few changes to the usual procedure have been implemented.

The planning of all assessments was done centrally and supported by a database. All the necessary assessment information (date, status, assessors, interviewees, context, etc) was entered in this database and could be viewed by all assessors. Reports were generated from the database for planning and progress tracking.

The total time needed for data collection was reduced by about 30% with respect to what has been experienced by SYNSPACE in the past. This was achieved by measures like using pre-assessment questionnaires where the answers were automatically imported into the assessment database, by using video projectors during the interviews for quick electronic access to relevant documents, and by giving free access to the assessors to the HVB Systems standard repository for application and project documentation.

For rating the Process Attributes, performance of each relevant base or management practice has been rated separately on the n/p/l/f scale defined by ISO/IEC TR 15504-5. To support the interview and rating process, the off-the-shelf tool SPICE 1-2-1 [HMS05] was selected and its generic contents adapted according to the tailored assessment model using the complementary

tool SYNEDIT. It turned out to be extremely effective, if a preliminary rating of the practices was already done in SPICE 1-2-1 by one of the assessors during the interviews. After the interviews, the rating was consolidated using the guidelines and rules as specified in the tailored assessment model. For each base practice or management practice that was not rated 'fully', an appropriate improvement recommendation was addressed to the process performer or to the relevant process manager and included in the assessment report. Systematic quality assurance measures included reviews of the assessment reports by a third assessor and assessor rotation among the assessor teams. Various algorithms have been considered for aggregation of the results. However, in the end the simple arithmetic mean of the measured Capability Levels was preferred by the HVB Systems management. It was, however, decided to apply an additional deduction if some of the processes were at capability level 0.

5 Assessment Outcomes

The key factors for the success of the assessments were full management support, high sample density throughout the organization and the combination of internal and external assessors.

About 15% of all HVB Systems employees were interviewed at least once during the two seven-month assessment periods in 2004 and 2005. Thus, the SPICE assessment process was highly visible throughout the organization and caused impressive awareness and motivation for improvement among the employees of HVB Systems. The result of each individual assessment was an assessment report that contained the process profiles and capability levels, a list of observed strengths, improvement recommendations addressed to the project and improvement recommendations addressed to those in charge of the standard process. The report was created using the central database which also served as a means to aggregate and track the improvement suggestions resulting from the assessments. Initially the only requirement for the assessments was to measure the actual capability level. However, top management objectives require all departments to improve the average capability level throughout the year. Therefore, the interest in the improvement recommendations from the individual assessment reports was tremendous and the most frequently asked question was 'what can we do to increase our level?'

In many departments, workshops were held shortly after completion of the assessments to derive clear improvement actions from the assessment reports that can be performed within the department. On the organizational level, workshops were held with those responsible for the

standard process to identify improvements to the current process definition, guidance and templates. In total, about 1000 improvement suggestions retained from individual assessments were grouped into about 50 general recommendations for the process performers and 40 recommendations for those in charge of the standard process.

Many of the recommendations for process performers were related to the systematic and consistent planning and tracking of a project and to the adherence of particular aspects of the standard process. Further recommendations focused on the traceability of the requirements for and the decisions made throughout the project. Recommendations for improving the standard process were mostly related to decreasing discrepancies and differences in planning and documentation across the organization and to providing a common basis for measurable processes. These recommendations were prioritized and mapped to a timeline for implementation.

Several departments and project managers have asked for special guidelines for achieving SPICE level 1, 2 and 3. These were developed as an outcome of the assessments taking into account the list of questions used in the assessment as well as the improvement recommendations. The result, a 'SPICE Guide' for process users has been published on the intranet and has been very positively accepted by the staff.

The particular strengths listed in the individual assessment reports, were consolidated and communicated as best practices that could be applied by other projects to improve the capability level of their processes.

Finally, top management held an event which was organized like a press conference, to inform the whole organization about the assessment results and to substantiate its expectations and objectives related to the required SPICE capability levels for the current calendar year.

6 Conclusions

The assessments performed at HVB Systems demonstrated that ISO/IEC TR 15504 (SPICE) can be effectively applied as a means to enforce the adherence to the standard process model of an IT organization which is focused on the financial sector.

The principal condition for the success of the assessments was firm support by top management which resulted in consequences for those that deviated from the standard process without prior approval and encouragement for those who focus on systematic implementation and feedback.

Further factors for the success and the acceptance of the assessments were

- the tailoring of the assessment model to the particular needs of the finance sector
- the improvement recommendations that have been provided with the individual assessment results
- the combination of internal and external assessors
- a list of questions related to base practices and management practices that used the organization's terminology
- guidance for rating the process attributes
- an optimized assessment process with minimal resource needs for both assessors and assessees.

The management of HVB Systems has clearly identified several key benefits from the assessments for its customers, for its employees and for the company as a whole. These benefits strongly support the commitment to continue with SPICE for the next years.

The experience made in application of SPICE and in particular the tailoring of the assessment model may be applied not only to other IT organizations in the finance sector but also to software development and maintenance in any sectors with similar IT architectures (incorporating large databases and a high number of simultaneous users) and processes such as public administration, insurance or logistics.

References

- [BS02] BS 15000-1: 2002: IT Service Management: Specification for Service Management. British Standard Institute, 2002.
- [BS03] BS 15000-1: 2003: IT service management. Code of practice for service management. British Standard Institute, 2003.
- [Ca01] Cass, A. et. al.: SPiCE for SPACE: A Process Assessment and Improvement Method for Space Software Development. ESA Bulletin 107, European Space Agency, 2001; pp. 112-119.

- [Ca04] Cass, A. et. al.: SPICE for SPACE trials, risk analysis, and process improvement. In: Software Process: Improvement and Practice Vol. 9, Issue 1, 2004; pp. 13-21.
- [HIS04] HIS-Herstellerinitiative Software: HIS Assessments: Process Scope. IMH - Institut für Motorenbau Prof. Huber GmbH, München 2004; available on: http://www.automotive-his.de/download/HIS_Process-Scope_SPICE_v04.pdf.
- [HMS05] HM&S IT-Consulting GmbH: SPICE 1-2-1. HM&S IT-Consulting GmbH, Graz, 2005; available on: <http://www.synspace.com/EN/Products/spice121.html>.
- [ISO98] ISO/IEC TR 15504:1998(E), Information Technology - Software Process Assessment, Parts 1-9, Type 2 Technical Report. International Organization for Standardization, 1998.
- [ISO95] ISO/IEC 12207:1995(E): Information Technology – Software Life Cycle Processes. International Organization for Standardization, 1995.
- [ISO03] ISO/IEC TR 15504-2:2003(E), Information Technology - Process Assessment, Part 2: Performing an Assessment. International Organization for Standardization, 2003.
- [ISO04a] ISO/IEC TR 15504-1:2004(E), Information Technology - Process Assessment, Part 1: Concepts and vocabulary. International Organization for Standardization, 2004.
- [ISO04b] ISO/IEC TR 15504-3:2004(E), Information Technology - Process Assessment, Part 3: Guidance on Performing an Assessment. International Organization for Standardization, 2004.
- [ISO04c] ISO/IEC TR 15504-4:2004(E), Information Technology - Process Assessment, Part 4 Guidance on use for process improvement and process capability determination. International Organization for Standardization, 2004.
- [ISO04d] ISO/IEC 12207:1995/Amd.2:2004(E): Information Technology – Software Life Cycle Processes. International Organization for Standardization, 2004.

- [Ku93] Kuvaja, P. et al.: Bootstrap: Europe's Assessment Method. IEEE Software, Vol. 10, May 1993; pp. 93-95.
- [Pa93] Paulk, M. et. al.: Capability Maturity Model for Software, Version 1.1. CMU/SEI-93-TR-24, Software Engineering Institute, February 1993.
- [Vö03] Völcker, C.: SPICE in Automotive, Talk given at IIR 2003 Stuttgart, 2003.

Towards a Model-driven Development of Monitored Processes

Christof Momm, Robert Malec, Sebastian Abeck

Cooperation & Management
Institut für Telematik

Universität Karlsruhe (TH)
76128 Karlsruhe
{momm,malec,abeck}@cm-tm.uka.de

Abstract

An integrated management of business processes demands a strictly process-oriented development of the supporting IT. Process-orientation is especially promoted by Service-Oriented Architectures (SOA), where loosely coupled business services are being composed to executable processes. In this paper we present a model-driven methodology for a top-down development of a process-oriented IT support based on a SOA. In contrary to existing approaches we also include the monitoring required for business process controlling and introduce meta-models for the specification of process performance indicators in conjunction with the necessary monitoring. Furthermore, we show how these models are transformed to executable process definitions extended by the required monitoring activities.

1 Introduction

Today, companies demand IT support that is strongly aligned with their business processes, which in turn are compliant with their (strategic) business goals. For achieving this, goal-driven approaches to an integrated Business Process Management (BPM) have been proposed [AaHW03; MuRo04]. For controlling goal achievement in business processes, solutions are required that allow a continuous, “real-time” monitoring of the performance within the IT support based on quantitative process performance indicators (PPI). This aspect is also referred to as “Business Activity Monitoring” (BAM) in relevant literature. The discussed BAM

architectures [JeSC03; McSc04] have in common that they abstract from the IT systems implementing the business processes by introducing management relevant business events. The PPIs are defined and evaluated on the basis of these business events, which are generated by event adapters triggered by an instrumentation of the underlying IT systems. Because the business events and their associated PPIs as well as the existing IT support are extremely company-specific, the implementation of the required instrumentation is itself very time consuming and is elongated if facing an extremely heterogeneous IT support.

The heterogeneity of the IT support can be significantly reduced by establishing a company-wide Service-Oriented Architecture (SOA) in which the business processes are consequently realized through orchestrations on the business process layer of a SOA [Le03; LeRS02]. This is accomplished by either using the Business Process Execution Language (BPEL) [ACDG03] and web services or other SOA platforms, for example CORBA and an appropriate workflow engine. Unfortunately, a uniform methodology for realizing the required monitoring – including the instrumentation - does not as yet exist. It is still very specific to the employed SOA platform. Accordingly, a solution for specifying the monitoring in a platform-independent way and a clear methodology for breaking down PPIs into appropriate measuring points within the orchestrations or queries on logging data is necessary [HaRa01].

Taking these drawbacks into account, in this paper we propose a top-down approach for developing a uniform IT support based on a SOA in conjunction with the monitoring aspects required for processing the PPIs. To enable the support of different SOA platforms as well as an automated generation of the required instrumentation and monitoring infrastructure, we decided to build the approach on the principles of the Model Driven Architecture (MDA) proposed by the OMG [MiMu01]. The approach is demonstrated using a concrete business process taken from the field of higher education. So far, the target platform is limited to most common SOA platform based on BPEL and web services.

2 Related Work

For developing a service-oriented IT support tightly aligned with the underlying business processes, various model-driven approaches have been proposed [BaMR04; KHSW05]. Thereby, the business processes are specified by means of computation-independent business process models (CIM), for instance based on Petri nets, Event-driven Process Chains (EPCs) or the

Business Process Modeling Notation (BPMN). These process models are systematically refined and transformed into platform-independent models (PIM) of executable business processes (i.e. orchestrations). Finally, these PIMs are transformed to platform-specific models (PSM), in particular to executable process definitions which are mainly based on BPEL. So far, the presented approaches deliver solutions to the development of the functional aspects but do not consider monitoring and control (i.e. management) aspects.

For component-based software development this aspect has already been addressed. [PAVB04] present an approach which integrates Quality of Service (QoS) aspects into a model-driven development process of component-based applications and allows for an automated generation of the required monitoring infrastructure and component instrumentation. As the service-orientation leads to a significant reduction of complexness, some essential adaptations are necessary to seamlessly integrate monitoring aspects into a model-driven SOA development process.

To enable BAM on the orchestrations, the generic solution proposed by [JeSC03] can be employed. As already pointed out, in this case the instrumentation of the involved IT systems would be very time consuming as it has to be accomplished manually for each orchestration and execution engine. Furthermore, the presented implementation is based on proprietary technologies, which particularly complicates the monitoring of cross-enterprise business processes. Taking especially this shortcoming into account [McSc04] propose a framework for analyzing and measuring business performance on the basis of web services. In doing so, the whole BAM system is encapsulated in a Solution Manager Service providing a standardized interface for the instrumented IT systems. The information required for evaluating the PPIs is transferred to the Solution Manager Service by extending the BPEL process definition by management calls. [Mc03] amplifies this BPEL instrumentation, but does not address the questions how to systematically develop such instrumented orchestration in a top-down fashion and how to automate the generation of these artifacts.

3 Approach to a Model-Driven Orchestration Development

Within this section we introduce our general idea for a model-driven development of monitored orchestrations. Figure 1 provides an overview of our approach.

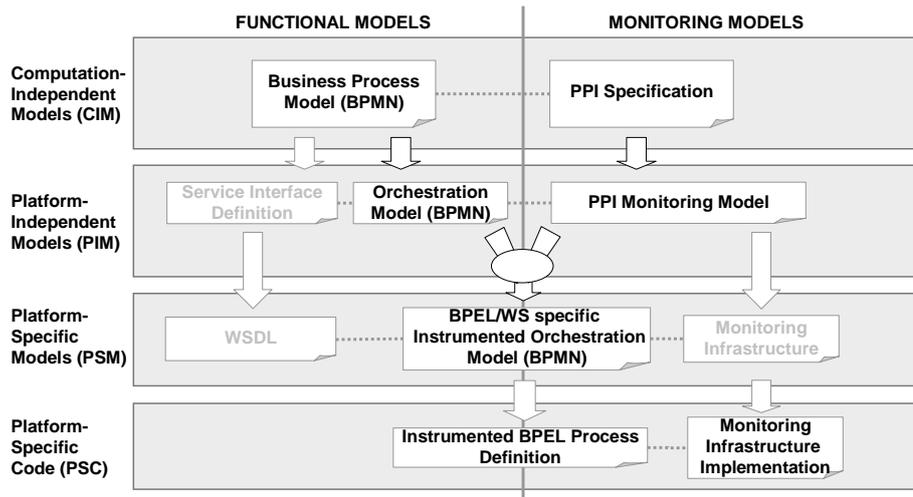


Figure 1: Model-Driven Orchestration Design

According to this we distinguish between functional and monitoring models on three different layers of abstraction (CIM, PIM and PSM). The target platform [MiMu01] for the functional parts is a specific SOA platform. These specifics are abstracted by the PIM. The same holds for the monitoring. There are several existing solutions, how monitoring can be realized. The instrumentation for instance can be implemented by adding sensors to the orchestrations or querying the audit trail. In practice, the vendors of a BPEL or other workflow engine provide their own, platform-specific monitoring solutions. Our goal is to support the different approaches to implementing the monitoring of orchestrations by means of a platform-independent monitoring model, which can be transformed (automatically) into specific models. This paper focuses on the generation of the specific BPEL process definitions (or code) extended by sensors as a first step towards this objective.

The CIMs are used for specifying the business processes along with the goals in a way that is independent from the IT support. Thereby, various modeling notations are available for modeling business processes. As pointed out in [BaMR04] an MDA approach allows supporting all kinds of notations by transforming them into an appropriate meta notation like the Business Process Definition Metamodel (BPDM) [FrGJ04]. Due to the fact that the BPMN standard defines mapping rules for a BPMN-to-BPEL transformation and is already supported by a couple of development tools, it currently represents one of the most appropriate platform-independent models [EmWA06]. Thus, we decided to also employ it for modeling the functional aspects on the CIM level.

On the PIM layer the functional models of the CIM layer are transformed into BPMN-based orchestration models describing the (machine) executable business processes as well as the external services invoked within the orchestration.

The high-level PPI specifications on the other hand are transformed into PPI monitoring models. Basically, these monitoring models define components for measuring the specified PPIs on the basis of metrics and monitoring information, which are derived from the functional model.

The orchestration models along with the monitoring models are transformed into a platform-specific instrumented orchestration model. Concretely, the applied transformation adds sensors to the orchestration model which are required for evaluating the specified PPIs. And furthermore, the specifics of the selected SOA implementation are added to the instrumented orchestration models. Finally, the platform-specific code, namely the executable BPEL process definitions, is generated from platform-specific model. These process definitions are extended by monitoring sensors which pass the information on to a static monitoring infrastructure (MI). This MI provides a uniform interface to a BAM system and offers the required monitoring information for evaluating the specified business process goals.

The MI comprises several monitoring agents, which are possibly arranged in a hierarchical way. Furthermore, several existing technologies, like for instance Web-Based Enterprise Management (WBEM) could be employed for implementing this infrastructure. In this paper, we limit the scope to a simple and static MI consisting of one monitoring agent for each specified PPI. A more flexible design would require the creation of adequate models for describing the details of the MI and the usage of the employed platform.

4 Computation-independent Modeling of Business Processes and Mapping to Platform-independent Orchestration Models

For the formal modeling of business processes various notations are available. As pointed out in section 3, we decided on BPMN for modeling business processes in a computation-independent way. Thereby, the BPMN defines both the (graphical) notation and the semantics of a process through the definition of a so-called Business Process Diagram (BPD) [EmWA06].

Figure 2 provides an overview of the underlying meta-model. The full specification is available in [Wh04].

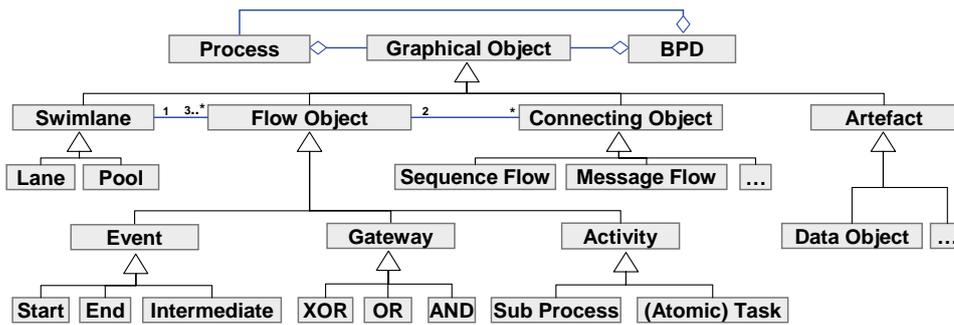


Figure 2: BPMN Meta-Model for Defining Business Processes in a Computation-Independent Way

In this section, the elements of a BPD which are important for understanding the monitoring models presented in the following sections are briefly introduced. In general, a BPD is comprised of activities performed by a certain organizational unit or role, a control flow between the contained activities, artifacts, like for instance data objects, which are processed within the activities, and events that may occur during process execution. The control flow is modeled by means of *Connecting Object* elements, especially the *Sequence Flow* along with *Gateway* elements for modeling parallel flows and conditioned branches. The process participants are modeled through the construct *Swimlane*. A *Pool* indicates that the containing process is owned by an independent organizational unit, whereas a *Lane* within a *Pool* specifies that a certain role is responsible for the covered process parts. An exchange of messages between two organizational units is described through a *Connecting Object* of type *Message Flow*. By means of the element *Sub Process* a process may be further segmented.

Using these modeling elements a business analyst is able to model business processes from a business perspective without regarding the involved IT. These models are refined and restructured to platform-independent orchestration models. Thereby, each activity is broken down to an executable task. In fact, the orchestration model must not contain a non-executable activity. The BPMN specifies the following concepts for defining orchestrations (

Figure 3).

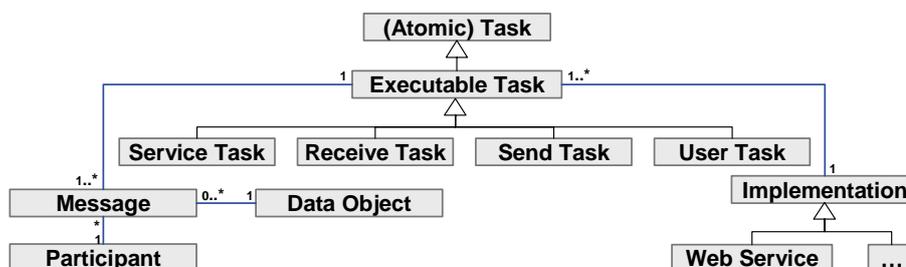


Figure 3: BPMN Meta-Model for Defining Orchestrations

Hence, an *Executable Task* generally involves the exchange of one or more *Messages* associated with a *Participant*. Furthermore, an *Implementation* (*Web Service* or other adequate implementations) is specified. The BPMN standard then distinguishes between four different kinds of executable tasks. A *Service Task* involves a *request-response* or *one-way* invocation of an operation provided by an external service. A *Receive Task* on the other hand awaits a message from an external client offered as a service operation by the orchestration itself. In case such an operation is of the type *request-response* a *Send Task* is used for returning the reply message to the requestor. A *User Task* comes into play if the orchestration involves human interaction. Within these tasks, a task message is assembled and delivered to an external task manager, which amongst other things allocates the tasks to a responsible employee, provides a user interface for the processing of the task and returns a task to the respective process as soon as it is finished. The standardization of this mechanism is currently being tackled by WS-BPEL4People initiative [KK+05]. The BPMN elements previously introduced for the CIM are also used within the PIM to model the orchestration's control flow. Note that the orchestration model may be very different from a computation-independent model. Therefore, the transformation can from our point of view not be automated.

5 Specification of the Process Performance Indicators and the PPI Monitoring Model

This section introduces newly developed meta-models for specifying PPIs for a process in a computation-independent way as well as a platform-independent PPI monitoring model which additionally defines how the specified PPIs are measured within the respective orchestration. The meta-models are based on existing approaches presented in [BKPS04; PAVB04]. In contrast to [BKPS04] we limited the scope to the specification of measurable, quantitative indicators and disregarded qualitative process or business goals. For the evaluation of the associated goals, an external BAM system could be employed

Figure 4 shows our meta-model for the specification of PPIs at the CIM level.

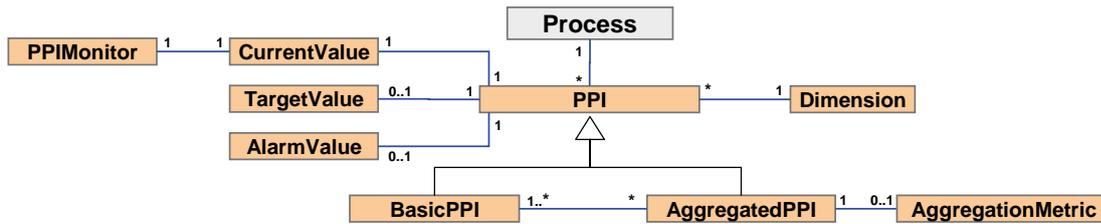


Figure 4: Meta-model for Performance Indicator Specification

A PPI is attached to the concept *Process* as part of the computation-independent process model. Optionally a *TargetValue* indicating the objective as well as an *AlarmValue* defining a threshold for an intervention may be specified. The PPI is further characterized by assigning a *Dimension*. Thereby information like the data type, the direction (e.g. ascending or descending) and the unit of the value are specified. The calculation of the mandatory *CurrentValue* on the basis of runtime information provided by the underlying orchestration is handled by the *PPIMonitor*. This aspect is tackled within the scope of the PPI monitoring model. Furthermore, we distinguish between basic and aggregated PPIs. A *BasicPPI* represents an atomic indicator, which can be measured within a single process instance whereas an *AggregatedPPI* spans multiple instances and is either evaluated through an *AggregationMetric* operating on basic PPIs (e.g. mean or variance) or directly calculated by the respective *PPIMonitor*.

Having the PPIs specified on the CIM level as a next step, the platform-independent *PPIMonitor* has to be defined in order to obtain a full PPI monitoring model tailored to the monitoring of orchestrations. The meta-model presented in Figure 5 has to be seen as an extension of the PPI specification meta-model.

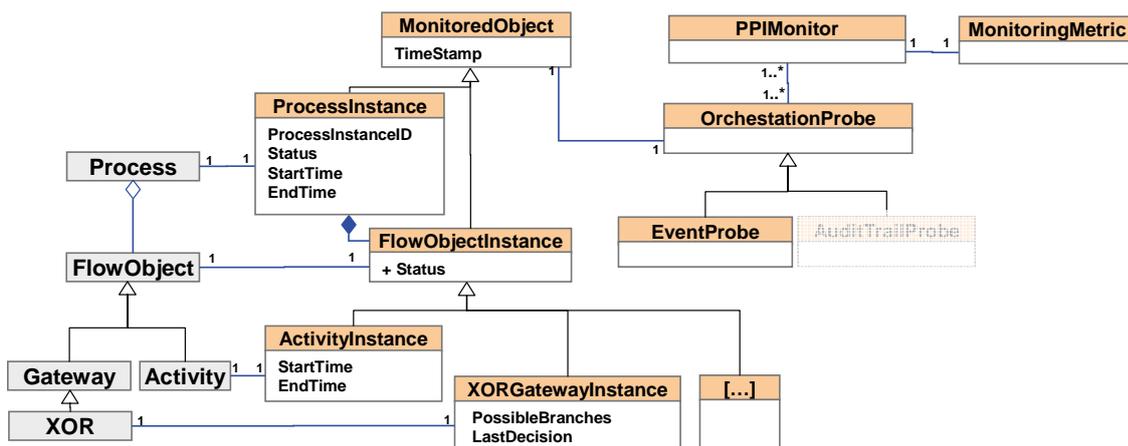


Figure 5: Meta-Model for Specifying the PPI Monitoring Model

Basically, a *PPIMonitor* operates on one or more managed objects of the respective orchestration and determines the desired PPI as-is value by means of a predefined *MonitoringMetric*. The

managed objects thereby represent a management view on the process and hence encapsulate information relevant for management [HeAN99]. As in our case the management functionality is limited to the monitoring of PPIs; the concept is termed *MonitoredObject*. The *MonitoredObject* of type *ProcessInstance* for example delivers information about the running process instance, like its current *Status* (e.g. “active” or “completed”), its cycle time (*StartTime*, *EndTime*) and the *ProcessInstanceID*, which is in orchestrations usually determined on the basis of a predefined *Correlation Set*. Within the scope of a process instance the monitoring can be further extended or refined to *FlowObjects* the process contains. Hence, a *MonitoredObject* of type *FlowObjectInstance* is introduced which may not exist without a *ProcessInstance*. The monitoring information required for a *FlowObject* depends on its concrete type. Hence, for each monitoring-relevant *FlowObject* a correspondent *MonitoredObject* is defined, as for example *ActivityInstance* or *XORGatewayInstance*. Whereas in the case of an *Activity* from a monitoring perspective the cycle time is of interest, for a *Gateway* of type *XOR* we would i.e. like to know the last decision. Depending on the PPIs that should be monitored, this information model for processes has to be further extended.

To retrieve the desired monitoring information (e.g. state updates) for a specified *MonitoredObject* from the underlying orchestration engine, an adequate instrumentation is required. The instrumentation is realized by means of *OrchestrationProbes*. Thereby, the information can be either gathered on the basis of the audit trail provided by the engine or events that are fired within the orchestration itself. Thus, a general distinction can be made between an *EventProbe* and an *AuditTrailProbe*. In the next section, the necessary BPMN extensions for defining and realizing an instrumentation based on *EventProbes* as well as the corresponding transformation (i.e. model merge) of the orchestration model along with the PPI monitoring model will be discussed. The generation of an *AuditTrailProbe* is not taken into consideration within this paper.

6 Transformation of the Orchestration and PPI Monitoring Model into an Instrumented Orchestration Model

To obtain monitoring information by means of *EventProbes* an extension of the BPMN-based orchestration meta-model is required which allows for the specification of the accordant instrumentation (Figure 6).

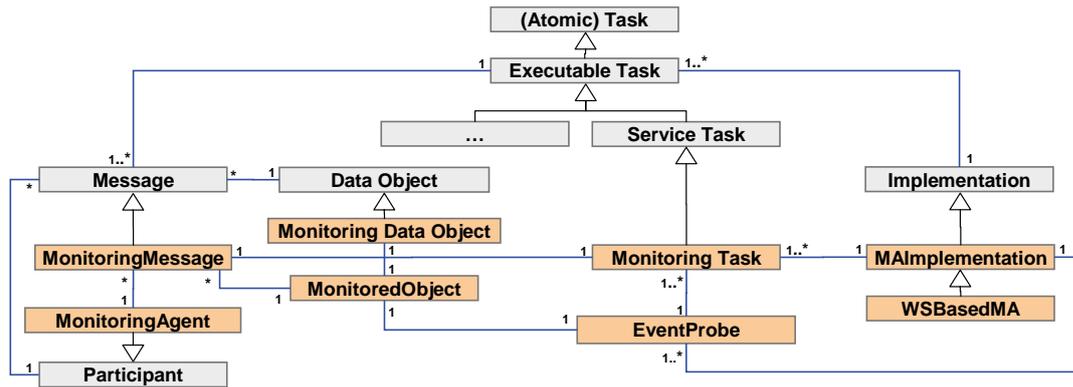


Figure 6: Extended BPMN Meta-Model for Specifying Instrumented Orchestrations

Thus, the monitoring information of a *MonitoredObject* provided by an *EventProbe* is gathered on basis of *MonitoringMessages* delivered by *MonitoringTasks*. A *MonitoringMessage* thereby contains a *Monitoring Data Object*, which only represents the BPMN version of the *MonitoredObject* (see Figure 5) and holds information about the current state. A *Monitoring Task* is a special kind of *Service Task*. But in contrast to those it only provides a *one-way* communication to the associated monitoring agent (MA), which is implemented by means of the *MAImplementation*. This implementation particularly realizes one or more *EventProbes*, meaning that it receives *MonitoringMessages* sent by a process instance through *MonitoringTasks* that belong to a distinct *EventProbe*. The probe update is then passed on to all associated *PPIMonitors*, which instantly calculate their *CurrentValue* on basis of the *MonitoredObject*'s state information provided by the probe by applying the predefined *MonitoringMetric*. The *MAImplementation* may rely on web services or other technologies. As we focus on a SOA implementation on the basis of BPEL and web services, we target a *WSBasedMA*.

The *MonitoringTasks* required for an *EventProbe* have to be placed at appropriate positions in the existing orchestration model. These positions depend on the concrete type of the *MonitoredObject* the probe is responsible for. In the following we will explain the basic idea of how the instrumented BPEL orchestration model is created from the orchestration model along with the PPI monitoring model (Figure 7). The approach is exemplified using the simple case of monitoring an activity.

The upper pool depicts a very simple orchestration model comprising of two activities of type *Executable Task* (*et1* and *et2*) which are executed in a sequence. The activity *et1* should be monitored on basis of an *EventProbe* providing the respective *MonitoredObject* of type *ActivityInstance*.

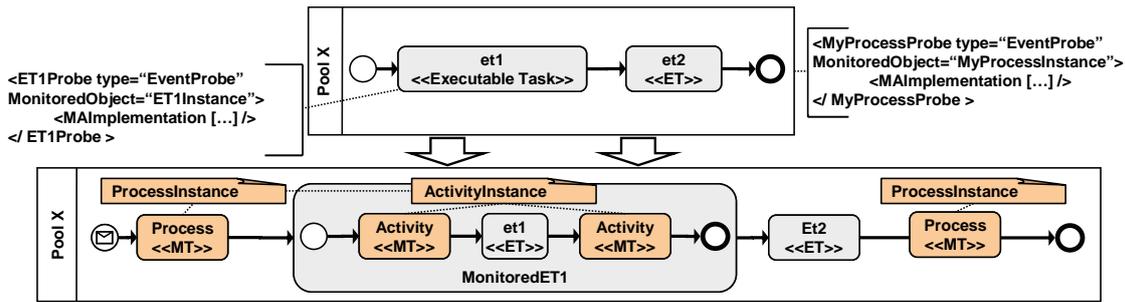


Figure 7: Mapping to Instrumented BPEL Orchestration Model

As an *ActivityInstance* may not exist without a *ProcessInstance* (see Figure 5) and the identifier of the running process instance is required for correlating the probes with the associated PPI, respective *EventProbes* for the whole process (*MyProcess*) as well as for the executable task *et1* have to be defined on basis of the PPI monitoring model and the instrumented orchestration model. As the evaluated modelling tools that support BPMN do not allow for an extension of the underlying BPMN meta-model, we decided to realize the association between concepts of the two models by means of BPMN annotations holding an XML-based definition of the *EventProbe*. This definition comprises all aspects needed for (automatically) creating the BPMN instrumentation on basis of *Management Tasks*. These annotations have to fully match the *EventProbes* specified within scope of the PPI monitoring model.

The transformation of the annotated orchestration model works as follows: In case of the *EventProbe* for the whole process two *MonitoringTasks* are added to the orchestrations model, namely one right after the *StartEvent* and one just before the *EndEvent*. The first *MonitoringTask* provides information about the determined process instance identifier and the starting time whereas the second one only adds the end time.

The instrumentation of an activity is performed in a similar manner. The only difference is that instead of the activity, a new sub process holding the activity itself along with the required *MonitoringTasks* is created and inserted into the orchestration model. Within the scope of the *MonitoringTasks* added before and after the actual activity, an *ActivityInstance* object is assembled or updated and in each case sent to the responsible MA. As indicated by the association between the *ActivityInstance* object and the *ProcessInstance* object, the process instance information is also delivered within the according *MonitoringMessage*. Furthermore, the associated *EventProbe* has to be included. This is accomplished by providing a fixed message part within each *MonitoringMessage* holding this meta-information.

It becomes clear, that for each *MonitoredObject* a fixed procedure for adding the necessary instrumentation can be identified. Hence, the automation of these procedures can be realized by applying adequate model transformations.

7 Case Study: Development of a Monitored Orchestration for the Management of Examinations

The approach put forward in this paper has been applied to a practical scenario developed in the context of the project “Karlsruher Integriertes InformationsManagement” (KIM) [JuMa05], which targets the process and service-oriented redesign of a university’s business processes along with the supporting IT. We particularly focused thereby on the business process within the scope of the examination management. Figure 8 shows a simplified computation-independent process model along with the refined platform-independent orchestration model for supporting the activities within the examination management lifecycle.

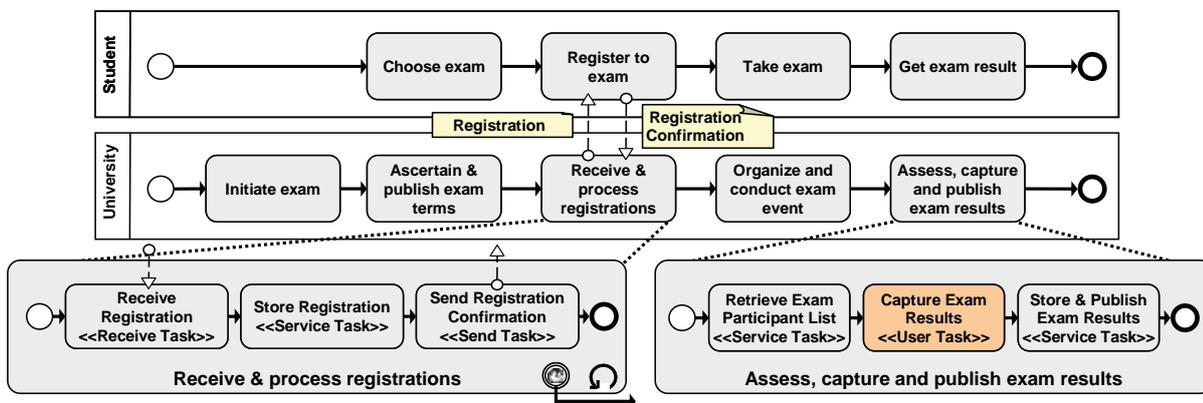


Figure 8: Process and Orchestration Model of the Examination Lifecycle Management

To demonstrate the approach, we limit the explanations to the activities relevant for the orchestration model. The orchestration is initiated after the university has decided to conduct an examination. In next step the terms for the exam are ascertained, transferred to the orchestration and published. Subsequently, registrations from students are received and processed by the orchestration. After the exam event has been organized and conducted, the exam results have to be assessed, captured and published. The capturing and publishing of the results is also supported by the orchestration. For this purpose, the final participant list is retrieved from the *EMService* within a *Service Task*. Afterwards, a *User Task* for the capturing of the results is

initiated. As soon as all results are available they are returned to the orchestration by the employed task management service and stored through a *ServiceTask*.

Due to the fact, that the exam results are required promptly for generating certificates and evaluating preconditions within the registration process for further exams, one key performance indicator is the students' waiting time for their results. Hence, a university wide policy defines that the capturing of the exam results must not exceed 3 weeks. If the results are still not available after 2 weeks, a reminder should be sent to the person in charge. Figure 9 shows the formalized PPI based on an UML profile for the presented PPI specification model.

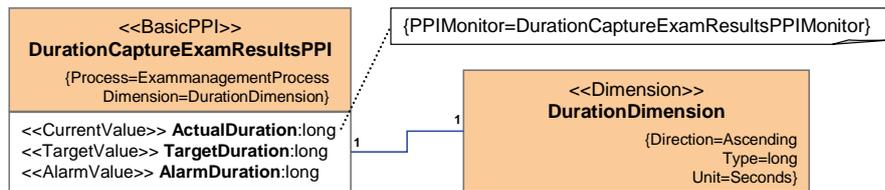


Figure 9: Specification of the PPI „DurationCaptureExamResults“

The assignments of a *Process*, a *Dimension* as well as the associated *PPIMonitor* for the attribute of stereotype *CurrentValue* to the specified PPI are realized through *TaggedValues*, either for on level of the stereotyped class or attribute. The target and the alarm value could also be realized through *TaggedValues*, but for the sake of flexibility we chose to define them as attributes. As soon as a concrete instance of the *DurationCaptureExamResultsPPI* is created (which has to be done for each newly created process instance) these values have to be assigned with three and two weeks.

As defined within the PPI specification the attribute *ActualDuration* is determined by the *DurationCaptureExamResultsPPIMonitor*. How this monitor works is specified by means of the PPI monitoring model is depicted on Figure 10.

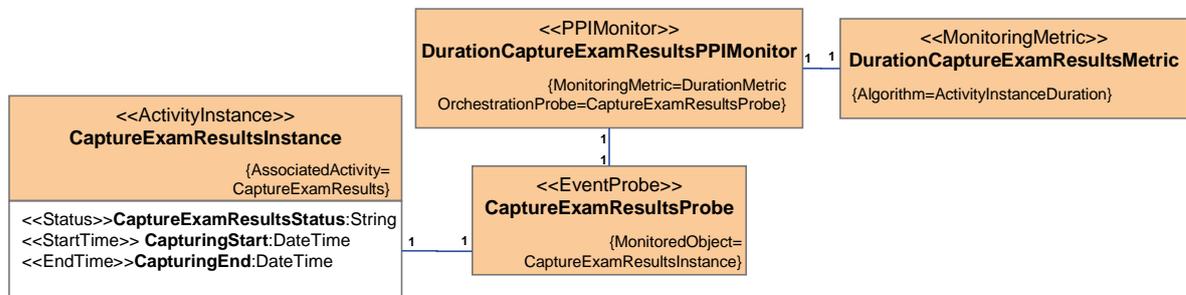


Figure 10: Monitoring Model for the PPI „DurationCaptureExamResults“

The *DurationCaptureResultsPPIMonitor* operates on a *CaptureExamResultsProbe*. This probe provides state information about the *MonitoredObject* of type *CaptureExamResultsInstance* which is associated with the process activity *Capture Exam Results* (see Figure 8). The actual value is calculated by executing the linked *MonitoringMetric*. In this case, the metric uses a generic algorithm for calculating the duration of an arbitrary *ActivityInstance*. This somewhat simple algorithm works as follows:

```

If (ActivityInstance.Status equals "Active")
    ActualDuration = TimeSpan(CurrentTime, ActivityInstance.StartTime)
Else if (ActivityInstance.Status equals "Completed")
    ActualDuration = TimeSpan(ActivityInstance.EndTime, ActivityInstance.StartTime)

```

To retrieve the state information about the *CaptureExamResultInstances* for all running process instances, the appropriate *MonitoringsTasks* are added to the orchestration model. As described in section 6 in case an *ActivityInstance* should be monitored, a new sub-process is created containing a sequence of the actual activity along with *MonitoringTasks* before and after the activity is performed. Figure 11 shows the instrumented orchestration model for the sample process.

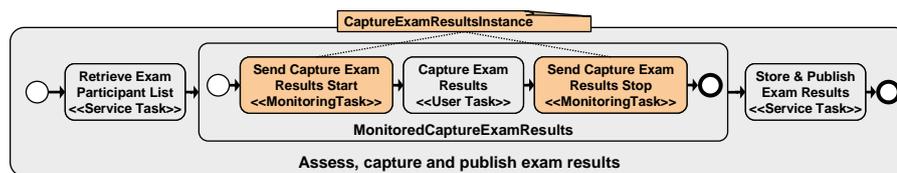


Figure 11: Instrumented Orchestration Model

For the added *MonitoringsTasks* some additional properties are specified, for instance the endpoint reference to the employed *MAImplementation*. To create the executable BPEL process definition we used the BPEL export functionality of the employed modelling tool (Borland Together 6.0). As, amongst other things, the required variable assignments are missing in the generated code and *UserTasks* are not supported at all, we had to manually add these aspects. For this purpose, we used the Oracle BPEL Designer along with the corresponding BPEL engine Oracle BPEL Manager. The final code for the sub process *MonitoredCaptureExamResults* is as follows:

```

<scope name="MonitoredCaptureExamsResult">
    [...]
    <assign name="setCaptureExamResultsInstanceStart" [...] </assign>
    <invoke name="sendCaptureExamResultsStartMessage" partnerLink="agentService"
    operation="processMonitoringMessage" inputVariable="captureExamResultsInstance" [...] "/>
    <!-- UserTask: CaptureExamResults-->

```

```

<scope name="CaptureExamResults" [...] xmlns:task="http://services.oracle.com/bpel/task">
  <partnerLinks>
    <partnerLink name="userTask" partnerLinkType="task:TaskManager"
      partnerRole="TaskManager" myRole="TaskManagerRequester" [...]/>
  </partnerLinks>
  [...]
</scope>
<assign name="setCaptureExamResultsInstanceCompleted"> [...] </assign>
<invoke name="sendCaptureExamResultsCompletedMessage" partnerLink="agentService"
  operation="processMonitoringMessage" inputVariable="captureExamResultsInstance" [...]/>
[...]
</scope>

```

Besides the XML representation of the *MonitoredObject* (here *CaptureExamResultsInstance*) the *MonitoringMessage* contains an additional message part holding information about the process instance ID along with the process ID. This information is required by the invoked monitoring agent for correlating the messages with the respective instances of the associated probe as well as the PPI monitor.

Our implementation of the monitoring infrastructure only consists of one monitoring agent for the presented PPI which handles both, the provision of probes and the calculation of the PPI. It would also be possible to decouple the provision of probes from the PPI monitoring. In doing so, the integration of an existing BAM system would be easier. The provided probes would have to be translated into events the BAM system understands within the scope of an appropriate adapter.

8 Conclusion & Outlook

In this paper, we presented the first steps towards a model-driven development of orchestrations along with the infrastructure for the monitoring of predefined PPIs. Thereby, the presented meta-model for the specification of the PPI monitoring along with the extension of the BPMN meta-model for modeling the required instrumentation and the sketched methodology for an automated generation of this instrumentation represent the main contribution of this work. In our future research we will try to achieve a fully automated generation of the orchestration instrumentation along with the monitoring infrastructure based on UML profiles for the meta-models and an adequate transformation language. Furthermore, we aim to extend the monitoring to a larger variety of *MonitoredObjects*, including more complex transactions with embedded sub transactions, and corresponding types of PPIs, especially aggregated PPIs. In this

case, the design of monitoring infrastructure would also have to be revised. In this context, we are planning on using WBEM standards (especially the Common Information Model (CIM) [BuST00]) in conjunction with WS-Management [DMTF06] for implementing the monitoring infrastructure. This would enable an integration of the underlying application management and hence allow for an integrated monitoring of business goals and the involved IT.

References

- [AaHW03] *van der Aalst, Wim M. P.; ter Hofstede, A. H. M.; Weske, M.:* Business Process Management: A Survey. In: Lecture Notes in Computer Science Band 2678. Springer-Verlag, Berlin 2003, S. 1-12.
- [ACDG03] *Andrews, T.; Curbera, F.; Dholakia, H.; Goland, Y.; Klein, J.; Leymann, F.; Liu, K.; Roller, D.; Smith, D.; Thatte, S.; Trickovic, I.; Weerawarana, S.:* Business Process Execution Language for Web Services 1.1.
<ftp://www6.software.ibm.com/software/developer/library/ws-bpel.pdf>, 2003-05-05, Abruf am 2006-07-21.
- [BKPS04] *Bayer, F.; Kühn, H.; Petzmann, A.; Schlossar, R.:* Service-oriented Architecture for Business Monitoring Systems. In: Panetto, H. (Hrsg.): Proceedings of the Workshop on Web Services and Interoperability (WSI'05). Hermes Science Publishing, Genf 2005, S. 127-134.
- [BaMR04] *Bauer, Bernhard; Müller, Jörg P.; Roser, Stephan:* A Model-Driven Approach to Designing Cross-Enterprise Business Processes. In: Lecture Notes in Computer Science Band 3292. Springer-Verlag, Berlin 2004, S. 544-555.
- [BuST00] *Bumpus, W.; Schweitzer, J. W.; Thompson, P.:* Common Information Model, John Wiley & Sons Ltd., New York, 2000.
- [DMTF06] *Distributed Management Task Force (DMTF):* Web Services for Management (WS-Management) Specification Version 1.0.0a,
http://www.dmtf.org/standards/published_documents/DSP0226.pdf, 2006-04-05, Abruf am 2006-07-21.

- [FrGJ04] *Frank, J. H.; Gardner, T. A.; Johnston, S. K.*: Business Process Definition Metamodel -Concepts and Overview.
<http://www.bpmn.org/Documents/BPDM/BPDM%20Whitepaper%202004-05-03.pdf>, 2004-04-08, Abruf am 2006-07-21.
- [EmWA06] *Emig, Christian; Weisser, Jochen; Abeck, Sebastian*: Development of SOA-Based Software Systems – an Evolutionary Programming Approach. In: International Conference on Internet and Web Applications and Services ICIW'06, Guadeloupe, 2006.
- [HaRa01] *Hauck, R.; Radisic, I.*: Service Oriented Application Management — Do Current Techniques Meet the Requirements?. In: New Developments in Distributed Applications and Interoperable Systems, Proceedings of the 3rd IFIP International Working Conference (DAIS 2001). Kluwer Academic Publishers, Krakau 2001, S. 295–304.
- [HeAN99] *Hegering, H.-G.; Abeck, Sebastian; Neumair, B.*: Integrated Management of Networked Systems: Architecture, Tools, Operational Implementation. Morgan-Kaufmann, San Francisco 1999.
- [JeSC03] *Jeng, J.-J.; Schiefer, J.; Chang, H.*: An Agent-based Architecture for Analyzing Business Processes of Real-Time Enterprises. In: Proceedings Seventh IEEE International Enterprise Distributed Object Computing Conference (EDOC'03). 2003, S 86-97.
- [JuMa05] *Juling, W.; Maurer, A.*: Karlsruher Integriertes InformationsManagement. In: Praxis der Informationsverarbeitung und Kommunikation (2005) 3, S. 169-175.
- [KHSW05] *Koehler, J.; Hauser, R.; Sendall, S.; Wahler, M.*: Declarative Techniques for Model-driven Business Process Integration. In: IBM System Journal (2005) 44, S. 47-65.
- [KKLP05] *Kloppmann, M.; Koenig, D.; Leymann, F.; Pfau, G.; Rickayzen, A.; von Riegen, C.; Schmidt, P.; Trickovic, I.*: WS-BPEL Extension for People (BPEL4People).

<ftp://www6.software.ibm.com/software/developer/library/ws-bpel4people.pdf>,
2005-07-01, Abruf am 2006-07-21.

- [Le03] *Leymann, Frank*: Web Services - Distributed Applications without Limits, Business, Technology and Web. Leipzig 2003.
- [LeRS02] *Leymann, Frank; Roller, Dieter; Schmidt, M.-T.*: Web Services and business process management. In: IBM Systems Journal (2002) 41, S. 198-211.
- [Mc03] *McGregor, Carolyn*: A Method to Extend BPEL4WS to Enable Business Performance Measurement.
<http://www.cit.uws.edu.au/research/reports/paper/paper03/TR-CIT-15-2003.pdf>,
2003-06-01, Abruf am 2006-07-21
- [McSc04] *McGregor, Carolyn; Schiefer, Josef*: A Web-Service based framework for analyzing and measuring business performance. In: Information Systems and e-Business Management. Springer-Verlag, Berlin 2004, S. 89-110.
- [MiMu01] *Miller, Joaquin; Mukerji, Jishnu*: Model Driven Architecture.
<http://www.omg.org/docs/ormsc/01-07-01.pdf>, 2001-07-01, Abruf am 2006-07-21.
- [MuRo04] *zur Muehlen, M.; Rosemann, M.*: Multi-Paradigm Process Management. In: Grundspenkis, Janis; Kirikova, Marite (Hrsg.): Proceedings of CAiSE'04 Workshops - 5th Workshop on Business Process Modeling, Development and Support (BPMDS 2004). Riga 2004, S. 169-175.
- [PAVB04] *Pignaton, R.; Asensio, J. I.; Villagra, V.; Berrocal, J. J.*: Developing QoS-aware Component-Based Applications Using MDA Principles. In: Eighth International Enterprise Distributed Object Computing Conference (EDOC 2004), 2004, S. 172-183
- [Wh04] *White, S. A.*: Business Process Modeling Notation. BPMN 1.0.
<http://www.bpmn.org/Documents/BPMN%20V1-0%20May%203%202004.pdf>,
2004-5-03, Abruf am 2006-07-21

Management of Portal Evolution

Introducing Evolution Management for the Corporate Financial Portal

Hong Tuan Kiet Vo, Helmuth Elsner

Institut für Informationswirtschaft und -management
Universität Karlsruhe (TH)
76131 Karlsruhe
{vo,elsner}@iism.uni-karlsruhe.de

Abstract

Software evolution is an essential concept underlying the engineering of corporate portals. Due to the complexity of such systems, it requires great effort and is not advisable to design and implement a feature-complete corporate solution. The concept of evolutionary portal development helps portal engineers to cope with design complexity. While the technical perspective on component based software development and evolution is widely discussed, little work addresses the actual management of software evolution, let alone in the portal engineering context. In this paper we focus on the management of portal evolution. On the basis of a portal engineering process model we discuss methods and practices that facilitate the management of portal evolution cycles. We further outline the evolution management procedure and tools that we have established for the corporate financial portal at Bayer and point out key lessons we have learned so far.

1 Introduction

Software evolution is an essential concept underlying the engineering of corporate portals. Due to the complexity of corporate portal systems that may even rival that of corporate ERP systems [Remu06], it is virtually impossible and not advisable to design and implement a feature complete corporate portal solution. The concept of evolutionary portal development helps portal engineers to cope with the design complexity and to foster user acceptance as iterative portal releases will incorporate actual user requirements gathered during portal operation. While the

technical perspective on software development and reuse is widely discussed (cf. component based software development), only a handful of work address the actual management of software evolution cycles, let alone in the portal engineering context.

In this paper we address the management of portal evolution i.e. we outline the processes and activities that are of importance to manage evolution cycles. On the basis of a portal engineering framework we discuss activities that help to specify direction for the next evolution cycles. We discuss the evolution management procedure and tools that we have established at the corporate financial portal project at Bayer and point out key lessons we have learned so far.

The paper is structured as follows. In the following section we discuss approaches for software evolution with focus on web applications. Section 3 outlines evolution management in the context of a portal engineering process model as proposed by the authors. Section 4 then illustrates the application of computer aided evolution management in the context of the corporate financial portal project. The paper closes with a summary and an outlook on future work.

2 Management of Software Evolution

The technical perspective of evolutionary software development is extensively discussed especially in the context of component based software engineering [BrWa96]. While these discussions focus on the question of how to facilitate reuse of software components, we are more interested in the actual management process that follows software evolution cycles. Yet, this question is often addressed only on a side note while the focus is on the iterative or evolutionary development process. Sommerville points out the importance of feedback or change proposals as the basis of software evolution and describes a generic change identification and evolution process cycle (cf. figure 1). Again, the focus is on the development process and Sommerville further states that the actual software evolution process will vary considerably depending on the type of software being maintained [Somm04].

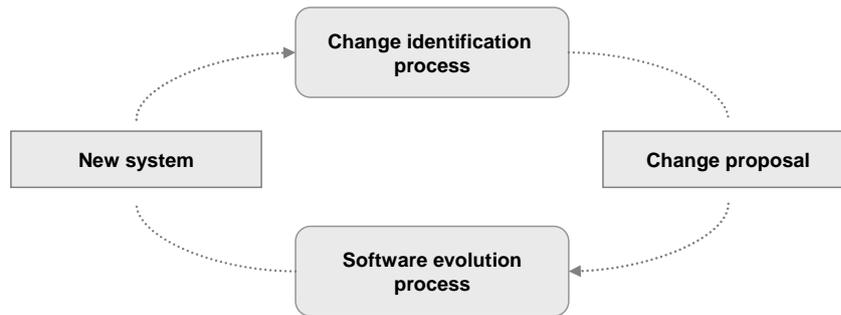


Figure 1: Change identification and evolution processes [Somm04]

With the current trend towards web application, software evolution is again becoming more important. One distinct feature of web applications is that it is no longer necessary to install the applications on the local machines, but all users access the current version of the software through their internet browser. Therefore aspects of software distribution and versioning are practically less important when developing for the web. That in return fosters iterative software evolution and development cycles. Software engineering for the web is discussed in the field of web engineering that emerged in the late 90s following the internet revolution and the increase of applications developed for this medium. In order to enable the development and maintenance of high quality Web-based system and applications Murugesan [MDHG99] proclaimed the need for web engineering that is *the establishment and use of sound scientific, engineering and management principles, and disciplined and systematic approaches to development, deployment and maintenance of Web-based systems* [MDHG99, p.6]. Web engineering distinguishes itself from traditional software engineering as aspects like software evolution, hardly predictable network behaviour, heterogeneous and usually unknown end-users with different cultural background are of greater importance when developing applications for the web [Powe98; Schw00]. With regard to software evolution, again current works primarily focus on the technical aspects of enabling the reuse thus evolution of web components [Gaed98]. Currently, dedicated discussions on the management of web application evolution are rare and often mentioned as a side note. That is surprising given the fact that the lifecycle of a web application not only comprises of develop, launch and maintain [Powe98] but also operation and iterative evolution. That strengthens our believe that for the engineering of web applications and especially web portals that are in a state of constant change, a structured management process that prepares and guides evolution cycles is necessary for a sustainable evolution.

3 Management of Portal Evolution

3.1 Management Evolution and Portal Engineering

As stated in section 2, research that particularly focuses on the management of evolution is rare both with regard to work on software engineering and web engineering. Following our literature review, this holds also true for current discussions on the engineering approaches for (corporate) portals.

Finkelstein and Aiken [FiAi00] present a framework for engineering enterprise portals. They especially focus on the design and implementation stage and illustrate how to implement a corporate portal using XML. The discussion centres around methods for data transformation and data modelling that facilitate the integration of structured and unstructured data. Portal evolution is discussed from a technical perspective only with regard on practices and architectures that enable future integration of data sources.

Alt et al. [ACLÖ04] outline a method for developing process portals. According to their definition, process portals are enterprise portals that especially focus on the integration of (external) customer processes and facilitate the collaboration between enterprises, customers and suppliers. Following the method engineering methodology, their method for developing process portals focuses on the portal strategy, portal design and portal technology stage. The notion of evolution is not within the scope of this work.

The Fraunhofer IAO portal analysis and design method – PADEM – follows the stages of a traditional software lifecycle: the definition of the strategy, subsequent requirement analysis, conceptual design, realisation and the introduction of the portal [GHKV04]. For each stage, the method defines checklists, questionnaire and offers practices that shall support companies in their portal selection and decision process and guide them through the portal implementation. PADEM focuses on providing guidance especially during the initial lifecycle of a portal implementation while taking a consulting and management level perspective on the heterogeneous decision problems. Evolution is not explicitly discussed although the authors state that there will be subsequent evolutionary portal releases.

Amberger et al. present a portal engineering process [AmRB04] that comprises of five stages: requirement analysis, analysis of profitability, detailed analysis of users, business processes and corporate IT, proof of concept prototyping and iterative evolution. However, the purpose of this

work is to give an overview of the field of portal engineering therefore the engineering stages – including evolution – lack a detailed further discussion.

In summary, existing frameworks on portal engineering agree on the importance and existence of iterative portal evolution. However, they lack an explicit discussion on how to prepare and guide such an evolutionary cycle. This is not surprising as portal engineering is quite a new field of research. Thus, in the next sections we specifically focus on the discussion of evolution management within the limits of a portal engineering process model that we outline in the following.

3.2 Evolution Cycles in Portal Engineering

Following traditional software engineering process models and lifecycles, figure 2 depicts a portal engineering process model. Dotted lines represent stage transitions while solid lines illustrate feedback relations among engineering stages.

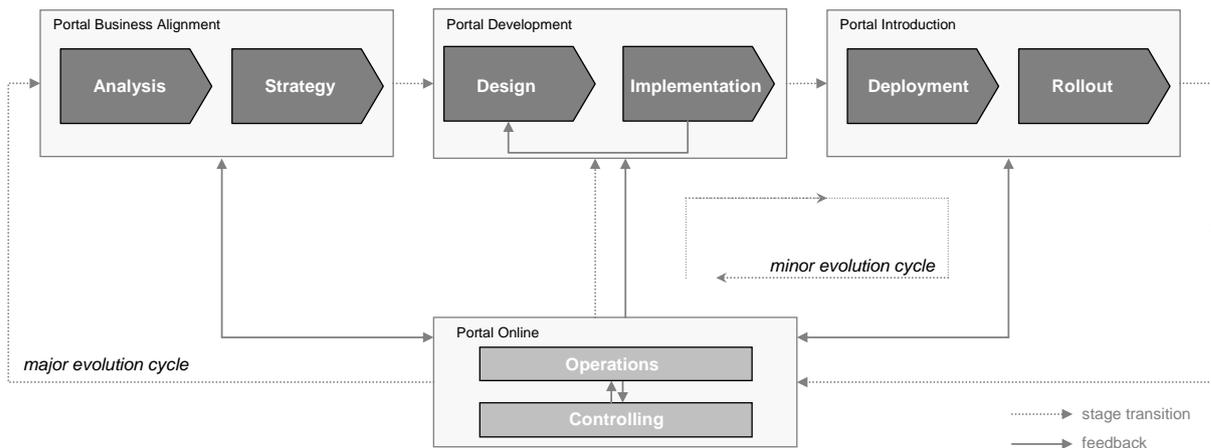


Figure 2: Portal engineering process model and evolution cycles

The portal engineering process model comprises of the *analysis* and *strategy* stages (portal business alignment), the *design* and *implementation* stages (development), the *deployment* and *rollout* stages (introduction) and finally the *operations* and *controlling* stages (online).

Following this process model, the notion of portal evolution is considered on two levels of magnitudes. On the one hand there are the major evolution cycles that require a thorough run through the engineering process from the refinement of the portal strategy to portal service rollout. Major evolution cycles are initiated according to the periodic project steering phases.

On the other hand there are minor evolution cycles nested within a major evolution cycles. A minor evolution is initiated whenever portal services are enhanced or improved in details only.

3.3 Evolution Management for Portal Engineering

According to the change identification and evolution process (cf. section 2, figure 1) every evolution process is triggered by feedback that comprises change proposals, bug reports or feature requests. For evolution management feedback has to be aggregated, categorized and evaluated to derive insights for the preparation of the next evolution cycle. While the majority of the feedback will be obtained during portal operation (i.e. user reported bugs or feature requests), feedback is also generated during portal development and of course following the rollout of a new portal release for example during training workshops. Note, that feedback should be regarded as desirable rather than as something that is unwanted. We believe that feedback is an indicator for actual system usage. In particular, feature requests can be regarded as indicators for the user's experience of the portal system as a *living* system that evolves over time according to the users need. Hence, to motivate the generation of user feedback, portal evolution management has to establish a process that first facilitates feedback reporting and second enables users to easily follow the status of their current requests.

Feedback that is accumulated over time has to be categorized and evaluated by the portal engineers. In general we distinguish between feature request for new applications, bug reports and change request. While we consider bug reports immediately in the next minor evolution cycle, feature and change requests must be evaluated in more detail. With regard to change requests, the developer of the application has to evaluate the redesign and implementation effort to decide on the priority of a change request and the initiator of this request should be notified on the decision. Small feature request are extensions to applications that depending on the request pipeline and the design and implementation effort can be addressed within minor evolution cycles. Major feature request often propose the development of completely new applications. The decision on the development and prioritisation is therefore a portal management level decision and handled respectively. Again the initiator of the request should be notified on the decisions. For the purpose of clear reference, in the following we use the term *issue* for any type of relevant feedback.

Figure 3 outlines the major stages of the portal evaluation management cycle with focus on the perspective of end-users as the primary source of relevant issues. Every evolution cycle starts

after the current batch of issues has been categorized and evaluated. Transparency as one key aspect to foster user participation in portal development requires the portal management to inform the portal users on the evaluation results of their reported issues. Relevant issues are processed in the subsequent development stage. During this stage, it is important to keep the users informed on the current processing status of their request. Once the requests have been implemented, the issue holder is involved in a limited introduction phase to evaluate the new developments. Upon approval, the issue is closed and the changes are introduced in a public release.

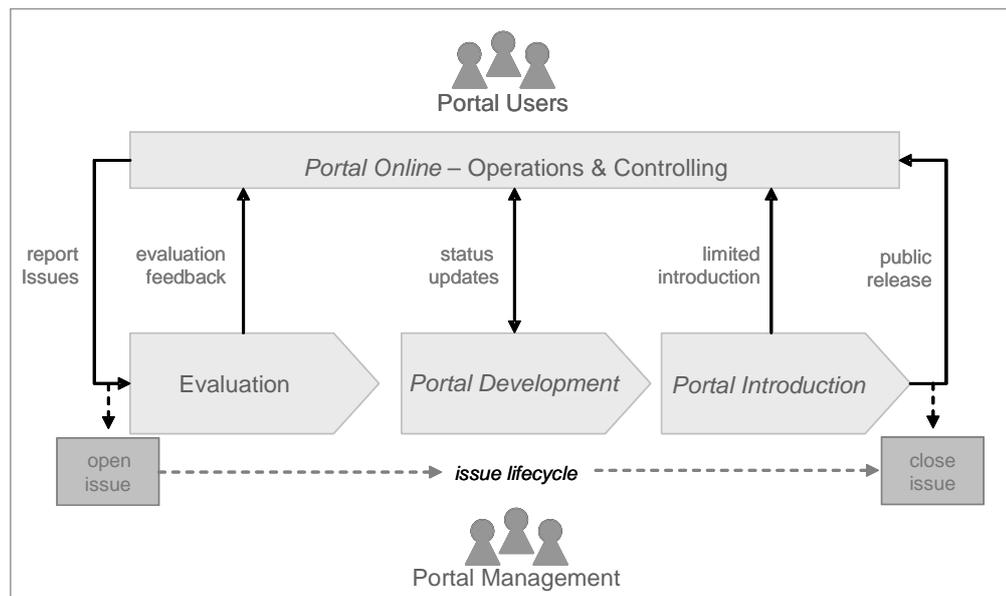


Figure 3: Portal evolution management cycle (issue life cycle)

In summary evolution management as proposed, centres around the management of an issue lifecycle starting with feedback collection, followed by issue categorization and evaluation to the final implementation of the requests. This process has to be supported by an issue management system that is tightly integrated with the issue processing stages. In the following sections, we present and discuss the implementation of the proposed evolution management process and the respective issue management system for Bayer's corporate financial portal.

4 Evolution and the Corporate Financial Portal

4.1 Overview: The Corporate Financial Portal Project

In this section we present the implementation of an issue management system in the Corporate Financial Portal (CoFiPot). The CoFiPot system is operated by the financial department of the Bayer AG. It was implemented with a few core financial services in 2001. Since then, the number of services provided by the system and the number of users has constantly grown. Today the portal is used by users from all over the world and is part of their daily work schedule. The goal for the following years is to extend the portal scope on further financial departments. For example the latest project focuses on providing services for the tax department partially based on their existing IT-solutions with the aim to greatly expand the range of services available to satisfy the information needs of the tax-accountants.

The portal system is based on Microsoft's .NET Framework and loosely on the IBuySpy¹ open source portal framework that at that time was introduced to demonstrate the development of personalized, component-based web applications using the just introduced ASP.NET technology. Portal development is coordinated by the financial department in cooperation with the IT-department. Right from the start the CoFiPot-project was accompanied by a research project offering the authors the possibility to follow the major stages of the corporate financial portal lifecycle.

4.2 The need for Evolution Management

In the initial development stage of the portal, the structure of the project and the number of users and developers were of manageable size. At maximum three developers were working simultaneously on the project and only a small number of services was offered to the users. The assignment of incoming development requests worked straightforward and did not require much administrative effort. The development itself took place on an ad-hoc basis. The communication between users and developers usually took place directly via phone or email. Due to the small development team there was no pressing necessity for a dedicated management of requests. Requests were simply stored and managed in an Excel file accessible for all team-members. New features and changes to the portal system were brought to attention of the development

¹ Current open source portal frameworks like DotNetNuke are derived from the IBuySpy framework, which currently is discontinued.

team via email or in meetings. However, a real documentation about the change proposals and their current status did not exist. Notifications about new features to the users were issued infrequently via mail and often only directly to the target user group.

When the portal system started to further extend its services the number of users, feature requests, change proposals and bug reports consequently began to rise. Due to the growth of the project the development team size increased as well. The previous Excel-based issue management system reached the limit of its scalability. The development-process required a more sophisticated approach that would allow handling a large-scale web application project with a growing number of users and developers. Last but not least with the new portal evolution management process we wanted to offer the portal users a new level of transparency with not only the possibility to report new change proposals but also to monitor the current status of their proposals. The key points for the process development thus were:

- a clearly defined process for the collection and management of requests
- a traceable log of all development steps
- transparency to the users

An issue-management system to support the process was designed and implemented in the portal. This system and the underlying process are presented in the following.

4.3 Corporate Financial Portal Issue Management

The main goal of the issue management system is to facilitate the process to collect and process incoming requests. On the one hand this requirement should enable the developers to keep an overview of all open issues and to minimize the effort to administrate them. Moreover, in order to facilitate planning for the whole project, the system should enable the team to categorize issues by their priority and impact. On the other hand the users of the portal should have the option to easily report issues. With the previous system at least they had to write an email or to call the developers to describe their issue. With the new system they are able to report an issue directly out of the portal system with little effort.

An important topic while designing the fundamentals of the issue management process was the question how to manage the status of a request throughout the development stages as shown in figure 3.

One of the initial questions when designing the issue management process for the Bayer system was how to handle multiple issues that refer to the same topic. In case of a bug in a component many users might report it as an issue. Therefore there is the requirement to link issues to each other so that the processing of multiple issues is combined in one process. The chosen solution contained the step to differentiate between incoming requests as issues and tickets as actual working orders. A ticket thereby combines any number of issues and has a clear key for identification. Issues only identify the original request and provide initial information about the task but they don't change any more, once they have been assigned to a ticket. While any new issue might deal with already recorded problems, each ticket handles a unique topic. If required the assignment of an issue to a ticket can change to another ticket.

The application of the core sequence as presented in figure 3 to the process for managing the CoFiPot system is shown in figure 4.

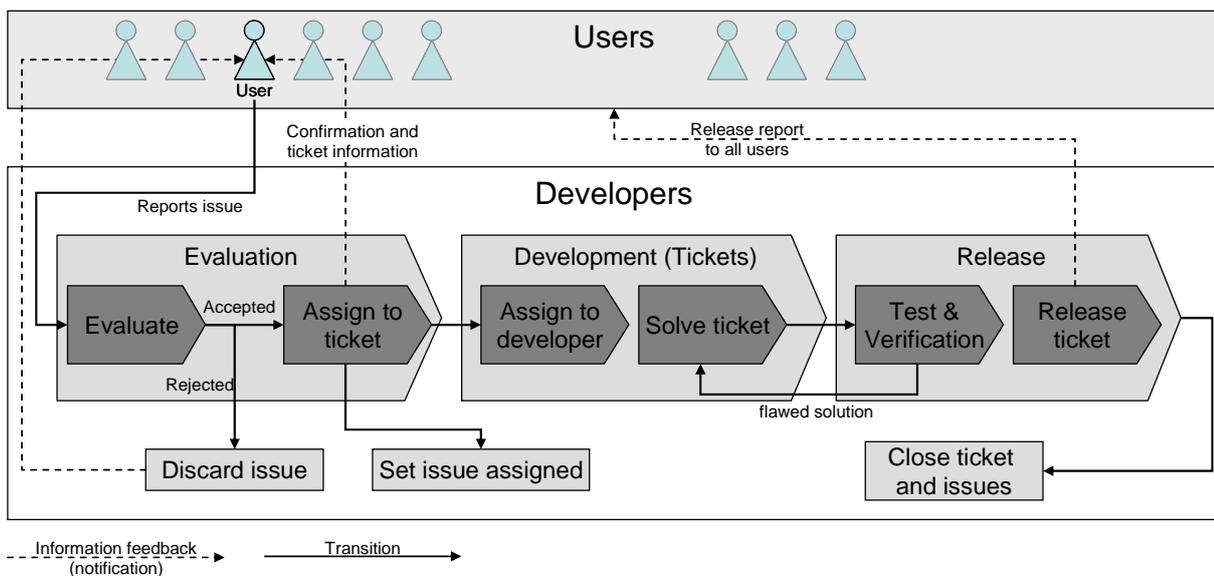


Figure 4: Issue life cycle with ticket handling

At the initial stage an issue is reported and stored in the database. The developers evaluate the request and decide whether to create a new ticket based on this issue, to assign it to an already existing ticket or to reject the topic. When the issue is accepted the resulting ticket will be categorized by priority and the required effort to solve it. In the development stage the ticket is assigned to a developer who will solve the case. When the developer sets the status of the ticket to *solved* the solution is to be tested and verified. Upon a successful evaluation the solution is released to the public and the ticket's status is set to *resolved* thus taking the ticket and all

related issues off the development list. Throughout all these stages the user who reported the initial issue will be informed upon major progresses.

In section 4.3.2. we will show the life-cycle of an issue as it is handled by the CoFiPot-portal in more detail.

Tickets are the actual foundation for the work of the development team. New tickets have the flag "open" and can be assigned to a responsible person. Throughout the working process, the information contained in a ticket can grow through the addition of comments and attachments. Once the developer marks it as resolved, preferably another member of the team verifies the solution and either reassigns the ticket for further work or finalizes the process by setting the ticket's status to *verified*. Every new release automatically will contain all verified tickets.

4.3.1 End user integration

A further goal of the issue management system was to further integrate the users in the development process. First of all they should easily have the option to report bugs or propose changes to the system. Further more they should be able to follow the ongoing development process and learn about the progress of any issues that they have reported or that caught their attention.

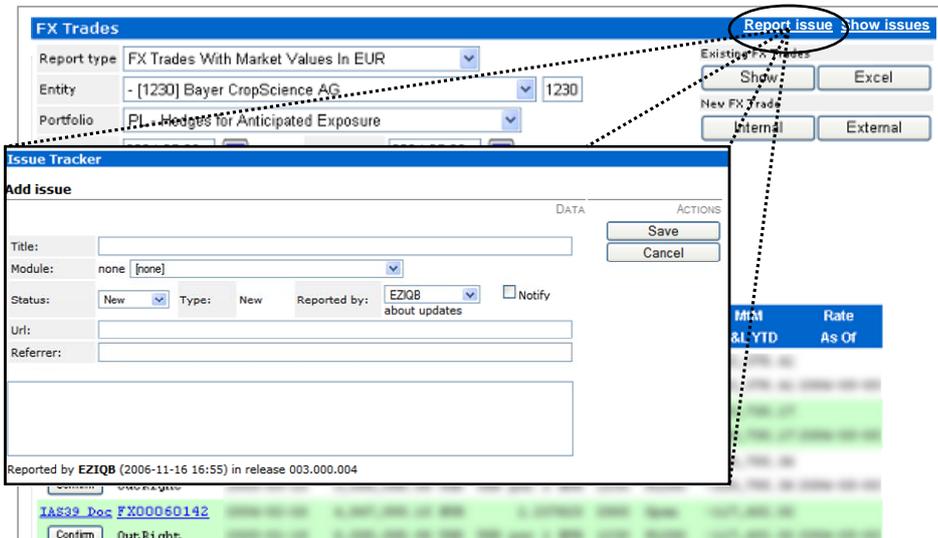


Figure 5: Report issue and module issues functionality on the CoFiPot live pages

The issue management system allows any user to directly report an issue through the portal system. Each module offers the option to directly report an issue or to view all current issues that apply to the module. Figure 5 shows a module with the described functionality and the form that is displayed to the user for reporting an issue.

Also, in case that an error occurs, an error report is automatically generated and stored as an issue. It is displayed to the user and provides input fields to add further information. For each reported issue the user can choose to be notified via email about its progress.

Notifications are sent upon the assignment of an issue to a ticket and when a resolved ticket is included in a release as shown in the process in figure 4. Through this notification service the users do not need to check the progress manually.

Further more the users have the option to view all Issues and Tickets of each component of the portal (see previous figure). Thus they can on the one hand learn which changes are planned and on the other hand follow up the past updates to the component. Thus the understanding of the components can increase. The users might even want to contribute more proposals for further improvements because they can see that issues are followed up.

4.3.2 Example case of an issue live-cycle

In this section we shortly want to follow the stages of an issue and the resulting ticket as they are handled with the issue management system.

Once an issue has been entered as described in the previous section, it will appear in the overview list shown in figure 6. One member of the development team evaluates the issue and assigns to a new or an existing ticket. For the ticket various fields can be set to categorize it.

Key	Title	Module	Type	Status	Created	Ticket
003.000.003-D0004	Missing chart content	Finance Charts	Defect	New	2006-11-14	---
003.000.003-D0003	[BY7W84] Unexpected Error (LSUNH): /Desktopdefault.aspx?tabid=123] System.OverflowException	Planning Monitor	Defect	Assigned	2006-11-13	Missing validation for...
003.000.003-D0002	Drill-Down fails	Financing Monitor	Defect	Assigned	2006-11-13	Drill-Down fails
003.000.003-D0001	Excel download fails for most rate types	Official Bayer Rates	Defect	Assigned	2006-11-10	Excel download fails f...
003.000.002-D0002	[BY7W84] Unexpected Error (KFCO): /DesktopDefault.aspx[System.ArgumentException	---	Defect	Assigned	2006-11-07	Fix minor bugs in Cofi...
003.000.002-D0001	Logoff.aspx may cause System.InvalidCastException	---	Defect	Assigned	2006-11-07	Fix minor bugs in Cofi...

Figure 6: Issue overview for the developer-team

These fields define the priority, severity and the type of the ticket. With these settings, the next development steps of the portal can be identified. While tickets that describe major changes that lead to a major development cycle, minor issues such as bug fixes can be handled directly.

After a ticket has been assigned to a developer and it has been fixed and the solution has been tested and approved by another person, its status is changed to *verified* and it can be included in the next release. When a new release is planned the responsible administrator prepares it using the issue management system. Each release is stored in the system as a collection of all verified tickets. Upon creation a list of all these tickets is displayed as show in figure 7 including information about changes that need to be taken care of before the portal code can be released. Once the release has been published a list of all changes can be obtained through the system and is also accessible for all users of the portal.

The screenshot shows a web form titled "Edit release" with an "INFO" tab. The form contains the following fields:

- Key:** 003.000.004
- Name:** Release
- Comment:** (empty text area)
- Release date:** 2006-07-14

Below the form is a section titled "Assigned tickets" with a "TICKETS" tab. It contains a table with the following data:

Key	Title	Type	Priority	Status	Responsible	Module
<input type="checkbox"/> 003.000.003-0004	Missing validation for as-of date	Fix	Minor	Verified	KFRWO	Planning Monitor
<input type="checkbox"/> 003.000.003-0003	Generating Limit Report takes over 2 minutes	Performance	Minor	Verified	KFRWO	Commodity Monitor
<input type="checkbox"/> 003.000.003-0001	Excel download fails for most rate types	Fix	Major	Verified	PSMBO	Official Bayer Rates

At the bottom of the table are links for "check all" and "uncheck all". Below the table is a section for "Deployment notes".

Figure 7: Release preparation form with assigned tickets

4.4 Lessons Learned

In the first months following the introduction of the issue management system, only a few selected users were allowed to access the features. But despite the restricted number, the new working process already had some significant influence on the development process.

Starting right from the beginning, the users filed all kinds of issues that they often had already had in mind for a longer time. Previously they would have needed to send an email or to directly tell it to one developer or even to several developers individually and than hope that one of them accepted or at least noted down the issue. Now they can add it to the list of open issues and know that the idea is at least in the list of open topics. Thus the barrier for reporting issues

turns out to be significantly lower than with the previous system. With the system, each issue can clearly be identified including information about its reporter and the creation date. This clear identification facilitated many talks between users and developers. They can refer to the issue and don't need to spend much effort on initially identify the problem and whether it was subject to discussion before.

Although the centralized collection of all issues significantly increased the number of open topics, it still had a positive effect for the work of the developers. First of all they do not need to enter each issue manually by themselves as they had to do with the previous Excel-based approach. Further more every user is now responsible for adding an issue to the system. The developers can ask every user that approaches them with a new issue, to report the issue in the issue management system themselves.

Comparing the number of resolved issues between the new system and the Excel-file, there is a significant difference. The old system required the developer to open the file and enter issue when it was resolved. Especially for minor changes the effort was seldom taken. There was also no immediate reason to enter each change. The new system requires each issue and resulting ticket to be processed. Thus each update is documented. For example if the developer forgets to set the status of a ticket to resolved, it will remain on his list of open tickets. Clearly he will try to reduce the number of tickets on the list and thus take care that he records each progress.

Not only did the quality of change-documentation improve but also could the work of the developers better be identified as for each ticket the actual time required to solve it is noted down. This also grants the possibility to analyse the development efforts and helps creating better estimates for new tickets.

The new process also improved the collaboration of the development team. Through the centralized system every team member can follow up all changes and open issues. The information gap between the members decreased. Through the clear assignment of tickets responsibilities became a more seldom topic of discussion. Each change and open task can now be triggered to one person. Time-consuming research to identify the right contact person is in most cases obsolete.

Future improvements will focus on further increasing the transparency of the development actions. The goal is to further encourage the users to observe ongoing issue developments. Thus they will learn more about features of the portal that might contribute to their work. Further on

they might be motivated to think about possible improvements to the system and thus contribute to its quality.

Another interesting topic for further analysis is the quality of issues who are added by the users through the portal. The question is what types of requests are made – whether they mainly describe minor issues or if they also include major change proposals. So far most reported issues did not contain the later but it must be considered, that the number of users was limited and that time for adaptation might be necessary before the users feel confident about adding more advanced issues.

In addition further analysis tools will be added to the issue management system to create reports about the development work that is performed. A long-term approach could lead to the integration of the ticket planning data into a project management application in order to improve the coordination of the developer team.

5 Conclusion

In this work we addressed evolution management in the context of portal engineering. On the basis of a portal engineering process model, that is derived from the traditional software lifecycle process we identify two different evolution cycles: minor evolution cycles that are nested in major evolution cycles. While major evolution cycles require a run through the complete portal engineering process, minor evolution cycles take places within two major cycles and by nature address smaller changes in the current portal iteration (e.g. bug fixes or the introduction of on breaking changes). Both, minor and major evolution cycles are based on feature requests, change proposals and bug reports, which must be managed and evaluated within the scope of a portal evolution management concept. We discuss the core concepts underlying an evolution management approach in theory and prove the feasibility of our claims in practice on the basis of the evolution management procedure established for the corporate financial portal at Bayer. Of course, the validity of our statements is subject to the general limitation of a case study. We cannot prove whether the framework will also be applicable on other corporate portal initiatives. Still, we believe that the core insights can also be transferred to other cases as the evolution management process and the supporting tool was not designed with any functional or organizational specifics in mind.

References

- [ACLÖ04] *Alt, Rainer; Cäsar, Marc A; Leser, Florian, Österle, Hubert; Puschmann, Thomas; Reichmayr, Christian; Zurmühlen, Rudolf*: Methode zur Entwicklung von Prozessportalen, Lösungen, Bausteine und Potenziale des Business Networking. Real-time Business. R. Alt und H. Österle., Berlin, Heidelberg, Springer 2004. p. 258-280.
- [AmRB04] *Amberger, Michael; Remus, Ulrich, Böhm, Martin*: Entwicklung von Unternehmensportalen. WISU 33(5), 2004, p. 658-665.
- [BrWa04] *Brown, Allen W.; Wallnau, Kurt C.*: Engineering of Component Based Systems. In: 2nd IEEE International Conference on Engineering of Complex Systems (ICECCS96), IEEE 1996.
- [FiAi00] *Finkelstein, Clive; Aiken, Peter*: Building Corporate Portals with XML. New York, McCraw-Hill 2000.
- [Gaed98] *Gaedke, Martin*: Web Composition: Ein Unterstützungssystem für das Web Engineering. In: GI Softwaretechnik-Trends 18(3) 1998, p. 20-25.
- [GHKV04] *Gurzki, Thorsten; Hinderer, Henning; Kirchhof, Anja; Vlachakis, Joannis*: Die Fraunhofer Portal Analyse und Design Methode (PADEM) - Der effiziente Weg vom Prozess zum Portal, Fraunhofer IAO 2004.
- [MDHG99] *Murugesan, San; Deshpande, Yogesh; Hansen, Steve; Ginige, Athula*: Web engineering: A New Discipline for Development of Web-Based Systems. In: First ICSE Workshop on Web Engineering, Los Angeles, Springer 1999, p. 1-9.
- [Powe98] *Powell, Thomas A.*: Web Site Engineering, Prentice Hall, 1998.
- [Remu06] *Remus, Ulrich*: Critical Success Factors of Implementing Enterprise Portals. In: 39th Hawaii International Conference on System Sciences 2006.
- [Schw00] *Schwickert, Axel C.*: Web Site Engineering. Stuttgart, Teubner, 2000.
- [Somm04] *Sommerville, Ian*: Software Engineering, Addison-Wesley 2004.

Einführung in den Track

Supply Chain Management

Dr. Jens Gottlieb

SAP AG

Prof. Dr. Stefan Voß

Universität Hamburg

Prof. Dr. Dr. h. c. mult. Peter Mertens

Universität Erlangen-Nürnberg

Der Track fokussiert die folgenden Themengebiete:

- Erfahrungen mit Komponenten, z.B. SNP, ATP, CTP
- Entscheidungsunterstützende Systeme, Optimierung z.B. APS
- Modellierung von Supply Chains
- Agentenbasierte Koordination
- Electronic Procurement und Supply-Chain-Verträge
- Supply Chain Event Management
- Reverse Logistics
- Anbindung an Standardsoftware
- Vergleich von kommerzieller Software

Programmkomitee:

Prof. Robert Bixby, PhD, Rice University and ILOG Inc., Houston, USA

Prof. Dr. Andreas Fink, Helmut-Schmidt-Universität Hamburg

Dr. Oliver Klaus, Arthur D. Little AG Zürich, Schweiz

Prof. Dr. Gerhard F. Knolmayer, Universität Bern, Schweiz

Prof. Dr. Herbert Meyr, Wirtschaftsuniversität Wien, Österreich

Dr. Sven Spieckermann, Simplan AG, Maintal/Main

Prof. Dr. Hartmut Stadtler, Universität Hamburg

Prof. David L. Woodruff, PhD, University of California at Davis, USA

Regelbasierte Koordinierung von agentengestützten Transportprozessen

Mathias Petsch¹, Dirk Pawlaszczyk², Hagen Schorcht¹

Institut für Wirtschaftsinformatik
Fachgebiet Wirtschaftsinformatik für Dienstleistung¹
Fachgebiet Wirtschaftsinformatik für Industrie betriebe²
Technische Universität Ilmenau
98694 Ilmenau

{mathias.petsch, dirk.pawlaszczyk, hagen.schorcht}@tu-ilmenau.de

Abstract

Die Planung und Gestaltung effizienter unternehmensübergreifender Transportnetze rückt zunehmend in den Fokus betriebswirtschaftlicher Betrachtungen. Für eine flexible überbetriebliche Koordination und informationstechnische Integration von Güterflüssen in Lieferketten wird dabei immer häufiger der Einsatz von Softwareagenten diskutiert. Zur Lösung des Koordinationsproblems innerhalb von Transportnetzen werden Allokationsmechanismen, beispielsweise Auktionsprotokolle, vorgeschlagen. Regeln kommt an dieser Stelle eine zentrale Bedeutung zu, da sie zur Lenkung des Verhaltens und als Ausdruck der Ziele von Akteuren dienen. Im vorliegenden Beitrag wird ein Ansatz zur regelbasierten Koordination agentengestützter Transportprozesse vorgestellt. Um eine Überwachung zwischenbetrieblicher Geschäftsprozesse zu ermöglichen, ist es notwendig technische und organisatorische Rahmenbedingungen auszuhandeln und verbindlich zu kodifizieren. Als Grundlage dient eine entsprechende Regelontologie, die zur Aushandlung gemeinsam abgestimmter Geschäftsregeln bzw. Prozessbedingungen genutzt werden kann und Rechte bzw. Pflichten der Agenten abbildet.

1 Einführung

Die Transportdisposition im Güterverkehr, genauer die Planung der Transportkapazitäten und die Feinabstimmung von Fahrplänen und Touren unterliegt naturgemäß einer hohen Dynamik.

Erschwerend kommt hinzu, dass Transportnetze grundsätzlich offene Systeme bilden, die durch Unsicherheiten und Informationsasymmetrien gekennzeichnet sind. Bei der Disposition von Transportaufträgen muss entsprechend häufig ad hoc - aus der Situation heraus - entschieden werden. Die raumzeitliche Verteilung und weitgehende Autonomie aller Akteure in solchen Netzwerken sowie die resultierende Koordinationskomplexität und die Notwendigkeit zur Gestaltung hoch skalierbarer Prozesse über das gesamte Netzwerk hinweg, erfordern gänzlich neue, weit über derzeit verfügbare Ansätze hinausgehende IT-Lösungen.

Das globale Verhalten derartiger Logistiksysteme ergibt sich aus der Interaktion einer Vielzahl von lokal agierenden, über nicht lineare Wechselwirkungen miteinander vernetzten Individuen. Die aus derart komplexen Beziehungsgefügen resultierenden Dynamiken können nur in Einzelfällen mathematisch exakt verstanden werden. Um dennoch mögliche Ordnungszustände in diesen Systemen analysieren und emergente Phänomene abbilden und kontrollieren zu können, werden zunehmend agentenbasierte Modelle genutzt. Die Subjekte des abzubildenden Objekts sind dabei durch Agenten, eine Form autonom agierender Softwareentitäten, innerhalb des Modells repräsentiert. Derartige Modelle finden insbesondere im Rahmen der Konzeption, Planung sowie Steuerung logistischer Systeme Anwendung ([GrGr05; DaHe04]). Mit dem Einsatz von Softwareagenten wird überdies eine höhere Flexibilität angestrebt, beispielsweise im Zusammenhang mit einer verbesserten Reaktionsfähigkeit auf Störeinflüsse innerhalb des Supply Chain Managements ([DaHe04; BBGL⁺04]).

Die Entitäten bzw. Akteure der softwaretechnisch abzubildenden Domäne werden durch technische und nichttechnische Agenten repräsentiert. So werden neben möglichen Entscheidungsträgern der beteiligten Organisationen (Auftraggeber, Disponenten, Empfänger) zumeist Transportmittel (LKW, Bahn), Verlade- und Umschlagplätze, mitunter selbst die zu befördernden Stückgüter durch Softwareagenten abgebildet.

Um die eigentliche Dynamik des Systems nachbilden zu können, beispielsweise die gemeinsame Durchführung von Transportaufträgen, müssen Agenten in Form von Agentengruppen, auch als Multiagentensystem (MAS) bezeichnet, zusammenarbeiten. Diese Systeme werden für die Durchführung einer gemeinsamen Aufgabe gebildet und sind durch den stetigen Ein- und Austritt von Agenten in oder aus dem Verbund gekennzeichnet. Trifft beispielsweise ein Transportagent an einem Umschlagplatz ein, so muss er spätestens bei seiner Ankunft die Modalitäten bzgl. Entladung oder möglicher Anschlusstouren mit einem entsprechenden Disponenten klären. Bis zu seiner Weiterfahrt ist der Transportagent, genauso wie alle anderen

ankommenden Transportagenten somit Teil des Multiagentensystems, welches durch den Umschlagplatz repräsentiert wird.

Die überwiegende Mehrzahl bestehender Beiträge zum Thema agentenbasierter Planung und Steuerung von Transportnetzen fokussiert auf Koordinationsaspekte, die sich durch die Zusammenarbeit der unterschiedlichen Agenten ergeben, da diese das Systemverhalten und damit die Effizienz des Logistiksystems maßgeblich determinieren. Im Allgemeinen wird dies mit dem Einsatz entsprechender Allokationsmechanismen gleichgesetzt. Außer dem eigentlichen Verlade- und Transporttermin und der Zuweisung einer Route müssen die beteiligten Parteien gleichsam über verschiedene Rahmenbedingungen (Constraints) übereinkommen, deren Einhaltung überprüfbar sein muss. Dem Aspekt der Festschreibung dieser Constraints wird bisher wenig Beachtung geschenkt. Alle Mitglieder eines Multiagentensystems sind bestimmten Rechten und Pflichten unterworfen, die beispielsweise in der Ausführung von Aufgaben oder einer Beschränkung von Ressourcen bestehen können.

Regeln dienen hierbei zur Lenkung des Verhaltens und als Ausdruck der Ziele der Akteure. Sie haben auch die Funktion, andere Agenten auf ein gemeinsames Ziel auszurichten sowie Pflichten und Rechte zu vermitteln. Eventuelle Verletzungen der Regelmenge können systemintern durch Sicherheitsmechanismen, wie z.B. Reputationsmechanismen, überwacht und auftretende Verstöße durch Sanktionen geahndet werden. Regeln können somit das Verhalten des Agenten determinieren und koordinieren [TaWa02]. Durch das Festschreiben von verbindlichen Regeln können sich Agenten in offenen Systemen auf den verschiedensten Plattformen frei bewegen und mit „fremden“ Agenten kooperieren. Grundvoraussetzung hierfür bildet die Möglichkeit, Regeln automatisch zu prüfen, was eine geeignete maschinenverständliche Formalisierung voraussetzt.

In diesem Beitrag wird ein Konzept für die Aushandlung von Regeln betrachtet (Regelabgleich). Als Basis dient eine entsprechende Regelontologie, die als Grundlage für die Vereinbarung von Regeln fungiert. Nur wenn die Kooperationspartner diese kennen, verstehen und akzeptieren, ist eine erfolgreiche Zusammenarbeit möglich. Es werden verschiedene Regelarten unterschieden und mittels Object Constraint Language (OCL) formalisiert, mit dem Ziel, eine Überwachung agentenbasierter Geschäftsprozesse zu ermöglichen.

2 Grundlagen

2.1 Agentenbegriff

Agenten können als Softwareartefakte verstanden werden, die fähig sind, ihr Umfeld wahrzunehmen und zu verändern [RuNo03]. Sie können mit anderen Agenten Informationen austauschen, wobei ihr Lebenszyklus nicht direkt von dem eines anderen Agenten abhängt: „Agents are situated in an environment, act autonomously, and are able to sense and to react to changes“ [KnTi99]. Neben Ihrer Autonomie sind Softwareagenten häufig durch soziale Eigenschaften charakterisiert, was sich u.a. in der Fähigkeit widerspiegelt, benachbarte Systeme zu erkennen und mit diesen zu kommunizieren [WoJe95]. Multiagentensysteme bestehen wiederum aus Agenten, die zum Zweck der Lösung einer gestellten Aufgabe mit anderen im System vorhandenen Agenten kooperieren. Durch den Einsatz der Agententechnologie kann eine Verbesserung der betrieblichen Flexibilität und Reduzierung von softwarespezifischen Kosten realisiert werden (vgl. [CoJe96; Buss98; Paru00; Anth03]). Komplexe realwirtschaftliche Problemstellungen werden somit effizient lösbar. Der Einsatz der Agententechnologie erscheint insbesondere dann sinnvoll, wenn (1) der betrachtete Anwendungsbereich durch eine natürliche verteilte Struktur charakterisiert ist, die nicht erst künstlich aus einer zentralen Struktur erzeugt werden muss, (2) eine dynamische Umwelt vorliegt, in der strukturelle Änderungen berücksichtigt und flexibel auf Änderungen reagiert werden muss und (3) die abzubildenden Interaktionsbeziehungen derart umfangreich sind, dass eine komplexe Koordination zwischen den einzelnen Entitäten unverzichtbar ist [Müll97].

Eine inhärente Verteilung und dezentrale Steuerung über unterschiedliche wirtschaftlich oft selbständig handelnde autonome Organisationen hinweg prägen den Anwendungsbereich der Transportnetze ebenso wie ein hoher Interaktionsbedarf, der durch die Vielzahl an unterschiedlichen Randbedingungen und dem damit verbundenen hohen Kommunikations- bzw. Koordinationsaufwand bedingt ist. Agenten bieten hier eine geeignete Metapher, um Handlungsträger und deren Interaktionsbeziehungen innerhalb eines solchen Modells abzubilden. Im Vergleich zu konventionellen Ansätzen des Operations Researchs ist die Agententechnologie stärker an den praktischen Anforderungen der Transportdisposition ausgerichtet und bietet ein adäquates Modell für die Abbildung der Transportdomäne.

2.2 Regeln

In der Informatik wird häufig der Policy-Ansatz für die Formulierung von Rahmenbedingungen und sicherheitsrelevanten Einschränkungen verwendet (z.B. Java). Das Java Policy-Konzept zeichnet sich dadurch aus, dass im Gegensatz zu den meisten Sicherheitsmechanismen von Betriebssystemen, bei denen der Zugriff auf Ressourcen und Daten offen ist und durch den Nutzer beschränkt werden muss, Zugriffsbeschränkungen bestehen und explizit freigegeben werden müssen [FrMu96]. Dabei sind Policies (insbesondere Java Policies) darauf beschränkt, Ressourcenzugriffe zu limitieren, d.h. Freigaben auf Daten, externe Programme oder Ressourcen einzuschränken bzw. zu erlauben. Der vorgeschlagene Ansatz der Verwendung von Geschäftsregeln beschränkt sich jedoch nicht auf die alleinige Freigabe von Ressourcen. Es ist vielmehr möglich zusätzlich organisatorische Anforderungen bzw. Regelungen, wie Zugehörigkeiten, institutionelle Beziehungen, Aufgabenanforderungen oder Kommunikations- und Interaktionsprotokolle einzuschließen. Folglich werden nicht nur reine Zugriffsbeschränkungen definiert, sondern es können auch organisatorische und institutionelle Regelungen spezifiziert werden, die über das Policy-Konzept hinausgehen.

In der Organisationstheorie wird die Gesamtheit aller formalen Regeln zur Arbeitsteilung und Koordination als formale Organisationsstruktur bezeichnet [KiKu92, 18]. Diese Regeln dienen neben dem Aspekt der Leistungs- auch der Herrschaftssicherung, d.h. die Mitglieder müssen im Rahmen der vertraglich geregelten Mitgliedschaftsbedingungen auch vorgegebene "Verfahrensvorschriften und Regeln" [KiKu92, 18] akzeptieren, die letztendlich die Hierarchie bestimmen. In Multiagentensystemen kann es notwendig sein, dass Agenten während der Laufzeit das MAS wechseln, d.h. ein MAS verlassen, um sich aus Gründen der verfügbaren Ressourcen bzw. Problemlösungsmöglichkeiten einem anderen Multiagentensystem anzuschließen. An o.g. Szenarios wird deutlich, dass der Transportagent jeweils Mitglied des MAS Umschlagplatz wird, die sowohl über eigene technologische (beispielsweise verwendete Kommunikations- und Interaktionsprotokolle), als auch organisatorische Regeln (beispielsweise verantwortlicher Disponent ist folgender Agent, Touren werden nach einem gewissen Modus vergeben, bestimmte Umschlagplätze sind nur für gewisse Frachten zugelassen usw.) verfügen können. Folglich sind MAS in unterschiedliche organisatorische Kontexte eingebunden, bzw. unterstützen trotz aller spezifizierten und implementierten Standards unterschiedliche Technologien, Protokolle etc. Dabei kann es notwendig werden, dass Agenten sich den in dem entsprechenden MAS vorherrschenden Regeln (zur Verwendung bestimmter Protokolle, organisatorischer

Regelungen o.ä.) unterwerfen müssen. In der Folge bedeutet das, dass Agenten ebenso gegebene Verfahrensvorschriften und Regeln innerhalb eines Multiagentensystems akzeptieren und befolgen müssen, d.h. die Beziehungen des Agenten zu dieser Organisation (MAS) sind vertraglich zu regeln, ohne dass jedoch die Autonomie (Individualität) des Individuums Agent nachhaltig negativ beeinflusst wird. Die Gesellschaft kennt zu diesem Zweck unter verschiedene Vertragsformen, die den Grad der Mitgliedschaft mit den jeweiligen Rechten und Pflichten regeln (z.B. Arbeitsvertrag; Werkvertrag, etc.).

Organisatorische Regeln müssen jedoch nicht explizit vorgegeben werden, sondern können sich dynamisch durch Koordination der Organisationsmitglieder untereinander bilden bzw. Ergebnis eines kollektiven Lernprozesses sein [KiKu92, 21]. Laut Scherer [Sche99, 1] werden Organisationen als System von expliziten und impliziten Regeln beschrieben, die einem bestimmten Zweck dienen und Erwartungen kommunizieren, sich in einer bestimmten Art und Weise zu verhalten. Diese Regeln können die Individuen normalerweise nicht beeinflussen und unterwerfen oder entziehen sich ihnen. Dabei müssen die Regeln interpretiert und in verschiedene Handlungen umgesetzt werden [Kies99, 303]. Um eine einheitliche Übereinstimmung der Interpretation zu erhalten, kommunizieren die Organisationsmitglieder untereinander.

Im Softwareengineering werden seit einigen Jahren Geschäftsregeln (auch Business Rules) zur Beschreibung von Bedingungen und Restriktionen von Geschäftsprozessen verwendet. Hierbei liegt der Fokus vor allem auf der Darstellung von Strukturen, deren Abhängigkeiten sowie Möglichkeiten der Steuerung und Kontrolle von Geschäftsprozessen im Rahmen der Gestaltung von Informationssystemen. Geschäftsregeln können als Beschreibung von Vorgaben und Restriktionen von Geschäftsprozessen sowie von deren Zuständen und Abläufen verstanden werden [HeKn94, 1]. Gemäß der Definition der Business Rules Group zu Geschäftsregeln [BrGr05], können zwei Perspektiven eingenommen werden (siehe Tab. 1):

	Definition
Business perspective	“...a business rule is a directive, intended to influence or guide business behavior, in support of business policy that has been formulated in response to an opportunity, threat, strength, or weakness.”
I/S perspective	“...a business rule is a statement that defines or constrains some aspect of the business. It is intended to assert business structure, or to control or influence the behavior of the business.”

Tab1: Business Rules [BrGr2005]

Durch diesen Ansatz konnte erstmalig die lange diskutierte Zusammenführung von Geschäftslogik, Präsentation und Speicherung von Daten realisiert werden. Hierbei werden

Geschäftsregeln jedoch weniger als eine Erweiterung der Modellierungsmöglichkeiten verstanden. Sie dienen vielmehr der Beschreibung, Formalisierung und Spezifizierung der Regeln von Agenten und des Multiagentensystems selbst.

Sie werden folglich ähnlich dem Java Policy-Konzept verwendet, wenn auch die Business Rules in ihren Restriktionen als auch Freiheitsgraden über Policies hinausgehen. Im Zusammenhang mit der Modellierung von Agenten, basierend auf Business Rules, entwickelten Taveter und Wagner die Definition „Business Rules are statements that express (certain parts of) a business policy, such as defining business terms, defining deontic assignments (of powers, rights and duties), and defining or constraining the operations of an enterprise, in a declarative manner“ [TaWa02, 528]. Dabei weisen Regeln nach Herbst und Knolmayer [HeKn94, 2], angelehnt an die Datenbanktheorie, folgende Struktur auf:

- Ereignis: Wann soll eine Regel überprüft werden?
- Bedingung: Was soll überprüft werden?
- Aktion: Wie soll reagiert werden?

Entsprechend dieser Struktur werden die Regeln operationalisiert und in die Software (Agenten, MAS) implementiert. Es ist folglich eine Regelsprache notwendig, die solche Konstrukte unterstützt, die beschreiben lässt, wann eine solche Regel ausgeführt bzw. betrachtet werden muss (als Reaktion auf eine Aktion bzw. innerhalb einer vorgegebenen temporären Struktur), welche Bedingungen erfüllt sein müssen damit eine entsprechende Regel angewandt werden kann und wie letztlich auf die so entstandene Situation durch eine Aktion reagiert wird.

3 Regelbasierter Ansatz

Im Folgenden wird ein Ansatz beschrieben, der Werkzeuge zur Beschreibung von Regeln zur Verfügung stellt und in einer Verhandlung gegebenenfalls die Regeln zwischen den beteiligten Entitäten anzupassen. Im Ergebnis der Verhandlung wird die Mitgliedschaft eines Agenten zu einem MAS über Regeln in einer Art Vertrag fixiert.

3.1 Regelontologie

Notwendige Voraussetzung für die eindeutige Beschreibung dynamischen Regelwissens sind Ontologien. Eine Ontologie ist „...an explicit specification of a conceptualization“ [Grub93, 1].

Durch die Konzeptualisierung wird eine vereinfachte Sichtweise auf die Welt geschaffen, wobei mit Hilfe eines definierten Vokabulars eine Domäne oder ein Thema spezifiziert wird. Eine Ontologie ist im Kontext der Verteilten Künstlichen Intelligenz folglich eine Wissensbasis, die durch eine Menge von Begriffen definiert ist, indem Entitäten der Diskurswelt (Klassen, Objekte etc.) mit durch Menschen lesbarem Text bzw. Erklärungen verbunden werden. Gemäß Hage und Verheij [HaVe99] besteht Regelwissen aus Konstrukten der universellen Ontologie, die als Tatbestände bezeichnet werden. Eine Regel verbindet demnach zumindest zwei Sachverhalte: die Vorbedingung und das Ereignis. Ontologien bilden ein entscheidendes Konstrukt bei der Modellierung von Regeln, um die Interpretation von Regelwissen der beteiligten Agenten und Systeme zu ermöglichen.

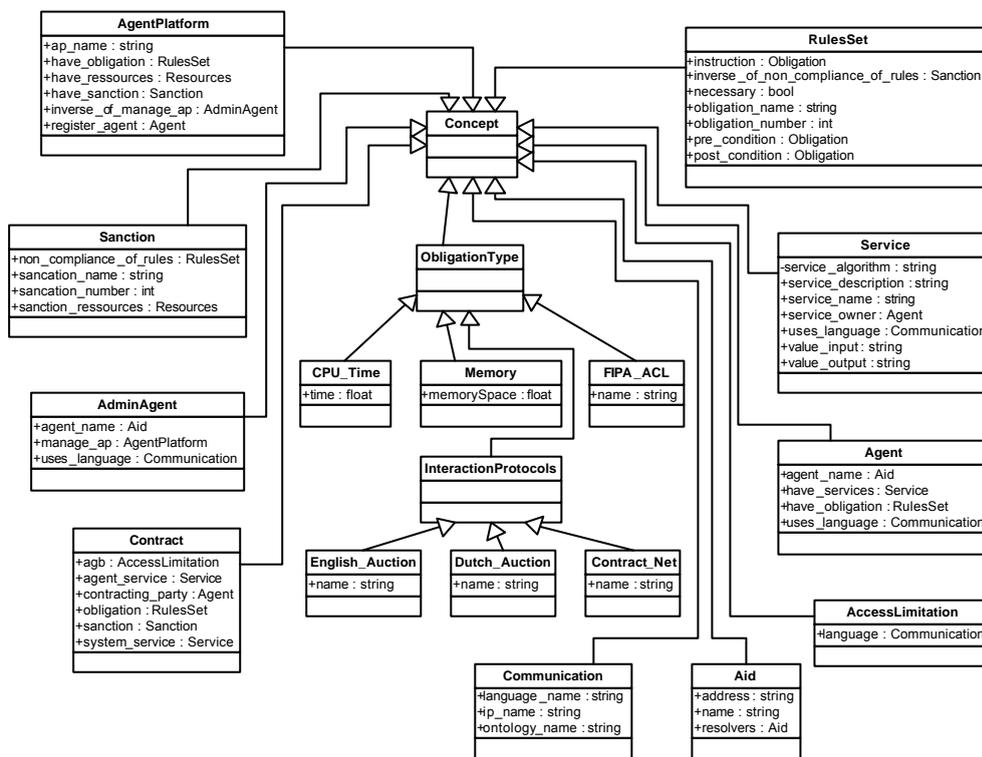


Abb. 1: Konzepte der Regelontologie

Um die semantische Korrektheit der regelbasierten Nachrichten zwischen den Agenten zu gewährleisten, wird eine Regelontologie eingeführt, die die wesentlichen Entitäten, Attribute und Aktionen zur Kommunikation und Interaktion über die Regelmenge definiert. Innerhalb der Ontologie wird zwischen Konzepten und Prädikaten unterschieden. Konzepte stellen dabei

Objekte dar, die die entsprechenden Daten der Kommunikation der Agenten beinhalten. In der praktischen Umsetzung (siehe Abb. 1) sind somit u.a. die Konzepte Contract (Vertrag zwischen dem Agenten und dem MAS), RulesSet (Menge der zu vereinbarenden Regeln), Obligation (die entsprechende Regel selbst) oder Sanction (Sanktion, welche bei einem Verstoß gegen eine Regel droht) definiert. Prädikate hingegen können als Methoden oder Funktionen betrachtet werden, die bestimmte Tätigkeiten eines Agenten beschreiben. So wurden u.a. die Prädikate searchForService (Suche nach einem Dienst, den ein potentielles zukünftiges Mitglied des MAS benötigt), askForRulesSet (Anfrage eines Agenten nach den Regeln eines Agenten), registerContract (Abschließen eines Vertrages zwischen Agent und System) oder leaveSystem (ein Agent beabsichtigt ein System zu verlassen) definiert.

Durch die oben beschriebenen Konzepte und Prädikate der Ontologie ist es nunmehr möglich, eine semantisch eindeutige Kommunikation zwischen Plattform und Agent über die Verhandlung einer Mitgliedschaft in einem Multiagentensystem zu gewährleisten. Die Ontologie ist generisch ausgelegt, beinhaltet folglich keine domänenspezifischen Elemente, die jedoch bei Bedarf noch hinzugefügt werden können. Die Ontologie bezieht sich allein auf die Regelverhandlung sowie Kontrolle der Einhaltung der Regeln und ist wenig oder nicht geeignet, anderweitige Problemklassen zu adressieren.

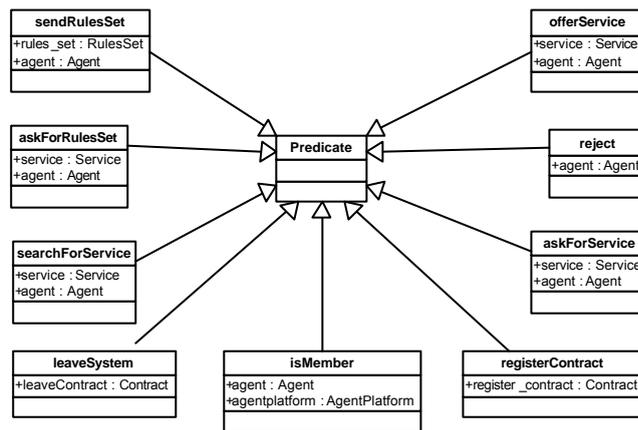


Abb. 2: Prädikate der Regelontologie

3.2 Regelmodell

Um eindeutige Regeln für Agenten und Multiagentensysteme modellieren und operationalisieren zu können, werden diese innerhalb eines Regelmodells abgebildet und definiert. Die Identifizierung und Modellierung von Regeln erfolgt mit Hilfe der Regelontologie, um die Eindeutigkeit der Regelkonstrukte zu gewährleisten. Die Formalisierung der Regeln erfolgt

durch die Object Constraint Language. OCL wurde durch die Object Management Group (OMG) zertifiziert und liegt derzeit in der Version 2.0 vor [OCL04]. Sie stellt eine Erweiterung der Modellierungssprache Unified Modeling Language (UML) dar. Notwendig wurde die Spracherweiterung, da herkömmliche Diagramme in UML keine Abhängigkeiten und Regeln abbilden können. OCL stellt eine Sprache dar, die neben der einfachen Verständlichkeit sehr eng an implementierungsnahen Methoden und Sprachen angesiedelt ist. Entsprechend codierte Regeln und Ausdrücke können daher ohne großen Anpassungsaufwand in objektorientierte Sprachen überführt werden. OCL stellt durch ihre bei der OMG erfolgte Spezifizierung eine verifizierte und in der Praxis erprobte Sprache dar.

Bei der Definition von Regeln kann zwischen domänenabhängigen Regeln, die in Abhängigkeit zur jeweils gegebenen Aufgabenstellung des Systems bzw. der Umwelt stehen und generischen Regeln, die keine Domänenspezifik besitzen und unabhängig von konkreten Aufgabenstellungen sind, unterschieden werden. Generische Regeln werden im Folgenden als statische Regeln bezeichnet, z.B.:

- Kommunikationsprotokolle im System (z.B. KQML, FIPA ACL);
- Koordinationsprotokolle (z.B. FIPA Dutch Auction, FIPA Contract Net);
- Inhaltsbeschreibungssprachen (z.B. Semantic Language (SL), KIF);
- Ontologien.

Domänenabhängige Regeln, die im Folgenden als dynamische Regeln bezeichnet werden, sind in Abhängigkeit von den gegebenen Aufgabenstellungen gestaltet, z.B.:

- Vertraulichkeit von Informationen;
- spezifische gesetzliche Bestimmung hinsichtlich Datenhaltung und Datenschutz;
- zu erbringender Anteil der Lösungskapazität eines Agenten in einem MAS und
- Anteil zu erbringender Kooperationsleistungen eines Agenten in einem System.

Weiterhin wird unterschieden zwischen Regeln, die unabdingbar (obligatorisch) und solchen, die optional für die Agenten und das Multiagentensystem sind (vgl. Tab. 2).

	Statisch	Dynamisch
Obligatorisch	Kommunikationssprache	Rollen, Ontologien
Optional	Interaktionsprotokolle	Zugriffsrechte, org. Regelungen

Tab. 2: Klassifizierung der Regeln von Multiagentensystemen

3.2.1 *Obligatorische statische Regeln*

Obligatorische statische Regeln definieren domänenunabhängige und zwingend notwendige Anforderungen, die das Funktionieren des Systems unmittelbar beeinflussen. Dazu gehören Anforderungen, die erfüllt sein müssen, um Kommunikation und Koordination in einem System überhaupt gewährleisten zu können. Diese werden für den Agenten und das MAS selbst festgelegt. Die Grundvoraussetzungen für die Interoperabilität von Agentensystemen sind innerhalb des FIPA-Standards definiert [FIPA06]. Dieser umfasst neben der Agentenkommunikationssprache ACL auch die Inhaltsbeschreibungssprachen (z.B. KIF und SLO). Obligatorische statische Regeln für Agenten im Kontext o.g. Szenarios können sein:

```
context agent inv:
  RulesSet.CommunicationLanguage = FIPA_ACL
  RulesSet.CommunicationLanguage.necessary = true
  RulesSet.ContentLanguage = FIPA_SLO
  RulesSet.ContentLanguage.necessary = true
```

Obligatorische statische Regeln für ein Multiagentensystem können sein:

```
context MAS inv:
  RulesSet.CommunicationLanguage = FIPA_ACL
  RulesSet.CommunicationLanguage.necessary = true
  RulesSet.ContentLanguage = FIPA_KIF
  RulesSet.ContentLanguage.necessary = true
```

Diese jeweils internen Regeln des Agenten sowie des MAS müssen hinsichtlich ihrer Vereinbarkeit überprüft werden. Bei obligatorischen Regeln ist eine Verhandlung über den Regelinhalt nicht möglich, d.h. eine Übereinstimmung ist zwingend notwendig, um die Grundvoraussetzungen für das Funktionieren des Systems zu schaffen.

Eine Überprüfung (oder in OCL die eigentliche Regel) würde wie folgt definiert werden:

```
MAS:: joinAgent(SystemRules: RulesSet, AgentRules: RulesSet) : boolean
  pre: SystemRules=AgentRules
  post: join=true
```

Dies würde einer trivialen Überprüfung bzw. einem Matching der Regeln entsprechen. Nur wenn die Regeln übereinstimmen, kann die Mitgliedschaft des Agenten bestätigt werden.

3.2.2 *Optionale statische Regeln*

Innerhalb optionaler statischer Regeln können vor allem Interaktionsprotokolle, wie das FIPA Kontrakt-Netz-Protokoll oder das FIPA Request-Protokoll definiert werden [FIPA06]. Interaktionsprotokolle sind zwar für die Kommunikation der Agenten nicht zwingend notwendig, sie sind aber unerlässlich um Koordination respektive Kooperation zwischen den Agenten zu gewährleisten. Gemäß dem eingeführten Szenario könnten optionale statische Regeln die unterschiedlichen Vergabekriterien von Transportaufträgen beinhalten, die beispielsweise über

Auktionen, oder aber über das FIPA Kontrakt-Netz Protokoll¹ durchgeführt werden können. Beim Abgleichen der Regelmenge des Agenten mit der des Multiagentensystems müsste das Eintrittskriterium so gestaltet werden, dass zumindest eines der durch den Agenten beherrschten Interaktionsprotokolle auch im System verfügbar ist und sich somit ein Transportagent überhaupt an der Vergabe von Transportaufträgen beteiligen kann. Eine OCL-konforme Formalisierung könnte wie folgt aussehen:

```

context agent inv:
  RulesSet.InteractionProtocol = FIPA_DUTCH_AUCTION;
  RulesSet.InteractionProtocol.necessary = false
RulesSet.InteractionProtocol = FIPA_ENGLISH_AUCTION;
  RulesSet.InteractionProtocol.necessary = false
RulesSet.InteractionProtocol = FIPA_CONTRACT_NET;
  RulesSet.InteractionProtocol.necessary = false

```

3.2.3 Obligatorische dynamische Regeln

Entsprechend dem Anwendungsszenario ist es erforderlich, domänenspezifische Regeln zu formulieren. So kann es geboten sein, spezielle Rollen innerhalb eines Multiagentensystems zu entwerfen (z.B. Agenten erfüllen spezielle Administrationsfunktionen oder fungieren als spezifische Schnittstellen zu existierenden IT-Systemen) oder aber Ontologien zu erstellen, die eine domänenspezifische Kommunikation unterstützen. Um ein inhaltliches Verständnis für die Kommunikation zu realisieren, kann als obligatorische Regel das Verwenden spezifischer Ontologien vorausgesetzt werden. Neben der vorgestellten Regelontologie, die generisch und somit domänenunabhängig verwendet werden kann, sind domänenspezifische Ontologien erforderlich, die die Begrifflichkeiten und Objekte der entsprechenden Domäne, im Beispiel eine Ontologie für logistische Sachverhalte (z.B. Agent.Enterprise Ontology²) definieren. Eine Formulierung der Regeln in OCL würde folgendermaßen gestaltet sein:

```

context agent inv:
  RulesSet.Ontology = RulesOntology;
  RulesSet.Ontology.necessary = true
RulesSet.Ontology = Agent.EnterpriseOntology;
  RulesSet.Ontology.necessary = true

```

Analog zu den obligatorischen statischen Regeln ist auch hier bei der Überprüfung der Regelmenge ein reines Matching, also Übereinstimmung der Regeln, erforderlich. Es wird also nur überprüft, ob die Regeln eines Agenten dem des Systems entsprechen bzw. ob der Agent bereit ist, falls er selbst nicht über die jeweiligen Regeln verfügt, die des Systems anzuerkennen.

¹ Innerhalb des FIPA Kontrakt-Netz Protokolls wird das Verfahren zur freihändigen Vergabe einer Aufgabe entsprechend der Fähigkeiten der verfügbaren Agenten spezifiziert.

² <http://www.agententerprise.net/>

3.2.4 Optionale dynamische Regeln

Optionale dynamische Regeln werden vorwiegend zur Formulierung organisatorischer Aspekte bzw. domänenspezifischer Anforderungen verwendet (z.B. Zugriffsrechte auf lokale Ressourcen wie Datenbanken, Speicherbereiche oder Dienste). Darüber hinaus können jedoch auch Regeln spezifiziert werden, die sich aus dem Anwendungskontext bzw. den organisatorischen Rahmenbedingungen ergeben können. Im Rahmen des Logistikszenarios ist es denkbar, dass die einzelnen Umschlagplätze bestehende Rahmenbedingungen definiert haben, die zum Beispiel den Zugriff auf lokale Ressourcen, wie Datenbanken, oder aber bei globalen Transportnetzen die jeweiligen gesetzlichen Bedingungen der entsprechenden Länder (beispielsweise Umgang mit Gefahrguttransporten, Ruhezeiten von Fahrern etc.) spezifizieren. Die Formulierung von eindeutigen, durch das System interpretierbaren Regeln erweist sich als verhältnismäßig komplex und lässt sich nur durch klar interpretierbare Ontologien erreichen. So könnte zum Beispiel eine Formulierung der Regeln in OCL wie folgt formuliert werden:

```
context MAS inv:
    RulesSet.Database.Login = "User";
    RulesSet.Database.Login.necessary = false
    RulesSet.Database.Password = "Guest";
    RulesSet.Database.Password.necessary = false
    RulesSet.DriverBreakAfter = "8h";
    RulesSet.necessary = true
```

Bei der Überprüfung wird nicht versucht, eine Übereinstimmung (Matching) der Regeln zwischen Agent und MAS zu erreichen, sondern die se müssen durch die jeweilig andere Entität akzeptiert und befolgt werden.

3.3 Regelabgleich/Mitgliedschaft

Der Anreiz für einen Agenten, Mitglied eines MAS zu werden, liegt entweder in den durch das System zur Verfügung gestellten Ressourcen, bzw. in Diensten, die durch Agenten des Systems bereitgestellt werden. Um Mitglied zu werden, muss der Agent mit dem System Regeln verhandeln, um die Interoperabilität und Koordinationsfähigkeit des Systems zu erhalten. Aus diesem Grund verfügen sowohl der Agent als auch das MAS über eine eigene Regelmenge, die neben technischen Spezifikationen (wie Protokolle und Sprachen) auch soziale und organisatorische Komponenten (Anforderungen an einen Agenten von Seiten seines Benutzers bzw. Anforderungen an das System in seinem sozialen Umfeld) beinhalten können.

Um einen Regelabgleich bzw. eine Verhandlung der Regelmenge durchführen zu können wird die Rete-Maschine [Forg82] von Jess (Java Expert System Shell [Jess06]) verwendet. Die

Regelmaschine von Jess besitzt eine Java-Schnittstelle, sodass Daten direkt aus dem Programm in Jess geschrieben und die Ergebnisse ausgelesen werden können.

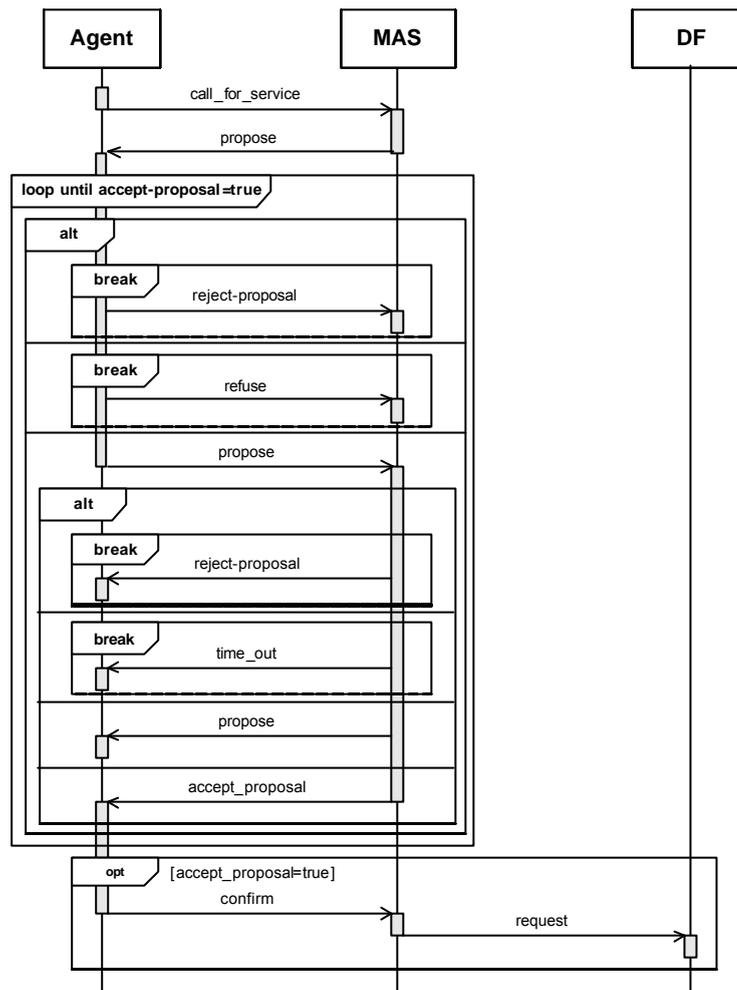


Abb. 3: Sequenzdiagramm des Interaktionsprotokolls zum Eintritt eines Agenten in ein MAS

Die Verhandlung zwischen dem System und dem Agent über eine mögliche Mitgliedschaft ist im Sequenzdiagramm (Abb. 3) dargestellt. Der Regelabgleich zwischen System und Agent ist im vorgestellten Prototyp gewichtungsabhängig realisiert. An einem verbesserten Algorithmus zum Regelabgleich auf Basis des Constraint Satisfaction Problem Verfahrens [RuNo03, 141], wird derzeit gearbeitet. Ist der Abgleich beendet, wird der Agent Mitglied des Systems, indem ein „Vertrag“ (Contract) durch das System und den Agenten vereinbart wird. Dieser Vertrag ist in einer zentralen Datenbasis gespeichert, um bei eventuell auftretenden späteren Verfehlungen darauf zurückgreifen zu können und gegebenenfalls Sanktionen zu verhängen, die von Ressourcenbeschränkungen bis hin zu einem Ausschluss eines Agenten aus einem System reichen

können. Für die Kontrolle der Einhaltung der Regelmenge durch den Agenten empfehlen sich Ansätze der s.g. Soft Security, wie zum Beispiel Reputationsmechanismen [Pado00].

4 Prototyp

Der im Folgenden vorgestellte Prototyp ist das Ergebnis der Entwicklung einer agentenbasierten Transportnetzwerk-Simulation. Der Fokus hierbei liegt in der Untersuchung des dezentralen Koordinationskonzeptes in dem Bereich der Logistik. Im modellierten logistischen Netzwerk werden Stückgüter, Umschlagplätze und Fahrzeuge jeweils durch Softwareagenten repräsentiert. Das Stückgut erreicht einen Umschlagplatz an Bord eines Fahrzeuges und kann auf diesem beliebig umgruppiert werden. Die Entscheidung über die Beladung der Fahrzeuge erfolgt nach festgelegten Regeln durch direkte Verhandlungen zwischen den Stückgütern. Hierbei können Entscheidungen über den Weitertransport noch während der Fahrt auf den Fahrzeugen getroffen werden. Die Implementierung des Prototyps setzt auf eine weitgehende Trennung von Ablauf und Darstellung (Abb. 4).

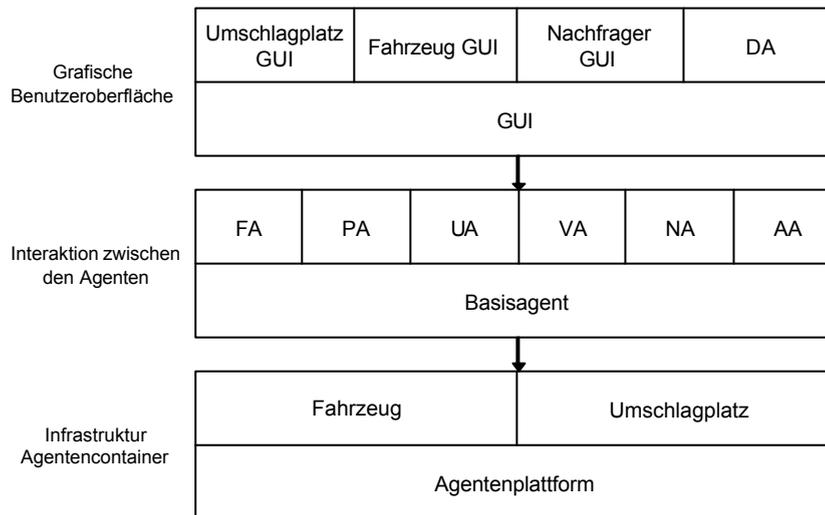


Abb. 4: Ebenen der Implementierung

Die unterste Ebene stellt die benötigte Infrastruktur und Basisdienste zur Verfügung. Auf der mittleren Ebene agieren und kommunizieren die Agenten untereinander. Hier sind die Funktionen und Eigenschaften der Agenten eingebunden. Graphical User Interfaces (GUI) für die einzelnen Agententypen werden auf der oberen Ebene repräsentiert.

Der Prototyp verfolgt die Strategie der dezentralen Koordination der logistischen Steuerelemente wie Fahrzeuge oder Stückgüter. Der Prototyp basiert auf dem *Java Agent Development Environment* (JADE) [JADE06]. JADE stellt eine Middleware auf Basis von Java dar und unterstützt die Entwicklung von Multiagentensystemen durch die Bereitstellung einer Agentenplattform den Rahmen für die Entwicklung von Agenten. Auf Grund der Konformität von JADE mit dem FIPA-Standard ist eine hohe Interoperabilität des Prototyps sichergestellt.

5 Zusammenfassung und Ausblick

Unternehmensübergreifende Transportnetzwerke zeichnen sich durch eine zunehmende Komplexität und vor allem eine natürliche Verteiltheit der Ressourcen aus. Diese Charakteristiken lassen den Einsatz von Agenten zur Planung und Koordination der logistischen Prozesse als geeignet erscheinen. Dabei wurde der Ansatz der Verwendung von Regeln zur Beschreibung und Definition von Anforderungen und Rahmenbedingungen des logistischen Netzwerks, der zur Verfügung stehenden Plattformen und deren Akteuren bislang nur unzureichend betrachtet. Durch deren Definition ist es nun möglich, auch in einer weitgehend heterogenen Systemlandschaft, mit unterschiedlich verwendeten Protokollen, Interoperabilität und vor allem Koordinationsfähigkeit herzustellen. Dabei spielen nicht nur technische Aspekte eine Rolle, sondern es können auch organisatorische und soziale Sachverhalte definiert werden. Der Ansatz schließt damit die Lücke zwischen dem Ziel einer hohen Flexibilisierung logistischer Prozesse und dem Wunsch, die Einhaltung vereinbarter Geschäftsregeln zu überwachen. Die Formalisierung der Regeln erfolgte mittels der Object Constraint Language, die eine direkte Überführung in Programmcode erlaubt. Durch die aufgezeigte Regelontologie wird die semantische Klarheit der Kommunikation zur Vereinbarung von Regeln zwischen den Akteuren verbessert. Weiterhin wurde ein Ansatz zur Aushandlung von Regeln vorgestellt. Im dargestellten Prototyp sind die Ansätze implementiert.

Für die zukünftige Weiterentwicklung sind weitere domänenspezifische Ontologien geplant. Der Prozess der Verhandlung und des Abgleichs der Regeln wird flexibilisiert und vereinfacht. Abschließend kann festgestellt werden, dass das Regelkonzept in vorliegendem Papier zwar in Hinblick auf agentenbasierte Anwendungen entwickelt wurde, jedoch ebenso für weitere nicht agentenspezifische Systeme, wie zum Beispiel elektronische Märkte, denkbar ist.

Literaturverzeichnis

- [Anth03] Anthes, G. H.: Agents of Change. www.computerworld.com, 2003.
- [BBGL+04] Benisch, M.; Greenwald, A.; Grypari, I.; Lederman, R.; Naroditskiy, V.; Tschantz, M.: Botticelli: A Supply Chain Management Agent. In: Proceedings of Third International Joint Conference on Autonomous Agents and Multi-Agent Systems, ACM Press, 2004.
- [BrGr05] Business Rules Group, <http://www.businessrulesgroup.org>, 2005.
- [Buss98] Bussmann, S.: Agent-Oriented Programming of Manufacturing Control Tasks. In: Proceedings of the 3rd International Conference on Multi-Agent Systems. IEEE CS Press, 1998, S. 57 - 63.
- [CoJe96] Cockburn, D.; Jennings, N.R.: Archon: A Distributed Artificial Intelligence System for Industrial Applications. In: O'Hare, G.M.P.; Jennings, N.R. (Hrsg.): Foundations of Distributed Artificial Intelligence. John Wiley, 1996, S. 319-344.
- [DaHe04] Davidsson, P.; Henesey, L et al: Agent-Based Approaches to Transport Logistics. AAMAS Workshop on Agents in Traffic and Transportation, 2004.
- [FIPA06] www.fipa.org, 2006, Abruf am 10. Juli 2006.
- [Forg82] Forgy, C.: Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem. Artificial Intelligence 13(1982)19, S 17-37.
- [FrMu96] Fritzing, J.; Mueller, M.: Java Security. White Paper, Sun Microsystems, 1996.
- [GrGr05] Gaudina, V.; Grundspenkis, J.: Technologies and Multi-Agent System Architectures for Transportation and Logistics Support: An Overview. In: International Conference on Computer systems and Technologies – CompSysTech IIIA.6-1 – IIIA.6-2, 2005.
- [Grub93] Gruber, T. R.: A Translation Approach to Portable Ontology Specifications. In: Knowledge Acquisition 2 (1993) 5, S.199-220.
- [HaVe99] Hage, J., Verheij, B.: The law as a dynamic interconnected system of states of affairs – a legal top ontology. In: International Journal of Human - Computer Studies 6 (1999) 51, S. 1043-1078.
- [HeKn94] Herbst, H.; Knolmayer, G.: Arbeitsbericht Nr. 46 - Ansätze zur Klassifikation von Geschäftsregeln. Institut für Wirtschaftsinformatik, Universität Bern, 1994.
- [Jade06] Jade Agent Development Environment: <http://jade.cselt.it/>, 2006.

- [Jess06] Java Expert System Shell, 2006, <http://www.sandia.gov>, Abruf am 10. Juli 2006.
- [KiKu92] Kieser, A.; Kubicek, H.: Organisation. Walter de Gruyter, Berlin 1992.
- [Kies99] Kieser, A.: Konstruktivistische Ansätze. In: Kieser, A. (Hrsg.): Organisationstheorien. Verlag W. Kohlhammer, Stuttgart 1999, S. 287-318.
- [KnTi99] Knirsch, P.; Timm, I. J.: Adaptive Multiagent Systems Applied on Temporal Logistics Networks. In: Proceedings 4th International Symposium on Logistics, 1999.
- [Müll97] Müller, J.P.: Towards Agents Systems Engineering. International Journal on Data and Knowledge Engineering, In: Special Issue on Distributed Expertise 23. 1997.
- [OCL04] UML 2.0 OCL Specification. <http://www.omg.org/docs/ptc/03-10-14.pdf>, 2004, Abruf am 10. Juli 2006.
- [Pado00] Padovan, B.: Ein Vertrauens- und Reputationsmodell für Multi-Agenten Systeme. Dissertation, Albert-Ludwigs-Universität Freiburg, 2000.
- [Paru00] V.D. Parunak, H.: A Practitioners' Review of Industrial Agent Application. In: Autonomous Agents and Multi-Agent Systems 3 (2000) 4, S. 389-407.
- [PLSP04] Perugini, D; Lambert, D.; Sterling, L.; Pearce, A.: Agent-Based Global Transportation Scheduling in Military Logistics. In: Proceedings of Third International Joint Conference on Autonomous Agents and Multi-Agent Systems, ACM Press, 2004.
- [RuNo03] Russel, S.; Norvig, P.: Artificial Intelligence : a modern approach. 2. ed. Prentice-Hall, New Jersey 2003.
- [Sche99] Scherer, A.G.: Kritik der Organisation oder Organisation der Kritik? – Wissenschaftstheoretische Bemerkungen zum kritischen Umgang mit Organisationstheorien. In: Kieser, A. (Hrsg.): Organisationstheorien. Verlag W. Kohlhammer, Stuttgart, 1999, S. 1-37.
- [TaWa02] Taveter, K.; Wagner, G.: Agent-Oriented Enterprise Modeling Based on Business Rules. In: Arisawa, H.; Kambayashi, Y. (Hrsg.): Conceptual modeling for new information systems technologies. Springer-Verlag, Berlin 2002.
- [WoJe95] Wooldridge, M.; Jennings, N.R.: Agent Theories, Architectures, and Languages: a Survey. In Wooldridge, M.; Jennings, N.R. (Hrsg.): Intelligent Agents. Springer-Verlag, Berlin 1995, S. 1-22.

Distribution network design with postponement

Frank Schwartz, Stefan Voß

Institut für Wirtschaftsinformatik

Universität Hamburg

20146 Hamburg

fs@econ.uni-hamburg.de, stefan.voss@uni-hamburg.de

Abstract

An important concern in supply chain management is about network design, involving factories, central warehouses, regional warehouses as well as customers. The best strategy has to be ascertained for distributing products within this network. The objective is to select the optimal numbers and locations of central and regional warehouses such that all customer demands are satisfied at minimum total costs of the network. An extension of existing approaches is to take into account aspects of postponement, in particular regarding the problem of postponing activities like assembling half-finished goods or packaging them. A mixed integer programming model is provided and it is demonstrated that the resulting formulation can be used to solve realistic problem instances with commercially available mathematical programming software.

1 Introduction

Designing the physical distribution network in a supply chain is an important strategic planning problem with undeniable implications towards tactical and operational planning success. In this paper we consider a distribution network made up of four tiers (factories, central warehouses, regional warehouses, and customer zones (demand points)), with the aim of defining the number and location of different types of facilities. Moreover, we consider certain postponement options motivated by some real-world cases. Aspects of postponement are a widely recognized approach to improve supply chains. However, up to now they have not yet been fully integrated in advanced planning tools [SkHa04], [Stad05]. Various postponement definitions can be found in literature. To some authors, postponement simply means delaying at least one differentiating step. To others, postponement rather means adding variety after receiving a customer order than

anticipating orders [ApGu05]. In the context of distribution network design postponing activities means that finishing products in factories is not mandatory. In lieu of this, performing activities like, e.g., assembling or packaging can be procrastinated. This requires the implementation of appropriate functions in central warehouses or regional warehouses for assembling or packaging goods, respectively.

In this paper a mixed integer linear programming (MILP) model for distribution network design is developed. This model permits decisions regarding the number and location of different types of potential facilities to be selected from a set of possible candidates. As an extension of existing models, decisions regarding the establishment of delayed functions to finish a product with respect to customer requirements are also taken into account. The objective in designing such a distribution network is to determine a least cost system design that satisfies the demands of all customers without exceeding the capacities of the factories, the warehouses including their implemented functions to finish a product as well as the capacities of the shipping routes between the facilities. The model belongs to the class of production-distribution allocation as well as facility-location allocation problems.

The remainder of the paper is organized as follows. In Section 2, a literature review on planning distribution networks as well as postponement is given. In Section 3, a specific distribution network incorporating postponement approaches is depicted in detail. Section 4 contains a mathematical formulation of the considered problem. In Section 5, some computational results for two test instances are reported indicating that respective models actually may prove useful in today's planning systems. A summary as well as an outlook to further research are given in Section 6.

2 Literature review

Designing a distribution network in a multi-echelon environment for locating distribution facilities and allocating functions for finishing products regarding customer needs requires strategic decisions (where to locate facilities and implement functions) as well as tactical/operational decisions (distribution strategy from factories to customers via central as well as regional warehouses). In Operations Research, models and methods for distribution planning are available since its early years, in particular for locating warehouses, but also for more comprehensive design problems regarding multiple products, limited capacities, single source constraints or

nonlinear transportation costs (see, e.g., the reviews of [Aike85], [OwDa98] or [KIDr05]). [GeGr74] were among the first to investigate the use of intermediate distribution. They present a model to solve the problem of designing a distribution system with an optimal location of intermediate distribution facilities between factories and customers. [BrGH87] depict an optimization-based algorithm for a decision support system used to manage complex problems involving facility selection, equipment location and utilization, as well as manufacturing and distribution of products. A MILP formulation with the objective of maximizing the total after-tax profit for manufacturing facilities and distribution centers is presented by [CoLe89]. The model determines the optimal deployment of resources associated with a particular policy option. The considered product structure in the model encompasses three levels (major components, subassemblies, finished products). Extending the model of [CoLe89], [CoMo90] investigate effects of various parameters on supply chain costs and determine which manufacturing facilities and distribution centers should be established. [Chan93] uses a model that plans besides vehicle routes deliveries to customers based upon inventories at warehouses and distribution centers. Later, [ChFi94] consider the coordination of production and distribution planning. [Flei93] developed a multi-commodity three-stage network flow model with arbitrary nonlinear transport and warehouse costs which may include fixed costs. In contrast to common models, location decisions are not determined by integer programming or add/drop procedures, but result from the solution of a network flow problem. A MILP model presented by [Pool94] allows for deciding where to locate factories and depots, allocating the production as well as how to serve customers. A MILP global supply chain model presented by [ABHT95] determines the number and location of distribution centers, customer-distribution center assignments, the number of tiers, and the product-factory assignment. [CCDE97] developed an integer model for finding the location of distribution centers and to assign those selected to customer zones. [Amir06] presents a model that takes into consideration different capacity levels. A tri-echelon multi-commodity system incorporating production, transportation and distribution planning is considered by [PiJa96]. In a succeeding work, [JaPi01] present a model that determines the location of a number of production plants and distribution centers with the objective of minimizing the total operating costs for the distribution network. Further models of distribution networks with several layers are also presented in, e.g., [Klin85], [TsSP01] or [AmSc05]. A multi-objective approach is pointed out in [SaBe00]. A distribution network model taking into account mode selection, lead times as well as capacitated vehicle distribution centers is proposed in [EUPB05].

[KaMN03] develop a generic strategic planning and design model for global supply chains which captures essential elements of many industrial environments. Further papers concerning distribution networks can be found in comprehensive reviews of, e.g., [ViGo97], [GoVD02] and [BiOz04].

Postponement is widely regarded as an approach that may result in superior supply chains (e.g., [JoRi85] or [Coop93]), and it has been recognized as a growing trend in manufacturing and distribution [SkHa04]. According to [YaBu03], much has been written in the literature on the benefits of postponement, yet little is known about its implementation.

Extensive investigations of benefits of postponement as well as postponement strategies have been carried out in the context of marketing and logistics as well as supply chain management. Some papers in this context are [Alde50], [Buck65], [ZiBo88], [Coop93], [FeLe97], [PaCo98], [vHoe98], [vHoe01] and [YaBu03] as well as [MiSk04]. Whereas these studies are primarily qualitative, some recent papers focus on quantifying the benefits and criteria of various postponement strategies. See, e.g., [LeBC93], [GaTa97], [SwTa99], [ErKa00], [MaWL02], [YeYa03], [SkHa04] or [ApGu05].

None of the papers discussed above explicitly deals with the implementation of postponement strategies in the context of planning a distribution network, and there are only few papers, e.g., [CoLe89], [CoMo90] or [ABHT95] that rudimentary combine aspects of postponement and distribution network design.

3 Designing distribution networks allowing for aspects of postponement

This work considers a distribution network with several facilities at different tiers of the network where different products are delivered from the plants to satisfy the requirements of several customer zones. Figure 1 shows a distribution network with factories, central warehouses, regional warehouses and customer zones. The arrows represent potential flows of the products from the factories up to the customer zones. Typically, the goods flow from factories to central warehouses, from central warehouses to regional warehouses and then to the customer zones. The locations of central and regional warehouses are unknown and have to be selected from a set of candidate locations. Furthermore, there is also the alternative to ship goods directly from factories to regional warehouses, from central warehouses to customer zones as well as from factories up to customer zones.

In the distribution network, different groups of finished or unfinished products are shipped to one or several central warehouses in order to be shipped in common from there. Regional warehouses serve as destinations of shipments from the factories or central warehouses, and as starting points for short-distance deliveries to customers. They permit to bundle the shipped goods over long distances, before splitting them into smaller quantities regarding customer orders. The customer zones comprise several customers within an enclosed area. Each customer zone has some demand for certain products, which has to be met by the distribution system.

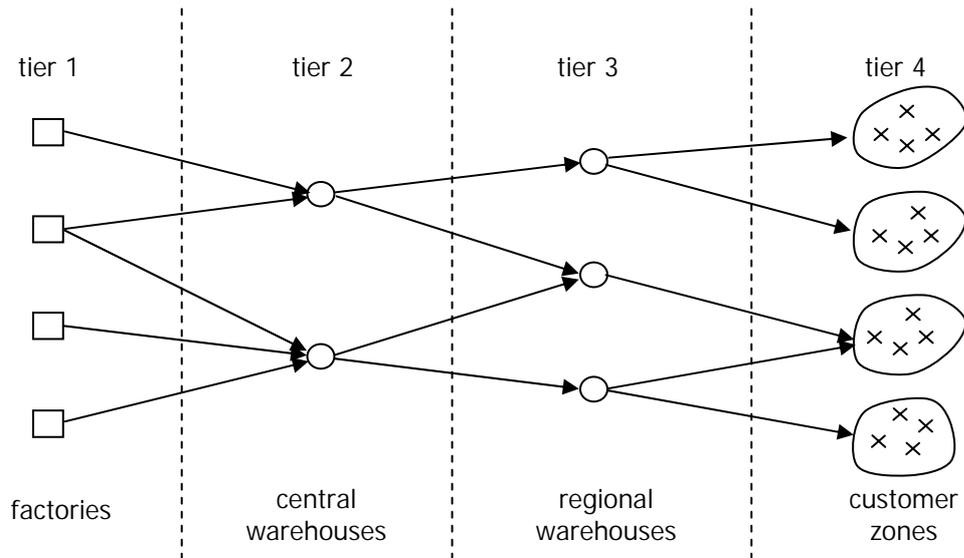


Figure 1: Distribution network with four tiers

Common models of distribution network planning assume that they comprise a flow of finished products from factories to customers. Here this assumption is relaxed such that activities to finish a product could be performed later, or rather postponed along the route from the factories up to the customers. Potential locations for postponed activities are both central and regional warehouses. [ZiBo88] describe five types of postponement that could be implemented in a distribution network: These types of postponement comprise labeling postponement, packaging postponement, assembly postponement, manufacturing postponement, and time postponement. Labeling postponement is an approach where standard products are stocked and labeled differently regarding the realized customer demand. In packaging postponement, products are not packaged into individual packs until final orders are received. Assembly and manufacturing postponement refer to situations where additional assembly or manufacturing may be performed in assembly facilities or at warehouses before shipping the products to customers after demand is realized. Finally, time postponement represents the concept that products are held at a central

warehouse and are shipped to customers directly, instead of shipping them in advance in virtue of corresponding forecasts to retail warehouses.

In the considered distribution network planning problem two types of postponement strategies are to be established: Packaging postponement as well as assembly postponement (see, e.g., [Coop93]). The motivation behind these postponements stems from the option to utilize respective degrees of freedom. That is, products may eventually be specified according to either the final assembly step or even the packaging. Especially for regions with large varieties in customer demands for one or the other product this may allow additional degrees of freedom. Regarding the implementation of the corresponding processes to assemble unfinished goods respectively to pack finished goods, the required resources can be installed on every tier of the distribution network, namely in the factories, in the central warehouses, or in the regional warehouses. Feasible routings through the distribution network are depicted in Figure 2.

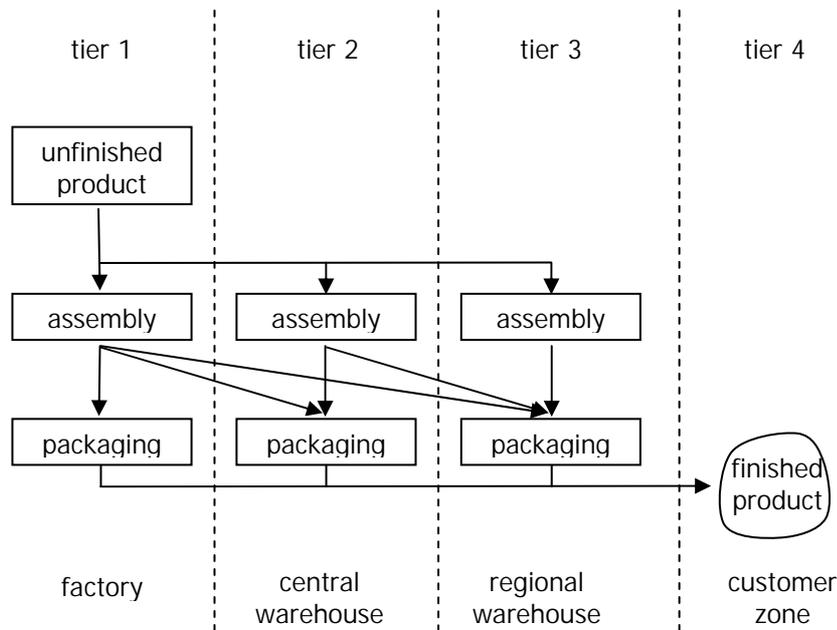


Figure 2: Feasible routings through the distribution network

Postponement may result in considerable cost tradeoffs. Regarding the considered assembly and packaging postponement strategies, processing costs for assembly respectively packaging increase and the transportation costs decrease. Increased per unit assembly costs as well as increased per unit packaging costs result from reduced economies of scale in the concerned warehouses in contrast to assembly respectively packaging in a central plant. Regarding unfinished products that have yet to be assembled, the reduction of transportation costs is typically due to a better density ratio of unassembled products compared to assembled products. In the context of

packaging products, a reduction of transportation costs results from bulk shipping finished, but not packed products from factories to central respectively regional warehouses or from central warehouses to regional warehouses.

The decisions to be determined in the presented distribution network design problem represent strategic decisions and include decisions concerning the number and location of potential central and regional warehouses to be established as well as the decision where to implement assembly and packaging functions in the distribution network. As mentioned above, possible locations for assembly and packaging functions are factories, central warehouses and regional warehouses. Furthermore, decisions regarding the quantity of products shipped between facilities have to be made. The objective is to minimize the combined total costs of the network for a given demand of several customers, taking into account both fixed infrastructure and variable operating costs. The developed model represents a steady-state form of the considered problem with time-invariant deterministic demands. The quantities determined by the optimization are considered to be time-averaged quantities.

Figure 3 represents a consolidated view of Figures 1 and 2. To ease comprehension of the mathematical model in the next section, the applied indices for different locations of facilities or customer zones (i), functions (f and l) and tiers (s) are incorporated in Figure 3. Index l denotes different functions that could be implemented in a facility. Index f also distinguishes different functions that could be implemented. Furthermore, index f interprets identical functions that deal with products with different completion status as different functions. For reason of clarity there is only one facility displayed at each tier. Further potential facilities in alternative locations at a tier as well as the corresponding flows are omitted. The displayed flows in Figure 3 comprise flows between facilities (shipping goods from one location to another one) as well as flows within facilities (processing goods). On the one hand, the arrows between facilities represent feasible process-determined sequences of functions implemented at facilities that have to be applied to finish a product in the production-distribution process. On the other hand, they indicate the changeover from a facility at one location to a facility at another one at a subsequent tier. There are also some arrows included that represent bypassing one and two tiers respectively.

In every facility displayed in Figure 3, various functions that are allowed to be implemented are shown. These functions include the function “no action” that simply refers to handling products without performing any activities like, e.g., assembly or packaging. At the second tier, the activ-

ity “no action” is represented by three nodes. This is due to the requirement to distinguish products with different completion status in a facility. The ability to distinguish products with different completion status is needed again both to determine the allowed succeeding activity in the distribution network, and to calculate the correct capacity consumption in the corresponding facility.

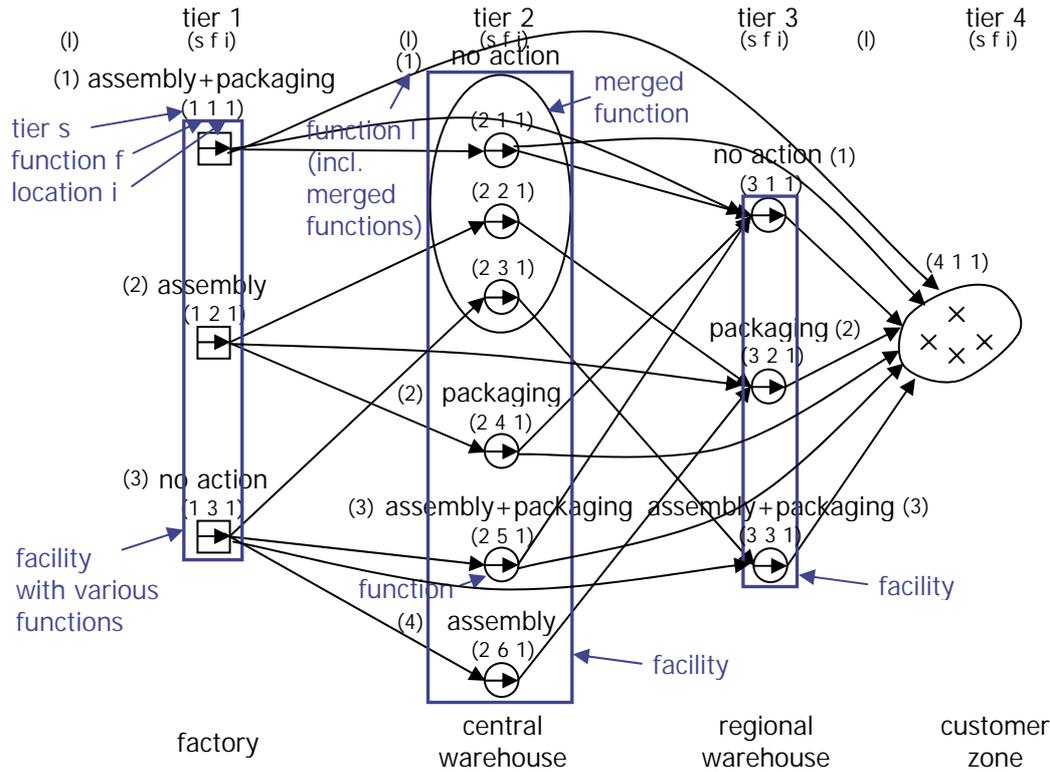


Figure 3: Distribution network with postponed activities

4 Model formulation

The integrated mathematical model for distribution network planning with postponement represents a cost minimization problem. It is a steady state model with deterministic demands, and it is represented by a mixed integer programming formulation. The following notation is used in the formulation of the model:

- S number of tiers
- Z index set of products
- L_s index set of different functions at tier s
- FCT_s index set of nodes of tier s that represents various functions for various degrees of a product’s completion

$LFCT_{sl}$ index set of nodes of tier s that represents the same function $l \in L_s$ for various degrees of a product's completion

LOC_s index set of potential locations of facilities at tier s

$LINKFCT1S_s / LINKFCT2S_s / LINKFCT3S_s$ allowed sequences of potential consecutive functions at tiers s and $s+1 / s+2 / s+3$

$LINKLOC1S_s / LINKLOC2S_s / LINKLOC3S_s$ allowed transportation links (facility/facility or facility/customer zone) at tiers s and $s+1 / s+2 / s+3$

The decision variables are:

$x_{zsjfg}^{c1S} / x_{zsjfg}^{c2S} / x_{zsjfg}^{c3S}$ quantities of product z shipped from a facility at candidate location i at tier s to a facility at candidate location j at tier $s+1$, performing function f in facility i and function g in facility j

x_{zsjf}^i quantity of product z processed in a facility at candidate position i at tier s by the implemented function f

y_{sil} 1, if function l is established in facility i at tier s ; 0 otherwise

The decision variables x_{zsjfg}^{c2S} and x_{zsjfg}^{c3S} represent a bypassing of facilities of one and two tiers respectively.

Furthermore, the following parameters are taken into account:

c_{zsf}^c shipping cost per unit and per unit distance for a product z processed by function f at tier s

$d_{sij}^{1S} / d_{sij}^{2S} / d_{sij}^{3S}$ distances between facility i at tier s and facility j at tier $s+1 / s+2 / s+3$

c_{zsjf}^i processing cost for product z of an implemented function f in a facility at candidate location i at tier s

c_{zsjf}^f fixed cost of establishing an implemented function f in a facility at candidate location i at tier s

U_{zsjf}^i maximum throughput quantity for product z (handling and inventory) at a facility in location i at tier s for the implemented function f

$U_{zsjfg}^{c1S} / U_{zsjfg}^{c2S} / U_{zsjfg}^{c3S}$ maximum shipping quantities for product z from a facility in location i at tier s to a facility in location j at tier $s+1 / s+2 / s+3$ after completing activity f implemented in i

Dem_{zi} demand of product z in customer zone i

Sup_{zi} supply of product z in factory i

In terms of the above notation, the problem can be stated as follows:

Problem P:

$$\begin{aligned}
 \min \quad & \sum_{\substack{z \in Z \\ s \in \{1, 2, \dots, S-1\} \\ (i,j) \in LINKLOC1S_s \\ (f,g) \in LINKFCT1S_s}} c_{zsf}^{c1S} d_{sij}^{c1S} x_{zsiifg}^{c1S} + \sum_{\substack{z \in Z \\ s \in \{1, 2, \dots, S-2\} \\ (i,j) \in LINKLOC2S_s \\ (f,g) \in LINKFCT2S_s}} c_{zsf}^{c2S} d_{sij}^{c2S} x_{zsiifg}^{c2S} + \sum_{\substack{z \in Z \\ s \in \{1, 2, \dots, S-3\} \\ (i,j) \in LINKLOC3S_s \\ (f,g) \in LINKFCT3S_s}} c_{zsf}^{c3S} d_{sij}^{c3S} x_{zsiifg}^{c3S} \\
 & + \sum_{\substack{z \in Z \\ s \in \{1, 2, \dots, S-1\} \\ i \in LOC_s \\ f \in FCT_s}} c_{zsiif}^i x_{zsiif}^i + \sum_{\substack{s \in \{1, 2, \dots, S-1\} \\ i \in LOC_s \\ l \in L_s}} c_{sil}^f y_{sil}^f
 \end{aligned} \tag{1}$$

subject to:

Material balance constraints

Flows entering a facility

$$x_{zsiifg}^i = \sum_{\substack{(i,j) \in LINKLOC1S_{s-1} \\ (f,g) \in LINKFCT1S_{s-1}}} x_{z(s-1)ijfg}^{c1S} \quad \forall z \in Z, s = 2, j \in LOC_s, g \in FCT_s \tag{2}$$

$$x_{zsiifg}^i = \sum_{\substack{(i,j) \in LINKLOC1S_{s-1} \\ (f,g) \in LINKFCT1S_{s-1}}} x_{z(s-1)ijfg}^{c1S} + \sum_{\substack{(i,j) \in LINKLOC2S_{s-2} \\ (f,g) \in LINKFCT2S_{s-2}}} x_{z(s-2)ijfg}^{c2S} \quad \forall z \in Z, s = 3, j \in LOC_s, g \in FCT_s \tag{3}$$

Demand

$$\begin{aligned}
 Dem_{zj} = & \sum_{\substack{(i,j) \in LINKLOC1S_{s-1} \\ (f,g) \in LINKFCT1S_{s-1}}} x_{z(s-1)ijfg}^{c1S} + \sum_{\substack{(i,j) \in LINKLOC2S_{s-2} \\ (f,g) \in LINKFCT2S_{s-2}}} x_{z(s-2)ijfg}^{c2S} + \sum_{\substack{(i,j) \in LINKLOC3S_{s-3} \\ (f,g) \in LINKFCT3S_{s-3}}} x_{z(s-3)ijfg}^{c3S} \quad \forall z \in Z, s = 4, j \in LOC_s, \\
 & g \in FCT_s
 \end{aligned} \tag{4}$$

Flows leaving a facility

$$x_{zsiifg}^i = \sum_{\substack{(i,j) \in LINKLOC1S_s \\ (f,g) \in LINKFCT1S_s}} x_{zsiifg}^{c1S} + \sum_{\substack{(i,j) \in LINKLOC2S_s \\ (f,g) \in LINKFCT2S_s}} x_{zsiifg}^{c2S} + \sum_{\substack{(i,j) \in LINKLOC3S_s \\ (f,g) \in LINKFCT3S_s}} x_{zsiifg}^{c3S} \quad \forall z \in Z, s = 1, i \in LOC_s, f \in FCT_s \tag{5}$$

$$x_{zsiifg}^i = \sum_{\substack{(i,j) \in LINKLOC1S_s \\ (f,g) \in LINKFCT1S_s}} x_{zsiifg}^{c1S} + \sum_{\substack{(i,j) \in LINKLOC2S_s \\ (f,g) \in LINKFCT2S_s}} x_{zsiifg}^{c2S} \quad \forall z \in Z, s = 2, i \in LOC_s, f \in FCT_s \tag{6}$$

$$x_{zsiifg}^i = \sum_{\substack{(i,j) \in LINKLOC1S_s \\ (f,g) \in LINKFCT1S_s}} x_{zsiifg}^{c1S} \quad \forall z \in Z, s = 3, i \in LOC_s, f \in FCT_s \tag{7}$$

Supply

$$Sup_{zi} \geq \sum_{f \in FCT_s} x_{zsiif}^i \quad \forall z \in Z, s = 1, i \in LOC_s \tag{8}$$

Capacity constraints

Transport capacity

$$\sum_{z \in Z} \frac{x_{zsjfg}^{c1S}}{U_{zsjf}^{c1S}} \leq 1 \quad \forall s \in \{1, 2, \dots, S-1\}, (i, j) \in LINKLOC1S_s, (f, g) \in LINKFCT1S_s \quad (9)$$

$$\sum_{z \in Z} \frac{x_{zsjfg}^{c2S}}{U_{zsjf}^{c2S}} \leq 1 \quad \forall s \in \{1, 2, \dots, S-2\}, (i, j) \in LINKLOC2S_s, (f, g) \in LINKFCT2S_s \quad (10)$$

$$\sum_{z \in Z} \frac{x_{zsjfg}^{c3S}}{U_{zsjf}^{c3S}} \leq 1 \quad \forall s \in \{1, 2, \dots, S-3\}, (i, j) \in LINKLOC3S_s, (f, g) \in LINKFCT3S_s \quad (11)$$

Facility capacity

$$\sum_{\substack{z \in Z \\ f \in LFCT_{st}}} \frac{x_{zsjf}^i}{U_{zsjf}^i} \leq y_{sil} \quad \forall s \in \{1, 2, \dots, S-1\}, i \in FCT_s, l \in L_s \quad (12)$$

Non-negativity constraints

$$x_{zsjfg}^{c1S} \geq 0 \quad \forall z \in Z, s \in \{1, 2, \dots, S-1\}, (i, j) \in LINKLOC1S_s, (f, g) \in LINKFCT1S_s \quad (13)$$

$$x_{zsjfg}^{c2S} \geq 0 \quad \forall z \in Z, s \in \{1, 2, \dots, S-2\}, (i, j) \in LINKLOC2S_s, (f, g) \in LINKFCT2S_s \quad (14)$$

$$x_{zsjfg}^{c3S} \geq 0 \quad \forall z \in Z, s \in \{1, 2, \dots, S-3\}, (i, j) \in LINKLOC3S_s, (f, g) \in LINKFCT3S_s \quad (15)$$

$$x_{zsjf}^i \geq 0 \quad \forall z \in Z, s \in \{1, 2, \dots, S-1\}, i \in LOC_s, f \in FCT_s \quad (16)$$

Binary constraints

$$y_{sil} \in \{0; 1\} \quad \forall s \in \{1, 2, \dots, S-1\}, i \in LOC_s, l \in L_s \quad (17)$$

Objective function (1) consists of three cost types. At first, the objective function encompasses variable shipping costs between the locations of facilities. The costs are assumed to be linear. Nonlinear shipping costs, taken into account, e.g., in [Flei93] or [TsSP01], could be incorporated into a more comprehensive model presented in a subsequent paper. A further type of costs in the objective function is induced by variable processing costs at the locations of the facilities. Activities taken into account that could be implemented at the considered facilities are assembly, packaging, both assembly and packaging in combination, as well as a simple handling of products in a facility without executing any further activities. The third cost type represents infrastructure costs with a fixed cost character for establishing the considered functions in the facilities. Constraint sets (2) and (3) represent material balances. The flow entering a facility must equal the quantity that is processed in this facility. Analogously, constraint sets (5), (6) and (7) declare material balances which ensure that the flow leaving a facility has to be as high as

the quantity that is processed there. Constraint set (4) assures that every customer's zone demand for each product type is met. Constraint set (8) restricts the material flows leaving the factories to their maximum quantity they are able to supply. With constraint sets (9) - (11), the maximum quantity of products that can be shipped from one location to another one is incorporated into the distribution network model. The total capacity consumption by shipping different products along the same link is calculated by a linear combination of the capacity consumption of individual products. Constraint set (12) represents a similar approach regarding the maximum quantity that can be processed by an implemented function within a facility. The binary variables in constraint set (12) indicate whether a specific function should be established in a facility or not. Furthermore, the constraints also permit calculating the capacity consumption in a facility if the products processed with the implemented function in the facility feature different completion status (see Figure 3 and its explanations in Section 3). Constraint sets (13) - (16) act for non-negativity constraints, constraint set (17) defines the binary variables.

5 Test problems

To illustrate the applicability of the mathematical formulation presented in this paper, two test instances for distribution networks with four tiers are examined. In the first instance representing a small-sized test instance, four manufacturing plants producing two products are incorporated. Product demands can be related to four customer zones. Furthermore, two potential central warehouses as well as three potential regional warehouses are taken into account. The second test instance represents a large problem incorporating two manufacturing plants, three locations of potential central warehouses, ten locations of potential regional warehouses as well as fifty customer zones. Forty products are considered. The decisions that have to be made in both instances regard the question where to implement assembly and packaging functions for enabling postponement in distribution networks. Potential locations are central and regional warehouses. Furthermore, implementations in factories are also incorporated. The latter case meets the traditional approach of distribution networks where only assembled and packaged products leave a factory.

The MILP problems were solved using CPLEX 8.1.0 in a reasonable time. After execution of a preprocessing, the first problem consists of 189 linear constraints with 458 linear and 29 binary variables as well as 1083 nonzeros. The linear objective function incorporates 487 nonzeros.

Using a Personal Computer (Pentium P4, 2.4 GHz, 1 GB main memory), the solution time is 0.016 seconds. The second problem includes 8506 linear constraints with 95200 linear and 48 binary variables as well as 284528 nonzeros. The linear objective function contains 93088 non-zeros, and the solution time on the same computer seems acceptable with 84.031 seconds.

The optimal solution of the first test instance is presented graphically. In Figure 4, the resulting distribution network is shown. The two numbers attached to the facilities as well as to the shipping links denote the quantity of flow regarding the two products taken into account. Both assembly and packaging (a+p) are postponed to the regional warehouses at the third tier. Thus, in virtue of the optimal solution semi-finished products have to be shipped from the factories to the central warehouses and from the central warehouses to the regional warehouses. However, this result depends highly on the applied data set. Some additionally tested data sets with different shipping, handling, assembly and packaging costs as well as different capacities result in totally different solutions regarding an optimal location of assembly and packaging activities.

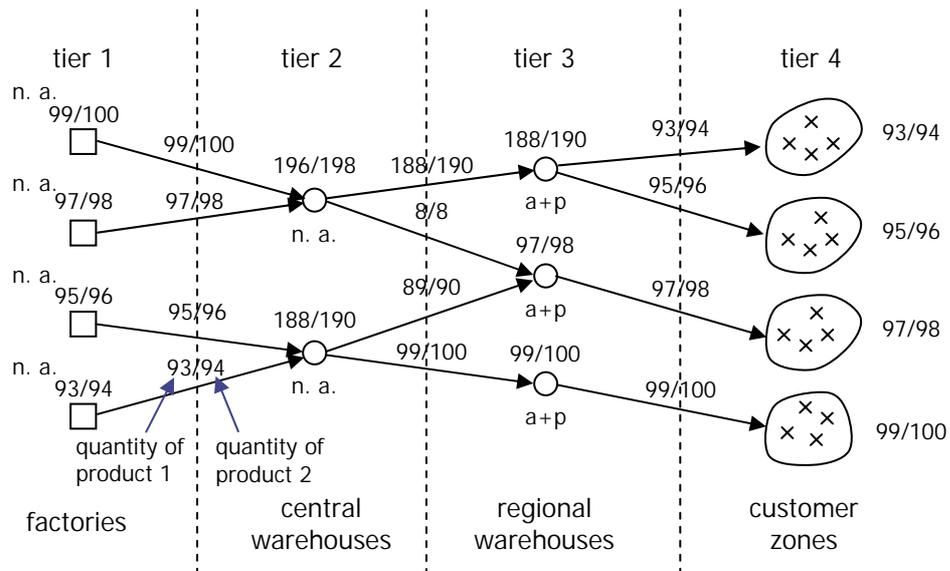


Figure 4: Optimal configuration of a distribution network (small-sized test instance)

6 Conclusions

In the presented paper the problem of designing a distribution network made up of four tiers taking into account aspects of postponement is studied. The developed mixed integer programming model enables the decision where to establish central and regional warehouses, where to implement potentially postponed functions for assembly and packaging in the distribution net-

work and which quantities should be shipped from one facility to another one. Two scenarios are conceived and the corresponding test instances are solved.

The proposed model can be regarded as a basis for further research. Thus, several aspects should be taken into account in future work, e.g., nonlinearities of costs, diverse capacity levels, alternative technologies for shipping as well as processing in the facilities, an observation across several periods (dynamic model) and thus considering inventories, explicitly taking into account means of transportation, or capital commitment, risk, as well as insurance contributions. Moreover, implications of incorporating our model into advanced planning systems need to be explored.

References

- [Aike85] Aikens, C.H.: Facility location models for distribution planning. In: *European Journal of Operational Research* 22 (1985), S. 263-279.
- [Alde50] Alderson, W.: Market efficiency and the principle of postponement. In: *Cost and Profit Outlook* 3 (1950) September, S. 15-18.
- [AmSc05] Ambrosino, D.; Scutellà, M.G.: Distribution network design: New problems and related models. In: *European Journal of Operational Research* 165 (2005), S. 610-624.
- [Amir06] Amiri, A.: Designing a distribution network in a supply chain system: Formulation and efficient solution procedure. In: *European Journal of Operational Research* 171 (2006), S. 567-576.
- [ApGu05] Appelqvist, P.; Gubi, E.: Postponed variety creation: Case study in consumer electronics retail. In: *International Journal of Retail & Distribution Management* 33 (2005) 10, S. 734-748.
- [ABHT95] Arntzen, B.C.; Brown, G.G.; Harrison, T.P.; Tafton, L.L.: Global supply chain management at Digital Equipment Corporation. In: *Interfaces* 25 (1995) 1, S. 69-93.

- [BiOz04] Bilgen, B.; Ozkarahan, I.: Distribution planning problem: A survey. In: Ahr, D.; Fahrion, R.; Oswald, M.; Reinelt, G. (Hrsg.): Operations Research Proceedings 2003. Springer, Berlin 2004, S. 39-46.
- [BrGH87] Brown, G.G.; Graves, G.W.; Honczarenko, M.D.: Design and operation of a multicommodity production distribution system using primal goal decomposition. In: Management Science 33 (1987) 11, S. 1469-1480.
- [Buck65] Bucklin, L.P.: Postponement, speculation and structure of distribution channels. In: Journal of Marketing Research 2 (1965), S. 26-32.
- [CCDE97] Camm, J.D.; Chorman, T.E.; Dill, F.A.; Evans, J.R.; Sweeney, D.J.; Wegryn, G.W.: Blending OR/MS, judgment, and GIS: Restructuring P&G's supply chain. In: Interfaces 27 (1997) 1, S. 128-142.
- [Chan93] Chandra, P.: A dynamic distribution model with warehouse and customer replenishment requirements. In: Journal of the Operational Research Society 44 (1993), S. 681-692.
- [ChFi94] Chandra, P.; Fisher, M.L.: Coordination of production and distribution planning. In: European Journal of Operational Research 72 (1994), S. 503-517.
- [CoLe89] Cohen, M.; Lee, H.L.: Resource deployment analysis of global manufacturing and distribution networks. In: Journal of Manufacturing and Operations Management 2 (1989), S. 81-104.
- [CoMo90] Cohen, M.A.; Moon, S.: Impact of production scale economies, manufacturing complexity, and transportation costs on supply chain facility networks. In: Journal of Manufacturing and Operations Management 3 (1990), S. 269-292.
- [Coop93] Cooper, J.C.: Logistics strategies for global business. In: International Journal of Physical Distribution and Logistics Management 23 (1993) 4, S. 12-23.
- [ErKa00] Ernst, R.; Kamrad, B.: Evaluation of supply chain structures through modularization and postponement. In: European Journal of Operational Research 124 (2000), S. 495-510.

- [EUPB05] Eskigun, E.; Uzsoy, R.; Preckel, P.V.; Beaujon, G.; Krishnan, S.; Tew, J.D.: Outbound supply chain network design with mode selection, lead times and capacitated vehicle distribution centers. In: *European Journal of Operational Research* 165 (2005), S. 182-206.
- [FeLe97] Feitzinger, E.; Lee, H.L.: Mass customization at Hewlett-Packard: The power of postponement. In: *Harvard Business Review* 75 (1997) 1, S. 116-121.
- [Flei93] Fleischmann, B.: Designing distribution systems with transportation economies of scale. In: *European Journal of Operational Research* 70 (1993), S. 31-42.
- [GaTa97] Garg, A.; Tang, C.S.: On postponement strategies for product families with multiple points of differentiation. In: *IIE Transactions* 29 (1997), S. 641-650.
- [GeGr74] Geoffrion, A.M.; Graves, G.W.: Multicommodity distribution system design by Benders decomposition. In: *Management Science* 20 (1974) 5, S. 822-844.
- [GoVD02] Goetschalckx, M.; Vidal, C.J.; Dogan, K.: Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. In: *European Journal of Operational Research* 143 (2002), S. 1-18.
- [JaPi01] Jayaraman, V.; Pirkul, H.: Planning and coordination of production and distribution facilities for multiple commodities. In: *European Journal of Operational Research* 133 (2001), S. 394-408.
- [JoRi85] Jones, T.C.; Riley, D.W.: Using inventory for competitive advantage through supply chain management. In: *International Journal of Physical Distribution and Materials Management* 15 (1985), S. 16-26.
- [KaMN03] Kalcsics, J.; Melo, M.T.; Nickel, S.: Mathematical programming models for strategic supply chain planning and design. In: Leopold-Wildburger, U.; Rendl, F.; Wäscher, G.: *Operations Research Proceedings 2002*. Springer, Berlin 2003, S. 108-113.
- [Klin85] Klincewicz, J.G.: A large-scale distribution and location model. In: *AT&T Technical Journal* 64 (1985) 7, S. 1705-1730.

- [KlDr05] Klose, A.; Drexler, A.: Facility location models for distribution system design. In: *European Journal of Operational Research* 162 (2005), S. 4-29.
- [LeBC93] Lee, H.L.; Billington, C.; Carter, B.: Hewlett-Packard gains control of inventory and service through design for localization. In: *Interfaces* 23 (1993) 4, S. 1-11.
- [MaWL02] Ma, S.; Wang, W.; Liu, L.: Commonality and postponement in multistage assembly systems. In: *European Journal of Operational Research* 142 (2002), S. 523-538.
- [MiSk04] Mikkola, J.H.; Skjøtt-Larsen, T.: Supply-chain integration: Implications for mass customization, modularization and postponement strategies. In: *Production Planning & Control* 15 (2004) 4, S. 352-361.
- [OwDa98] Owen, S.H.; Daskin, M.S.: Strategic facility location: A review. In: *European Journal of Operational Research* 111 (1998), S. 423-447.
- [PaCo98] Pagh, J.D.; Cooper, M.C.: Supply chain postponement and speculation strategies: How to choose the right strategy. In: *Journal of Business Logistics* 19 (1998) 2, S. 13-33.
- [PiJa96] Pirkul, H.; Jayaraman, V.: Production, transportation, and distribution planning in a multi-commodity tri-echelon system. In: *Transportation Science* 30 (1996) 4, S. 291-302.
- [Pool94] Pooley, J.: Integrated production and distribution facility planning at Ault Foods. In: *Interfaces* 24 (1994) 4, S. 113-121.
- [SaBe00] Sabri, E.H.; Beamon, B.M.: A multi-objective approach to simultaneous strategic and operational planning in supply chain design. In: *Omega* 28 (2000) 5, S. 581-598.
- [SkHa04] Skipworth, H.; Harrison, A.: Implications of form postponement to manufacturing: A case study. In: *International Journal of Production Research* 42 (2004), S. 2063-2081.

- [Stad05] Stadtler, H.: Supply chain management and advanced planning – basics, overview and challenges. In: *European Journal of Operational Research* 163 (2005), S. 575-588.
- [SwTa99] Swaminathan, J.M.; Tayur, S.R.: Managing design of assembly sequences for product lines that delay product differentiation. In: *IIE Transactions* 31 (1999), S. 1015-1026.
- [TsSP01] Tsiakis, P.; Shah, N.; Pantelides, C.C.: Design of multi-echelon supply chain networks under demand uncertainty. In: *Industrial & Engineering Chemistry Research* 40 (2001) 16, S. 3585-3604.
- [vHoe98] van Hoek, R.I.: Reconfiguring the supply chain to implement postponed manufacturing. In: *The International Journal of Logistics Management* 9 (1998) 1, S. 95-110.
- [vHoe01] van Hoek, R.I.: The rediscovery of postponement: A literature review and directions for research. In: *Journal of Operations Management* 19 (2001) 2, S. 161-184.
- [ViGo97] Vidal, C.J.; Goetschalckx, M.: Strategic production-distribution models: A critical review with emphasis on global supply chain models. In: *European Journal of Operational Research* 98 (1997), S. 1-18.
- [YaBu03] Yang, B.; Burns, N.: Implications of postponement for the supply chain. In: *International Journal of Production Research* 41 (2003), S. 2075-2090.
- [YeYa03] Yeh, C.; Yang, H.-C.: A cost model for determining dyeing postponement in garment supply chain. In: *The International Journal of Advanced Manufacturing Technology* 22 (2003) 1-2, S. 134-140.
- [ZiBo88] Zinn, W.; Bowersox, D.J.: Planning physical distribution with the principle of postponement. In: *Journal of Business Logistics* 9 (1988) 2, S. 117-136.

Order-driven planning in build-to-order scenarios

Thomas Volling, Thomas Spengler

Department of Production and Logistics Management
Technical University Braunschweig
38106 Braunschweig
{t.volling, t.spengler}@tu-braunschweig.de

Abstract

The adoption of build-to-order order fulfillment processes gives rise to a paradigm shift in production planning. Since all business is linked to customer orders, it is the order-driven planning activities that determine the success of operations. Therefore, a clear understanding of the associated planning tasks order promising and master production scheduling as well as their dynamic interaction is essential. Based on an analysis of the decision situation we provide a hierarchical framework comprising quantitative models for order promising and master production scheduling. The approach is evaluated using simulative analysis.

1 Introduction

The implementation of customer-oriented manufacturing strategies has frequently been associated with significant gains in profit and a competitive edge [GuNg05; ShLa05]. Yet, striving to serve high volume markets with individualized or customized products, new manufacturing concepts are needed. By switching to build-to-order (BTO) order fulfillment processes, companies seek to likewise benefit from characteristics like scale effects, standardized processes, and a high quality attributed to high volume make-to-stock strategies as well as from the high flexibility to adopt to diverse and changing market demand of make-to-order strategies [Pine99]. In doing so, mixed model assembly systems are the dominating production typology [BoFS06]. The constituting characteristic of this type is the coupling of assembly stations to flow-lines in a serial manner in accordance with the product structure. Mixed model assembly is thereby distinct from specialized flow-lines by the fact that a number of variants, or models respectively, is manufactured on the same production line. Facing heterogeneous production sequences, how-

ever, technological and organizational set-up efforts between consecutive models need to be reduced drastically. As a consequence, orders can be produced in lot sizes of one piece without loss of profitability [Scho99, 7].¹ Yet, considering the varying work content of different models at certain assembly stations, not every sequence is feasible given a line set-up. This problem of determining feasible production sequences is referred to as model-mix planning [WeKi64]. Sequence planning has attracted a lot of attention within the operations management community. The major implication of BTO, however, lies in the reduction of decoupling mechanisms against demand characteristics [GuNg05]. Production activities are thus subject to the induced variability and dynamics of the market. To be specific, these are characterized by the timing and characteristics of customer requests (i.e. the demand sequence), the resulting model-mix (demand structure), and the aggregated demand per period (demand level). At the same time, real-world production systems are characterized by a limited flexibility in terms of production and procurement capabilities. The synchronization of capacity with the volatility of the environment is therefore not viable. Instead, adequate control concepts are needed to match the supply of resources with the demand for products. These encompass policies for the determination of due dates in response to customer requests as well as for the consolidation of those promised orders into production plans. The associated planning tasks are referred to as order based planning. Commercial applications to support order based planning have been introduced by leading business software providers as integral part of advanced planning and scheduling systems (APS).² Yet, to our knowledge, there is no analysis available boldfacing the performance of such systems in dynamic environments such as BTO scenarios. The answer to this question is in particular important, since the decision situation in real-world settings can be described by the complex interaction of planning functions, a high number of potentially influential parameters as well as an unknown dynamic performance of the decision support systems. Against this background, the aim of this paper is to provide a hierarchical framework for order based planning in BTO scenarios, to develop mathematical programs for that framework and to evaluate the approach using simulation. The remainder is organized as follows: In section 2 we will highlight the decision situation of order based planning in BTO scenarios and review prior work on the subject. Based on that, models will be presented in section 3 and analyzed using simulation in section 4. Some concluding remarks are given in section 5.

¹ To be more specific, mixed-model flow lines typically require for an - on average - levelled load of their constituting stations. This prohibits extended sequences with similar capacity demand and therefore batch production.

² e.g. Real Time Positioning [SAP05], Global Order Promising [Orac05]

2 Decision Situation

Order based planning is distinct to its anticipative or push-based counterpart as to its reactive character (i.e. it regards orders placed instead of forecasted demand). The decision situation can thus be characterized as follows: In order to cope with the restricted capabilities to adjust capacities, the mid-term coordination on the basis of forecasts and orders at hand is essential. This results in aggregated production plans, which serve as cornerstones to synchronise further business functions like purchasing and capacity planning. The realisation of market demand in terms of customer orders requires for an adequate reaction. This function is usually denominated as *order promising* (OP) and comprises decisions about whether to accept or deny an order and which due date to promise [FIMe04]. Discrepancies between forecasting and orders placed as well as the potentially limited level of detail of OP might require for further adjustments to the plan. More specifically, promised due dates, instructions from the subordinate planning, characteristics of the production systems, and the ability to service new requests are to be taken into account. The resulting goal conflicts are to be resolved in the course of a second planning function called *master production scheduling* (MPS).

The complexity of the decision situation requires for an adequate structuring approach. In the following we will therefore present a hierarchical framework for order based planning in BTO scenarios. This lays the basis to reflect prior work in section 2.2 and to develop and evaluate decision models in sections 3 and 4.

2.1 Conceptual framework

Architectures of hierarchical planning systems are subject of numerous studies (e.g. [FIMe03; VBWJ05, 8]). However, these general frameworks need to be adopted in order to adequately fit decision models. Figure 1 depicts a framework for order-based planning in BTO scenarios and its interfaces to further planning tasks. The integral components are modules for OP and MPS.

Considering OP, requirements in terms of customer response time, the reliability and level of detail of the quoted offer as well as restrictions regarding the availability of resources have to be taken into account. Also, OP procedures usually have to cope with a high number of requests, which induce significant dynamics into planning. Finally, instructions from the subordinate demand planning (e.g. quotas regarding volumes dedicated to certain distribution channels) might need to be incorporated. The outcome is a pool of fully specified orders (in terms of product configuration and delivery date) which is transferred to MPS. More specifically, MPS

seeks to coordinate production, procurement, and sales on the short-run in order to facilitate efficient resource utilization [VBWJ05, 169; FlMe04]. Considering production and procurement, a leveled demand, i.e. a stable utilization of the resources (including demand for components) is to be achieved [EnZi98]. A specific objective could be to minimize deviations from specific levels of capacity utilization, which determine efficient operating points of the production line [Bola03]. The same holds true for component demand. Usually minimal and maximal levels are agreed on by means of mid-term supply contracts. With respect to sales, the ability to service newly arriving orders is of great importance. That means that MPS ought to minimize bottlenecks. Planning objects are specified orders, which are to be assigned to production lines and periods. Capacities are usually treated as fixed, as provided by the master production planning [Scho99, 111]. Executing MPS results in aggregated plans that are passed on to sequencing and material requirements planning for further detailing.

With an increasing number of orders becoming known as time progresses, the question of identifying adequate re-planning frequencies arises. A high frequency results in a high number of changes to the plan, which contradicts the aim of a stable coordination. A low frequency, on the contrary, potentially results in a delayed transfer of planning information. The integrated assessment of OP and MPS therefore seems to be inappropriate.

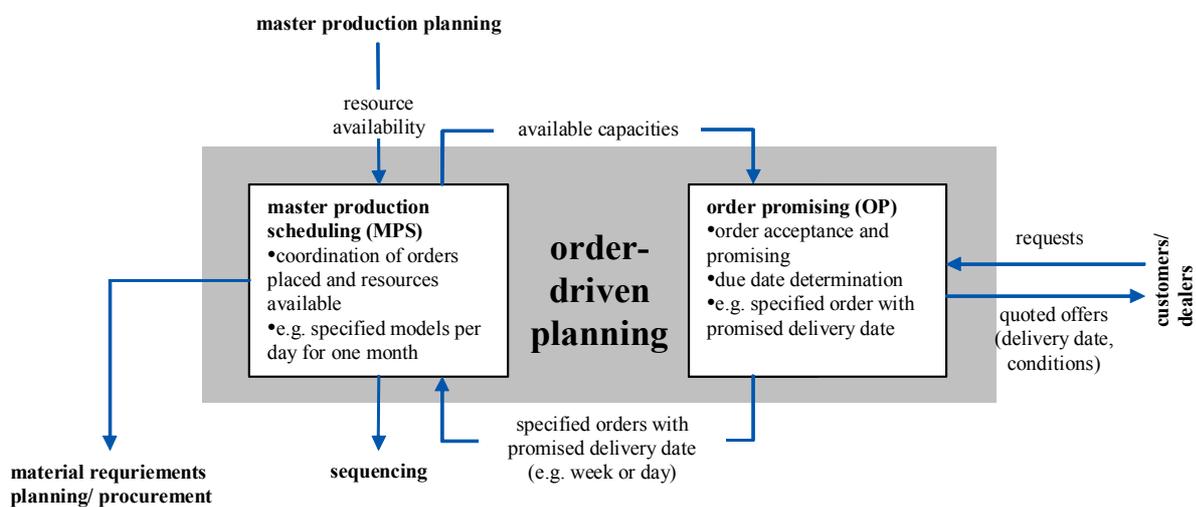


Figure 1: Framework for order based planning in BTO scenarios

Reconsidering the proposed framework, the decision situation can be interpreted as a hierarchical system composed of OP and MPS [SpVR06]. Referring to the work from [Schn03] this setting is called tactical-operational. More specifically, OP determines due dates for newly arriving orders (the so called factual instruction) considering the information available at the times of its execution (e.g. at each order arrival). In a second frequency, but usually at a later point in time,

MPS is executed, given the prior instructions, i.e. all promised orders up to that point in time, with the aim to assign these to planning periods (the final instruction). The decision of MPS is thereby dependent on that of OP, such that it constitutes the operational level within the horizontal hierarchy while OP demarks the tactical level. The final instruction, i.e. the aggregated production plan, depends on both decisions.

2.2 Related work

Approaches to order based planning in BTO scenarios are subject of two streams of research. These have on the one hand evolved from work on mixed-model assembly line balancing and sequencing and were on the other developed under the denomination available-to-promise (ATP) as part of work on hierarchical production planning and control systems (e.g. APS).

Regarding the former, aspects of order based planning have been incorporated into sequencing, i.e. the determination of optimal production sequences by means of model-mix planning. These approaches are, however, based on the assumption, that orders are exogenously given [Scho99, 107]. With respect to the framework introduced above, this means, that only the MPS-functionality is being investigated. Also, considering planning horizons needed for the reasonable leveling of production, the computational burden to solve the detailed sequencing problems is prohibitive.³ Therefore aggregated models have been presented in recent contributions.

[Bola03] provides solution procedures for the problem to select orders to be produced in the upcoming period out of a pool of promised orders. The objective is to minimize delivery date dependent costs. Leveling aspects are incorporated in setting ranges for acceptable capacity utilization levels. Potential gains of a branch & bound procedure to solve the resulting multi-dimensional knapsack problem are illustrated by a numerical analysis. In a recent publication [Boys05] presents a model that explicitly incorporates desired due dates into MPS for mixed-model assembly. The objective is to minimize delivery dependent costs. Also extensions to line assignment and leveling are provided. Both approaches do not distinguish between OP and MPS. The requirements of order based planning are therefore only partly satisfied.

ATP approaches, on the contrary, focus on decision support for OP. In analogy to the approaches discussed before, *optimization based batch ATP* aims at processing a pool of specified but not yet promised orders at hand (i.e. either rejecting or quoting each order). The objective is to maximize profit. Results include promised due dates and delivery conditions. Optimizing

³ Even the sequencing problem for single stations is NP-hard [Tsai95].

batch models have been discussed by [ChZB02] as well as [Pibe05]. These models seek to maximize contribution margin with respect to real (e.g. holding, production, and procurement costs) and ‘soft costs’ (e.g. costs for late delivery or for low capacity utilization). A central premise of those approaches is, that in the course of sequential approach, promised due dates and quantities are treated as fixed restrictions for subsequent planning cycles. Accordingly, good results have been numerically proven for rather large batching intervals. A reduction of the intervals, however, results in the significant deterioration of the planning performance. A similar model structure is introduced by [FlMe04]. To avoid the illustrated drawbacks of short batching intervals, the authors propose the coupling with MPS. They however do not elaborate on the consequences in terms of modeling and performance. *Rule based batch ATP* aims at defining decision rules for the acceptance of orders [KiSc05]. A hierarchical approach is provided by [JSJK02] for the case of LCD production. The authors differentiate between a rule-based OP routine and a heuristic to determine the unused capacity based on a detailed schedule in a job-shop setting. The approach, however, does not incorporate the specific characteristics of BTO as described above. This in particular holds true for the requirements of high variance flow-shop production. A drawback of batch approaches is that they do not support the interaction between customer and company. *Rule-based real-time approaches* in contrast build on an ad-hoc assessment of the ability to deliver at a certain point in time. A recent review is given by [MGGP04]. Various extensions incorporating sophisticated statistics as well as stochastic influences have been published (e.g. [WGHH04]). The aggregated assessment seems adequate for decoupled production systems. It in turn does not reflect aspects like leveling, which are special to BTO settings. *Optimization based real-time approaches* have to our knowledge only been presented for academic examples (e.g. [Kate94]). The objective is to quickly determine detailed schedules. Due to the complexity of the problem the scalability of the approach is limited. To sum up, knowledge on the performance of planning schemes, in particular regarding the interplay of OP and MPS, is limited. In the following we will therefore provide a mathematical program for order based planning in BTO scenarios.

3 Modeling

An adequate model to support order based planning in BTO scenarios has to cope with the divergent requirements of the customer interaction on the one hand and resource related aspects

on the other. We therefore chose a hierarchical approach, comprising a real-time ATP model as well as a MPS model for mixed-model assembly planning on rolling horizons.

3.1 The order acceptance problem

Time is discretized in equidistance planning periods $t=1,2,\dots,T^{\max}$. Orders i ($i=1,\dots,I$) arrive randomly and are evaluated individually upon their arrival. Evaluation thereby refers to the determination of a production period τ ($\tau = t, \dots, T^{\max}$), considering capacities available and starting from the first disposable period t . Production coefficients a_{ir} are used to transform each order's specific product configuration into capacity requirements for each resource r . They likewise reflect the demand for resource capacity of the production system ($r \in \Omega^1$) as well as those of component availability usually referred to as ATP ($r \in \Omega^2$). The cumulated index set of all resources will be referred to as Ω ($\Omega = \Omega^1 \cup \Omega^2$). We assume that all orders have been economically evaluated by means of a subordinate demand management and have positive contribution margins. Orders have been back scheduled with respect to their production and distribution lead times, such that the analyses can be restricted to the (planned) start of production. Alternative production facilities are not considered.

A bucketized model for the binary assignment of order i to planning period τ is therefore given by (1)-(4). In the course of OP, each order is being assigned to a period within the planning horizon $[t, T^{\max}]$, such that the associated costs are minimized (1). Due to the real-time execution mode, each order is treated separately. This decision is modeled by the vector \mathbf{x}_i with the elements $x_{i\tau}$, which are set to 1 if the order is assigned to the particular period and 0 if otherwise. The relevant costs of an assignment $c_{i\tau}$ reflect the customer preferences and are a function of the requested period and the promised period.⁴ Inequalities (2) incorporate resource constraints into the model. $ctp_{r\tau}$ thereby refers to the available capacity at the time of the order arrival and is updated with each order being processed. Equations (3) assure that the order is assigned to a period. The result of the OP procedure is a promised planning period, or the corresponding delivery date respectively, for a certain product configuration.

$$\text{Minimize} \quad C^{OP}(\mathbf{x}_i) = \sum_{\tau=t}^{T^{\max}} c_{i\tau} \cdot x_{i\tau} \quad (1)$$

$$x_{i\tau} \cdot a_{ir} \leq ctp_{r\tau} \quad \forall r \in \Omega, \tau = t, \dots, T^{\max} \quad (2)$$

⁴ For an in-depth discussion of time dependent costs refer to [Bola03].

$$\sum_{\tau=t}^{T^{\max}} x_{i\tau} = 1 \quad (3)$$

$$x_{i\tau} \in \{0,1\} \quad \forall \tau = t, \dots, T^{\max} \quad (4)$$

3.2 The master production scheduling problem

The decision to come to in the course of the MPS procedure is to determine final planning periods for the set of promised but not finalized orders Ψ . The MPS is executed on rolling horizons once every planning period with a planning horizon of T periods. Accordingly only the first planning period is put into practice, while the others are being updated with respect to the new information available (i.e. the newly promised orders). Since both, restrictions of the production system as well as those regarding component availability have to be incorporated into the decision making to assure feasibility, we use a differentiated definition of capacity: considering a flow-shop, the maximum capacity per period cap_{τ}^{\max} is given by the quotient of the overall production time Q_{τ} and the cycle time W_{τ} . In addition to that, model-mix restrictions have to be incorporated into the calculus to adequately reflect the sequencing limitations of the prevailing mixed-model production system. Therefore we reduce the overall capacity of the production system by means of a coefficient α_r for each capacitated resource $r \in \Omega^1$.⁵ The maximum aggregated capacity available per resource and period is hence given by $\alpha_r \cdot \text{cap}_{\tau}^{\max}$. Further resource restrictions ($r \in \Omega^2$) regard maximum levels for the component availability in a planning period $\text{cap}_{r\tau}^{\text{SM}}$. These do not depend on the subsequent sequencing and can consequently be assumed exogenously given (e.g. by the subordinate planning). Accordingly, the maximum capacity of resource r in period τ is given by:

$$\text{cap}_{r\tau}^{\max} = \begin{cases} \alpha_r \cdot \text{cap}_{\tau}^{\max} & \text{if } r \in \Omega^1, \\ \text{cap}_{r\tau}^{\text{SM}} & \text{if } r \in \Omega^2. \end{cases} \quad (5)$$

In order to reflect the divergent requirements discussed above we implemented two intervals for the objective function: the first interval ranges from t to k and the second one from $k+1$ to the end of the planning horizon, with $t \leq k \leq t + T - 1$. Let further $\text{ctp}_{r\tau}^{-}$ denote the standardized

⁵ To determine α_r so called $H_0:N_0$ -rules have been used. The interpretation is, that no more than H_0 units out of a sequence with length N_0 may require an option o or a corresponding resource r respectively [DrKi01].

deviations to the target capacity consumption level $cap_{r\tau}^{\min}$ and $ctp_{r\tau}^+$ the capacity available $ctp_{r\tau}$ standardized with respect to the maximum capacity level $cap_{r\tau}^{\max}$. We furthermore used the weighting functions $P_{r\tau}^{\text{leveling}}(\cdot)$ to assess deviations to the targeted capacity utilization of resource r in period τ (leveling aspects) and $P_{r\tau}^{\text{service}}(\cdot)$ to assess the available capacity of resource r in period τ (measure of the ability to service new orders). Finally, $\bar{c}_{i\tau}$ denotes the cost of assigning order i to period τ .⁶ The program MPS is thus given as:

$$\text{Minimize } C^{MPS}(\bar{x}) = \sum_{\tau=tr \in \Omega}^{t+k} P_{r\tau}^{\text{leveling}}(ctp_{r\tau}^-) - \sum_{\tau=t+k+1}^{t+T-1} \sum_{r \in \Omega} P_{r\tau}^{\text{service}}(ctp_{r\tau}^+) + \sum_{\tau=ti \in \Psi}^{t+T-1} \bar{c}_{i\tau} \cdot \bar{x}_{i\tau} \quad (6)$$

w.r.t.

$$\frac{cap_{r\tau}^{\min} + ctp_{r\tau} - cap_{r\tau}^{\max}}{cap_{r\tau}^{\min}} \leq ctp_{r\tau}^- \quad \forall r \in \Omega; \tau = t, \dots, t+k \quad (7)$$

$$\frac{ctp_{r\tau}}{cap_{r\tau}^{\max}} \geq ctp_{r\tau}^+ \quad \forall r \in \Omega; \tau = t+k+1, \dots, t+T-1 \quad (8)$$

$$\sum_{i \in \Psi} \bar{x}_{i\tau} \cdot a_{ir} = cap_{r\tau}^{\max} - ctp_{r\tau} \quad \forall r \in \Omega; \tau = t, \dots, t+T-1 \quad (9)$$

$$\sum_{\tau=t}^{t+T-1} \bar{x}_{i\tau} = 1 \quad \forall i \in \Psi \quad (10)$$

$$ctp_{r\tau}, ctp_{r\tau}^-, ctp_{r\tau}^+ \geq 0 \quad \forall \tau = t, \dots, t+T-1 \quad (11)$$

$$\bar{x}_{i\tau} \in \{0,1\} \quad \forall i \in \Psi, \tau = t, \dots, t+T-1 \quad (12)$$

The assignment is modeled by matrix \bar{x} with T-dimensional column vectors \bar{x}_i . Doing so, the binary vector elements $\bar{x}_{i\tau}$ are set to 1 if an order i is assigned to period τ in analogy to the OP procedure described above. The objective of the MIP (6)-(12) is then to minimize the relative deviations to the targeted utilization levels $ctp_{r\tau}^-$ weighted by $P_{r\tau}^{\text{leveling}}(\cdot)$ throughout the interval $[t, t+k]$ and to maximize the weighted capacity available $ctp_{r\tau}^+$ as given by $P_{r\tau}^{\text{service}}(\cdot)$ throughout the interval $[t+k+1, t+T-1]$. In addition to that, service objectives are to be incorporated by means of the cost of assigning order i to period τ . The figures for the first interval result from standardizing shortfalls on the targeted capacity utilization $cap_{r\tau}^{\min}$ according to inequali-

⁶ These are in contrast to $c_{i\tau}$ a measure of bringing forward or delaying an order given a specified promised delivery date and therefore reflect e.g. holding costs or penalties respectively.

ties (7) (figure 2). Those of the second interval are calculated as given by inequalities (8). Constraints (9) assure feasibility with respect to resource capacities, whereas ctp_{rt} specifies the prevailing slack. Set coverage constraints (10) assure that each order is assigned to one period while constraints (11)-(12) define non-negativity and binary coding for the variables. The two models are coordinated by means of the coupling conditions (9) (bottom-up) and the service costs \bar{c}_{it} (top-down), since they depend on the promised period.

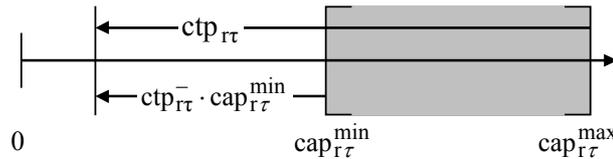


Figure 2. Determination of the standardized deviation to the targeted minimal capacity utilization

To demonstrate the meaning of the segmented objective function, refer to the illustrative example depicted in figure 3. The figure shows the distribution function and the expected number of customer orders placed y as a function of the period x within the planning horizon, given a demand of 20 orders per planning period and normal distributed ordering lead times with a mean of 10 and a standard deviation of 3 periods (i.e. considering the 10th period of the planning horizon, on average 50% of all customers demanding a product for that particular period will have placed their orders at the time of the MPS execution). Given a targeted level for the minimal capacity utilization of 80% (i.e. 16 orders per period) two intervals can be identified. For period 1 to 7, the expected orders transcend the critical level, while the opposite is true for periods 8 to 21. By segmenting the objective function, a stable resource utilization can be pursued in the first interval for which (on average) most orders have been placed, while at the same time bottlenecks can be avoided in the second one with a high (expected) inflow of new orders.

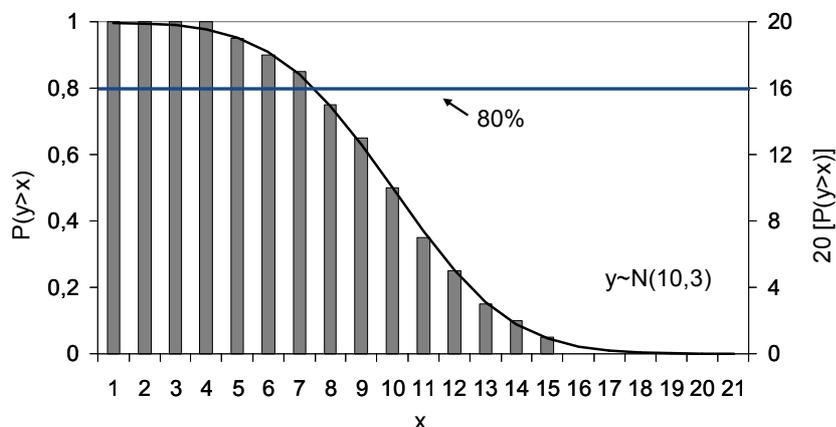


Figure 3: Illustrative order arrival pattern for $y \sim N(10;3)$

The implementation of the objective function requires to economically assess the particular terms. Monetary consequences of an unbalanced model-mix include costs for extra staff, overtime, or a reduced productivity (e.g. due to slack) as well as additional procurement costs. Monetary consequences of a reduced responsiveness include lost revenues (i.e. opportunity costs) since newly arriving orders cannot be served (lost sales). The explicit determination of monetary consequences often raises difficulties in real-world settings, since data is not available in the sufficient quality or the computational complexity does not allow for an explicit determination. This essentially leads to a multi-criteria decision situation. For such settings, additive weighting functions have been used to identify compromise solutions [RoWa05].

The models introduced above are characterized by a high number of parameters influencing the static but in particular the dynamic performance, i.e the interplay of OP and MPS. For similar settings, simulation studies have been used successfully. In the remainder, we will evaluate the presented framework using discrete-event simulation.

4 Simulation experiments

The scope of the following evaluation is twofold: at first it is to be shown, that the structure of the MPS objective function controls the simulation response and, in doing so, to numerically proof the meaning of the two objective function intervals. A second objective is to analyze the performance of the presented approach. Therefore we will compare the performance of the proposed framework to a policy restricted to OP (i.e. no MPS execution). The analysis is performed in three steps. At first, a demand sequence is generated for each replication. More specifically, each sequence comprises a set of orders with a specific configuration, an order date, and a preferred delivery date. In a second step, this demand sequence is subject to the order based planning approach to be analyzed. Accordingly, each order is at first promised individually and secondly assigned to a production period in the course of a MPS procedure executed on rolling horizons. In a third step, the results are evaluated. For the implementation we used eM-Plant as simulation tool and the commercial solver Lingo to solve the mathematical programs.

4.1 Experimental Data and Performance Measures

We considered an autoregressive moving-average demand process to derive the aggregated demand level d_τ of period τ according to $d_\tau = 20 + 0.8 \cdot (d_{\tau-1} - 20) - 0.1 \cdot \varepsilon_{\tau-1} + \varepsilon_\tau$ with

$\varepsilon_i \sim N(0;2)$. In a second step, we calculated normal distributed lead times with a mean of 10 and a standard deviation of 3 periods and subtracted them from the designated production period to achieve individual order arrival times.⁷ These arrival times were then used to synthesize the consolidated demand stream. Finally, capacity coefficients a_{ir} were derived according to take-rates $take_r$ for each of the two resource r using a uniformly (0,1)-distributed random number rnd as suggested by [Boys05]:

$$a_{ir} = \begin{cases} 1 & \text{if } rnd \leq take_r, \\ 0 & \text{else.} \end{cases} \quad (13)$$

For the analysis we assumed $take_1 = 0.7$ and $take_2 = 0.3$. Each simulation run covered 40 periods such that on average 800 orders were processed. MPS was executed once every period. In order to evaluate the effect of the capacity/demand ratio λ , we set the maximum capacity per resource and period to $cap_{r\tau}^{\max} = \lambda \cdot take_r \cdot E[d_\tau]$. The minimal targeted capacity utilization $cap_{r\tau}^{\min}$ was set to 80% of the maximal capacity. We assumed a linear cost function to determine c_{it} : each period earlier than the requested added 1 unit to the costs while each period later than the requested added 5 units. \bar{c}_{it} was assumed to be the same as c_{it} for earliness; lateness was not allowed for. Accordingly, the following term was added to the MPS model.

$$\sum_{\tau=t}^{t+T-1} \tau \cdot (\bar{x}_{i\tau} - x_{i\tau}) \leq 0 \quad \forall i \in \Psi \quad (14)$$

(Piecewise) linear terms were implemented to evaluate the effect of the first two terms of the MPS objective function using the coefficients p_1 and p_2 ((15) and (16)). The interval parameter k was set to 6.

$$P_{r\tau}^{\text{leveling}}(\text{value}) = p_1 \cdot \text{value} \quad (15)$$

$$P_{r\tau}^{\text{service}}(\text{value}) = \begin{cases} 0 & \text{for value} < 0.5 \cdot cap_{r\tau}^{\max}, \\ p_2 \cdot \text{value} & \text{for value} \geq 0.5 \cdot cap_{r\tau}^{\max}. \end{cases} \quad (16)$$

We incorporated four objectives into the analysis. At first, the total under-utilization and the standard deviation of the utilization were evaluated as measures for the compliance with resource-oriented objectives. In addition to that, we analyzed the customer-oriented performance

⁷ The reason for the differentiated approach is that the order behavior might be explained with the central limit theorem, since it reflects the behavior of a large number of independent entities (i.e. the customers). The demand level, however, underlies various influences which might cause auto correlation (e.g. hockey stick effect, marketing campaigns). Ongoing research seeks to build a more thorough understanding of the arrival process.

be means of the average costs of the assignment in the course of OP and the standard deviation of this figure. Figures were computed as given in the appendix.

4.2 Experimental Design

A full factorial design was used to evaluate the structure of the MPS objective function. The aim was to deduce the importance of the two objective function intervals in terms of their main effects and interactions. Since the performance of order based planning was expected to depend on the capacity/demand ratio, we added it as another experimental factor. Using two levels for each factor (i.e. 2^k -factor design) cumulated in $2^3=8$ configurations (called scenarios in the following). 80 replications were run for each scenario. Since we expected all performance measures to strongly correlate with the demand scenario (i.e. number of orders, preferred lead time, and model-mix), we used a common random number approach. Table 1 summarizes the factors and levels used for the analysis.

factor	denomination	levels	degrees of freedom
weighting of the leveling term	p_1	10-10,000	1
weighting of the service term	p_2	10-10,000	1
capacity/demand ratio	lambda	1.0-1.2	1

Table 1: Factorial design

4.3 Numerical Results

The results of the analysis are shown in the subsequent profile plots (figure 4 and 5). Depicted are the estimated means for the particular scenarios as compared to the performance of the policy restricted to OP (baseline policy in the following).⁸ The slope can be interpreted as the (main) effect of changing p_2 from its low level to its high level, while the offset is attributed to the level of p_1 . A change in the offset indicates interaction effects between the factors. Hence, the findings can be summarized as follows. p_1 affects all performance measures positively. The effect is in particular relevant for the resource-oriented objectives and a high capacity/demand ratio. This finding seems intuitively clear, since the first interval explicitly seeks to minimize deviations to the minimal targeted capacity. p_2 positively affects both the customer and the resource-oriented performance measures for tight capacity scenarios while the opposite is true for the customer-oriented figures and a high capacity/demand ratio, yet on a very low (absolute)

⁸ Doing so, we standardized the difference between each scenario's results and the baseline policy's with respect to the baseline policy. Accordingly, higher values indicate a better performance.

level. There is no effect of p_2 regarding the resource-oriented objectives for a high capacity/demand ratio. The analysis does not show any major interaction effects.

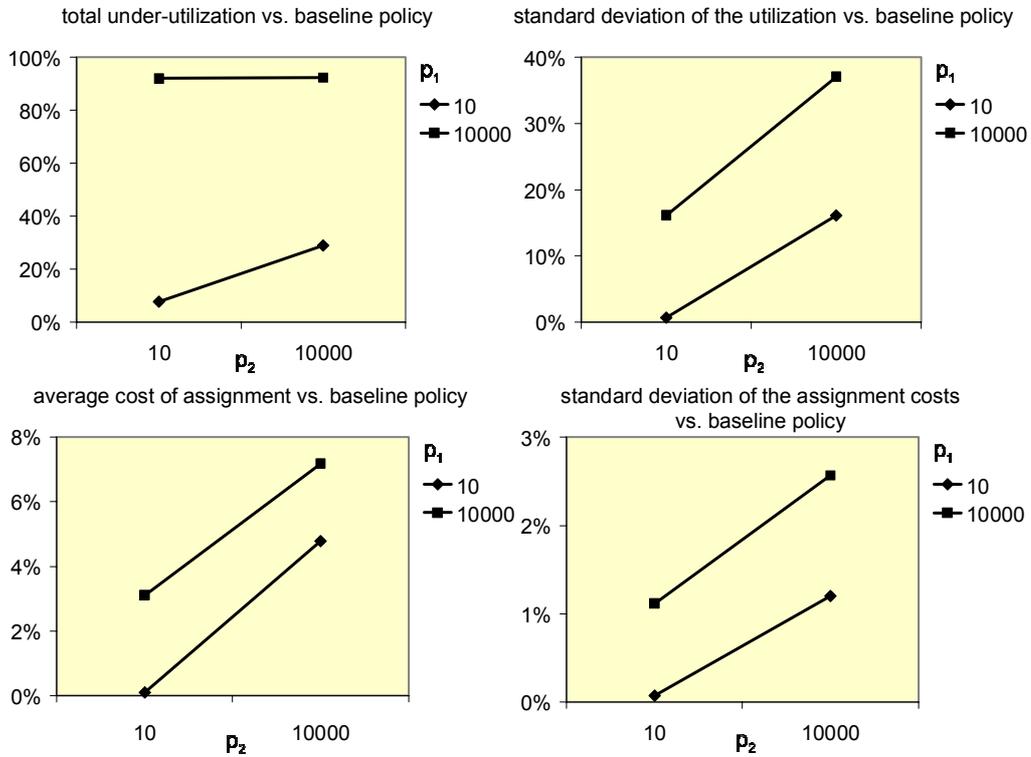


Figure 4: Profile plots for capacity/demand ratio 1.0 (all figures are standardized with respect to the baseline policy)

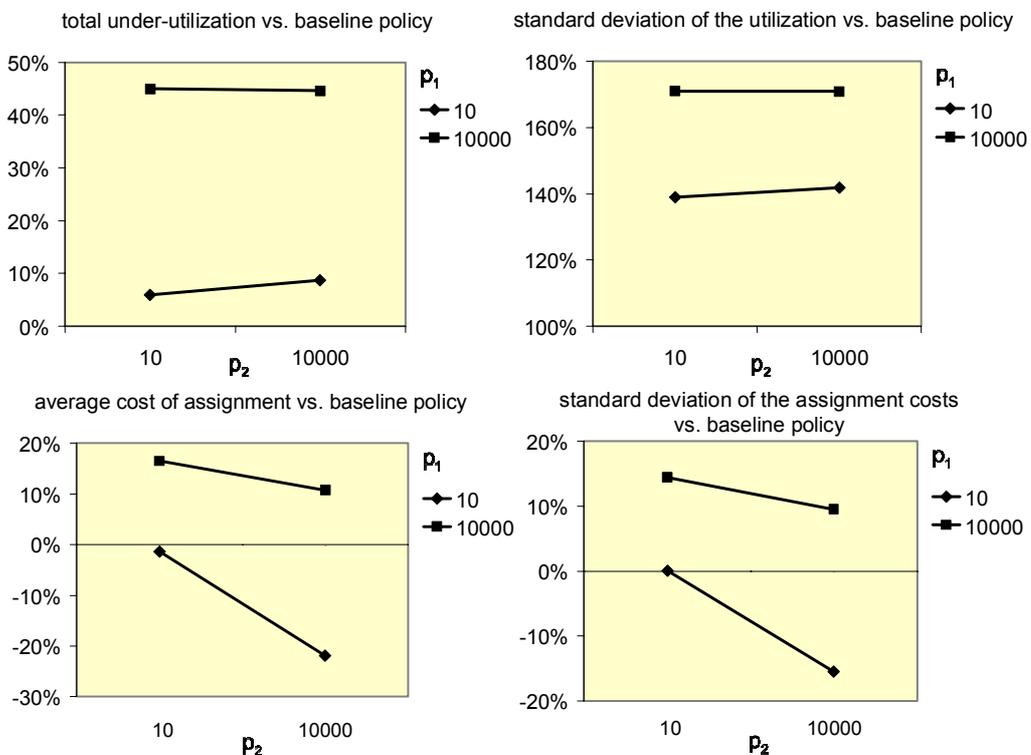


Figure 5: Profile plots for capacity/demand ratio 1.2 (all figures are standardized with respect to the baseline policy)

Overall, the average absolute levels of the customer-oriented objectives for the low capacity/demand ratio by far exceed those for the high capacity/demand ratio (18 times for the average costs and 6 times for the standard deviation) while the opposite is true for the resource-oriented ones (10 times for the under-utilization and 2 times for the standard deviation). This indicates an increased relevance of customer-oriented objectives if capacity is tight and of resource-oriented for excess capacity settings vice versa. Considering the factor combinations with p_1 and p_2 likewise being on their low and high levels respectively, discrepancies in the performance measures are caused by the relative importance as compared to the third term of the MPS objective function, i.e. the cost of the assignment.

5 Conclusions

Changing to BTO strategies, companies face an increased exposure to market dynamics. Since all business is linked to customer orders, it is the order-driven planning activities that determine the success of operations. Therefore, a clear understanding of the associated planning tasks OP and MPS as well as their dynamic interaction is essential. In this paper we provided an analysis of the decision situation of order based planning in BTO scenarios by means of a hierarchical framework, developed distinct models for OP and MPS and evaluated these using simulation.

Following the analysis, the presented approach seems promising to improve order based planning in BTO scenarios. The performance in both, customer and resource-oriented objectives is significantly controlled by the interplay of MPS and OP as well as the structure of the MPS objective function. As compared to a policy restricted to OP, improvements of 7% regarding the average cost of assignment and 45% regarding the total under-utilization can be achieved. Also, by decomposing OP and MPS, customers benefit from an instantaneous response to their requests and can thus evaluate different configurations with respect to their delivery date before placing their order. The production system on the contrary benefits from a stable coordination which is facilitated by MPS.

The scope of this study was on rather general insights into the dynamic behavior of order based planning. Consequently, the models are thought to be just as specific as to reflect the general characteristics of BTO scenarios. More work is needed to improve the empirical basis of the approach and to deepen the understanding of the dynamic performance as to provide guidance for the configuration of the presented models for specific settings.

Appendix

Table 2: Performance measures (Asteroids indicate the factual or final assignment (i.e. the promised period and the final date of production); T^* denominates the simulation run length)

	total under-utilization	$\sum_{r \in \Omega} \sum_{\tau=1}^{T^*} \left(\max \left(\text{cap}_{r\tau}^{\min} + \sum_{i \in \Psi} \bar{x}_{i\tau}^* \cdot a_{ir} - \text{cap}_{r\tau}^{\max}; 0 \right) / \text{cap}_{r\tau}^{\min} \right)$
resource-oriented	standard deviation of the utilization	$\sum_{r \in \Omega} \sqrt{\sum_{\tau=1}^{T^*} \left(\text{ctp}_{r\tau}^+ - \left(\text{ctp}_{r\tau}^+ \right)^* \right)^2 / T^* - 1} / \Omega $, with $\left(\text{ctp}_{r\tau}^+ \right)^* = \left(\sum_{i \in \Psi} \bar{x}_{i\tau}^* \cdot a_{ir} \right) / \text{cap}_{r\tau}^{\max} \quad \forall r \in \Omega; \tau = 1, \dots, T^*$ $\left(\text{ctp}^+ \right)^* = \left(\sum_{\tau=1}^{T^*} \sum_{r \in \Omega} \left(\text{ctp}_{r\tau}^+ \right)^* \right) / \Omega \cdot T^*$
	average cost of assignment	$\left(C^{OP} \right)^* = \sum_{i \in \Psi} \sum_{\tau=1}^{T^*} c_{i\tau} \cdot x_{i\tau}^* / \Psi $
customer-oriented	standard deviation of the assignment costs	$\sqrt{\left(\left(\sum_{i \in \Psi} \sum_{\tau=1}^{T^*} c_{i\tau} \cdot x_{i\tau}^* \right) - \left(C^{OP} \right)^* \right)^2 / \Psi - 1}$

References

- [Bola03] *Bolat, Ahmet*: A mathematical model for selecting mixed models with due dates. In: International Journal of Production Research 41 (2003) 5, p. 897-918.
- [BoFS06] *Boysen, Nils; Fliedner, Malte; Scholl, Armin*: A classification of assembly line balancing problems. Wirtschaftswissenschaftliche Fakultät, Jena 2006.
- [Boys05] *Boysen, Nils*: Produktionsprogrammplanung bei Variantenfließfertigung. In: Zeitschrift für Planung und Unternehmenssteuerung 16 (2005) 1, p. 53-72.
- [ChZB02] *Chen, Chien-Yu; Zhao, Zhenying; Ball, Michael O.*: A model for batch advanced available-to-promise. In: POM 11 (2002) 4, p. 424-440.
- [DrKi01] *Drexel, Andreas; Kimms, Alf*: Sequencing JIT mixed-model assembly lines under station-load and part-usage constraints. In: Mgt. Sci. 47 (2001) 3, p. 480-491.

- [EnZi98] *Engel, Christoph; Zimmermann, Jürgen*: Objectives for order-sequencing in automobile production. In: Drexl, A.; Kimms, A. (eds.): Beyond manufacturing resource planning, Springer, Berlin 1998, p. 307-331.
- [FlMe03] *Fleischmann, Bernhard; Meyr, Herbert*: Planning hierarchy, modeling and advanced planning systems. In: de Kok, A.; Graves, S. (eds.): Supply chain management, Elsevier, Amsterdam 2003, p. 457-523.
- [FlMe04] *Fleischmann, Bernhard; Meyr, Herbert*: Customer orientation in advanced planning systems. In: Dyckhoff, H.; Lackes, R.; Reese, J. (eds.): Supply chain management and reverse logistics, Springer, Berlin 2004, p. 297-321.
- [GuNg05] *Gunasekaran, A.; Ngai, Eric W.T.*: Build-to-order supply chain management. In: Journal of Operations Management 23 (2005) 5, p. 423-451.
- [JSJK02] *Jeong, Bongju; Sim, Seung-Bae; Jeong, Ho-Sang; Kim, Si-Won*: An available-to-promise system for TFT LCD manufacturing in supply chain. In: Computers Industrial Engineering 43 (2002) 1, p. 191-212.
- [Kate94] *ten Kate, Hans A.*: Towards a better understanding of order acceptance. In: International Journal of Production Economics 37 (1994) 1, p. 139-152.
- [KiSc05] *Kilger, Christoph; Schneeweiss, Lorenz*: Demand fulfilment and ATP. In: Stadler, H.; Kilger, C. (eds.): Supply chain management and advanced planning, Springer, Berlin 2005, p. 179-195.
- [MGGP04] *Moses, S.; Grant, H.; Gruenwald, L.; Pulat, S.*: Real-time due-date promising by build-to-order environments. In: IRPR 42 (2004) 20, p. 4353-4376.
- [Orac05] Oracle: Fact Sheet Global Order Promising 10i.10. Redwood Shores, CA, 2005
- [Piber05] *Pibernik, Richard*: Advanced available-to-promise. In: International journal of production economics 93-94 (2005), p. 239-252.
- [Pine99] *Pine, B. Joseph*: Mass customization: the new frontier in business competition. Harvard Business School Press, Boston, Mass. 1999.

- [RoWa05] *Rohde, Jens; Wagner, Michael*: Master planning. In: *Stadtler, H.; Kilger, C.* (eds.): *Supply chain management and advanced planning*, Springer, Berlin 2005, p. 159-177.
- [SAP05] *SAP*: “Managing the e-supply chain in der Automobilindustrie mit mySAP Automotive”, Walldorf, 2005.
- [Schn03] *Schneeweiß, Christoph A.*: *Distributed decision making*. Springer, Berlin 2003.
- [Scho99] *Scholl, Armin*: *Balancing and sequencing of assembly lines*. Physica, Heidelberg 1999.
- [ShLa05] *Sharma, Arun; Laplaca, Peter*: Marketing in the emerging era of build-to-order manufacturing. In: *Industrial Marketing Management* 34 (2005) 5, p. 476-486.
- [SpVR06] *Spengler, Thomas; Volling, Thomas; Rehkopf, Stefan*: Taktisch-operative Planung bei variantenreicher Serienfertigung. In: *Jacquemin, M.; Pibernik, R.; Sucky, E.* (eds.): *Quantitative Methoden der Logistik und des Supply Chain Management*, Dr. Kovac, Hamburg 2006, p. 183-210.
- [Tsai95] *Tsai, Li-Hui*: Mixed-model sequencing to minimize utility work and the risk of conveyor stoppage. In: *Management Science* 41 (1995) 3, p. 485-495.
- [VBWJ05] *Vollmann, Thomas E.; Berry, William L.; Whybark, D. C.; Jacobs, F. Robert*: *Manufacturing planning and control systems for supply chain management*. McGraw-Hill, New York, NY 2005.
- [WeKi64] *Wester, L.; Kilbridge, M.*: The assembly line mixed model sequencing problem. In: *Proceedings of the 3rd International Conference on OR* (1964) p. 247-260.
- [WGHH04] *Wullink, G.; Gademann, A.J.R.M.; Hans, E.W.; van Harten, A.*: Scenario-based approach for flexible resource loading under uncertainty. In: *International Journal of Production Research* 42 (2004) 24, p. 5079-5098.

Einführung in den Track

Techniken des Operations Research in der Wirtschaftsinformatik

Dr. Peter Korevaar

IBM Deutschland GmbH

Prof. Dr. Leena Suhl

Universität Paderborn

Prof. Dr. Dirk Mattfeld

TU Braunschweig

Quantitative Methoden und Techniken des Operations Research bilden den Kern vieler betrieblicher Anwendungssysteme, insbesondere wenn es um den effizienten Einsatz von beschränkten Ressourcen oder Minimierung von Kosten bzw. Maximierung von Gewinn geht. Optimierungs- und Simulationstechniken kommen häufig bei der Produktionsplanung, Supply Chain Management, Transport- und Tourenplanung, Verschnitt und Verpackung, Energie- und Wassernetzen, Telekommunikation, Projektplanung, Medienwirtschaft etc. zum Einsatz. Im Rahmen stochastischer Modelle können Unsicherheiten der realen Welt in betrieblichen Anwendungssystemen modelliert werden. Clusteringverfahren strukturieren große Datenmengen, um klare Grundlagen zur Entscheidungsunterstützung zu erzeugen. Heute können OR-Modelle in der Anwendung auch dynamisch konfiguriert werden, um flexibel auf Änderungen reagieren zu können.

Programmkomitee:

Prof. Dr. Michael Breitner, Universität Hannover

Prof. Dr. Andreas Fink, HSU Hamburg

Dr. Armin Fügenschuh, TU Darmstadt

Jun.-Prof. Dr. Natalia Kliewer, Universität Paderborn

Prof. Dr. Taieb Mellouli, Universität Halle-Wittenberg

Dr. Torsten Reiners, Universität Hamburg

Dr. Franz Rothlauf, Universität Mannheim

A Fast Look-ahead Heuristic for the Multi-depot Vehicle Routing Problem

J. Branke¹, C. Schmidt², M. Withopf^{1,2}

¹ Institute AIFB, University of Karlsruhe, Germany
branke@aifb.uni-karlsruhe.de

² Locom Software GmbH, Karlsruhe, Germany
christian.schmidt@locom.de

Abstract

The multi-depot vehicle routing problem (MDVRP) is a very challenging part of supply chain optimization. We propose here a simple yet powerful heuristic for the MDVRP with an integrated look-ahead. Compared to other state-of-the-art approaches, our heuristic is significantly faster, but yields competitive results and even found several new best solutions on a set of standard benchmark problems.

1 Introduction

The share of logistics and transportation in the overall cost of a product is increasing. Subsequently, there is a growing pressure in the logistics industry for optimization and cost reduction. In this paper, we look at the multi-depot vehicle routing problem as a typical optimization problem in modern logistics: Given some customers and their demands, three questions have to be answered:

1. From which depot a customer is served
2. On which route/truck a customer is served
3. The sequence in which the customers are served on the route.

All these aspects are interdependent. Nevertheless, many approaches split the problem into two stages to reduce the complexity: first, customers are assigned to depots, and then a classical vehicle routing problem (VRP) is solved for each depot.

The heuristic we propose in this paper also works in stages. It uses a simple construction heuristic for assigning customers to depots, followed by a savings algorithm to solve the resulting VRP for each depot. The savings algorithm uses a look-ahead procedure to approximate the effect a decision has on the final solution quality. Finally, our method applies some simple local improvement heuristics to fine-tune the solution. The overall heuristic is not only competitive to other state-of-the-art heuristics and finds several new best solutions, but it is also very fast.

The paper is structured as follows. First, we provide a more detailed definition of the MDVRP in Section 2. A short overview of related work is provided in Section 3. Section 4 describes the proposed heuristic in detail. Empirical results are summarized in Section 5. The paper concludes with a summary and some ideas for future work.

2 The multi-depot vehicle routing problem

The standard VRP is described as follows: n customers must be serviced from a unique depot. Each customer i has a given demand d_i for goods. Vehicles used for delivery have a maximal capacity C and can drive a maximal distance D before they have to be back at the depot. The goal is to minimize the total transportation cost which is often equated with the total distance driven. A solution to the VRP is a collection of tours starting and ending at the depot, where each customer is assigned to exactly one tour, each tour has length at most D , and the total demand for each tour is at most C .

The multi-depot vehicle routing problem (MDVRP) considered here differs from the standard VRP by having several depots. Tours have to be assigned to depots, i.e. they can start at any depot, but have to return to the same depot they started from. This is a common scenario in logistics. The VRP as well as the MDVRP are NP-complete problems [TV01].

Note that the benchmarks most commonly used in the literature [HEC] have an additional constraint on maximal the number of vehicles per depot.

3 Related work

Since the MDVRP is NP-complete, it is unlikely that exact polynomial time algorithms exist. For this reason, we concentrate on heuristic approaches in this section. The heuristics found in the literature can be roughly categorized into three groups: simple heuristics that attempt to solve the whole problem simultaneously, and heuristics that first assign customers to depots, and then solve a classical VRP problem for each depot in a second stage. As a third group, metaheuristics are discussed.

3.1 Single stage methods

For single-depot VRPs, the savings method is one of the best fast heuristics available [CW64]. Several authors have adapted this method to multi-depot VRPs. In this case, savings additionally depend on the depot the tour is assigned to. A particular difficulty is that the savings values may change when tours are combined. Different authors have addressed this problem either by only allowing combinations that do not change the savings values [Til69], or by re-calculating the savings values when necessary [Weu83, Mat78]. As expected, the latter yields better results but requires more computational effort.

In [TH71], the multi-depot savings method is extended by a look-ahead procedure: For every decision, all possible alternatives are considered on the first level, the 3 best (according to savings values) for the second level, and the single best alternative for the third level. Then the decision on the first level is fixed which allows to realize the highest sum of savings values on the considered three levels. The authors conclude that the improvements obtained by integrating the look-ahead “are marginal and these occur at the expense of greatly increased computer times”.

3.2 Two-stage and multi-stage methods

The simplest two-stage procedure assigns each customer to the closest depot, and then solves a VRP for each depot [Mat78]. Gillet and Johnson [GJ76] supplement this idea with local improvement steps moving a customer from one depot to another. Ashour and Bhatt [AB75] construct a minimum spanning tree over all depots and customers and use

this to assign customers to depots.

Salhi and Sari [SS97] proposed a rather complex multi-stage heuristic. We discuss this heuristic in more detail, as it is the basis for the assignment phase of our approach. First, customers are divided into core customers and borderline customers by looking at the ratio between the distance to their nearest depot and the distance to the second-nearest depot. If this ratio is smaller than some parameter ϵ (i.e. the nearest depot is much closer than the second-nearest), the customer is terminally assigned to the nearest depot as core customer. Otherwise, the two closest depots have a similar distance, and the assignment is left open. Note that by varying ϵ between 0 and 1, the percentage of customers assigned to depots as core customers may be varied from 0% to 100%. In a second step, VRPs are solved for each depot with the core customers only. In Step 3, borderline customers are inserted one by one in the order of decreasing opportunity cost, where opportunity cost is the absolute difference between inserting the customer into a tour starting at the closest depot, and inserting the customer into a tour starting at the second closest depot (thereby, the customer can be inserted into an existing tour or a new tour can be created to serve the customer directly from the depot). The customer is inserted where the additional cost is minimal, and the tours are updated. Note that only insertions are allowed that keep the tour valid.

After all customers have been inserted, a multitude of local improvement heuristics are applied, including the removal of tours and reinsertion of its customers elsewhere, the move of a single customers from one tour to another, the swapping of customers between tours etc.

3.3 Metaheuristics

The probably best methods for the MDVRP to date are Tabu Search (TS) metaheuristics (for a general introduction to TS see e.g. [GL98]). The different variants mainly differ in the neighborhood used: In [RLB96], a λ -interchange neighborhood is used. [CGL97] move customers from one tour to another. This may result in infeasible solutions which are then penalized. Finally, the variant proposed in [Sch04] allows to move or swap subsets of customers from one tour to another. The last two variants are responsible for all of

the best known solutions of the 23 standard benchmark problems from [HEC] ([CGL97] found 17 best known solutions, [Sch04] found 6). Evolutionary algorithms have been used in [TS01] and [OH04].

4 Look-ahead Assignment and Routing Heuristic

In this section, we present a new fast two-stage heuristic with look-ahead. In the following, we denote this heuristic as Look-ahead Assignment and Routing (LAR).

The first stage is closely inspired by the complex multi-stage heuristic from Salhi and Sari [SS97]. We start by dividing the customers into core customers and borderline customers depending on the parameter ϵ as described in Section 3 and generate initial tours with only the core customers. The insertion of the borderline customers differs in two major ways: first, we not only consider the two closest depots as potential sources for a customer, but all depots j with $w_c/w_j > \epsilon$, where w_c is the distance to the closest depot, and w_j is the distance to depot j . Second, and more importantly, we allow tours to become infeasible with respect to tour length and/or capacity restrictions when greedily inserting customers.

After all customers have been assigned to depots, we discard the constructed tours and solve a VRP for each depot from scratch, using a savings algorithm with look-ahead. The use of look-ahead is inspired by the use of look-ahead in the multi-depot savings method proposed in [TH71].

In order to make a decision in the current iteration, the look-ahead procedure constructs a decision tree, looking several stages into the future for several alternative choices. Our look-ahead procedure has four configuration parameters:

1. a_1 , the number of alternatives considered on the first level,
2. a_2 , the number of alternatives considered on the second level,
3. r , the decrease in the number of alternatives considered from one level to the next after stage two. Together with a_2 , this determines the number of alternatives considered in level i , as $a_i = \max\{1, a_2 - (i - 2)r\}$.

4. m , the maximum number of levels considered.

So, while the standard savings algorithm in each iteration connects the two tours with the largest savings value, the look-ahead savings algorithm considers the a_1 alternatives with the largest savings value. For each of these, it tries the a_2 alternatives with then largest savings value, and so on for m levels. Finally, out of the a_1 alternatives, it selects the one that allowed to achieve the largest total savings. An example is given in Figure 1.

After the VRPs are solved, we use three simple local improvement heuristics to generate the final solution. These improvement heuristics are 2-opt and 3-opt applied to individual tours, and an inter-tour swap operator that attempts to swap some customers between two tours. Thereby, all possible sub-tours of length at most 3 are examined and it is tested, whether an insertion into any other tour at any possible location would be feasible and better. These local improvement steps are iterated three times.

As mentioned above, the most commonly used MDVRP benchmark problems [HEC] additionally restrict the number of vehicles per depot. LAR can not handle such a constraint explicitly, because the assignment may assign more customers to a depot than can be handled by the vehicles, and because the savings heuristic does not allow to restrict the resulting number of vehicles. However, a simple twist allowed us to generate feasible solutions for all cases: We simply limit the capacity of a depot in the assignment step to a fraction p of its maximal capacity (i.e. the maximal number of vehicles times vehicle capacity). Once the assigned customers reach the depot's capacity, it is no longer considered as a potential source for the remaining customers. The fraction p is an additional parameter, but a default setting of $p = 0.975$ seemed to work well in most cases.

5 Empirical results

5.1 Experimental setup

In order to evaluate the quality of our LAR procedure, we tested it on 33 standard benchmarks from the literature, all publicly available at [HEC].

The first 23 test instances are classical benchmarks from the literature. Instances 1-7 are based on some single-depot VRPs described in [CE69], instances 8-11 are proposed in

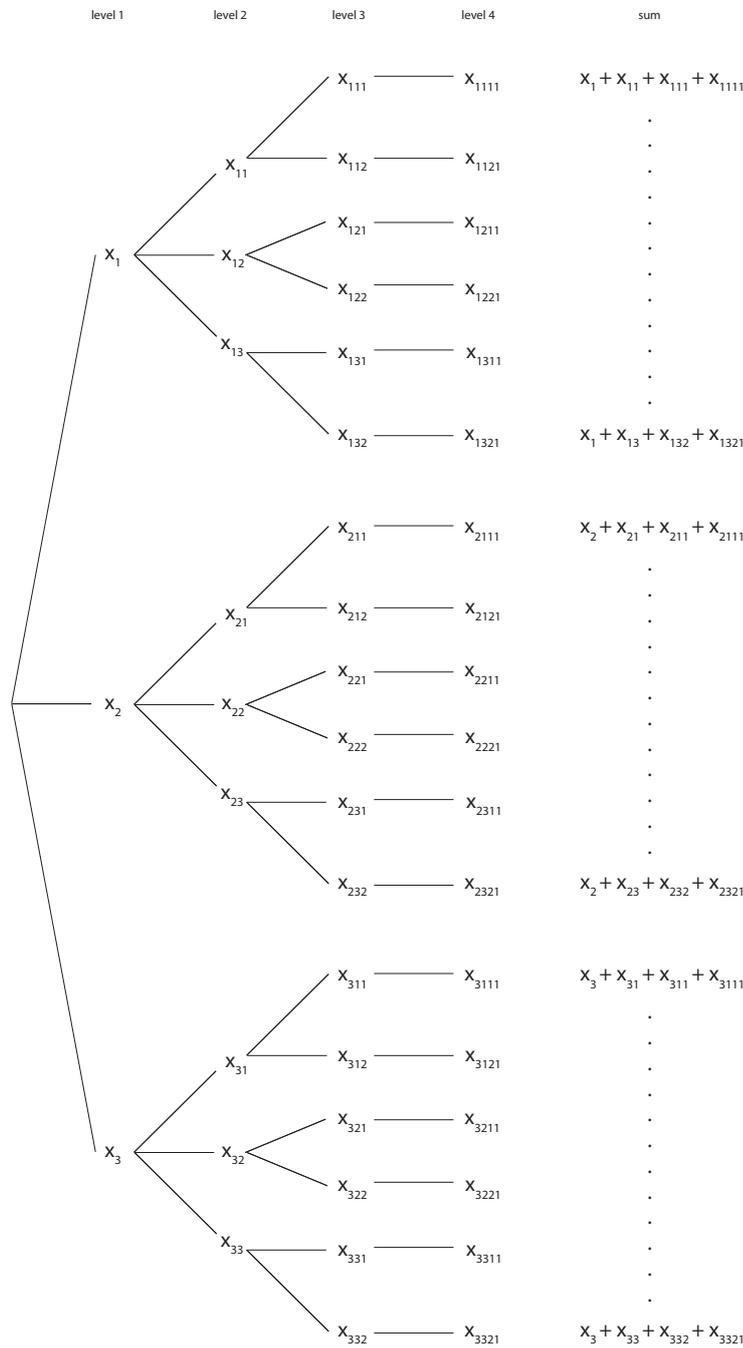


Figure 1: Decision tree for parameters $a_1 = 3, a_2 = 3, r = 1, m = 3$. The variable x denotes the savings realized by certain choices, e.g., x_{ij} denotes the savings realized if the i th highest savings value is selected in the first step, and the j th highest savings value is selected in the second step.

[GJ76], and instances 12-23 are from [CGW93]. The last ten instances have been proposed more recently in [CGL97]. Some characteristics of the test problems are summarized in Table 1. All instances are based on Euclidean distances.

Our parameter settings for the algorithm were as follows: For parameter ϵ used during assignment, all values in $\{0.4, 0.5, 0.6, \dots 1.0\}$ were used (i.e. we ran the algorithm 7 times and report on the best found value here). For look-ahead, we set $a_1 = 20, a_2 = 1, r = 0, m = 50$. To keep the number of vehicles within the allowed limits, we set $p = 0.975$.

5.2 Results on standard benchmarks

The results for the first 23 instances are summarized in Table 2, reporting on the running time (T) and the deviation from the best known solution (Δ) for different algorithms from the literature. The last column (“Best”) reports on the quality of the best known solution, and where this solution has been published. Note that [CGL97] report on two types of results: The runs with default parameter setting listed in columns 4/5 of the table, and results obtained by “playing around” with algorithm parameters, e.g. increasing the number of iterations, often resulting in the best known solution listed in the last column. Of all the approaches reported, only those from [CGL97, Sch04, TS01] take into account the constraint on the number of vehicles per depot. In [RLB96, SS97, OH04], this constraint is simply ignored.

As can be seen, the solution found by LAR is generally very close to the best known solution. For problem instance 8, we even found a new best solution, which is quite remarkable given that these problems have been thoroughly tested by many researchers. Although a direct comparison of the runtimes of the different heuristics is difficult due to differences in hardware (see footnote in Table 2 for underlying hardware specifications), it is obvious that LAR is very fast (5 seconds runtime on average).

Table 1: Characteristics of test instances.

Problem	# Customers	# Depots	Vehicle restrictions			Utilization ^c
			Capacity	Distance ^a	#Vehicles ^b	
1	50	4	80	∞	4	60,70%
2	50	4	160	∞	2	60,70%
3	75	2	140	∞	3	64,95%
4	100	2	100	∞	8	91,13%
5	100	2	200	∞	5	72,90%
6	100	3	100	∞	6	81,00%
7	100	4	100	∞	4	91,13%
8	249	2	500	310	14	86,47%
9	249	3	500	310	12	67,26%
10	249	4	500	310	8	75,66%
11	249	5	500	310	6	80,71%
12	80	2	60	∞	5	72,00%
13	80	2	60	200	5	72,00%
14	80	2	60	180	5	72,00%
15	160	4	60	∞	5	72,00%
16	160	4	60	200	5	72,00%
17	160	4	60	180	5	72,00%
18	240	6	60	∞	5	72,00%
19	240	6	60	200	5	72,00%
20	240	6	60	180	5	72,00%
21	360	9	60	∞	5	72,00%
22	360	9	60	200	5	72,00%
23	360	9	60	180	5	72,00%
24	48	4	200	500	1	82,13%
25	96	4	195	480	2	78,21%
26	144	4	190	460	3	78,42%
27	192	4	185	440	4	83,68%
28	240	4	180	420	5	93,08%
29	288	4	175	400	6	87,40%
30	72	6	200	500	1	79,00%
31	144	6	190	475	2	87,98%
32	216	6	180	450	3	84,44%
33	288	6	170	425	4	94,36%

^a ∞ : no tour length restriction

^b max. number of vehicles per depot

^c $Utilisation = \frac{total\ demand\ of\ all\ customers}{\#Depots \cdot \#Vehicles \cdot Capacity}$

Table 2: Performance comparison of several methods

Problem	[RLB96]		[CGL97]		[SS97]		[TS01]		[OH04]		LAR		Best ^a c(s*)
	Δ	T ^b											
1	0.00%	03:12	0.00%	03:14	1.89%	00:06	2.58%	05:36	4.77%	00:01	0.87%	00:01	576.87
2	0.00%	04:48	0.07%	03:28	2.34%	00:06	1.79%	03:06	1.37%	00:01	1.37%	00:01	473.533
3	0.00%	05:48	0.62%	05:40	0.74%	00:06	8.31%	04:24	10.25%	00:01	1.61%	00:01	641.186
4	0.24%	11:24	0.52%	07:47	4.64%	01:18	6.08%	08:24	2.33%	00:03	4.28%	00:03	1001.47
5	0.03%	12:48	0.44%	08:13	3.62%	00:18	0.64%	10:18	4.68%	00:03	2.33%	00:03	750.029
6	0.00%	08:24	0.15%	07:39	1.38%	00:18	11.35%	09:42	3.69%	00:01	1.38%	00:01	876.5
7	0.89%	06:48	0.82%	07:43	3.87%	01:54	10.38%	07:24	3.77%	00:01	4.25%	00:01	884.664
8	1.17%	09:24	1.11%	25:26	1.80%	14:06	8.55%	41:48	5.79%	-0.14%	00:21	00:21	4433.31
9	1.56%	41:12	1.12%	26:44	3.30%	07:06	10.50%	31:24	9.35%	1.64%	00:15	00:15	3877.37
10	0.39%	43:00	1.63%	25:30	4.64%	04:30	17.41%	45:30	9.02%	2.78%	00:12	00:12	3655.18
11	2.67%	36:24	0.75%	25:55	4.51%	04:06	15.15%	37:24	9.19%	3.09%	00:07	00:07	3554.18
12	0.00%	05:24	0.00%	05:34	0.60%	00:18	7.81%	04:48	0.00%	0.00%	00:02	00:02	1318.95
13	0.00%	04:48	0.00%	05:35	0.00%	00:18	0.00%	05:18	0.00%	0.00%	00:01	00:01	1318.95
14	0.41%	02:36	0.00%	05:26	0.00%	00:18	0.00%	05:48	0.41%	0.41%	00:02	00:02	1360.12
15	1.84%	15:30	1.15%	14:04	3.24%	01:06	22.10%	07:54	2.95%	2.52%	00:03	00:03	2505.42
16	0.00%	11:06	0.00%	14:03	0.48%	01:06	5.74%	09:54	0.61%	0.19%	00:03	00:03	2572.23
17	0.82%	05:48	0.41%	13:42	0.41%	00:42	6.85%	11:48	0.82%	0.82%	00:03	00:03	2709.09
18	2.11%	23:12	0.21%	24:51	4.06%	01:54	47.53%	20:18	5.43%	3.11%	00:05	00:05	3702.84
19	0.00%	22:00	0.00%	25:12	1.07%	01:54	3.39%	28:18	1.92%	0.21%	00:05	00:05	3827.06
20	0.96%	10:00	0.00%	24:43	0.41%	01:24	7.07%	31:48	0.96%	0.96%	00:05	00:05	4058.07
21	3.32%	48:42	1.12%	48:10	5.73%	03:36	25.52%	24:42	8.25%	3.73%	00:08	00:08	5474.83
22	0.28%	33:30	0.24%	48:54	0.71%	03:42	4.97%	27:24	3.71%	0.14%	00:08	00:08	5702.15
23	0.82%	17:18	0.73%	47:52	0.18%	03:18	3.16%	29:42	0.82%	0.82%	00:08	00:08	6095.46
Avg.	0.76%	19:16	0.48%	18:30	2.16%	02:20	9.86%	17:57	3.92%	1.58%	00:05	00:05	

^a Origin of best known solution: C [CGL97] with tuned parameters and longer runtime, S [Sch04]

^b Runtime in minutes on the following hardware: Sun 4/370 [RLB96], Sun Sparcstation 10 [CGL97], VAX 4000-500 [SS97], DEC Alpha [TS01], and Pentium4/2.8 GHz for LAR

Table 3: Performance comparison of LAR with the Tabu search algorithm proposed by Cordeau et al. [CGL97] on the benchmarks proposed there.

Problem	[CGL97]	LAR		
	c(s)	c(s)	Δ	T
24	861,32	861,319	0,00%	00:00
25	1307,61	1349,92	3,24%	00:01
26	1806,60	1805,7	-0,05%	00:02
27	2072,52	2172,05	4,80%	00:05
28	2385,77	2428,23	1,78%	00:09
29	2723,27	2710,39	-0,47%	00:13
30	1089,56	1116,19	2,44%	00:00
31	1666,60	1804,08	8,25%	00:01
32	2153,10	2217,94	3,01%	00:04
33	2921,85	3228,52	10,50%	00:08
Avg.			3,35%	00:04

The results for the 10 instances proposed in [CGL97] can be found in Table 3. On these test problems, we improved on the best known solutions for 2 out of the 10 problem instances. On average, LAR performed 3.35% worse than the Tabu search algorithm used in [CGL97], which is mainly due to the bad performance on Problem 31 and 33.

Over all 33 test instances, LAR has a deviation from the best known solution of 2.11%. While this is worse than the 0.33% obtained by the TS used in [CGL97], LAR is significantly faster (based on the running times observed in Table 2 – the running time for Tabu search on problem instances 24-33 was not reported in [CGL97]).

5.3 Design choices

In this section, we examine for some algorithm design choices their effect on solution quality. For this purpose, we compare three algorithmic variants:

1. Our proposed LAR
2. LAR where we require feasibility of tours when assigning borderline customers to depots (as in the heuristic from Salhi and Sari [SS97])

3. A method which requires feasibility of tours when assigning borderline customers to depots, and then takes the resulting tours and applies local improvement. This is similar to the heuristic proposed by Salhi and Sari [SS97], although they had some more local improvement heuristics.

All these algorithms are followed by the usual three iterations of our local improvement heuristics 2-opt, 3-opt and Swap.

As summarized above, in these tests, LAR achieved an average deviation from the best known solution of 2.11%. If we insist on feasible tours during the assignment stage (as in [SS97]), the deviation increases to 2.77%. If we additionally omit the re-planning of tours with the look-ahead savings algorithm and only apply local improvement (as in [SS97]), the deviation increases further to 5.27%. On the other hand, omitting the re-planning of tours reduces the running time by approximately 50%.

5.4 Parameter sensitivity

In this subsection, we have a closer look at the influence of the parameter settings. Figure 2 looks at the influence of the number of first-level alternatives in the look-ahead procedure on the solution quality. As can be seen, considering only slightly more than just the best alternative improves performance significantly. After about 5 considered alternatives, additional alternatives improve the results only slightly, while the running time keeps increasing linearly.

Similarly, looking some stages ahead can help a lot, while after about 50 stages, the additional effect becomes negligible, see Figure 3

Overall, the heuristic seems to be fairly robust with respect to parameter settings. Also, note that we always used the same parameter settings for all 33 problems, which differ significantly in size and structure.

We also tested what solution quality we could obtain by tuning the parameters for each problem separately. To this end, we tested all combinations of the following parameter settings: $a_1 \in \{1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100\}$, $a_2 \in \{1, 2, 3, 4, 5\}$, $m \in \{1, 5, 10, 20, 30, 40, 50\}$, $p \in \{0.8, 0.9, 0.95, 0.975, 1.0\}$. The best parameter settings for

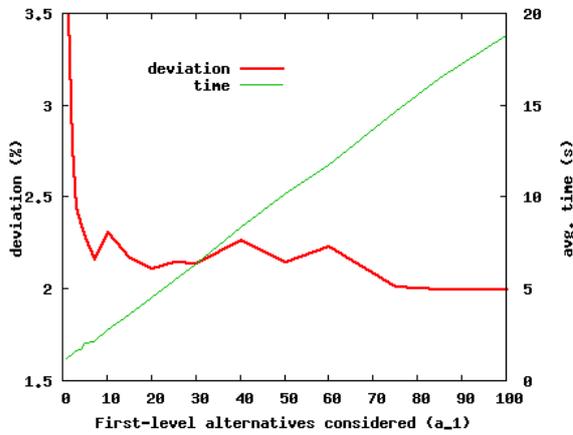


Figure 2: Average deviation over all 33 test problems depending on maximal number of alternatives considered by look-ahead procedure on first level, a_1 . $m = 50$.

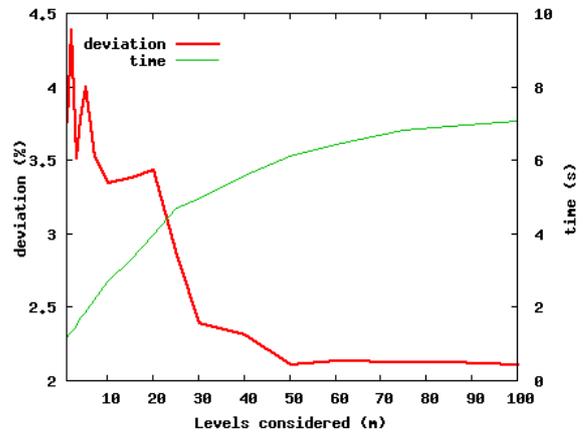


Figure 3: Average deviation over all 33 test problems depending on maximal number of levels for look-ahead, m . $a_1 = 20$.

each problem, together with the best solution found, are summarized in Table 4. Parameter fine-tuning further improved performance from an average error of 2.11% with standard parameter settings to an average error of 1.15%. Furthermore, in addition two of the three new best solutions reported above could be further improved, and for two more problems new best solutions were found (making a total of 5 new best solutions).

Table 4: Best results obtained with different parameter settings

Problem	Best		LAR							
	c(s*)		ϵ	a_1	a_2	r	m	p	c(s)	Δ
1	576,87	C	0,5	10	1	1	20	0,80	577,46	0,10%
2	473,533	C	1,0	20	1	1	50	1,00	480,04	1,37%
3	641,186	C	0,9	20	1	1	50	1,00	651,479	1,61%
4	1001,47	S	0,4	100	5	1	50	1,00	1023,35	2,18%
5	750,029	C	0,3	100	5	1	50	1,00	762,58	1,67%
6	876,5	C	0,5	70	4	1	50	1,00	885,91	1,07%
7	884,664	S	0,8	100	1	1	100	0,98	912,198	3,11%
8	4433,31	S	1,0	20	1	1	50	1,00	4427,16	-0,14%
9	3877,37	S	0,8	100	1	1	100	0,98	3922,21	1,16%
10	3655,18	S	0,8	20	1	1	50	1,00	3749,17	2,57%
11	3554,18	C	0,8	5	3	1	50	1,00	3627,00	2,05%
12	1318,95	C	1,0	2	1	1	5	1,00	1318,95	0,00%
13	1318,95	C	1,0	2	1	1	5	1,00	1318,95	0,00%
14	1360,12	C	1,0	1	1	1	1	1,00	1365,69	0,41%
15	2505,42	C	1,0	2	1	1	5	1,00	2568,50	2,52%
16	2572,23	C	1,0	20	1	1	5	1,00	2577,12	0,19%
17	2709,09	C	1,0	1	1	1	1	1,00	2731,37	0,82%
18	3702,84	S	0,1	30	1	1	50	0,90	3818,05	3,11%
19	3827,06	C	1,0	2	1	1	2	1,00	3835,28	0,21%
20	4058,07	C	1,0	1	1	1	1	1,00	4097,05	0,96%
21	5474,83	C	0,4	20	1	1	50	1,00	5681,03	3,77%
22	5702,15	C	1,0	2	1	1	5	1,00	5710,38	0,14%
23	6095,46	C	1,0	1	1	1	1	1,00	6145,58	0,82%
24	861,32	C	0,4	100	1	1	100	0,98	861,32	0,00%
25	1307,61	C	0,1	150	1	1	150	0,98	1302,77	-0,37%
26	1806,6	C	0,7	100	1	1	100	0,98	1792,62	-0,77%
27	2072,52	C	0,6	100	1	1	100	0,95	2091,01	0,89%
28	2385,77	C	0,7	150	1	1	150	0,98	2374,41	-0,48%
29	2723,27	C	0,6	100	1	1	50	0,95	2697,85	-0,93%
30	1089,56	C	0,5	100	1	1	100	0,85	1102,45	1,18%
31	1666,6	C	0,6	100	1	1	100	0,98	1731,77	3,91%
32	2153,1	C	0,8	150	1	1	150	0,98	2165,64	0,58%
33	2921,85	C	0,8	150	1	1	150	0,98	3048,79	4,34%
1,15%^b										

^a avg. deviation on problem instances 1 to 23^b avg. deviation over all 33 problem instances

6 Conclusion

In this paper, we have considered the multi-depot vehicle routing problem (MDVRP), which extends the standard vehicle routing problem (VRP) by allowing servicing a customer from one of several depots, which is a common scenario in practical transportation problems. To solve this problem, we have proposed a new heuristic which first assigns customers to depots, and then uses a savings heuristic with integrated look-ahead to solve the VRP for each depot. An extensive empirical comparison on 33 standard benchmark problems showed that our method is not only competitive with other state-of-the-art procedures, but also much faster. For 5 out of the 33 benchmark problems, new best solutions have been found.

Overall, we conclude that our proposed method is a very fast and simple yet very effective heuristic for the MDVRP. As a next step, we are working on hybridizing this heuristic with an evolutionary algorithm to obtain even better results.

References

- [AB75] S. Ashour and P. Bhatt. A graph-theoretical approach for delivery problems. Technical report, unpublished, 1975.
- [CE69] N. Christofides and S. Eilon. An algorithm for the vehicle-dispatching problem. *Operational Research Quarterly*, 20(3):309–318, 1969.
- [CGL97] J.F. Cordeau, M. Gendreau, and G. Laporte. A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks*, 30:105–119, 1997.
- [CGW93] I-M. Chao, B. Golden, and E. Wasil. A new heuristic for the multi-depot vehicle routing problem that improves upon best-known solutions. *American Journal of Mathematical and Management Sciences*, 13(3):371–406, 1993.
- [CW64] G. Clarke and J.W. Wright. Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12:568–581, 1964.

- [GJ76] B.E. Gillet and J.G. Johnson. Multi-terminal vehicle-dispatch algorithm. *Omega*, 4:711–718, 1976.
- [GL98] F. Glover and M. Laguna. *Tabu Search*. Kluwer, 1998.
- [HEC] <http://www.hec.ca/chairedistributique/data/>.
- [Mat78] F. Matthus. *Tourenplanung - Verfahren zur Einsatzdisposition von Fuhrparks*. S.Toeche-Mittler Verlag, Darmstadt, 1978.
- [OH04] B. Ombuki and F. Hanshar. An effective genetic algorithm for the multi-depot vehicle routing problem. Technical report, Brock University Ontario, 2004.
- [RLB96] J. Renaud, G. Laporte, and F. F. Boctor. A tabu search heuristic for the multi-depot vehicle routing problem. *Computers & Operations Research*, 23:229–235, 1996.
- [Sch04] S. Scheuerer. *Neue Tabusuche-Heuristiken für die logistische Tourenplanung bei restringierendem Anhängereinsatz, mehreren Depots und Planungsperioden*. PhD thesis, Universität Regensburg, 2004.
- [SS97] S. Salhi and M. Sari. A multi-level composite heuristic for the multi-depot vehicle fleet mix problem. *European Journal of Operational Research*, 103:95–112, 1997.
- [TH71] F. A. Tillman and R. W. Hering. A study of a look-ahead procedure for solving the multiterminal delivery problem. *Transportation Research*, 5:225–229, 1971.
- [Til69] F.A. Tillman. The multiple terminal delivery problem with probabilistic demands. *Transportation Science*, 3:192–204, 1969.
- [TS01] S.R. Thanghia and S. Salhi. Genetic clustering: An adaptive heuristic for the multidepot vehicle routing problem. *Applied Artificial Intelligence*, 15:361–383, 2001.

- [TV01] P. Toth and D. Vigo. *The vehicle routing problem*. Society for Industrial and Applied Mathematics, 2001.
- [Weu83] H.-K. Weuthen. *Tourenplanung - Lösungsverfahren für Mehrdepot - Probleme*. PhD thesis, Universität Karlsruhe TH, 1983.

Progress in solving large scale multi-depot multi-vehicle-type bus scheduling problems with integer programming

Uwe H. Suhl, Swantje Friedrich and Veronika Waue

Institut für Produktion, Wirtschaftsinformatik und Operations Research

Freie Universität Berlin
D-14195 Berlin

suhl@wiwiss.fu-berlin.de

Abstract

This paper discusses *solution methods* of multi-depot, multi-vehicle-type bus scheduling problems (MDVSP), involving multiple depots for vehicles and different vehicle types for timetabled trips. All models are from real-life applications. Key elements of the application are a mixed-integer model based on time-space-based networks for MDVSP modeling and a customized version of the mathematical optimization system MOPS to solve the very large mixed integer optimization models. The modeling approach was already described in several other publications. This paper focuses on the solution methods critical to solve the very large integer optimization models. We discuss aspects to solve the initial LP and the selection of the starting heuristic. Real life applications with over one million integer variables and about 160000 constraints were solved by the optimizer MOPS. A key role plays also the architecture of new Windows workstations with Intel 64 bit processors. We present numerical results on some large scale models and a brief comparison to another state-of-the-art mathematical programming system. The presented research results have been developed in cooperation with the Decision Support & OR Lab of the university Paderborn which was responsible for the development of the MDVSP-model and the application software.

1 Introduction

Public transportation companies are under high cost pressures, because market prices for suburban transportation are not cost covering and traditional subsidies from municipalities are decreasing. The companies must therefore focus on efficient use of resources, especially vehicles and drivers.

We consider the scheduling of vehicles under constraints and objectives arising in urban and suburban public transport. Each timetabled trip can be served by a vehicle belonging to a given set of vehicle types. Each vehicle has to start and end its work day in one of the given depots. After serving one timetabled (loaded) trip, each bus can serve one of the trips starting later from the station where the vehicle is standing, or it can change its location by moving unloaded to another station (deadhead trip) in order to serve the next loaded trip starting there.

The cost components include fixed costs for required vehicles as well as variable operational costs. The variable costs consist of distance-dependent travel costs and time-dependent costs for time spent outside the depot – the case where a driver is obliged to stay with the bus. All cost components depend on vehicle type. Since the fixed vehicle cost components are usually orders of magnitude higher than the operational costs, the optimal solution always involves the minimal number of vehicles.

The combinatorial complexity of the multi-depot bus scheduling problem (MDVSP) is determined by numerous possibilities to assign vehicle type to each trip, to build sequences of trips for particular buses, and to assign buses to certain depots. To represent these sequences of trips, exact modeling approaches known in the literature consider explicitly all possible connections - pairs of trips that can be served successively.

It is well known that the general MDVSP-model is NP-hard [BeCG87]. The practical complexity of instances of the MDVSP depends on factors such as:

- the number of timetabled trips,
- the number of depots, or more precisely, the average number of depot-vehicle type combinations per timetabled trip,

- the number of possible unloaded trips, which can vary depending on the completeness of the distance matrix for stop points.

Since real life MDVSP-models result in a very large number of integer variables and constraints there have been numerous modeling and optimization approaches to solve such models. There are three different approaches among the existing modeling techniques:

- Path-oriented - leading to set partitioning formulation [RiSo94],
- Arc-oriented - leading to multi-commodity flow formulation [FoHW94],
- Combinations of these two approaches [CDFT89].

For an overview of the various modeling and optimization techniques see [KIMS06].

In all these models the possible trip connections are considered explicitly and the number of such connections, corresponding to the number of integer variables, grows quadratic as a function of the number of loaded trips. Therefore, models with several thousand scheduled trips become too large to be solved in a reasonable amount of time by standard integer programming software. Various techniques to reduce the number of possible connections have been proposed in the literature. Some approaches discard arcs with too long waiting times; others generate arcs applying the column generation idea to the network flow representation.

The model used in this paper is based on a time-space network based modeling approach described in [MeIKI02, GiKS05, KIMS06]. For completeness we will describe in the following section the underlying mathematical model in its simplest form.

2 The multi-depot multi-vehicle type scheduling problem (MDVSP)

We define the vehicle scheduling problem (VSP), arising in public bus transportation, as the task of building an optimal set of rotations (vehicle schedule), such that each trip of a given timetable is covered by exactly one rotation. For each trip the timetable specifies a departure time and an arrival time with start and end stations respectively.

Within a bus tour consisting of several (loaded) service trips chained with each other, the use of deadhead trips (unloaded trips between two end stations) often provides an improvement in

order to serve all trips of a given timetable by a minimum number of buses. Thus a work day for a given bus is defined as a sequence of trips, deadheads, waiting times at stations (parking stops) and pull-out/pull-in trips from/to the assigned depot. Since deadhead trips mean an additional cost factor, minimization of this cost and minimization of waiting time cost are important optimization goals.

There are several variations of the bus scheduling problem involving different side constraints or numbers of depots and / or of bus types. The constraints and optimization criteria may differ from one problem setting to another. The presented model can be modified such that several practical side constraints such as outsourcing of the parts of timetables to private bus companies or return trips to different depots can be handled.

The multi-depot vehicle scheduling problem involves several depots, so that a vehicle has to return in the evening to the same depot from which it started in the morning.

A multi-vehicle-type VSP copes with a heterogeneous fleet of vehicles. For a given trip we define a group of vehicle types this trip can be served by. In a feasible solution each rotation is assigned to exactly one depot and to one vehicle type. Furthermore, it is possible to state capacity constraints for depots to consider the number of parking slots for buses. Other kinds of capacity constraints set a limit for the number of available vehicles of certain bus types and for the number of certain type vehicles in a given depot.

Time-space network (TSN) models have been proposed for routing problems in airline scheduling [HBJM95], because they are advantageous in modeling possible connections between arriving and departing flights. In a time-space network, connections within a location are realized by using a time line that connects all possible landing and takeoff events within the location. Thus, there is no need to explicitly model connections for each feasible pair of events within a location. Time-space network models were not used for bus scheduling problems until now, because, compared to airline scheduling where deadheading is generally not allowed, bus scheduling permits unrestricted deadheading. Thus the advantages given by TSN remained negligible, because of too many deadhead arcs.

However, as was shown in [MelKI02, GiKS05, KIMS06] a new modeling technique exploits the advantages of TSN models for bus scheduling problems. A crucial modeling technique is aggregation of possible trip matches, which allows a drastic reduction in model size. This

modeling technique allows the solution of large practical MDVSP models with exact solution algorithms such as integer programming.

Let $N = \{1,2,\dots,n\}$ be the set of trips, and let D be the set of depots. The depot is here defined as a combination of a depot and a vehicle type. The vehicle scheduling network $G^d = (V^d, A^d)$ corresponding to depot d is defined as an acyclic directed graph with nodes V^d and arcs A^d . Let c_{ij}^d be the vehicle cost of arc $(i,j) \in A^d$, which can be a function of travel and idle time. The vehicle cost of arcs representing idle time activity in the depot is 0. Furthermore, a fixed cost for using a vehicle is set on the circulation arc. Let $N^d(t) \in A^d$ be the arc corresponding to trip t in the vehicle scheduling network G^d . Decision variable x_{ij}^d indicates whether an arc (i,j) is used and assigned to the depot d or not. An upper bound u_{ij}^d is defined for each decision variable as follows:

$$u_{ij}^d = \begin{cases} 1, & \text{if } x_{ij}^d \text{ corresponds to a timetable trip} \\ u^d, & \text{if } x_{ij}^d \text{ corresponds to a circulation arc where } u^d \text{ is the capacity for depot } d \\ M, & \text{otherwise, where } M \text{ is the maximum number of available vehicles.} \end{cases}$$

The MDVSP model can now be formulated as

$$\min \sum_{d \in D} \sum_{(i,j) \in A^d} c_{ij}^d x_{ij}^d \quad (1)$$

$$\sum_{\{j:(i,j) \in A^d\}} x_{ij}^d - \sum_{\{j:(j,i) \in A^d\}} x_{ji}^d = 0, \quad \forall i \in V^d, \forall d \in D \quad (2)$$

$$\sum_{d \in D, (i,j) \in N^d(n)} x_{ij}^d = 1, \quad \forall n \in N \quad (3)$$

$$0 \leq x_{ij}^d \leq u_{ij}^d, \quad \forall (i,j) \in A^d, \forall d \in D \quad (4)$$

$$x_{ij}^d \text{ integer}, \quad \forall (i,j) \in A^d, \forall d \in D \quad (5)$$

The objective (1) is to minimize the sum of total vehicle costs. Constraints (2) are flow conservation constraints, indicating that the flow into each node equals the flow out of each

node, while constraints (3) assure that each trip must be covered by exactly one vehicle. This is a time-space network model based on a multi-commodity flow formulation. It should be mentioned that some integer variables in the MDVSP model can be declared as continuous variables. They obtain automatically integer values in a feasible integer solution.

3 Solving the Linear Programming relaxation of MDVSP-models

Real life MDVSP-models are usually very large. The following table represents real life models of the PTV AG and the Decision Support & OR Lab of the university Paderborn. The number of depots is actually depot-vehicle-type combinations, as mentioned above, and thus corresponds to the terminology in [Löbe99].

MDVSP model instances					Integer model sizes			
City	#trips	#stations	#v. types	#depots	rows	variables	Int. vars.	nonzeros
Halle1	2047	21	3	2	14997	53249	40387	113069
Halle2	2047	21	3	3	21939	87128	67367	184111
Halle3	2047	21	3	4	29031	118768	91957	250675
Mun1	1808	76	1	19	52303	478823	429088	981163
Mun2	3054	49	1	9	61254	573300	515530	1174095
Mun3	11062	161	12	19	163142	1479833	1330580	3031285

Table 1: MDVSP model instances and corresponding integer model sizes

All successful solution methods to solve general integer optimization methods are based on a branch-and-bound / cut approach. A key role plays the solution of the LP-relaxations: the initial LP and the LPs corresponding to subproblems (nodes) in the branch-and-bound / cut tree.

There are three competitive solution algorithms to solve general LP-models [Bixb02]:

- primal simplex method, the oldest simplex solution algorithm [Dant63]
- dual simplex method, which has become a strong contender over the years [Lemk54, Bixb02, Kobe05, KoSu07]
- Interior point algorithms, also called barrier algorithms [Karm84, Mehr92, Mész96].

It is well known, that there are problem classes where each of those algorithms works best. Furthermore each method has fundamental advantages and disadvantages:

A fundamental advantage of the simplex methods is that an optimal LP solution is a basic solution. This means that only basic variables may have values between their bounds. Nonbasic variables (except free variables) have values at their lower or upper bound. Most lower bounds of practical models are zero. Therefore, an optimal basic solution has typically a smaller number of nonzero activities than an optimal solution produced by an interior point method. A basic solution exploits furthermore a tight linear programming relaxation of an integer model. The simplex method has in addition very good warm start capabilities if an LP-model is slightly modified and a nearly optimal basic solution is available. The simplex method is therefore the method of choice for solving LPs during branch and bound / cut algorithms. State-of-the-art dual simplex codes are in general superior to primal simplex codes. However both codes are dependant on each other that is the dual requires frequently the primal to remove a cost perturbation and the primal requires frequently the dual simplex code to remove a bound perturbation [KoSu07].

Recent interior point technology is based on primal-dual methods [Mehr92, Mész96]. There are large LP-models which can be solved much faster than with the best simplex codes. The number of iterations for an interior point method is typically relatively small (20-80) and independent of the size of the problem. The main work of an iteration is the solution of a symmetric, positive definite system of linear equations. A symbolic Cholesky factorization can be computed once by using an ordering algorithm [GeLi81]. The number of nonzeros in the Cholesky factorization is a key factor for the performance of the interior point method and is strongly influenced by the ordering algorithm used to compute the pivot sequence. It is therefore important, in particular for very large LP-models, to experiment with the different ordering heuristics for a given model class. The Cholesky factorization has in general much more nonzeros than LU-factorizations in a simplex type algorithm. The use of an interior point method therefore requires a much larger amount of main memory than the simplex method. This behavior rules out the use of interior point methods for very large models on some system platforms such as a classical Windows XP-system with 2 GB address limit (see Table 2).

Another disadvantage is a very limited warm start capability which is required during branch-and-bound / cut because the similarity of LP subproblems can be exploited by the simplex method but not by the interior point method.

Interior point methods do generally not produce a basic solution. In many situations (integer programming, tight linear programming relaxations, save / restore of basis) basic solutions are necessary. There are “cross over” algorithms, also called (optimal) basis identification which can be used to produce an optimal basic solution from an optimal interior point solution [Ande99]. These algorithms are specialized simplex algorithms. The crossover method used in MOPS uses the numerical kernels of the simplex method.

The optimization system MOPS [MOPS06, Suhl94] contains three state-of-the-art engines which were tested on the MDVSP models. The interior point method in MOPS is based on the work of C. Mészáros [Mész98]. The dual simplex method was recently completely new designed and implemented [Kobe05, KoSu07] and is one of the best implementation according to our benchmarks. A key question was initially which engine is best suited to solve the initial LP. It was clear from the beginning that the primal simplex method will probably not be competitive to the other methods on the MDVSP problems. The numerical experiments with the smallest model Halle1 in Table 1 shows the expected results (see Table 2). We ran a comparison against Cplex 9.1 [ILOG06] and it shows the same behavior. As a consequence the larger models were only tested with dual simplex and barrier with crossover (x-over). The smallest models Halle1 and Halle2 were solved faster with the dual simplex. When model sizes get larger the barrier code outperforms the dual simplex. This observation is in line with the fundamental advantage in computational complexity of the barrier code compared to a simplex code. The following numerical results are based on a typical Windows XP Workstation with a 32-Bit Pentium Processor. This machine has a maximum virtual address space limit of 2 Gigabytes (GB). Under certain conditions an address limit of 3 GB is possible. This type of workstation is frequently used in practice for running such applications. As can be seen in Table 2 model Mun3 cannot be solved with the barrier code on such a machine, because the required virtual memory exceeds 3 GB. We were also not able to solve that model with the 32-Bit barrier code of Cplex 9.1 on this machine.

Name	MOPS times (secs) on PIV (3,4 GHz, WinXP) to solve the initial LP		
	Primal simplex	Dual simplex	Barrier + x-over
Halle1	281.69	21.28	33.19
Halle2	Nt	61.38	73.06
Halle3	Nt	176.05	108.81
Mun1	Nt	3051.58	1519.33
Mun2	Nt	4266.36	798.24
Mun3	Nt	15012.08	nem

nt: not tested, nem: not enough (virtual) memory, i.e. > 2 GB

Table 2: LP Solution times on a 32 Bit Windows workstation with MOPS (32)

Model Mun3 was also solved by Cplex 9.1 on the same 32 bit workstation. The barrier code of Cplex 9.1 was also not able to solve this model due to insufficient memory. The dual simplex engine of Cplex 9.1 solved the initial LP of Mun3 in 14832.17 secs. The purpose of this paper is not to make a comparison between Cplex and MOPS. The test just shows that the current state-of-the-art system Cplex required also several hours computing to solve this model.

A recent development for Windows / Intel workstations are processor and memory architectures which allow 64 bit addressing and integer arithmetic. Microsoft provides the operating system WindowsXP (x64). Intel offers C++ and FORTRAN compilers which generate 64 bit code for such machines. This development is very important from a practical point of view because Intel / Windows system platforms are used predominantly in industry. Virtually all 32-bit software systems run unchanged on such 64 bit systems allowing the parallel use of 32 bit and 64 bit software systems.

It was a straightforward task to recompile MOPS using the Intel Compilers and generating a 64 bit library. The following numerical results with MOPS are based on a workstation with Intel Xeon processor (3.4 GHz) with Intel 64 bit memory technology 64MT, 4 GB of main memory. Both CPUs are Xeon Processors with a clock speed of 3.4 GHz. The internal data caches are identical with 16 KB. The 32 Bit CPU has an on board L2 cache with a size of 1 MB ECC whereas the 64 Bit CPU has an on board L2 cache with a size of 2 MB ECC resulting in a much higher memory bandwidth of the 64 Bit CPU. Furthermore the 64 Bit CPU has more registers and additional instructions. Despite of the same compiler releases one can expect some differences in the compiled code.

Name	MOPS (64) times (secs) on Xeon (3,4 GHz, Win x64 to solve initial LP	
	Dual simplex	Barrier + x-over
Mun1	2895.09	1496.81
Mun2	4106.36	778.24
Mun3	11336.45	3490.20

Table 3: LP Solution times on a 64 Bit Windows workstation with MOPS64

One surprise was the result with the dual on Mun3. The time was nearly 4000 secs faster than the result for the 32 bit version. It is not clear which of the possible influence factors (compiler, cache size and architecture) was responsible for this result.

A key influence on the running time of the interior point code has the ordering heuristic used for the Cholesky factorization. There are several well known ordering heuristics such as minimum degree, minimum local fill-in, and nested dissection [GeLi81] and more recent orderings such as multisection [AsLi98, Mész98] which are used in MOPS. In the default ordering we perform the minimum degree and the nested dissection ordering and compare the computed number of nonzeros; then we select the better ordering i.e. with the fewer number of nonzeros. However the best results for the MDVSP-models are based on the multisection ordering which was used throughout in the benchmarks. The following table contains a comparison of two ordering heuristics with respect to the number of nonzeros and solution times of three models of Table 1 on the Xeon (3.4) and Winx64. Note, that most other LP-models are solved faster with the default ordering.

Name	default ordering		multisection ordering	
	Nonzeros in Cholesky	Barrier time (sec)	Nonzeros in Cholesky	Barrier time (sec)
Mun1	63,648,566	3567.72	45,470,777	1496.81
Mun2	26,887,134	1206.81	22,688,525	778.24
Mun3	84,710,597	5689.22	71,388,781	3490.20

Table 4: LP Solution times and Cholesky nonzeros with the MOPS barrier code and two ordering heuristics

4 Solving the MDVSP-models

The time-space network based models presented above have automatically a very tight LP-relaxation. The relative gap between the value of the LP-relaxation after IP-Preprocessing and an optimal integer value is extremely small, sometimes zero. Almost all variables have integer values in the optimal basic solution of the LP-relaxation. Due to the aggregation of possible connections, the mathematical model tends to use one general integer variable instead of several

binary variables. The optimal vehicle schedule is computed in the post-processing phase from the optimal network flow via flow decomposition [KIMS06].

The normal IP-Preprocessing [SuSz94] to tighten the LP-relaxation does not produce any significant improvements. Neither lifted cover cuts [SuWa04] nor clique or implication cuts are violated in the LP-relaxation(s). Only Gomory mixed integer cuts are able to tighten the LP-relaxation on some MDVSP-models, reducing the fractionality of the LP-solution. However the Gomory cuts can be quite dense. The number of nonzeros depends on the density of row k of the inverse. Therefore inserting the cut may produce significant fill in the following LU-factorizations of the modified basis matrices reducing the iteration speed in the branch-and-bound / cut algorithm. Therefore the decision whether a cut is actually appended to the original model is crucial, in particular for very large models. This aspect is under further investigation and is not discussed here.

An initial heuristic is used to find good integer solutions quickly. MOPS contains different heuristics prior to the branch-and-cut algorithm. We use the relaxation-based search space (RSS) heuristic for solving the MDVSP-models which produces the overall best results.

The RSS heuristic distinguishes between basic and nonbasic integer variables of the current LP solution after the initial IP-Preprocessing. Let nb the number of nonbasic variables in the LP-solution and δ a parameter between 0 and 1 (default is 0.7). The $\delta \cdot nb$ nonbasic variables with the largest magnitude of their reduced costs d_j are fixed to the corresponding lower or upper bound depending on the sign of d_j .

We define two rounding intervals $[0,rl]$ and $[ru,1]$ where $0 \leq rl < ru \leq 1$. The default values are $rl = 0.1$ and $ru = 0.9$. For a basic integer variable $j \in J_I$ with a value $\bar{x}_j = \lfloor \bar{x}_j \rfloor + f_j$, f_j specifies its fractional part, where $0 \leq f_j < 1$. Variable x_j is rounded to $\lfloor \bar{x}_j \rfloor$ if $f_j \in [0,rl]$ and to $\lceil \bar{x}_j \rceil$ if $f_j \in [ru,1]$. In other words “quasi integer” basic variables are rounded to the next integer value. The LP relaxation is solved after rounding all quasi integer variables. Several rounding iterations can be done as long as the LP solution is feasible, not integer and variables are rounded. In case of infeasibility the last rounding step is undone. The rounding intervals can be enlarged if no variable can be fixed in the first pass of rounding and reduced, if the LP-relaxation is infeasible in the first rounding pass or the LP-relaxation is integer.

Basic Algorithm of rounding process

```
1  For i=1 to number of rounding iterations do
2    For j=1 to number of all variable do
3      If variable is fixed or continuous cycle
4       $f_j = \bar{x}_j - \lfloor \bar{x}_j \rfloor$ 
5      If  $f_j \leq rl$  then
6        Fix variable to  $\lfloor \bar{x}_j \rfloor$ 
7      Else if  $f_j \geq ru$  then
8        Fix variable to  $\lceil \bar{x}_j \rceil$ 
9      End if
10     Perform bound reduction on all variables
11     If problem is infeasible then
12       Clear settings of last rounding pass
13       If first rounding pass then reduce rounding intervals
14       Exit
15     End if
16   End for
17   If problem is infeasible then
18     If first rounding pass then
19       Cycle
20     Else
21       Exit
22     End if
23   End if
24   If no variables are rounded in this pass then
25     If first rounding pass then
26       Enlarge rounding interval
27     Cycle
28     Else
29       Exit
30     End if
31   End if
32   Solve LP
33   If problem is infeasible then
34     Clear settings of last rounding pass
35     If first rounding pass then
36       Reduce rounding intervals
37     Cycle
38     End if
39     Exit
40   Else if problem is integer then
41     Clear all settings
42     Reduce rounding intervals
43   End if
44 End for
```

The branch-and-bound engine of MOPS is called after rounding. If the search in the restricted search space is ended before one of the termination criteria (see below) is satisfied, the rounding intervals are reduced and the rounding procedure is repeated. The RSS heuristic is terminated if

- a given node limit is reached (default 50 nodes)
- the relative gap between the value of an integer solution found in the heuristic and the value of the LP relaxation after IP-Preprocessing is less than a threshold (default is 5%, i.e. 0.05)
- a time limit is reached (default is model size dependant).

Basic Algorithm of RSS heuristic prior to the branch-and-bound-algorithm

- 1 Solve LP after IP-Preprocessing
- 2 Fix the δ *nb variables with the maximum magnitude of reduced costs to the corresponding lower or upper bound
- 3 **Do**
- 4 Perform Basic Algorithm of rounding process
- 5 Use branch-and-bound algorithm until node limit, time limit or gap is reached
- 6 Clear settings
- 7 **If** termination criterion is reached **Exit**
- 8 Reduce size of rounding intervals
- 9 **Enddo**

5 Numerical results on real life models

Table 5 summarizes the computational results of the test problems presented in Table 1. Since the heuristic is also a specialized branch-and-bound-algorithm where the main work is to solve an LP at given node the nodes are not distinguished between heuristic and branch-and-bound algorithm. The heuristic is executed at most 50 nodes. The branch-and-bound algorithm is used thereafter to prove optimality.

Name	initial LP time (sec)	Nodes in heuristic + b&b	Total time (sec)
Halle1	33.19	0	33.77
Halle2	73.06	0	73.19
Halle3	108.81	0	115.81
Mun1	1527.81	0	1665.34
Mun2	798.24	10	879.13
Mun3	3490.20	0	3586.45

Table 5: Solution times on a 64 Bit Windows workstation with MOPS64

With the proposed modeling approach in [KIMS06] we were able to solve quite large problems in an acceptable amount of time. This required the selection of the: proper LP-engine, ordering heuristic for the Cholesky factorization, starting heuristic, branching and node selection strategies.

One remark on “acceptable” solution times is in order. Running times of a couple of hours do not seem ideal. However, the MDVSP-models are not solved on a daily basis. It is therefore acceptable to run such models over night.

6 Conclusion

MDVSP models from real life applications as modeled by time-space network flow models [MeiK102, GiKS05, KIMS06] can now be solved efficiently by a customized version of the optimizer MOPS. Customization requires only the setting of a few parameters. The progress in solution times is based on several improvements of the computational engines, an improved heuristic and the use of 64 bit platforms (Windows XP x64). Many of these instances were not solvable with the existing approaches or the running time was too long. It should be mentioned that the improvements in algorithms and implementation are also beneficial to many other applications based on linear mixed-integer programming models.

References

- [Ande99] Andersen, E.D.: On exploiting problem structure in a basis identification procedure for linear programming, In: *INFORMS Journal on Computing* 11 (1999) 1, S. 95-103.
- [AsLi98] Ashcraft, C. and Liu, J.W.: Robust Ordering of Sparse Matrices using Multisection, In: *SIAM Journal on Matrix Analysis and Applications* 19 (1998) 3, S. 816 – 832.
- [BeCG87] Bertossi, I., Carraresi, P., Gallo, G.: On some matching problems arising in vehicle scheduling models, In: *Networks* 17 (1987), S. 271-281.
- [Bixb02] Bixby, R. E.: Solving real-world linear programs: a decade and more of progress, In: *Operations Research* 50 (2002) 1, S. 3-15.
- [CDFT89] Carpaneto, G., Dell’Amico, M., Fischetti, M., Toth, P.: A branch and bound algorithm for the multiple depot vehicle scheduling problem, In: *Networks* 19 (1989), S. 531-548.
- [Dant63] Dantzig, G.: *Linear Programming and Extensions*, Princeton University Press, Princeton (1963).
- [FoHW94] Forbes, M. A., Holt, J. N., Watts, A. M.: An exact algorithm for multi-depot bus scheduling, In: *European Journal of Operational Research* 72 (1994), S. 115-124.
- [GeLi81] George, A. and Liu, J.W.: *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall Inc. (1981).
- [GiKS05] Gintner V., Kliewer N. and Suhl L.: Solving large multiple-depot multiple-vehicle-type bus scheduling problems in practice, In: *OR Spektrum* 27 (2005), S. 507-523.

-
- [HBJM95] Hane, C., Barnhart, C., Johnson, E.L., Marsten, R.E., Nemhauser, G.L., Sigismondi, G., The Fleet Assignment Problem: Solving A Large Integer Program, In: Mathematical Programming, 70 (1995) 2, S. 211-232.
- [ILOG06] ILOG homepage: <http://www.ilog.com/>
- [Karm84] Karmakar, N.: A new polynomial time algorithm for linear programming, In: Combinatorica 4 (1984), S. 373-395.
- [Kliw05] Kliewer, N.: Optimierung des Fahrzeugeinsatzes im öffentlichen Personennahverkehr, PhD-Thesis at the University of Paderborn, II - Fakultät für Wirtschaftswissenschaften / Department Wirtschaftsinformatik, (2005)
- [KIMS06] Kliewer, N.; Mellouli T. and Suhl L.: A time-space network based exact optimization model for multi-depot bus scheduling, In: European Journal of Operational Research 175 (2006), S. 1616-1627
- [Kobe05] Koberstein, A.: The Dual Simplex Method: Techniques for a fast and stable implementation, PhD-Thesis at the University of Paderborn, II - Fakultät für Wirtschaftswissenschaften / Department Wirtschaftsinformatik, (2005)
- [KoSu07] Koberstein, A. and Suhl U.H.: Progress in the Dual Simplex Algorithm for solving large scale LP problems: Practical dual phase 1 algorithms, to appear 2007 in Computational Optimization and Applications
- [Lemk54] Lemke, C.E.: The Dual Method of Solving the Linear Programming Problem, In: Naval Research Logistics Quarterly 1 (1954), S. 36-47.
- [Löbe99] Löbel, A.: Solving Large-Scale Multiple-Depot Vehicle Scheduling Problems, In: Computer-Aided Transit Scheduling 471 (1999), S. 193-220.
- [Mehr92] Mehrotra, S.: On the implementation of a primal-dual interior point method, In: SIAM Journal on Optimization 2 (1992) 4, S. 575-601.
- [MelK102] Mellouli, T. und Kliewer, N.: Umlaufplanung im öffentlichen Verkehr mit mehreren Depots und Fahrzeugtypen: Neue Lösungsmodelle und praktische

Aspekte, In: Tagungsbericht der HEUREKA'02 (Optimierung in Verkehr und Transport), FGSV-Verlag, Köln (2002), S. 63-76.

- [Mell03] Mellouli T.: Scheduling and Routing Systems in Public Transport Systems: Modeling, Optimization, and Decision Support, Habilitationsschrift, Universität Paderborn (2003).
- [Mész96] Mészáros, Cs.: Fast Cholesky Factorization for Interior Point Methods of Linear Programming, In: Computers & Mathematics with Applications 31 (1996), S. 49-51.
- [Mész98] Mészáros, Cs.: Ordering heuristics in interior point LP methods, In: New Trends in Mathematical Programming, (Eds.: F Gianessi, S. Komlósi und T. Rapcsák), Kluwer Academic Publishers (1998), S. 203-221.
- [MOPS06] MOPS - Mathematical Optimization System homepage: <http://www.mops-optimizer.com>.
- [RiSo94] Ribeiro, C., Soumis, F.: A column generation approach to the multiple-depot vehicle scheduling problem, In: Operations research 42 (1994) 1, S. 41-52.
- [Suhl94] Suhl, U.H.: MOPS - Mathematical OPTimization System, In: European Journal of Operational Research 72 (1994), S. 312-322.
- [SuSz94] Suhl, U.H. and Szymanski R.: Supernode Processing of Mixed-Integer Models, In: Computational Optimization and Applications 3 (1994), S. 317-331.
- [SuWa04] Suhl, U.H. und Waue V.: Fortschritte bei der Lösung gemischt-ganzzahliger Optimierungsmodelle, In: Quantitative Methoden in ERP und SCM, (Eds.: L. Suhl und S. Voss), DSOR Beiträge zur Wirtschaftsinformatik (2004), S. 35-53.

Optimierungssysteme für die Dienstplanung im ÖPNV

Leena Suhl, Natalia Kliewer, Ingmar Steinzen

Universität Paderborn, Decision Support&OR Lab und
International Graduate School for Dynamic Intelligent Systems
Warburger Str. 100, D-33098 Paderborn
E-Mail: {suhl, kliewer, steinzen}@dsor.de

Abstract

Die Einsatzplanung von Bussen und die Dienstplanung von Busfahrern im städtischen öffentlichen Personennahverkehr (ÖPNV) sind sehr komplex und erfordern eine Unterstützung durch computerbasierte Planungssysteme. Softwarepakete für die Planung im ÖPNV bieten traditionell eine graphisch-interaktive Benutzeroberfläche, die eine interaktive, prinzipiell manuelle Planung ermöglicht. In den letzten Jahren ist es jedoch durch die rasante Entwicklung der mathematischen Optimierungssoftware möglich geworden, komplexe Planungsmodelle im ÖPNV optimal zu lösen. Solche Modelle sollten daher in Planungssysteme für die praktische Arbeit integriert werden.

In diesem Beitrag werden Optimierungsmodelle für die Dienstplanung im ÖPNV und deren Einbettung in Planungssysteme diskutiert. Insbesondere werden Möglichkeiten und Grenzen von Set-Partitioning-Modellen sowie deren Erweiterung für mehrere Zielkriterien unter Nutzung von Metaheuristiken betrachtet. Einige der beschriebenen Algorithmen sind in der kommerziellen Software Interplan der PTV AG integriert.

1 Computergestützte Planungsprozesse im ÖPNV

Der Planungsprozess im öffentlichen Personennahverkehr lässt sich grundlegend in strategische und operative Planung unterteilen (s. Abb. 1). Die Planung ist aufgrund zahlreicher Anforderungen, die es zu berücksichtigen gilt, ein komplexer Prozess. Das Ziel dieser Planung ist eine möglichst kostengünstige und trotzdem qualitativ hochwertige Bedienung der Reisewünsche von Fahrgästen im Nahverkehr.

Nach einer Bedarfsermittlung werden im Rahmen der strategischen ÖPNV-Planung die Passagierströme geschätzt und es wird eine grobe Angebotsplanung vorgenommen. Die anschließende Netzplanung legt die Infrastruktur des Nahverkehrsnetzes fest. In der ersten Stufe der operativen Planung werden die zu bedienenden Linien festgelegt, die anschließend im Rahmen der Fahrplanerstellung mit genauen Zeitangaben für jede Linienfahrt versehen werden. Diese Aufgaben sowie die Steuerung des Vergabeverfahrens von Linien und Fahrten des ÖPNV liegen normalerweise in der kommunalen Hand.

Der Planer im Verkehrsunternehmen steht einer schwierigen Aufgabe gegenüber. Die von ihm erstellten Pläne müssen in sich schlüssig, also fahrbar sein, und dabei einen möglichst effizienten Ressourceneinsatz ermöglichen. Die Fahrzeugumläufe und Fahrdienste sind dabei so zu planen, dass alle Linienfahrten des Unternehmens mit möglichst geringen Fahrzeug- und Personalkosten bedient werden.

Verkehrsunternehmen müssen somit im Rahmen des Planungsprozesses die genauen Umlaufverläufe ihrer Fahrzeuge festlegen und die Arbeitspläne für die benötigten Fahrer erstellen. Bei der Planung spielen jedoch nicht nur die Kostenaspekte eine Rolle. Es gibt viele zu beachtende Nebenbedingungen, wie z.B. gesetzliche sowie tarifvertragliche Regelungen, welche die Länge der Pausen oder der Arbeitszeiten der Fahrer festsetzen, oder technische Bedingungen, beispielsweise Wartungsintervalle und Geschwindigkeit der Transportmittel.

Um eine gesamtkostenoptimale ÖPNV-Planung unter Beachtung aller Nebenbedingungen durchzuführen, wäre es nötig, alle Teilaufgaben des Planungsprozesses, wie Linien-, Fahrplan-, Umlauf- und Dienstplanung simultan zu lösen. Da dies aber auf Grund der komplexen Datenlage nicht möglich ist, wird der Planungsprozess in kleinere Teilprobleme zerlegt, die dann sequentiell gelöst werden (s. Abb. 1). In diesem Prozess kann allerdings eine Wiederholung eines vorherigen Planungsschrittes nötig werden. Dies geschieht dann, wenn das Ergebnis von einem Teilprozess in Bezug auf die Gesamtplanung nicht den Wünschen entspricht, oder gar eine Lösung des nachfolgenden Teilproblems hierdurch unmöglich wird. Wird z.B. durch die Festlegung der Fahrzeugumläufe eine zu große Anzahl an Fahrern benötigt, um diese zu bedienen, so kann es sinnvoll sein, die Fahrzeugplanung zu revidieren, so dass weniger Fahrer ausreichen würden. Zwar ist dann die Lösung des Fahrzeugumlaufplanungsproblems nicht mehr optimal im Sinne des für diese Aufgabe ursprünglich festgelegten Optimalitätskriteriums, aber die damit entstandenen Kosten können durch Einsparungen bei der Fahrereinsatzplanung kompensiert werden.

Eine mögliche grobe Zerlegung der gesamten operativen Planung beinhaltet die Teilaufgaben Linienplanung, Fahrplanerstellung sowie Fahrzeugumlauf- und Dienstplanung. Letztere wird in der Regel in zwei Stufen durchgeführt: zunächst werden anonyme Tagesschichten gebildet und erst in der nachfolgenden Dienstreihenfolgeplanung der eigentliche (personalisierte) Dienstplan. In der Praxis wird üblicherweise hierarchisch in der genannten Reihenfolge geplant.

2 Schritte der operativen Planung

Als Schwerpunkt der operativen Planung wird in diesem Beitrag die Dienstbildungsaufgabe für Busfahrer angesehen. Diese Planungsaufgabe ist natürlich immer im Kontext der gesamten operativen Planung zu sehen, da die einzelnen Planungsschritte aufeinander aufbauen und somit nicht losgelöst voneinander betrachtet werden können. Nachfolgend werden in diesem Zusammenhang die operativen Planungsschritte kurz vorgestellt.

2.1 Linienplanung

Im Rahmen der Linienplanung werden die Grundrouten und Fahrzeitprofile (Zeitbedarfe für die einzelnen Fahrtabschnitte einer Linie) der zu bedienenden Linien bestimmt. Das Ziel dieses Planungsschrittes ist es, mit einer begrenzten Anzahl von Linien ein Maximum an Direktverbindungen (d.h. Verbindungen ohne Umsteigen) anzubieten.

Unter dem Begriff *Linie* wird eine Folge von Haltestellen bzw. Wegpunkten verstanden, die nacheinander bedient werden sollen. Unter dem Begriff *Liniennetz* wird die Gesamtheit der erzeugten Linien mit ihren Haltestellen, Pausenräumen und Depots verstanden. In einem Liniennetz werden alle Linien erfasst, die zu bedienen sind. Abbildung 2 beinhaltet einen Ausschnitt des Liniennetzes des PaderSprinter in Paderborn.

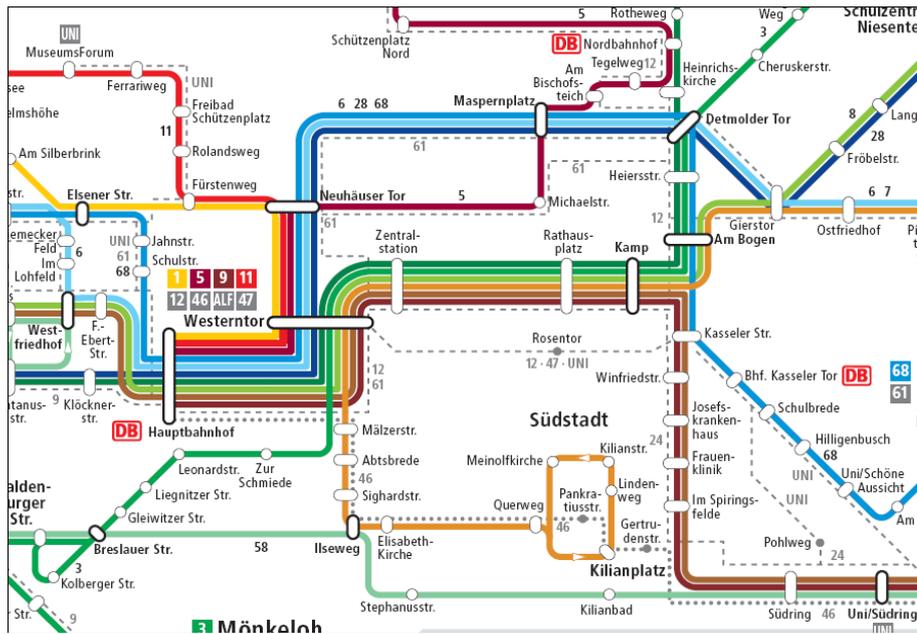


Abb. 2: Ausschnitt aus einem Liniennetz von PaderSprinter, Paderborn

Ausgangspunkt für diesen Planungsschritt sind die Ergebnisse der strategischen Planung, insbesondere die Schätzungen über die Nachfrage an benötigten Verkehrsverbindungen. Basierend auf dem geschätzten Beförderungsbedarf und auf einem vorgegebenen Streckennetz wird der Verlauf einzelner Linien festgelegt. Die Anzahl der Linien ist in der Regel begrenzt. Sie sollen so geplant werden, dass möglichst viele Direktverbindungen zustande kommen. Falls erforderlich, wird für jede Linie eine *Fahrzeugtypgruppe* bestimmt, die alle für die Fahrten dieser Linie geeignete Fahrzeugtypen enthält.

Die Aufgabe der Linienplanung kann mit Optimierungsmethoden unterstützt werden. Einige Modellierungsansätze findet man z.B. in [BoLW04] und speziell zu der Planung der Linienverläufe im schienengebundenen Verkehr in [BuZi97].

2.2 Fahrplanerstellung

Durch die Festlegung des Linienverlaufs ist zwar der Zeitbedarf für die einzelnen Linien festgelegt, die Taktfrequenz der Linie und die Abfahrtszeitpunkte für jede Linienfahrt müssen jedoch noch bestimmt werden. Dies geschieht im Rahmen der Fahrplanerstellung. Die möglichen Wartezeiten der Fahrgäste beim Umsteigen werden dabei minimiert, so dass möglichst nur „verträgliche“ (also weder zu lange noch zu kurze) Wartezeiten entstehen. An dieser Stelle im Planungsprozess kommen wieder die zugrunde gelegten Daten der geschätzten Verkehrsverbindungen zum Tragen. In dieser Planungsaufgabe wird beispielsweise darauf geachtet, dass es bei Straßenbahnen nicht möglich ist, ein Gleis zur gleichen Zeit doppelt zu

nutzen. Bei der Festlegung der Abfahrtszeiten müssen auch solche Nebenbedingungen zusätzlich beachtet werden.

Bei der Fahrplanerstellung wird also die räumliche Struktur eines Liniennetzes mit der zeitlichen Struktur des Fahrplans überlagert. Dabei werden Linien in Personenbeförderungsfahrten zwischen zwei Endhaltestellen zerlegt. Das Ziel der Fahrplanerstellung ist die Vermeidung bzw. Minimierung der gegenseitigen Behinderung von Fahrplanfahrten auf gemeinsam benutzten Straßen und eine optimale Ergänzung der Fahrten im Gesamtverkehrssystem. In dieser Planungsphase werden auch die Mengen der zulässigen Fahrzeugtypen für die einzelnen Fahrten festgelegt, falls sie nicht für alle Fahrten einer Linie gleich sein sollen.

2.3 Fahrzeugumlaufplanung

In der Fahrzeugumlaufplanung (auch als Umlaufplanung bezeichnet) wird entschieden, welche Fahrzeugtypen bzw. Fahrzeuge auf den zuvor definierten Linien verkehren. In dieser Teilaufgabe der Planung werden die vom Fahrplan vorgegebenen Personenbeförderungsfahrten den vorhandenen Fahrzeugen zugeordnet. Bei der Zuordnung können ein oder mehrere Ziele angestrebt werden. Es kann eine Minimierung der Anzahl einzusetzender Fahrzeuge oder die Minimierung der Leerfahrtzeiten und/oder die Einhaltung bestimmter Vorgaben, wie die Begrenzung der maximalen Anzahl von Linienwechsel pro Umlauf, sein.

Für eine mathematische Formulierung des Umlaufplanungsproblems mit mehreren Depots (MDVSP) verweisen wir auf [KIMS06], [Kliew05] und [GiKS05], die eine Mehrgüterflussmodellierung vorschlagen und damit reale Instanzen mit über 10.000 Fahrten und 18 Depots lösen. Das MDVSP ist NP-hart [BeCG87].

2.4 Dienstplanung

Bei der Dienstplanung geht es um die Planung des Personaleinsatzes zur Bedienung der im Rahmen der Umlaufplanung festgelegter Fahrzeugumläufe. Dabei werden mehrere Teilaufgaben, insbesondere Tagesdienstbildung (engl.: *crew scheduling*) und Dienstreihenfolgeplanung (engl.: *crew rostering*), sequentiell durchgeführt. Die Tagesdienstplanung liefert eine Menge anonymer Dienste, die in der Dienstreihenfolgeplanung zu persönlich zugeordneten Wochendiensten zusammengefasst werden. Nachfolgend werden diese beiden Teilaufgaben der Dienstplanung beschrieben.

Für die Tagesdienstplanung werden die Fahrzeug-Umläufe an gegebenen *Ablösepunkten* geteilt und die so entstandenen *Dienstelemente* Diensten kostenminimal zugeordnet. Eine Lösung des Dienstplanungsproblems ist genau dann zulässig, wenn jedes Dienstelement ausgeführt wird und jeder Dienst die gesetzlichen sowie betrieblichen Regelungen erfüllt.

Ein *Dienststück* ist eine Abfolge von Dienstelementen, die ein Fahrer an einem Fahrzeug ohne Pausenunterbrechung bedient. Ein Dienst besteht aus einem oder mehreren solcher Dienststücke unterteilt durch anrechenbare Pausen.

2.5 Tagesdienstplanung

Die ermittelten Dienste werden zunächst als „anonyme“ Tagesdienste erstellt, die nicht auf bestimmte Personen bezogen sind. Die Dienstaktivitäten (oder Dienstelemente) werden durch ein manuelles oder automatisches Schneiden der Fahrzeugumläufe in Teilumläufe gebildet. Diese Aktivitäten werden anschließend auf Einzeldienste bestimmter Dienstklassen grafisch- interaktiv oder automatisch zugeordnet.

Beim Schneiden von Diensten oder Dienststücken aus den Fahrzeugumläufen müssen unterschiedliche Randbedingungen berücksichtigt werden. Dies sind beispielsweise die minimale bzw. maximale Länge eines Dienstes oder Dienststückes, die Zeitintervalle für Dienstbeginn und -ende oder die Länge der entstehenden Reststücke, um so unverplanbare Reststücke zu vermeiden. Eine wichtige Nebenbedingung ist die Einhaltung tarifvertraglicher Anforderungen.

Die meisten Entscheidungsunterstützungssysteme zur Dienstplanung beinhalten ein grafisch- interaktives Bilden von Dienststücken und Diensten mit anschließendem automatischem Kombinieren der Dienststücke zu zulässigen Diensten. Beim Generieren von Einzeldiensten wird ein Ausgangsumlaufstück mit sinnvoll kombinierbaren weiteren Stücken zu einem Dienstvorschlag erweitert und bewertet. Der Planer entscheidet, welchen dieser alternativen Vorschläge er als Dienst akzeptiert.

Bei vorgegebenen Dienstklassen, wie Früh-, Mittel- oder Spätdienste, muss, soweit möglich, eine definierte Anzahl unterschiedlicher Dienste einer Klasse generiert und bewertet werden, die den Rahmenbedingungen entsprechen. Falls ein Optimierungsverfahren eingesetzt wird, wird aus der Gesamtmenge möglicher Dienste eine Teilmenge ausgewählt, die alle Dienstaktivitäten abdeckt und bestimmte Zielsetzungen, wie die Kostenminimierung, am besten erfüllt. Diese Tagesdienste werden an die Dienstreihenfolgeplanung weitergegeben.

2.6 Dienstreihenfolgeplanung

Anschließend werden die Dienstpläne der verschiedenen Wochentage unter Berücksichtigung von Mindestruhezeiten (auch für Feiertagsübergänge), durchschnittlichen Wochenarbeitszeiten und vorgebbaren Turnusfolgen von Arbeits- bzw. arbeitsfreien Tagen erstellt. Dabei werden mehrere Tagesdienste zu Wochen-Dienstplänen zusammengefasst, so dass sie hintereinander ausgeführt werden können, ohne dass Bedingungen wie die minimale Ruhezeit zwischen zwei Diensten verletzt werden. Danach erfolgt schließlich die Zuweisung der Dienstpläne zu den einzelnen Fahrern. Erst dann wird die Anonymität der Dienstpläne aufgegeben.

2.7 Das Forschungsproblem

Als Ziel der Planung im Rahmen des öffentlichen Personennahverkehrs kann eine möglichst günstige Bedienung des Beförderungsbedarfs der Kunden angesehen werden. Allerdings darf der Kostenpunkt dabei nicht allein betrachtet werden, da es viele andere – sowohl gesetzliche als auch tarifvertragliche – Rahmenbedingungen gibt. Dazu zählen die Arbeitszeitrestriktionen für Fahrer, die technischen Bedingungen für Fahrzeuge oder auch eine Vorschrift zur Bedienung unrentabler Verbindungen.

In den nachfolgenden Abschnitten wird speziell das Dienstplanungsproblem betrachtet. Prinzipiell soll dabei eine Menge von Diensten konstruiert werden, so dass alle Fahrten durch einen Fahrer bedient werden, alle Restriktionen erfüllt sind und die Gesamtkosten dabei minimiert werden. Wegen der hohen kombinatorischen Komplexität sind Dienstplanungsmodelle praktischer Größenordnung heute nicht optimal lösbar. Im Folgenden werden Methoden vorgestellt, die helfen, möglichst gute – wenn auch nicht optimale – Lösungen zu generieren. Ein solcher Ansatz basiert auf einem Column-Generation-Verfahren und nutzt bei der Lösung der entstehenden Unterprobleme eine neue effiziente Time-Space-Formulierung. Ein weiterer Ansatz berücksichtigt nicht nur die reine Kostenminimierung als Zielfunktion, sondern auch weitere Ziele und nutzt Metaheuristiken zur Lösung der multikriteriellen Modelle.

3 Optimierungsmodelle für das Dienstplanungsproblem

Nachdem exakte Optimierungsmodelle zur Lösung des Dienstplanungsproblems in Fluggesellschaften seit über drei Jahrzehnten genutzt werden, haben diese in den letzten Jahren

den Einzug auch in die Praxis des ÖPNV gefunden. Einige Optimierungssoftwarepakete (vgl. Hastus, Interplan) beinhalten heute Optimierungskomponenten, wobei zur Lösung sowohl exakte Methoden als auch Heuristiken eingesetzt werden.

Das Dienstplanungsproblem wird meist als ein *Set Partitioning Problem (SPP)* formuliert und mit Hilfe eines *Column Generation*-Verfahrens gelöst (z.B. [DeSo89], [GrBL03]).

Das Basismodell des SPP ist in (3.1)-(3.3) dargestellt, wobei die binären Entscheidungsvariablen x_k den berücksichtigten Diensten als Spalten entsprechen: der Wert $x_k = 1$ bedeutet, dass der Dienst ausgewählt wurde und $x_k = 0$, dass der Dienst nicht ausgewählt wurde. Die Menge I beschreibt alle Dienstelemente, die durchgeführt werden müssen und $K(i)$ die Menge der Dienste die das Dienstelement i beinhalten.

$$\sum_{k \in K} c_k x_k \longrightarrow \min \tag{3.1}$$

$$\sum_{k \in K(i)} x_k = 1 \quad \forall i \in I \tag{3.2}$$

$$x_k \in \{0,1\} \tag{3.3}$$

Das Optimierungsverfahren sucht unter den vorgegebenen Diensten nach einer kostenoptimalen Kombination (3.1), so dass alle Dienstelemente mit adäquater Besetzung durchgeführt werden (3.2). Weil die Anzahl der zulässigen Dienste bereits für kleine Modelle Millionen beträgt, können nicht alle möglichen Dienste als Spalten der Koeffizientenmatrix von vornherein generiert werden. Zuerst wird eine begrenzte Anzahl von Diensten generiert und das entsprechende Optimierungsmodell als lineares Programm gelöst. Danach werden solche neuen Dienste als Spalten generiert, die negative reduzierte Kosten besitzen und somit tendenziell die Lösung verbessern können. Das Verfahren beinhaltet somit abwechselnd die Lösung eines Masterproblems als LP und die Lösung eines Unterproblems (Pricing) zur Generierung vorteilhafter neuer Spalten. Das LP-Verfahren endet, wenn keine vorteilhaften Spalten mehr existieren oder wenn die vorgegebene Lösungszeit verbraucht ist. Danach muss noch eine ganzzahlige Lösung bestimmt werden.

Zur Lösung der des Master-Problems wird nach unserer Vorgehensweise zuerst eine primale und duale LP-Lösung berechnet. Statt des klassischen Simplex-Verfahrens wird hier der sogenannte Volume-Algorithmus eingesetzt [BaAn00], der schnell approximative Lösungen berechnet und somit insgesamt Rechenzeit einspart.

In einem zweistufigen Pricing-Prozess bestimmen wir mit der so errechneten dualen Lösung neue Spalten mit negativen reduzierten Kosten und fügen sie dem Master-Problem hinzu.

Weiterhin verwenden wir die primalen und dualen Informationen zum Spaltenmanagement im beschränkten Master-Problem. Wenn keine neuen Spalten im Pricing-Problem gefunden werden können, berechnen wir eine ganzzahlige Lösung.

Das Unterproblem im Pricing-Prozess kann als ein Resource-Constrained-Shortest-Path-Modell (RCSP) formuliert werden, wobei unterschiedliche Netzwerkrepräsentationen möglich sind. Unsere Formulierung basiert auf dem gerichteten Graphen $G=(N,A)$, in dem die Knoten N den Ablösepunkten J , der Quelle s und Senke t entsprechen. Kanten in A repräsentieren Leerfahrten (Fahrten ohne Passagiere), Dienstelemente, Ein- und Ausrücken aus dem Depot sowie Pausen. Da jeder Ablösepunkt einen Zeitpunkt darstellt und wir die Knoten nach (aufsteigenden) Zeitpunkten sortieren, ist das Netzwerk azyklisch. Jede Kante $(i,j) \in A$ ist mit Kosten c'_{ij} und einem Ressourcenverbrauch $d_{ij}^r \geq 0$ für jede Ressource $r \in R$ assoziiert. Die Kosten sind so definiert, dass die Kosten eines Pfads mit den reduzierten Kosten des entsprechenden Dienstes übereinstimmen. Sei λ_i der mit Restriktion (3.2) assoziierte duale Wert für Dienstelement i , dann sind die reduzierten Kosten c'_{ij} für alle Dienstelement-Kanten $c'_{ij} = c_{ij} - \lambda_i$. Für alle übrigen Kanten in A gilt $c'_{ij} = c_{ij}$. Liegt der Ressourcenverbrauch eines Pfads $P = \sum_{(i,j) \in P} d_{ij}^r$ von s nach t für alle $r \in R$ und alle Kanten in dem Intervall $[u^r, o^r]$, so handelt es sich um einen zulässigen Dienst. u^r gibt dabei den minimalen und o^r den maximalen erlaubten Verbrauch der Ressource r an. Einen Überblick über verschiedene Netzwerkformulierungen für das Dienstplanungsproblem und die Modellierung komplexer Regeln geben [DGSS92], [StGS06]. Die Formeln (3.4)-(3.7) zeigen die mathematische Formulierung von RCSP.

$$\sum_{(i,j) \in A} c'_{ij} x_{ij} \longrightarrow \min \quad (3.4)$$

$$\sum_{\{i|j \in A\}} x_{ij} - \sum_{\{i|(j,i) \in A\}} x_{ji} = 0 \quad \forall i \in N \setminus \{s, t\} \quad (3.5)$$

$$u^r \leq \sum_{(j,i) \in A} d_{ij}^r x_{ji} \leq o^r \quad \forall r \in R \quad (3.6)$$

$$x_{ij} \in \{0,1\} \quad \forall (i, j) \in A \quad (3.7)$$

Die binären Flussvariablen x_{ij} (3.7) geben an, ob ein Fluss auf Kante $(i,j) \in A$ existiert. Die Flusserhaltungsbedingungen (3.5) erzeugen einen zulässigen Fluss mit Ausnahme von Quelle und Senke. Die Ressourcenbedingungen (3.6) stellen sicher, dass jeder Dienst zulässig ist. Es ist zu beachten, dass nicht jede beliebige in der Praxis auftretenden Restriktion mit den Nebenbedingungen (3.6) modelliert werden kann. Es ist bekannt, dass RCSP schon für $|R| \geq 1$

ein NP-hartes Problem ist [HaZa80]. RCSP muss dabei nicht in jeder Column-Generation-Iteration bis zur Optimalität gelöst werden, da es ausreicht, entweder eine Menge von Pfaden negativer Länge zu erzeugen oder zu zeigen, dass solche Pfade nicht existieren.

4 Berücksichtigung mehrerer Zielfunktionen

Weil die klassische Zielsetzung von Betrieben im ÖPNV die Minimierung von Kosten unter einem gegebenen Servicegrad ist, berücksichtigt das oben erläuterte Optimierungsmodell die Kosten als alleiniges Hauptziel in der Zielfunktion. Neben der reinen Kostenminimierung müssen in der Praxis jedoch auch weitere Ziele berücksichtigt werden, die heute in der Regel entweder als Restriktionen formuliert werden oder bei der Optimierung außer Acht gelassen werden. Im Folgenden werden solche praxisrelevante Ziele kurz charakterisiert.

Ein wichtiger Aspekt ist die Zahl der Linienwechsel eines Fahrers pro Tag, d.h. wie oft wechselt der Fahrer auf eine andere Linie, die eine andere Route im Stadtgebiet hat, so dass Kenntnisse aus mehreren geographischen Bereichen erforderlich sind. Es wird häufig gewünscht, dass die Zahl der Linienwechsel pro Fahrer und Tag gering ist bzw. eine gegebene Zahl (z.B. 3) nicht überschreitet [KIGS06], [Kliew05]. Allerdings ist dies nicht in allen Verkehrsbetrieben der Fall. Es gibt aber auch Städte, deren Verkehrsbetriebe Linienwechsel sogar begrüßen, weil der Tagesablauf eines Fahrers dadurch interessanter und vielseitiger wird.

Die gesamte Anzahl von unterschiedlichen Diensten, die gefahren werden sollen, soll bei einigen Anwendern minimiert bzw. nach unten beschränkt werden. Man verspricht sich davon einen einfacheren Planungsablauf und Umsetzung [LoPP01].

Dienste, die nur aus einem Dienststück bestehen, sind in der Regel unerwünscht, weil der Busfahrer in diesem Fall nur für einen Teil des Tages im Einsatz ist. Allerdings ist es häufig nicht erforderlich, diese Restriktion explizit zu formulieren, weil sie über Kostenaspekte bereits in der Zielfunktion berücksichtigt werden.

Beim ÖPNV ist es heute üblich, im Rahmen von Outsourcing einen Teil der Fahrten an einen externen Dienstleister zu vergeben. In diesem Fall müssen nicht alle Fahrten durch einen Bus und Fahrer des eigenen Betriebs abgedeckt werden. Man sucht somit nach einer optimale Kombination der Fahrten, die extern vergeben werden. Weiterhin ist es (im Rahmen von sogenannten Set-Covering-Modellen) möglich, dass ein Dienstelement mehr als einem Busfahrer zugewiesen wird.

Die veröffentlichten Fahrpläne im ÖPNV sind normalerweise täglich identisch, mit Ausnahmen von Wochenenden und Feiertagen. Ein Unternehmen führt jedoch häufig auch weitere Fahrten aus (Sonderfahrten), die nicht in den öffentlichen Fahrplänen enthalten sind.

Nach unserer Erfahrung können bis zu 5% der Fahrplanmasse eines Tages unregelmäßige Fahrten sein, wie z.B. Schul-, Bäder- oder Werksfahrten. Durch Feiertage oder Schulferien werden diese Ausnahmefahrten wiederum oft nicht über die gesamte Fahrplanperiode durchgängig ausgeführt. Fahrpläne weisen somit oft auch keine wöchentliche Regelmäßigkeit auf. Mit traditionellen Optimierungsansätzen für die Umlauf- und Dienstplanung entstehen zwar kostenminimale, aber in der Praxis unerwünschte unregelmäßige Dienstpläne. Regelmäßige Dienstpläne sind in der Praxis wichtig, da diese einfacher einzuführen und zu handhaben sind. Busunternehmen versuchen bislang heuristisch, Regelmäßigkeit durch das Lösen eines zweistufigen Problems zu erreichen. Der Planer identifiziert erst alle regelmäßigen Fahrten einer Periode und berechnet mit einem traditionellen Ansatz einen zulässigen Umlauf- und Dienstplan. Danach fixiert er manuell eine Teilmenge der regelmäßigen Dienste und löst ein zweites Problem mit allen verbleibenden Fahrten. Das zweite Problem beinhaltet im Allgemeinen sehr verstreute Fahrten, die nur mit Hilfe vieler Leerfahrten abgedeckt werden können und daher sogar in der optimalen Lösung hohe operative Kosten besitzen.

5 Ein bikriterielles Optimierungsmodell

Im Folgenden betrachten wir daher das Umlauf- und Dienstplanungsproblem im Regionalverkehr für unregelmäßige, sich wöchentlich nicht wiederholende Fahrpläne. Wir zeigen, dass durch das hier vorgeschlagene zweistufige Verfahren im Bezug auf Regelmäßigkeit bessere Dienstpläne bei geringer Erhöhung der operativen Gesamtkosten erzeugt werden können.

Um die Regelmäßigkeit bewerten zu können, ist zuerst ein Maß für die Distanz zwischen zwei Dienstplänen zu definieren. Das Maß soll für zwei identische Dienstpläne gleich 0 sein und soll weiterhin gemeinsame Fahrtsequenzen zwischen den Plänen berücksichtigen: zwei Pläne sind sich ähnlich, wenn diese viele identische Fahrtsequenzen (Folgen von Fahrten) beinhalten. Das folgende Maß für die Distanz eines Dienstes D_i zu einem Referenzplan R berücksichtigt diese Anforderung und hat sich für Praxisdaten bewährt:

$$\delta_{D_i}^R = |F_i| - \sum_{S \in \bar{S}(R, D_i)} |S| \quad (3.8)$$

wobei

- R : Referenz-Dienstplan,
- S : Fahrten, die Teil der gemeinsame Sequenz sind,
- \bar{S} : Menge aller gemeinsamen Sequenzen,
- D_i : Dienst i und
- F_i : Menge der Fahrten, die durch Dienst i abgedeckt werden,

ist. Das Maß gibt also an, wie viele Fahrten eines Dienstes nicht Teil einer Sequenz sind. Es ist Null, wenn der Dienst D_i komplett im Referenzplan enthalten ist.

Seien c_k die operativen Kosten eines Dienstes $k \in K$ mit K als Menge aller zulässigen Dienste und definiere $K(i) \in K$ als Menge der Dienste, die Dienstelement $i \in I$ abdecken. Die binären Entscheidungsvariablen x_k geben an, ob ein Dienst k für die Lösung ausgewählt wurde oder nicht.

Die Betrachtung der Ziele Kosten- und Distanzminimierung führt zu dem bikriteriellen Optimierungsproblem (3.10)-(3.12) (2ICSP):

$$\sum_{k \in K} c_k x_k \longrightarrow \min \quad (3.10)$$

$$\sum_{k \in K} \delta_k^R x_k \longrightarrow \min \quad (3.11)$$

s.t.
$$\sum_{k \in K(i)} x_k = 1 \quad \forall i \in I \quad (3.12)$$

$$x_k \in \{0,1\}.$$

Die binäre Entscheidungsvariable x_k bekommt den Wert 1, wenn der Dienst k ausgewählt wird, und 0 sonst. Die Distanz eines Dienstes k zum Referenzplan R wird mit δ_k^R bezeichnet (s. oben).

Bei einem bikriteriellen Optimierungsproblem existiert in der Regel keine eindeutige optimale Lösung. Vielmehr gibt es eine Menge von effizienten Lösungen, die nicht verbessert werden können, ohne dass mindestens eine der Zielfunktionen sich verschlechtert. Die Menge der effizienten Lösungen kann prinzipiell mit der ϵ -constraint-Methode bestimmt werden (vgl. (Ehrgott 2000)). Dabei wird zunächst das Optimum c^{opt} bzgl. eines Zielkriteriums nach einer gegebenen Rangordnung bestimmt. Danach wird im Modell ϵ ICSP (3.13)-(3.15) das Optimum

für das zweite Kriterium bestimmt, so dass das erste Zielkriterium unter einer Schwelle $\varepsilon = (1+\alpha) c^{opt}$ mit $\alpha \geq 0$ bleibt.

$$\sum_{k \in K} \delta_k^R x_k \longrightarrow \min \quad (3.13)$$

s.t.
$$\sum_{k \in K(i)} x_k = 1 \quad \forall i \in I \quad (3.14)$$

$$\sum_{k \in K} c_k x_k \leq \varepsilon \quad (3.15)$$

$$x_k \in \{0,1\}.$$

Für das Dienstplanungsproblem mit Daten aus der Praxis konnte das Zielkriterium Regelmäßigkeit unter der Verschlechterung von höchstens $\alpha=2\%$ des Zielkriteriums Kosten wesentlich verbessert werden (s. Tabelle 1).

instance	ICSP				ε ICSP			
	#cg_iter	ip_obj	gap	cpu_time	#cg_iter	ip_obj	gap	cps_time
real421	94	72510	0.1	03:36:25	3	49	0.0	00:03:33
real436	92	71121	0.1	03:44:29	2	32	0.0	00:00:43
avg.				03:40:27				00:02:08

Tabelle 1: Ergebnisse zum Lösungsverfahren mit der ε -Constraint-Methode

instance	#EQDR		%EQDR		\emptyset SEQLEN		%SEQLEN	
	ICSP	ε ICSP	ICSP	ε ICSP	ICSP	ε ICSP	ICSP	ε ICSP
real421	4	13	8.5	27.7	2.04	3.72	26.2	47.8
real436	5	24	11.4	54.6	2.44	5	28.5	58.5
avg.			9.9	41.1			27.4	53.1

Tabelle 2: Ergebnisse zur Regelmäßigkeit der Lösungen von ICSP und ε ICSP

Ziel der zweiten Optimierungsstufe nach Regelmäßigkeit ist es, möglichst viele Sequenzen der kostenoptimalen Lösung beizubehalten, so dass der Kostenwert sich nicht mehr als 2% verschlechtert. Tabelle 2 zeigt, dass sowohl die Anzahl der unveränderten Dienste (#EQDR) als auch deren prozentuale Anteil (%EQDR) für Testmodelle aus der Praxis wesentlich erhöht werden konnten. Dasselbe gilt für die Anzahl und die durchschnittliche Länge der gemeinsamen Sequenzen (\emptyset SEQLEN bzw. %SEQLEN).

Durch systematische Untersuchungen mit unterschiedlichen ε -Werten können prinzipiell alle effizienten Lösungen bestimmt werden. Welche Lösung für die Realisierung ausgewählt wird, sollte der Entscheidungsträger interaktiv entscheiden. Dafür ist es hilfreich, eine graphische Darstellung in Form eines Decision Support Systems zu erzeugen (Abbildung 3 zeigt ein Beispiel für zwei Kriterien).

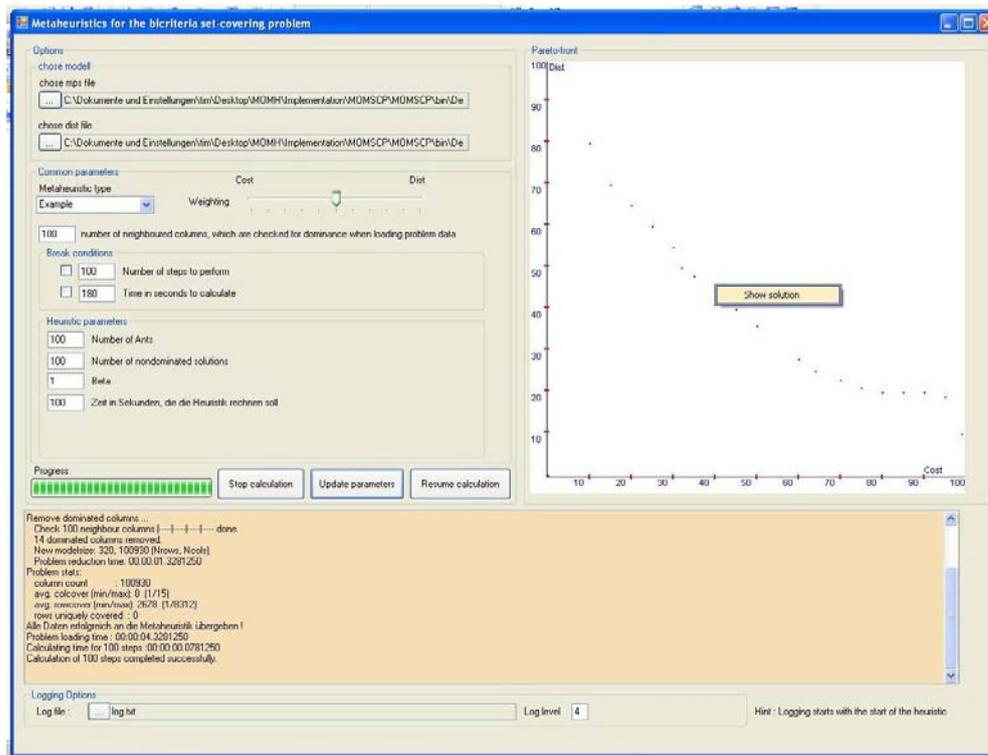


Abb. 3: Visuelle Darstellung der effizienten Lösungen

6 Kombination von exakten und heuristischen Lösungsmethoden

Bei der Lösung des Pricing-Problems beim Column-Generation-Verfahren wird in der Regel eine große Anzahl Spalten explizit oder implizit im Rahmen des Pricing-Prozesses generiert. Nur solche Spalten werden in das Mastermodell aufgenommen, die negative reduzierte Kosten im Hinblick auf die ursprüngliche Zielfunktion haben. Weil jedoch auch andere Ziele von Bedeutung sind, können die generierten Spalten mit wenig Aufwand im Sinne der weiteren Zielfunktionen bewertet werden. Da die exakte Lösung von kombinatorischen Optimierungsmodellen mit Hilfe der mathematischen Programmierung i.d.R. sehr zeitaufwändig ist, wäre es nicht sinnvoll zu versuchen, das Modell bzgl. weiterer Zielfunktionen exakt zu lösen. Durch die Nutzung der generierten Dienste als Ausgangspunkt für Metaheuristiken kann somit eine Menge an möglichen guten Lösungen schnell generiert und dem Planer zur Auswahl gestellt werden.

Unter Metaheuristiken versteht man eine Familie von allgemeinen Problemlösungs- und Suchmethoden, die insbesondere für kombinatorische Optimierungsmodelle geeignet sind. Eine für bestimmte Aufgabenstellung angepasste Metaheuristik findet i.d.R. schnell gültige

Lösungen, die im weiteren Verlauf verbessert werden können. Es kann keine Garantie der Lösungsoptimalität gegeben und weiterhin auch keine Aussage über die Lösungsqualität gemacht werden.

In einem laufenden Forschungsprojekt werden folgende Metaheuristiken für das bikriterielle Dienstplanungsproblem 2ICSP implementiert und mit dem Ziel getestet, möglichst gute Pareto-Fronten zu bestimmen:

- der genetische Algorithmus SPEA2 nach [ZiLT01],
- Simulated Annealing u.a. nach [CzJa98],
- Tabu Search nach [LoPP01] und
- Ameisenalgorithmen (ACO) nach [GuMi03].

Für einen ersten Vergleich der Heuristiken wurden nach dem Verfahren von [HuFW05] Fahrpläne mit 320 bis 800 Fahrten generiert. Sie beinhalten jeweils ein Depot, das gleichzeitig einziger Ablösepunkt ist, um ein typisches Szenario aus dem Regionalverkehr abzubilden. Mit Hilfe des beschriebenen Column Generation Verfahrens werden Dienste generiert, die den Ausgangspunkt für die Optimierung durch die Metaheuristiken darstellen. Wir erlauben dabei genauso wie im Vergleich der Heuristiken die Mehrfachüberdeckung von Dienstelementen.

Tabelle 3 stellt die Größe der verwendeten Testinstanzen dar.

Name	Zeilen	Variablen (Dienste)	Nicht-Null-Elemente
320-1	320	100.944	958.159
320-2	320	60.128	444.606
400-1	400	72.673	532.579
400-2	400	57.769	410.361
640-1	640	156.044	1.383.364
640-2	640	104.595	747.708
800-1	800	135.572	987.909
800-2	800	155.146	1.230.629

Tabelle 3: Größe der Testinstanzen zum Vergleich der Metaheuristiken

Die vier Metaheuristiken wurden anhand des sog. Hypervolumens (vgl. [ZiTh98]) bewertet. Das Hypervolume ermöglicht es, eine Menge von Lösungen (bspw. die Pareto-Front) auf ein Skalar abzubilden, welches ein Maß für die Qualität der gefundenen Lösungen ist. Je größer das Hypervolume, desto höher ist auch die Qualität der Pareto-Front. Es handelt sich allerdings um ein ordinales Kriterium, das keine Aussage über die Stärke der Merkmalsausprägung oder die Größe der Merkmalsunterschiede erlaubt. Tabelle 4 stellt die Hypervolumen der gefundenen Pareto-Fronten der vier Metaheuristiken für die Testinstanzen nach 10 Minuten Rechenzeit auf einem PC (Intel P4 3.0 GHz, 2GB) dar.

Die Ergebnisse zeigen, dass mit dem evolutionären Algorithmus SPEA2 für alle Testinstanzen die besten Pareto-Fronten bestimmt werden konnten. Die gefundenen Fronten bieten viele Lösungen und verteilen sich gut über den Suchraum. Der ACO Algorithmus konnte zwar nur wenige gute Lösungen finden, diese verteilten sich jedoch gut über den Suchraum. Im Gegensatz dazu konnten mit Tabu Search und Simulated Annealing nur wenige, nah beieinander liegende Lösungen gefunden werden.

Instanz	Simulated Annealing	Tabu Search	SPEA2	Ant Colony Optimization
320-1	18.201.164	18.262.251	20.072.003	19.120.189
320-2	29.887.737	29.809.664	31.589.976	30.007.990
400-1	71.836.269	71.740.288	75.745.072	74.281.816
400-2	89.681.239	89.746.885	94.141.782	90.036.482
640-1	199.034.704	199.575.139	216.579.504	203.016.853
640-2	271.548.362	270.471.248	286.592.522	275.709.660
800-1	511.536.389	517.875.331	548.948.336	528.818.610
800-2	350.059.120	346.716.776	376.969.363	361.932.510

Tabelle 4: Hypervolumes der Metaheuristiken

Weiterhin hat der Entscheider im Rahmen eines interaktiven Entscheidungsunterstützungssystems (vgl. Abbildung 3) die Möglichkeit, in den Lösungsprozesses steuernd einzugreifen.

7 Zusammenfassung

Der Beitrag beschreibt den Stand der Technik und eigene Entwicklungen für die multikriterielle Optimierung bei der Dienstplanung im ÖPNV. Durch die jüngsten Entwicklungen bei der Software zur Lösung von exakten Optimierungsmodellen können sehr schwierige Praxismodelle mit Hilfe der Set-Partitioning-Formulierung und Nutzung eines Column-Generation-Verfahrens nahezu optimal gelöst werden. Ein entscheidender Aspekt ist dabei die Modellierung des Unterproblems als ein Resource-Constrained-Shortest-Path-Modell.

Die beim Column-Generation-Verfahren generierten zahlreichen gültigen Dienste können mit wenig Aufwand im Hinblick auf weitere Ziele untersucht werden. Durch den Einsatz von Metaheuristiken werden diese in mehrere Zielrichtungen verbessert, so dass eine Menge gültiger Lösungen entsteht. Diese werden im Rahmen eines interaktiven Entscheidungsunterstützungssystems dem menschlichen Entscheidungsträger visualisiert.

Literaturverzeichnis

- [BaAn00] Barahona F., Anbil R. (2000) The volume algorithm: producing primal solutions with a subgradient method. *Mathematical Programming* 87(3):385-399.
- [BeCG87] Bertossi A., Carraresi P., Gallo G. (1987) On some matching problems arising in vehicle scheduling models. *Networks* 17:271-281.
- [BoLW04] Borndörfer R., Löbel A., Weider S. (2004) A bundle method for integrated multi-depot Vehicle and Duty Scheduling in Public Transit. ZIB Report 04-14, Berlin.
- [BuZi97] Bussieck, M., Zimmermann, U. (1997). Optimale Linienführung und Routenplanung in Verkehrssystemen. Technical Report, TU Braunschweig.
- [CzJa98] Czyzak P., Jaskiewicz A. (1998) Pareto Simulated Annealing – a metaheuristic technique for multiple-objective combinatorial optimisation. *Journal of Multi-Criteria Decision Analysis* 7:34-47.
- [DGSS92] Desrochers M., Gilbert J., Sauvé M., Soumis F. (1992) Crew-Opt: Subproblem Modeling in a Column Generation Approach to Urban Crew Scheduling. In: Desrochers M., Rousseau J. (Hrsg) *Computer-Aided Scheduling, Lecture Notes in Economics and Mathematical Systems* 386, Springer, Berlin.
- [DeSo89] Desrochers M., Soumis F. (1989) A Column Generation Approach to the Urban Transit Crew Scheduling Problem. *Transportation Science* 23:1–13.
- [Ehrg00] Ehrgott M. (2000) *Multicriteria Optimization*. Springer, Berlin.
- [GiKS05] Gintner V., Kliewer N., Suhl L. (2005) Solving large multiple-depot multiple-vehicle-type bus scheduling problems in practice. *OR Spectrum* 27(4):507-523.
- [GrBL03] Grötschel M., Borndörfer R., Löbel A. (2003) Duty Scheduling in Public Transit. In: *Mathematics - key technologies for the future*. Springer, Berlin.

- [GuMi03] Guntzsch M., Middendorf M. (2003) Solving Multi-Criteria Optimization Problems with Population-based ACO. In: Proc. of the Second Intern. Conf. on Evolutionary Multi-Criterion Optimization (EMO03), Springer, Berlin.
- [HaZa80] Handler G., Zang I. (1980) A Dual Algorithm for the Constrained Shortest Path Problem. *Networks* 10:293-310.
- [HuFW05] Huisman D., Freling R., Wagelmans A. (2005) Multiple-Depot Integrated Vehicle and Crew Scheduling. *Transportation Science* 39:491-502.
- [KIMS06] Kliewer N., Mellouli T., Suhl L. (2006) A time-space network based exact optimization model for the multi-depot bus scheduling. *European Journal of Operational Research*, 175(3): 1616-1627.
- [Kliew05] Kliewer N. (2005) Optimierung des Fahrzeugeinsatzes im öffentlichen Personennahverkehr. Dissertation, Universität Paderborn, DS & OR Lab.
- [KIGS06] Kliewer N., Gintner V., Suhl L. (2006) Line change considerations within a time-space network based multi-depot bus scheduling model. *Lecture Notes in Economics and Mathematical Systems (LNEMS), Computer-Aided Scheduling of Public Transportation (CASPT 2004)*, Springer, Berlin.
- [LoPP01] Lourenço H.R., Paixão J.O., Portugal R. (2001) Multiobjective Metaheuristics for the Bus Driver Scheduling Problem. *Transportation Science*, Vol. 35(3): 33.
- [StGS06] Steinzen I., Gintner V., Suhl L. (2006) Network models for a decomposed pricing problem. 10th International Conference on Computer-Aided Scheduling of Public Transport, Leeds, UK, 21.-23. Juni 2006.
- [ZiLT01] Zitzler E., Laumanns M., Thiele L. (2001) SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multiobjective Optimization. In: *Evolutionary Methods for Design, Optimisation and Control*, Barcelona, Spain.
- [ZiTh98] Zitzler E., Thiele L. (1998) Multiobjective optimization using evolutionary algorithms - a comparative case study. In: *Proc. of the Fifth International Conference on Parallel Problem Solving from Nature*, Berlin, Germany.

Barwertorientierte Projektplanung mit mehreren Akteuren mittels eines verhandlungsbasierten Koordinationsmechanismus

Andreas Fink

Professur für BWL, insbes. Wirtschaftsinformatik
Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg
Holstenhofweg 85, 22043 Hamburg
andreas.fink@hsu-hamburg.de

Abstract

In dem Beitrag werden Projektplanungsprobleme mit mehreren autonomen Entscheidungsträgern betrachtet. Es wird davon ausgegangen, dass die Aktivitäten eines hinsichtlich der Struktur bereits determinierten Projektes mit Ressourcenbeschränkungen durch eine Menge von Akteuren durchgeführt werden. Die Akteure verfolgen eigene Interessen hinsichtlich der zeitlichen Lage der ihnen zugeordneten Aktivitäten, die jeweils mit Zahlungen bei Aktivitätsabschluss verbunden sind. Jeder Akteur verfolgt eine individuelle Barwertmaximierung. Für das resultierende Koordinationsproblem mit autonomen Akteuren und Informationsasymmetrie wird ein generischer verhandlungsbasierter Koordinationsmechanismus ausgestaltet, angewendet und analysiert. Die beschriebenen Ergebnisse belegen die Effektivität des verwendeten Verfahrens hinsichtlich der Ermittlung hochwertiger gemeinsamer Projektpläne.

1 Einleitung

Die Wirtschaftsinformatik befasst sich mit soziotechnischen Informationssystemen und entsprechenden Informationsverarbeitungsaufgaben in Wirtschaft und Verwaltung [WKWI94]. Hinsichtlich der Lenkung des betrieblichen Geschehens nehmen hierunter Planungsaufgaben einen großen Anteil ein, deren Lösung primär Gegenstand der Operations Research ist. Im Kontext der von Mertens propagierten Langfristzielsetzung „sinnhafte Vollautomation“ [Mert95] ist für die Wirtschaftsinformatik die (Teil-)Automatisierung der Erfüllung von Planungsaufgaben im Rahmen softwaretechnischer Entscheidungsunterstützungssysteme ebenfalls von Relevanz. Damit ergeben sich deutliche Überschneidungen und fruchtbare Ergänzungen zwischen Erkenntnissen und Methoden von Wirtschaftsinformatik und Operations Research.

In diesem Beitrag werden Projektplanungsprobleme mit mehreren autonomen Entscheidungsträgern betrachtet. Es wird davon ausgegangen, dass die Aktivitäten eines hinsichtlich der Struktur bereits determinierten Projektes durch eine Menge von Akteuren durchgeführt werden (z.B. ein Bauprojekt, in dem verschiedene Gewerke bestimmten Bauunternehmen zugeordnet sind). Die Akteure verfolgen eigene Interessen hinsichtlich der zeitlichen Lage jeweiliger Aktivitäten. Demzufolge ist ein zentralistisches Entscheidungsmodell mit symmetrischer Informationsverfügbarkeit nicht unbedingt angemessen, da von Entscheidungsträgern mit konkurrierenden Zielen und asymmetrischer Informationsverteilung auszugehen ist. Klassische Verfahren der multikriteriellen Optimierung sind damit nicht direkt anwendbar, was unter Berücksichtigung der Randbedingungen autonomer Entscheidungsträger in diesem Beitrag zur Ausgestaltung, Anwendung und Analyse eines verhandlungsbasierten Koordinationsmechanismus führt.

Im Folgenden werden zunächst die Problemstellung und Anforderungen an entsprechende Lösungskonzepte diskutiert (Abschnitt 2). Anschließend wird ein generischer verhandlungsbasierter Koordinationsmechanismus beschrieben und für die betrachtete Problemstellung konkretisiert (Abschnitt 3). Das verfolgte Lösungskonzept wird mittels experimenteller Berechnungen analysiert, die in Abschnitt 4 dokumentiert sind. Die beschriebenen Ergebnisse belegen die Effektivität des verwendeten Verfahrens hinsichtlich der Ermittlung hochwertiger gemeinsamer Projektpläne, was abschließend in einem Fazit zusammengefasst wird (Abschnitt 5).

2 Problemstellung

2.1 Barwertorientierte Projektplanung mit mehreren Akteuren

Das in diesem Beitrag betrachtete Projektplanungsproblem ist folgendermaßen definiert: Das Projekt besteht aus einer endlichen Menge von zeit- und ressourcenbeanspruchenden Aktivitäten mit jeweils einem fixen Modus, der die Dauer und Ressourcenbeanspruchung einer Aktivität bestimmt. Jede Aktivität $j \in A = \{1, \dots, J\}$ ist mit einer Ausführungsdauer d_j verbunden; die Ausführung einer Aktivität darf nicht unterbrochen werden. Aufgrund technischer oder organisatorischer Restriktionen bestehen zwischen bestimmten Aktivitäten zeitliche Zusammenhänge, wobei hier lediglich schlichte Vorgänger-Nachfolger-Beziehungen angenommen werden derart, dass für eine Aktivität j alle Aktivitäten aus der jeweiligen Vorgängermenge v_j abgeschlossen sein müssen, bevor die Aktivität j durchgeführt werden kann. Der Projektbeginn und das Projektende werden durch zwei Sonderaktivitäten 1 bzw. J repräsentiert mit $d_1 = d_J = 0$, $v_1 = \emptyset$ und $v_J = \{j \mid j = 1, \dots, J-1\}$. Für den Projektbeginn wird eine untere Schranke von 0 angenommen, während gegebenenfalls eine obere Schranke das Projektende nach oben begrenzt.

Zur Durchführung von Aktivitäten werden Ressourcen bestimmter Ressourcentypen benötigt. Hier werden nur erneuerbare Ressourcen berücksichtigt, die für jeden Zeitpunkt im Planungshorizont in einer festen Menge zur Verfügung stehen (z.B. verschiedene Maschinen). Ausgehend von einer Menge R von Ressourcentypen und einer Verfügbarkeit K_m einer Ressource $m \in R$ pro Zeiteinheit repräsentiert $k_{j,m}$ den Ressourcenbedarf einer Aktivität j bezogen auf die Ressource m pro Zeiteinheit.

Ein zulässiger Projektplan weist jeder Aktivität einen Startzeitpunkt zu und berücksichtigt dabei die zeitlichen und ressourcenbezogenen Nebenbedingungen. Die Menge der zulässigen Projektpläne konstituiert den Lösungsraum, der im Zusammenhang mit mehreren beteiligten Akteuren im Folgenden auch als Vertragsraum bezeichnet wird.

Ziel der Projektplanung ist die Ermittlung eines zulässigen Projektplans im Hinblick auf eine von den Startzeitpunkten abhängige Zielsetzung. In der Literatur wird häufig die Minimierung der Projektdauer zugrunde gelegt (Resource Constrained Projekt Scheduling Problem (RCPSP); vgl. z.B. [BDM+99]), während bei einer ökonomischen Betrachtung insbesondere barwertorientierte Zielsetzungen relevant sind, wovon auch im Folgenden ausgegangen wird. Hierbei ist jede Aktivität j mit einem Zahlungswert z_j verbunden, wobei die jeweilige Ein- oder Auszahlung mit Abschluss der jeweiligen Aktivität anfällt. Unter Annahme eines Abzinsungsfaktors p werden

die Zahlungen in Abhängigkeit von der zeitlichen Einplanung auf den Zeitpunkt 0 abgezinst. Bei einer Gesamtplanung ist das Ziel die Maximierung der Summe der abgezinsten Zahlungen aller Aktivitäten (Resource Constrained Projekt Scheduling Problem with Discounted Cash Flows (RCPS-DC), vgl. z.B. [HeDD97; Koli97; VaDH01]). Die entsprechende einkriterielle Problemstellung stellt als Verallgemeinerung des RCPS-DC ein *NP*-schweres Problem dar.

Das Projekt wird durch eine Menge $P = \{ 1, \dots, n \}$ von Akteuren kooperativ durchgeführt. Die Aktivitäten sind eindeutig bestimmten Akteuren zugeordnet. Das heißt, ein Akteur $i \in P$ ist für die Durchführung einer jeweiligen Menge A_i zugeordneter Aktivitäten zuständig, wobei die Aktivitätenmenge A vollständig und disjunkt auf die Akteure aufgeteilt ist. Jeder Akteur verfolgt im Rahmen zulässiger Projektpläne gewinnorientiert das Ziel, die Barwertsumme der ihm zugeordneten Aktivitäten zu maximieren.

2.2 Anforderungen an einen Koordinationsmechanismus

Im Zusammenhang mit überbetrieblichen Transaktionen ist im Allgemeinen von autonomen Akteuren mit konkurrierenden Zielen und gewissen Informationsasymmetrien auszugehen. So wird auch hier für die oben beschriebene Problemstellung angenommen, dass zwar die Projektstruktur einschließlich der zeitlichen und ressourcenbezogenen Bedingungen allen Projektbeteiligten bekannt ist, andererseits aber allein die einzelnen Akteure Informationen zu den Zahlungswerten der ihnen zugeordneten Aktivitäten besitzen. Hieraus ergibt sich im Sinne der Institutionenökonomie die Aufgabe, die Ausschöpfung möglicher, aber gefährdeter Kooperationsvorteile durch zweckmäßige Koordinationsmechanismen zu unterstützen. Das heißt, es stellt sich die Frage, inwieweit ein zu entwickelndes Verfahren, das den Eigensinn der Akteure und deren private Informationen berücksichtigt, das Ergebnis einer hypothetischen integrierten Projektplanung mit einer Maximierung des Gesamtbarwerts erreichen kann und wie Kooperationsvorteile auf die Akteure verteilt werden.

Die generellen Anforderungen an einen problemunabhängigen Mechanismus zur Koordination autonomer Akteure, deren Entscheidungsinterdependenzen im Rahmen einer operativen Planungsebene formal beschrieben werden können, sind in [Fink06, Abschnitt 2] näher beschrieben. Es wird davon ausgegangen, dass die Akteure einen Koordinationsmechanismus dann akzeptieren, wenn eine hohe „Ergebnisqualität“ erreicht wird und das Verfahren durch eine gewisse „Fairness“ ausgezeichnet ist. Eine Konkretisierung dieser zunächst nur begrifflichen Krite-

rien kann unter Rückgriff auf entsprechende Erkenntnisse der Entscheidungstheorie bzw. kooperativen Spieltheorie [Fren88; Nash50; Nash53] erfolgen (vgl. bzgl. Fairness die ersten beiden folgenden Kriterien sowie das dritte Kriterium bzgl. Qualität):

- Das erzielte Ergebnis sollte invariant hinsichtlich positiv-affiner Transformationen der Nutzenfunktion sein; das heißt, ein Akteur sollte nicht durch eine hinsichtlich Ursprung und Skalierung willkürliche Veränderung seiner Nutzenfunktion im Rahmen seiner Präferenzen profitieren können.
- Das erzielte Ergebnis sollte unabhängig von der Nummerierung der Akteure sein (Symmetrie); das heißt, dass allein die Präferenzen der Akteure relevant sein sollten.
- Das erzielte Ergebnis sollte Pareto-optimal sein; das heißt, kein Akteur sollte besser gestellt werden können, ohne dabei einen anderen Akteur schlechter zu stellen.

Das dritte Kriterium kann als Optimierungsziel formuliert werden, indem die Minimierung einer zu definierenden Distanz von dem effizienten Rand (der Menge aller Pareto-optimalen Ergebnisse) angestrebt wird.

Neben diesen grundsätzlichen Anforderungen erscheinen als Verfahrensanforderungen außerdem von Relevanz (vgl. [Sand99]): Einfachheit und Universalität, effiziente Berechenbarkeit, geringe Ansprüche an die Offenlegung von Informationen der Akteure sowie hinsichtlich der Strategiewahl der Akteure das Vorhandensein eines eindeutigen Nash-Gleichgewichts und Anreizkompatibilität.

3 Verhandlungsbasierte Ermittlung von Projektplänen

Zur Lösung klassischer multikriterieller Problemstellungen mit einem zentralen Entscheidungsträger und Informationssymmetrie wurden vielfältige Verfahren entwickelt (vgl. z.B. [Ehrg05]). Hierzu zählen Erweiterungen von Metaheuristiken, die sich im Kern als iterative Suchprozesse auffassen lassen (vgl. z.B. [GSST05]). Das Konzept der iterativen Suche wurde auch auf Koordinationsprobleme mit mehreren Entscheidungsträgern und Informationsasymmetrie erweitert, wobei jeder Schritt eines auf lokaler Suche basierenden Ablaufs auf einer Abstimmung zwischen den Akteuren basiert [KFSB03a; KFSB03b]. Dieses Konzept, welches als ein vollautomatischer Verhandlungsmechanismus mit komplexen Vertragsräumen aufgefasst werden kann, wurde erweitert und differenzierter ausgestaltet und dabei für Problemstellungen aus dem Supply Chain Management validiert [Fink04; Fink06]. In den folgenden beiden Unterabschnit-

ten werden der entsprechende Koordinationsmechanismus zunächst allgemein kurz rekapituliert und anschließend die Konkretisierung für die hier betrachtete barwertorientierte Projektplanung mit mehreren Akteuren beschrieben.

3.1 Generischer Koordinationsmechanismus

Der grundsätzliche Koordinationsmechanismus, welcher als ein Verhandlungsprotokoll aufgefasst werden kann, ist in Abb. 1 beschrieben. Dieser Mechanismus ist generisch, wobei neben einem problemspezifischen Vertragsraum C eine problemspezifische Nachbarschaftsrelation N vorausgesetzt wird, die für jeden Vertrag eine Menge als benachbart geltender weiterer Verträge bestimmt.

Eingabe: Akteure $1, \dots, n$; Vertragsraum C ; Nachbarschaftsrelation $N : C \rightarrow 2^C$

Zufallsbasierte Erzeugung eines ersten aktiven Vertrags $c := c_0 \in C$.

Wiederhole für eine festgelegte Anzahl von Verhandlungsrunden:

Zufallsbasierte Erzeugung eines Vertragsvorschlags $c' \in N(c)$.

Jeder Akteur $i \in P$ beurteilt c' und äußert „Ja“ oder „Nein“.

Falls alle Akteure mit „Ja“ antworten:

Anpassung des aktiven Vertrags $c := c'$.

Ausgabe: Letzter aktiver Vertrag c als Ergebnis.

Abb. 1: Generischer Koordinationsmechanismus

In jedem Schritt wird ausgehend von dem jeweils vor diesem Schritt allseits vorläufig akzeptierten Vertrag (aktiver Vertrag) ein Vertragsvorschlag aus der jeweiligen Nachbarschaft zufallsbasiert ermittelt und den beteiligten Akteuren vorgelegt. Nur falls alle Akteure dem Wechsel vom bislang aktiven Vertrag zu dem Vertragsvorschlag zustimmen, wird aus dem Vertragsvorschlag der nun allseits vorläufig akzeptierte aktive Vertrag nach dem entsprechenden Schritt. Der nach dem letzten Schritt akzeptierte aktive Vertrag stellt als finale Übereinkunft das Ergebnis des Verfahrens dar.

Die Erzeugung von Vertragsvorschlägen hat in unparteiischer Form und unter Berücksichtigung der Privatheit von Informationen zur individuellen Qualität von Verträgen für die Akteure zu erfolgen. Demzufolge bietet sich die gleichverteilt zufällige Erzeugung sowohl von der Startlösung als auch von Nachbarlösungen (als jeweilige Vertragsvorschläge) an. Dies kann mittels

eines entsprechenden offen gelegten Algorithmus in Verbindung mit einem offen gelegten Zufallszahlengenerator erfolgen.

Sofern so genannte gierige Akteure vorgelegte Vertragsvorschläge eigensinnig nur dann akzeptieren, falls sich keine individuelle Verschlechterung des jeweiligen Barwerts ergibt, ist mit einem schnellen Stocken des Verfahrensablaufs und damit in der Regel einem relativ schlechten Endergebnis zu rechnen. Erfahrungen zu iterativen Suchprozessen für klassische ein- sowie multikriterielle Problemstellungen zeigen, dass im Verfahrensablauf die gelegentliche Inkaufnahme von Verschlechterungen vorteilhaft sein kann, um aus so genannten lokalen Optima und entsprechenden Regionen im Suchraum auszubrechen.

Dementsprechend wird das Verhandlungsprotokoll um eine Regel ergänzt, die die Akteure zwingt, in bestimmten Phasen des Verfahrensablaufs gewisse Mindestakzeptanzraten (bezüglich des Anteils der individuell akzeptierten Vertragsvorschläge) zu gewährleisten; diese Mindestakzeptanzraten nehmen zweckmäßigerweise im Verfahrensablauf ab. Da die Akzeptanzentscheidungen allseits beobachtbar sind, ist diese Regel verifizierbar und damit praktikabel. Akteure, die die gesetzten Mindestakzeptanzraten einhalten, werden als kooperativ bezeichnet.

Indessen stellt sich für die Akteure die Frage nach einer individuell vorteilhaften Entscheidungsstrategie, die einer vorgegebenen Mindestakzeptanzrate gerecht wird. Hierfür bietet sich der Rückgriff auf das probabilistische Akzeptanzkriterium der Metaheuristik Simulated Annealing an („Metropolis-Kriterium“), in dem die Wahrscheinlichkeit $P_T(\Delta)$ für eine Akzeptanz einer Verschlechterung um den Wert Δ exponentiell vom Ausmaß der Verschlechterung abhängt [KiGV83]:

$$P_T(\Delta) = e^{-\Delta/T}$$

Dieses Akzeptanzkriterium beinhaltet einen so genannten Temperaturparameterwert T . Mittels einer simulativen Verfahrensweise können die Akteure einen individuellen Temperaturparameterwert näherungsweise so bestimmen, dass eine entsprechende Anwendung zum Einhalten einer bestimmten Mindestakzeptanzrate führt. Für eine detaillierte Beschreibung und Diskussion des Verfahrens sei auf [Fink06] verwiesen.

Im Zusammenhang mit den in Abschnitt 2.2 formulierten Anforderungen ist hinsichtlich Fairness anzumerken, dass für alle Akteure die gleichen Mindestakzeptanzraten gelten und mögliche positiv-affine Transformationen der Nutzenfunktion durch das Akzeptanzkriterium in Verbindung mit dem ermittelten Temperaturparameter ausgeglichen werden. Die Ergebnisqualität ist Gegenstand der experimentellen Untersuchungen in Abschnitt 4.

3.2 Konkretisierung des Koordinationsmechanismus für die barwertorientierte Projektplanung mit mehreren Akteuren

Die Menge der zulässigen Projektpläne konstituiert den zugrunde gelegten Vertragsraum für die Projektplanung. Ein Vertrag kann grundsätzlich als ein Tupel zulässiger Startzeitpunkte für die einzelnen Aktivitäten aufgefasst werden. Da die direkte Ableitung zulässiger Vertragsvorschläge als Modifikationen des Startzeitpunkt-Tupels des aktiven Vertrags nicht einfach möglich ist, wird hier mit einer indirekten Repräsentation gearbeitet: Eine Permutation der Aktivitäten definiert eine Reihenfolge oder Priorisierung, die im Rahmen eines Generierungsschemas die Auswahl einplanbarer Aktivitäten bestimmt. Hierbei werden ein serielles und ein paralleles Generierungsschema verwendet (vgl. z.B. [Koli96; KoHa99]).

Bei dem seriellen Generierungsschema wird schrittweise eine einplanbare Aktivität (mit bereits eingeplanten direkten Vorgänger-Aktivitäten), zu einem für diese Aktivität frühesten zeit- und ressourcenzulässigen Zeitpunkt eingeplant. Hierbei wird in der Reihenfolge gemäß der indirekten Repräsentation mittels einer Permutation vorgegangen, wobei in der Permutation die vorliegenden Vorgänger-Nachfolger-Beziehungen eingehalten sein müssen.

Beim parallelen Generierungsschema wird zeitlich aufsteigend geplant. Zu einem jeweils betrachteten Zeitpunkt wird aus allen Aktivitäten mit bereits abgeschlossenen direkten Vorgänger-Aktivitäten, die ressourcenzulässig einplanbar sind, ausgewählt. Die Auswahl aus einer gegebenenfalls nicht einelementigen Menge einplanbarer Aktivitäten erfolgt hierbei gemäß der Priorisierung entsprechend der Permutation der Aktivitäten als indirekter Repräsentation.

Die mit dem seriellen und parallelen Generierungsschema erzeugten Projektpläne gehören zur Menge der so genannten aktiven bzw. unverzögerten Pläne, wobei die Menge der unverzögerten Pläne eine Teilmenge der Menge der aktiven Pläne ist [SpKD95]. Damit ist der Suchraum, der beim parallelen Generierungsschema zugrunde gelegt wird, in der Regel kleiner als beim seriellen Generierungsschema. Es ist weiterhin zu anmerken, dass aus einer Permutation der Aktivitäten durch ein Generierungsschema zwar ein eindeutiger Projektplan abgeleitet wird, während durchaus verschiedene Permutationen der Aktivitäten zu einem identischen Projektplan führen können.

Im Zusammenhang mit der Vertragsrepräsentation durch eine Permutation der Aktivitäten werden auf Verschiebungen von Aktivitäten basierende Nachbarschaftsrelationen verwendet. Für das parallele Generierungsschema wird schlicht eine zufällige Verschiebung einer Aktivität an eine andere Stelle verwendet, wodurch sich die Priorisierung entsprechend verändert. Das se-

rielle Generierungsschema kann ohne Modifikation nur im Zusammenhang mit einer Permutation angewendet werden, in der Vorgänger-Nachfolger-Beziehungen eingehalten werden. Deshalb wird hierfür nach einer zufälligen Verschiebung einer Aktivität an eine andere Stelle gegebenenfalls eine „Reparatur“ verletzter Vorgänger-Nachfolger-Beziehungen gemäß [FlHi04] angewendet, wobei betroffene Vorgänger und Nachfolger der verschobenen Aktivität wieder vor bzw. hinter die verschobene Aktivität verschoben werden; sofern hierbei die Ausgangspermutation wieder hergestellt wird, wird eine andere zufällige Verschiebung verwendet (gegebenenfalls wiederholt).

4 Experimentelle Untersuchung

Die Implementierung und Untersuchung der oben beschriebenen Verfahren erfolgte mittels C++ und ist teilweise in [Mora05] beschrieben. Hierbei wurde auf dem Metaheuristik-Framework HotFrame [FiVo02] aufgebaut (etwa bezüglich Komponenten zur Repräsentation des Vertragsraums und Nachbarschaftsrelationen). Im Folgenden werden zunächst die verwendeten Probleminstanzen sowie dann die erzielten Rechenergebnisse dargestellt.

4.1 Probleminstanzen

Die Probleminstanzen bauen auf den RCPSP-Benchmark-Datensätzen der PSPLIB auf (für nähere Information sei auf [KoSp96] sowie <http://129.187.106.231/psplib/> verwiesen). Die hier verwendeten 36 Probleminstanzen zeichnen sich durch folgende Merkmale aus:

- Anzahl der Aktivitäten: 30, 60, 120
- Netzwerk-Komplexität (mittlere nicht-redundante Anzahl von Vorgänger-Nachfolger-Beziehungen einer Aktivität): 1,5; 1,8; 2,1
- 4 Ressourcentypen
- Ressourcen-Faktor (mittlere Nachfrage nach Ressourcen durch eine Aktivität): 0,25; 0,5; 0,75; 1
- Ressourcen-Stärke (mittlere Ressourcenverfügbarkeit pro Zeiteinheit): 0,5

Die Probleminstanzen mussten um die Aufteilung von Aktivitäten auf mehrere Akteure sowie um Zahlungswerte je Aktivität ergänzt werden. Hierbei wurde von zwei Akteuren ausgegangen. Die Aktivitäten wurden zufällig auf die beiden Akteure aufgeteilt. Zahlungswerte wurden so

ermittelt, dass Auszahlungen eher am Anfang und Einzahlungen eher am Ende des Projekts anfallen sowie insgesamt je Akteur ein positiver Barwert resultiert. Hierzu wurden je Akteur die Aktivitäten nach ihrer Nummer, die die Position im Rahmen des Projekts grob widerspiegelt, sortiert und dann der ersten Hälfte negative sowie der zweiten Hälfte positive Zahlungswerte zugewiesen. Die Auszahlungen wurden zufällig gleichverteilt aus dem Intervall $[-50; -100]$ ausgewählt. Der doppelte absolute Mittelwert der entsprechenden Auszahlungen wurde als untere Schranke für das Intervall für die zufällig gleichverteilte Auswahl der Einzahlungswerte verwendet; für die obere Schranke wurde 50 zu dieser unteren Schranke addiert. Der Abzinsungsfaktor wird auf 1% je Zeiteinheit festgelegt.

4.2 Ergebnisse

Im Folgenden wird die individuelle Qualität eines Vertrags (Projektplans) für einen Akteur durch die Abweichung von einem approximativ ermittelten Bestwert für einen Projektplan unter Berücksichtigung von allein den Zahlungswerten des entsprechenden Akteurs angegeben. Diese hypothetischen Referenzwerte wurden durch die Anwendung von Simulated Annealing in der in [JAMS89] beschriebenen Form auf die entsprechenden einkriteriellen Probleme ermittelt. Die auf dieser Grundlage angegebenen prozentualen Abweichungen stellen jeweils Mittelwerte einer zehnfach wiederholten Simulation entsprechender Verhandlungsprozesse (mit unterschiedlichen Zufallszahlenfolgen) je Konfiguration dar. Die Berechnungen nehmen jeweils wenige Sekunden auf einem Standard-PC in Anspruch, wobei die Verhandlungen sowohl für 5000 als auch für 15000 Runden simuliert wurden.

Um zu untersuchen, ob und gegebenenfalls in welchem Ausmaß die eingeführten Mindestakzeptanzraten (mit entsprechend kooperativen Spielern) zu besseren Ergebnissen führen als das Zusammentreffen von Spielern, die nur individuell jeweils nicht verschlechternde Vorschläge akzeptieren (gierige Spieler), wurden beide Spielstrategien in unterschiedlichen Kombinationen untersucht. Kooperative Strategien wurden mit Mindestakzeptanzraten von 20% und 4% für die anfängliche bzw. letzte Verhandlungsphase verbunden; diese Parameterwertkombination hat sich bereits für verschiedene Problemstellungen als vorteilhaft herausgestellt hat (vgl. [Fink04; Fink06]).

In Tabelle 1 wird die Anwendung der beiden verwendeten Generierungsschemas im Zusammenhang mit einer homogenen Strategiewahl der Akteure verglichen. Das serielle Generierungsschema führt hier zu besseren Ergebnissen als das parallele Generierungsschema, was sich

mit ähnlichen Erkenntnissen für das klassische RCPSP deckt [KoHa99]. Dementsprechend wird im Folgenden auf die Darstellung von Ergebnissen für das parallele Generierungsschema verzichtet.

		5000 Runden		15.000 Runden	
		Akteur 1	Akteur 2	Akteur 1	Akteur 2
Seriell- es Generie- rungsschema	gierig	6,11 %	5,50 %	5,71 %	5,16 %
	kooperativ	5,36 %	4,56 %	4,69 %	4,23 %
Paralleles Generie- rungsschema	gierig	7,51 %	7,40 %	7,11 %	7,07 %
	kooperativ	7,02 %	6,76 %	6,68 %	6,40 %

Tab. 1: Ergebnisse für unterschiedliche Generierungsschemas

Die Tabellen 2 und 3 zeigen die Ergebnisse beim Aufeinandertreffen zweier gieriger Akteure bzw. zweier kooperativer Akteure für die verschiedenen Problem-Instanzen. Die durchschnittlichen Verhandlungsergebnisse für die kooperative Strategie und die hiermit verbundene Mindestakzeptanzratenregel erscheinen signifikant besser als bei gierigen Agenten. Dies wird durch einen statistischen t-Test mit einer Irrtumswahrscheinlichkeit von 1% belegt.

Bei einem hypothetischen Zusammentreffen zweier Akteure mit unterschiedlichem Verhalten profitiert der gierige Akteur erwartungsgemäß deutlich vom kooperativen Verhalten des anderen Akteurs, wobei sich das Ungleichverhältnis mit steigender Rundenzahl verstärkt. Beispielsweise ergeben sich bei 5000 und 15.000 Runden durchschnittliche Abweichungen von 3,46% (gierig) vs. 8,08% (kooperativ) bzw. 2,58% (gierig) vs. 8,47% (kooperativ).

Die in [Fink04] dargestellte und auch hier zutreffende spieltheoretische Analyse würde bei einer freien Strategiewahl der Akteure zur Kombination zweier gieriger Akteure als eindeutiges Nash-Gleichgewicht führen. Damit würde in gleicher Weise wie beim so genannten „Gefangen-Dilemma“ die einzige dominierte Lösung als das Verhandlungsergebnis resultieren und sich keine effektive Kooperation eigennütziger Akteure herausbilden [Axel84]. Dieses Dilemma wurde durch die oben beschriebene Einführung von Mindestakzeptanzraten aufgelöst, was die Akteure zu einem in diesem Sinne kooperativen Verhalten zwingt und damit beiden Akteuren bessere Resultate erbringt. Die entsprechenden Ergebnisse sind hochwertig, wie ein Vergleich mit den Ergebnissen einer Anwendung von Simulated Annealing für das hypothetische integrierte Gesamtproblem zeigt; es ergeben sich im Mittel Abweichungen der Barwertsummen von weniger als einem Prozent.

Anzahl Aktivitäten	Netzwerk- Komplexität	Ressourcen- Faktor	5000 Runden		15.000 Runden	
			Akteur 1	Akteur 2	Akteur 1	Akteur 2
30	1,5	0,25	0,00%	0,00%	0,00%	0,00%
		0,50	4,28%	5,16%	4,28%	5,16%
		0,75	1,12%	5,63%	1,04%	5,63%
		1,00	1,33%	2,93%	1,33%	2,93%
	1,8	0,25	0,45%	1,61%	0,45%	1,61%
		0,50	0,87%	1,55%	0,81%	1,09%
		0,75	2,18%	4,68%	2,18%	4,68%
		1,00	5,08%	3,48%	5,08%	3,48%
	2,1	0,25	5,28%	1,12%	5,28%	1,12%
		0,50	2,21%	2,80%	2,21%	2,80%
		0,75	0,64%	0,92%	0,64%	0,92%
		1,00	2,60%	3,06%	2,52%	2,95%
60	1,5	0,25	3,26%	1,37%	3,26%	1,37%
		0,50	5,14%	7,25%	4,72%	6,47%
		0,75	12,15%	8,24%	10,98%	7,65%
		1,00	6,35%	6,25%	5,98%	6,04%
	1,8	0,25	2,37%	3,48%	2,37%	3,48%
		0,50	2,37%	5,13%	2,37%	5,13%
		0,75	8,51%	8,25%	8,48%	8,13%
		1,00	10,28%	4,52%	9,64%	4,35%
	2,1	0,25	2,56%	1,98%	2,56%	1,98%
		0,50	2,72%	4,63%	2,60%	4,38%
		0,75	5,15%	7,14%	5,12%	7,09%
		1,00	11,68%	6,48%	11,32%	6,37%
120	1,5	0,25	3,51%	1,74%	3,49%	1,69%
		0,50	9,35%	9,43%	8,13%	8,64%
		0,75	12,90%	8,55%	11,84%	7,86%
		1,00	18,35%	13,04%	17,46%	12,29%
	1,8	0,25	3,36%	1,10%	3,14%	1,06%
		0,50	8,82%	12,06%	6,83%	9,41%
		0,75	8,66%	7,03%	7,24%	6,18%
		1,00	11,07%	12,61%	10,27%	12,08%
	2,1	0,25	6,33%	3,83%	6,01%	3,66%
		0,50	2,94%	3,83%	2,81%	3,62%
		0,75	21,15%	10,10%	19,78%	8,75%
		1,00	14,76%	17,10%	13,51%	15,84%
Durchschnitt			6,11%	5,50%	5,71%	5,16%

Tab. 2: Ergebnisse bei beiderseitig gieriger Strategie

Anzahl Aktivitäten	Netzwerk- Komplexität	Ressourcen- Faktor	5000 Runden		15.000 Runden	
			Akteur 1	Akteur 2	Akteur 1	Akteur 2
30	1,5	0,25	0,00%	0,00%	0,00%	0,00%
		0,50	4,01%	4,49%	4,92%	3,72%
		0,75	1,69%	2,47%	1,68%	2,58%
		1,00	1,39%	2,20%	1,58%	1,77%
	1,8	0,25	1,05%	0,69%	1,49%	0,00%
		0,50	1,50%	0,66%	1,00%	0,75%
		0,75	2,04%	3,81%	1,92%	2,55%
		1,00	4,78%	4,02%	4,26%	3,46%
	2,1	0,25	1,82%	0,39%	1,82%	0,39%
		0,50	1,55%	3,35%	0,14%	4,34%
		0,75	0,62%	0,94%	0,62%	0,93%
		1,00	1,82%	3,34%	1,93%	3,12%
60	1,5	0,25	2,23%	1,83%	1,68%	1,85%
		0,50	6,23%	5,37%	6,33%	5,79%
		0,75	11,00%	7,51%	11,66%	5,13%
		1,00	6,32%	5,03%	4,43%	6,38%
	1,8	0,25	2,17%	2,49%	2,42%	2,20%
		0,50	2,55%	3,59%	2,64%	3,10%
		0,75	6,97%	8,14%	7,27%	6,98%
		1,00	7,85%	4,34%	8,50%	3,09%
	2,1	0,25	2,45%	0,92%	2,57%	0,95%
		0,50	3,23%	2,58%	2,72%	2,57%
		0,75	6,35%	6,04%	5,58%	6,55%
		1,00	11,93%	4,46%	8,49%	7,23%
120	1,5	0,25	3,01%	1,99%	2,54%	2,19%
		0,50	7,92%	10,74%	7,29%	7,25%
		0,75	11,43%	7,11%	8,44%	6,89%
		1,00	16,87%	11,29%	12,49%	12,10%
	1,8	0,25	1,64%	1,23%	1,45%	0,95%
		0,50	5,79%	8,40%	5,83%	5,77%
		0,75	6,74%	6,42%	4,74%	6,29%
		1,00	8,26%	11,44%	7,47%	10,10%
	2,1	0,25	5,32%	3,53%	4,05%	3,31%
		0,50	3,10%	2,92%	3,07%	2,67%
		0,75	16,22%	7,36%	14,25%	7,65%
		1,00	14,97%	12,93%	11,62%	11,57%
Durchschnitt			5,36%	4,56%	4,69%	4,23%

Tab. 3: Ergebnisse bei beiderseitig kooperativer Strategie

Im Folgenden werden die Verhandlungsergebnisse in Abhängigkeit vom Ressourcen-Faktor näher betrachtet. Es ist zu erwarten, dass bei einem höheren Ressourcen-Faktor, welcher mit der Nachfrage nach Ressourcen wächst, zwischen den Akteuren eine stärkere Konkurrenz um individuell günstige Einplanungszeiten im Projektplan entsteht und damit die Abweichungen insgesamt steigen. Dies wird durch die in den Tabellen 4 und 5 dargestellten Ergebnisse, die für den Fall einer beiderseitig kooperativen Strategie bei 5000 Verhandlungsrunden in Abb. 2 visualisiert sind, bestätigt.

Ressourcen-Faktor	5000 Runden		15.000 Runden	
	Akteur 1	Akteur 2	Akteur 1	Akteur 2
0,25	3,01 %	1,80 %	2,95 %	1,77 %
0,50	4,30 %	5,76 %	3,86 %	5,19 %
0,75	8,05 %	6,73 %	7,48 %	6,32 %
1,00	9,06 %	7,72 %	8,57 %	7,37 %

Tab. 4: Ergebnisse in Abhängigkeit vom Ressourcen-Faktor bei beiderseitig gieriger Strategie

Ressourcen-Faktor	5000 Runden		15.000 Runden	
	Akteur 1	Akteur 2	Akteur 1	Akteur 2
0,25	2,19 %	1,45 %	2,00 %	1,32 %
0,50	3,99 %	4,68 %	3,77 %	3,99 %
0,75	7,01 %	5,53 %	6,24 %	5,06 %
1,00	8,24 %	6,56 %	6,75 %	6,54 %

Tab. 5: Ergebnisse in Abhängigkeit vom Ressourcen-Faktor bei beiderseitig kooperativer Strategie

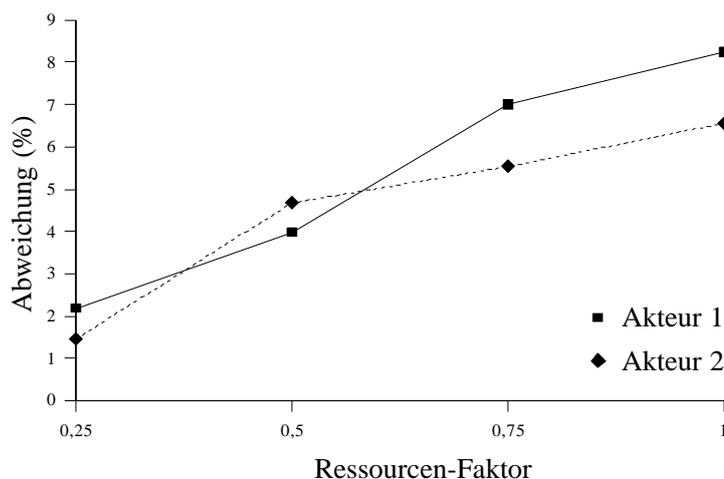


Abb. 2: Ergebnisse in Abhängigkeit vom Ressourcen-Faktor bei beiderseitig kooperativer Strategie

5 Fazit

In diesem Beitrag wurde für eine komplexe Projektplanungsproblemstellung mit eigensinnigen Akteuren und Informationsasymmetrie ein automatischer Koordinationsmechanismus beschrieben und angewendet. Dieser ermöglicht eine ertragsorientierte Ablaufgestaltung von Projekten mit mehreren beteiligten Akteuren. Die experimentellen Ergebnisse belegen die Effektivität des Verfahrens im betrachteten Kontext. Eine Anwendung und Validierung des Mechanismus für Problemstellungen aus dem Supply Chain Management mit auch mehr als zwei Akteuren ist in [Fink06] beschrieben.

Für die betrachtete Projektplanungsproblemstellung bietet es sich an, näher zu untersuchen, in welchem Maße der beschriebene und ausgestaltete Koordinationsmechanismus die vollständige Ausschöpfung theoretisch möglicher, aber gefährdeter Kooperationsvorteile leistet. Hierzu wäre einerseits etwa der effiziente Rand (die Menge aller Pareto-optimalen Ergebnisse) zu ermitteln und eine entsprechende Abweichung zu bewerten. Aufgrund der Tatsache, dass das zugrunde liegende einkriterielle Problem bereits *NP*-schwer ist, erscheinen hierfür nur approximative Vorgehensweisen anwendbar. Andererseits erscheint die Verbesserung der Generierungsschemas im Hinblick auf die Barwertmaximierungszielsetzung denkbar, insofern einzelne Aktivitäten mit unabhängigen Pufferzeiten verbunden sein können, was eine beschränkte individuelle Verschiebung einzelner Aktivitäten ermöglichen kann; allerdings sind hierbei die Wechselwirkungen mehrerer Verschiebungen unter Berücksichtigung der Ressourcenbeschränkung zu berücksichtigen.

Der beschriebene verhandlungsbasierte Koordinationsmechanismus ist prinzipiell für alle verteilten Entscheidungsprobleme anwendbar, bei denen der Verhandlungsgegenstand und damit der entsprechende Suchraum formal definierbar sind. Durch die vollautomatische Umsetzbarkeit des Verfahrens ist dann eine laufzeiteffiziente und damit praxistaugliche Anwendung möglich. Die individuelle Bewertung einzelner Verträge liegt allein bei den beteiligten Akteuren. Dabei erfolgt keine Offenlegung der individuellen Nutzenfunktionen und es wird kein interpersoneller Nutzenvergleich vorausgesetzt. Der Koordinationsmechanismus erzwingt damit unter Berücksichtigung praxisbezogener Randbedingungen nur in geringem Maße eine Offenlegung von Informationen.

Literaturverzeichnis

- [Axel84] *Axelrod, R.*: The Evolution of Cooperation. Basic Books, New York 1984.
- [BDM+99] *Brucker, P.; Drexl, A.; Möhring, R.; Neumann, K.; Pesch, E.*: Ressource-constrained Project Scheduling: Notation, Classification, Models, and Methods. European Journal of Operational Research 112 (1999), S. 3–41.
- [Ehrg05] *Ehrgott, M.*: Multicriteria Optimization, 2. Auflage. Springer, Berlin 2005.
- [Fink04] *Fink, A.*: Supply Chain Coordination by Means of Automated Negotiations. Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS-37). IEEE 2004.
- [Fink06] *Fink, A.*: Supply Chain Coordination by Means of Automated Negotiation Between Autonomous Agents. In: *Chaib-draa, B.; Müller, J.* (Hrsg.): Multiagent-Based Supply Chain Management, Studies in Computational Intelligence Vol. 28. Springer, Berlin 2006, S. 351–372.
- [FiVo02] *Fink, A.; Voß, S.*: HotFrame: A Heuristic Optimization Framework. In: *Voß, S.; Woodruff, D.L.* (Hrsg.): Optimization Software Class Libraries, Kluwer, Boston 2002, S. 81–154.
- [FlHi04] *Fleszar, K.; Hindi, K.S.*: Solving the Resource-constrained Project Scheduling Problem by a Variable Neighbourhood Search. European Journal of Operational Research 155 (2004), S. 402–413.
- [Fren88] *French, S.*: Decision Theory: An Introduction to the Mathematics of Rationality. Ellis Horwood, Chichester 1988.
- [GSST05] *Gandibleux, X.; Sevaux, M.; Sörensen, K.; T'kindt, V.* (Hrsg.): Metaheuristics for Multiobjective Optimisation. Springer, Berlin 2005.

- [HeDD97] *Herroelen, W.S.; Dommelen, P.V.; Demeulemeester, E.L.*: Project Network Models with Discounted Cash Flows – A Guided Tour Through Recent Developments. *European Journal of Operational Research* 100 (1997), S. 3–41.
- [JAMS89] *Johnson, D.S.; Aragon, C.R.; McGeoch, L.A.; and Schevon, C.*: Optimization by Simulated Annealing: An Experimental Evaluation; Part 1, Graph Partitioning. *Operations Research* 37 (1989), S. 865–892.
- [KiGV83] *Kirkpatrick, S.; Gelatt Jr., C.D.; Vecchi, M.P.*: Optimization by Simulated Annealing. *Science* 220 (1983), S. 671–680.
- [KFSB03a] *Klein, M.; Faratin, P.; Sayama, H.; Bar-Yam, Y.*: Negotiating Complex Contracts. *Group Decision and Negotiation* 12 (2003), S. 111–125.
- [KFSB03b] *Klein, M.; Faratin, P.; Sayama, H.; Bar-Yam, Y.*: Protocols for Negotiating Complex Contracts. *IEEE Intelligent Systems* 18 (6) (2003), S. 32–38.
- [Koli96] *Kolisch, R.*: Serial and Parallel Resource-constrained Project Scheduling Methods Revisited: Theory and Computation. *European Journal of Operational Research* 90 (1996), S. 320–333.
- [Koli97] *Kolisch, R.*: Investitionsplanung in Netzwerken. *Zeitschrift für Betriebswirtschaft* 67 (1997), S. 1057–1071.
- [KoHa99] *Kolisch, R.; Hartmann, S.*: Heuristic Algorithms for Solving the Resource-constrained Project Scheduling Problem: Classification and Computational Analyses. In: *Weglarz, J. (Hrsg.): Handbook on Recent Advances in Project Scheduling*. Kluwer, Boston 1999, S. 147–178.
- [KoSp96] *Kolisch, R.; Sprecher, A.*: PSPLIB – A Project Scheduling Library. *European Journal of Operational Research* 96 (1996), S. 205–216.
- [Mert95] *Mertens, P.*: Wirtschaftsinformatik: Von den Moden zum Trend. In: *König, W. (Hrsg.): Wirtschaftsinformatik '95*, Physica, Heidelberg 1995, S. 25–64.

- [Mora05] *Morales, M.*: Anwendung eines automatisierten Verhandlungsmechanismus zur Koordination der Projektplanung autonomer, gewinnmaximierender Beteiligter. Diplomarbeit, Universität Hamburg, 2005.
- [Nash50] *Nash, J.F.*: The Bargaining Problem. *Econometrica* 18 (1950), S. 155–162.
- [Nash53] *Nash, J.F.*: Two-person Cooperative Bargaining Games. *Econometrica* 21 (1953), S. 128–140.
- [Sand99] *Sandholm, T.W.*: Distributed and Rational Decision Making. In: Weiss, G. (Hrsg.): *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge MA 1999, S. 201–258.
- [SpKD95] *Sprecher, A.; Kolisch, R.; Drexel, A.*: Semi-active, Active, and Non-delay Schedules for the Resource-constrained Project Scheduling Problem. *European Journal of Operational Research* 80 (1995), S. 94–102.
- [VaDH01] *Vanhoucke, M.; Demeulemeester, E.; Herroelen, W.*: On Maximizing the Net Present Value of a Project Under Renewable Resource Constraints. *Management Science* 47 (2001), S. 1113–1121.
- [WKWI94] *WKWI (Wissenschaftliche Kommission Wirtschaftsinformatik im Verband der Hochschullehrer für Betriebswirtschaft)*: Profil der Wirtschaftsinformatik. *Wirtschaftsinformatik* 36 (1994), S. 80–81.

Konzeption und Implementierung eines Planungssystems für Schnittstellentests

Franz Rothlauf

Lehrstuhl für ABWL und Wirtschaftsinformatik, Universität Mannheim
rothlauf@uni-mannheim.de

Abstract

Bei der Erstellung von komponentenbasierten, betrieblichen Informationssystemen ist es eine Aufgabe der Qualitätssicherung sicherzustellen, dass die Schnittstellen zwischen den einzelnen Komponenten keine Fehler aufweisen. Der vorliegende Beitrag beschreibt die Konzeption und Umsetzung eines Entscheidungsunterstützungssystems (EUS), welches die Qualitätssicherungsabteilung eines Softwareherstellers bei der Aufstellung eines Zeitplans für die zentrale Durchführung von Schnittstellentests unterstützt. Wichtigstes Element des EUS ist ein metaheuristikbasiertes Planungssystem, welches dem Entscheider qualitativ gute Zeitpläne vorschlägt. Das Planungssystem berücksichtigt sowohl unterschiedliche Planungsziele, als auch relevante Problemrestriktionen durch eine geeignete Wahl der Problemkodierung, einer geschickten Erzeugung des Zeitplans aus der Problemkodierung sowie einer geeigneten Initialisierung.

1 Einleitung

Bei der Erstellung von komponentenbasierten, betrieblichen Informationssystemen stellt sich das Problem, dass im Rahmen der Qualitätssicherung sowohl die Funktionalität der einzelnen Softwarekomponenten, als auch deren Zusammenspiel überprüft werden müssen. Die Funktions- und Qualitätskontrolle der Komponenten wird hierbei in der Regel durch die einzelnen Fachabteilungen selbst durchgeführt und nur die Endkontrolle durch eine zentrale Qualitätssicherungsabteilung vorgenommen. Bei der Überprüfung des korrekten Zusammenspiels sind eine Vielzahl von unterschiedlichen Fachabteilungen involviert, welche durch eine dedizierte Qualitätssicherungsabteilung organisiert und koordiniert werden. Hierbei stellt sich die Aufgabe, dass Schnittstellentests an zentraler Stelle durchgeführt werden und ein entsprechender Zeit-

plan für die Durchführung der Schnittstellentests erstellt werden muss. Das Aufstellen eines Zeitplans ist aufwendig, da eine Menge unterschiedlicher Ziele existieren und zusätzliche Restriktionen bezüglich der zeitlichen Verfügbarkeit von Ressourcen und Personen sowie Unverträglichkeiten zwischen einzelnen Softwarekomponenten berücksichtigt werden müssen.

Das Problem der Erzeugung von Zeitplänen für die Durchführung von Schnittstellentests gehört zur Klasse der resource-constrained project scheduling Probleme [Bruc99] und ist vergleichbar mit Maschinenbelegungs- oder Stundenplanbelegungsproblemen. Derartige Probleme gehören zu den traditionellen Forschungsgebieten der Betriebswirtschaftslehre und es existieren eine Vielzahl von exakten Methoden, welche hierfür optimale Lösungen ermitteln können [Brü95]. Aufgrund der Komplexität der zugrundeliegenden Planungsprobleme sind diese allerdings auf viele realistische Problemstellungen oft nicht anwendbar, so dass in zunehmenden Maße Metaheuristiken zur Lösung komplexer Planungsprobleme eingesetzt werden [Zäpf94, KäTe04]. Mit deren Hilfe können auch manche Planungsprobleme gelöst werden, welche klassischen, exakten Methoden nicht zugänglich sind [SWM95, ScMe00].

Der vorliegende Beitrag stellt die Konzeption und Umsetzung eines EUS für die Planung von Schnittstellentests vor. Das EUS wurde in Zusammenarbeit mit einem großen deutschen Softwarehersteller im Bereich Steuer- und Buchführungssoftware konzipiert und implementiert und wird mittlerweile mit Erfolg in der Qualitätssicherung eingesetzt. Das im EUS integrierte Planungssystem erstellt mit Hilfe einer Metaheuristik Zeitpläne für die Durchführung von Schnittstellentests und berücksichtigt hierbei unterschiedliche Ziele und Nebenbedingungen. Die verschiedenen Ziele und einzuhaltenden Restriktionen werden durch die Konzeption einer geeigneten Methode zur Kodierung und Bewertung von Zeitplänen sowie durch eine geschickte Erzeugung von Startlösungen für die Metaheuristik berücksichtigt.

2 Erstellung von Zeitplänen für Schnittstellentests

Bei der komponentenbasierten Erstellung von Informationssystemen wird die Funktionalität der zu erstellenden Software in einzelne Komponenten zerlegt. Wichtig bei der Entwicklung komponentenbasierter Software ist, dass verbindliche und klar definierte Schnittstellen zwischen den Softwarekomponenten festgelegt werden. Es ist Aufgabe der Qualitätssicherung die Einhaltung der entsprechenden Schnittstellenspezifikationen zu überprüfen und Fehler im Zusammenspiel unterschiedlicher Komponenten aufzudecken.

Bei der vorliegenden Problemstellung wird der Schnittstellentest durch die Qualitätssicherung organisiert und entsprechende Spezialisten und Entwickler aus den einzelnen Fachabteilungen hinzugezogen. Die eigentlichen Tests werden zentral bei der Abteilung für Qualitätssicherung zeitlich parallel auf einer Menge von Testsystemen (ca. 5-8), welche jeweils aus zwei Rechnern bestehen, durchgeführt. Der Test einer Schnittstelle wird immer durch einen oder zwei Ansprechpartner (Tester) aus den jeweiligen Fachabteilungen, welche für die beteiligten Komponenten zuständig sind, durchgeführt. Für die Durchführung der Tests ist ein Zeitplan aufzustellen, welcher festlegt, wann welche Schnittstellentests auf welchen Testsystemen durchgeführt werden sollen. Durch den Zeitplan wird auch bestimmt, welche Ansprechpartner (ca. 20-50 verschiedene) wann welche Schnittstellentests durchführen. Obwohl die einzelnen Schnittstellentests in recht kurzer Zeit durchgeführt werden können (30 bzw. 60 Minuten) ist aufgrund der Vielzahl der zu testenden Schnittstellen (üblicherweise zwischen 300 und 500) der gesamte Zeitaufwand (ca. 4-7 Tage) für die Durchführung aller Tests hoch. Üblicherweise werden die Schnittstellentests erst kurz vor dem nächsten Softwareveröffentlichungstermin durchgeführt, sodass der vorgegebene Zeitrahmen für die Durchführung der Tests beschränkt ist.

2.1. Ziele des Planungsprozesses

Die folgenden konfligierenden Zielsetzungen wurden im Rahmen des Projektes in Zusammenarbeit mit der Qualitätssicherungsabteilung ausgearbeitet.

Zusammenhängende Testzeiträume für einzelne Tester: Bei der Durchführung von Schnittstellentests muss für jede Schnittstelle je ein Ansprechpartner (Tester) für jede der beteiligten Komponenten vor Ort sein. Ziel der Planung soll es sein, dass jeder Tester jeweils eine möglichst große Menge von Tests am Stück, das heißt zeitlich nacheinander, durchführen kann. Die Tests können hierbei auch an unterschiedlichen Testsystemen durchgeführt werden, aber es sollten keine Lücken im Testplan des jeweiligen Ansprechpartners entstehen.

Frühzeitiger Test von Komponenten mit Basisfunktionalität: Die zu testenden Komponenten lassen sich in zwei Gruppen unterteilen: Basiskomponenten realisieren grundlegende Basisfunktionalitäten des Softwaresystems und werden von anderen Komponenten wiederverwendet. Auf diese Basiskomponenten (z. B. Stammdatenverwaltung) setzen dann Zusatzkomponenten (z. B. Auskunftssystem) auf, welche die durch Basiskomponenten zur Verfügung gestellte Funktionalitäten nutzen und zusätzliche Funktionalität zur Verfügung stellen.

Bei der Durchführung der Tests sollen Basiskomponenten vor Zusatzkomponenten getestet werden. Falls Basiskomponenten vorher getestet werden, kann eindeutig bestimmt werden, welche Fehler auf Basiskomponenten und welche auf Zusatzkomponenten zurückzuführen sind.

Minimierung der Installationszeiten: Für die Durchführung von Schnittstellentests ist es notwendig, dass die jeweils zu testenden Softwarekomponenten auf den Testsystemen installiert werden. Üblicherweise werden die Installationen durch die Abteilung Qualitätssicherung vor der Testphase durchgeführt und für die eigentlichen Tests werden die schon installierten Komponentenversionen verwendet. In der Regel müssen allerdings bei der Aufstellung eines Zeitplans zusätzliche Installationszeiten für Komponenten berücksichtigt werden. Beim erstmaligen Test einer Schnittstelle auf einem der vorhandenen Testsysteme muss dann die entsprechende Komponente installiert werden. Die dafür notwendige Zeit muss bei der Erstellung des Zeitplans berücksichtigt werden. Ziel des Planungsprozesses ist es, die Gesamtinstallationszeiten zu minimieren, was z. B. dadurch realisiert wird, dass die entsprechenden Komponenten auf einer möglichst geringen Anzahl an Testsystemen installiert werden.

2.2. Bei der Planung einzuhaltende Nebenbedingungen

Neben der Vorgabe von Planungszielen existieren auch Nebenbedingungen, welche durch ein automatisiertes Planungssystem berücksichtigt werden müssen.

(1) Einhaltung des vorgegebenen Gesamtzeitraums für die Durchführung der Schnittstellentests: Eine wichtige Nebenbedingung bei der Aufstellung eines Zeitplans ist die Einhaltung des vorgegebenen Gesamtzeitraums (in Tagen gemessen), welcher für die Durchführung aller Schnittstellentests angesetzt wird. Üblicherweise werden nur wenige Tage für die Durchführung der Tests eingeplant, da die Tests möglichst schnell abgeschlossen werden sollen, um genügend Zeit für die Behebung der bei den Tests aufgedeckten Probleme zu haben.

(2) Keine gleichzeitige Durchführung von verschiedenen Schnittstellentests auf einem Testsystem: Auf jedem der unterschiedlichen Testsysteme kann immer nur genau eine Schnittstelle getestet werden, und es ist keine gleichzeitige Durchführung von unterschiedlichen Tests auf einem Testsystem möglich. Darüber hinaus sind die einzelnen Schnittstellentests nicht aufteilbar, sondern müssen immer am Stück durchgeführt werden.

(3) Tester können nur jeweils eine Schnittstelle gleichzeitig testen: Es ist nicht möglich, dass ein Tester mehrere Schnittstellen auf unterschiedlichen Testsystemen zur gleichen Zeit testet, sondern jeder Tester beschäftigt sich für die gesamte Dauer eines Tests nur mit dieser einen Schnittstelle. Die Einhaltung dieser Restriktion ist schwierig, da ein Tester i.d.R. für mehrere

Schnittstellen zuständig ist (einzelne Tester sind für bis zu 50 Schnittstellen zuständig) und sichergestellt werden muss, dass nicht zwei von ihm zu testende Schnittstellen zur gleichen Zeit auf unterschiedlichen Testsystemen eingeplant werden.

(4) Zeitrestriktionen für Tester: Manche Tester weisen zeitliche Restriktionen bezüglich ihrer Verfügbarkeit auf und können nicht beliebig für Schnittstellentests eingeplant werden.

(5) Interdependenzen zwischen Testsystem und zu testender Komponente: Die einzelnen Schnittstellentests werden auf unterschiedlichen Testsystemen durchgeführt. Für einen Teil der Schnittstellentests können allerdings nur ausgewählte Testsysteme verwendet werden, da z. B. ein Teil der Softwarekomponenten hohe Anforderungen an die Leistungsfähigkeit der verwendeten Rechner stellen bzw. manche Komponenten auch auf leistungsschwächeren Testsystemen noch zuverlässig und mit akzeptabler Geschwindigkeit funktionieren müssen. Dies führt dazu, dass die zugehörigen Schnittstellen nicht auf beliebigen Testsystemen getestet werden können, sondern Interdependenzen zwischen Testsystem und installierter Komponente bestehen.

(6) Inkompatibilitäten zwischen einzelnen Komponenten: Eine ähnliche Nebenbedingung liegt in der Unverträglichkeit von Komponenten. Im Produktportfolio des Softwareherstellers existieren unterschiedliche Produktlinien, welche aber teilweise zueinander inkompatibel sind. Daher können bestimmte Komponenten nicht gleichzeitig auf dem gleichen Testsystem installiert und getestet werden, sondern Tests von zueinander inkompatiblen Komponenten müssen auf jeweils unterschiedlichen Testsystemen durchgeführt werden.

(7) Zeitliche Restriktionen: Für die einzelnen Tester müssen die vorgegebene Stundenanzahl pro Tag, Mittagspausen sowie Wochenenden und Feiertage berücksichtigt werden.

3 Mögliche Lösungsansätze für das Planungsproblem

Aufgabe des zu erstellenden Planungssystems ist es, den menschlichen Planer bei der Erstellung eines Zeitplans für die Durchführung der Schnittstellentests zu unterstützen. Hierbei ist die Erstellung eines zulässigen Testplans, welcher die meisten der Nebenbedingungen und Restriktionen berücksichtigt, trivial. Ein Testplan kann z.B. dadurch erzeugt werden, dass zufällig alle Schnittstellentests auf den einzelnen Testsystemen eingeplant werden. Anschließend werden iterativ die Nebenbedingungen (2)-(7) berücksichtigt und Schnittstellentests, welche gegen bestimmte Restriktionen verstoßen, einfach am Ende des Planungszeitraums eingeplant.

Das Problem einer derartigen Vorgehensweise ist allerdings, dass der so ermittelte Testplan i.d.R. sowohl gegen Restriktion (1) verstößt als auch die vorgegebenen Planungsziele außer Acht lässt. Bei einem derartigen Zeitplan existieren keine zusammenhängende Testzeiträume für die einzelnen Tester und auch der Gesamtzeitraum für die Durchführung der Tests ist inakzeptabel. In der Praxis einsetzbare und von den Testern akzeptierte Testpläne müssen alle Restriktionen erfüllen, als auch möglichst gut bezüglich der aufgeführten Zielfunktionen sein. Im Folgenden wird ein kurzer Überblick über mögliche Lösungsansätze gegeben.

3.1. Gegenwärtiger Lösungsansatz

Beim Partnerunternehmen wurde die Aufstellung eines Zeitplans bisher durch einen menschlichen Planer mithilfe eines Tabellenkalkulationsprogramms (Excel) durchgeführt. Der Planungsprozess besteht aus der Ermittlung aller zu planenden Schnittstellen und der manuellen Einplanung aller Tests. Der menschliche Planer geht so vor, dass er zu Beginn des Planungsprozesses Komponenten mit vielen Schnittstellen so am Stück eingeplant, dass alle Restriktionen berücksichtigt werden und darüber hinaus der entsprechende Tester möglichst am Stück testet. Anschließend plant er die anderen Schnittstellentests iterativ durch Versuch und Irrtum ein. Hierbei wird immer wieder manuell überprüft, ob nicht eine Restriktion verletzt wird und gleichzeitig versucht, die Planungsziele aus Abschnitt 2.1 möglichst gut zu erreichen.

Das Hauptproblem bei der gegenwärtigen Planungsmethode liegt im hohen Aufwand, welcher mit der Erstellung eines Testplans verbunden ist. Im Moment benötigt ein erfahrener Planer für die Aufstellung eines durchschnittlichen Testplans (ca. 300-500 zu planende Schnittstellen) ca. zwei bis drei Tage, wobei der resultierende Plan in der Regel immer gegen einige wenige Restriktionen verstößt. Da gegenwärtig ca. 10 Testpläne pro Jahr erstellt werden, ist der Aufwand alleine für die Erstellung der entsprechenden Testpläne recht hoch.

3.2. Exakte Verfahren

Es existiert eine Reihe von unterschiedlichen exakten Planungsverfahren, welche nach entsprechenden Anpassungen auch für das hier vorliegende Schnittstellenplanungsproblem eingesetzt werden können [Zäpf94,KäTe04]. Das Schnittstellenplanungsproblem kann hierbei als eine Variante des Maschinenbelegungsproblems formuliert werden. Hierbei müssen Jobs (Schnittstellen) so auf unterschiedlichen Maschinen (Testsystemen) eingeplant werden, dass sowohl die Restriktionen eingehalten als auch die vorgegebenen Zielfunktionen minimiert werden. Als Lösungsverfahren für derartige Probleme können z. B. vollständige Enumeration, Entschei-

dungsbaumsuchverfahren, begrenzte Enumeration, Cut-Verfahren oder Branch-and-Bound Ansätze verwendet werden. Das Hauptproblem bei der Verwendung von exakten Verfahren liegt in der Komplexität und Größe des vorliegenden Schnittstellenplanungsproblems. Da das Problem auf ein resource-constrained project scheduling Problem [Bruc99] zurückgeführt werden kann, können exakte Verfahren oft nur für kleine Probleminstanzen mit akzeptablem Aufwand eingesetzt werden.

3.3. Lösungsheuristiken

Ein anderer Weg wird durch Lösungsheuristiken beschritten. Lösungsheuristiken (oft auch als Heuristiken bezeichnet) sind Verfahren, bei denen explizit auf die Sicherstellung der Ermittlung der optimalen Lösung verzichtet wird. Heuristiken werden oft in Anlehnung an praktisch handhabbare oder plausibel erscheinende Vorgehensweisen so konstruiert, dass sie Planungsprobleme schnell und mit vertretbarer Lösungsqualität lösen. Das Hauptproblem derartiger Verfahren liegt in der Ermittlung geeigneter heuristischer Vorgehensweisen, welche in akzeptabler Zeit eine gute Lösung bezüglich vorgegebener Zielkriterien liefern (vgl. z. B. [Brü95]).

Ein möglicher Ansatzpunkt für die Ableitung einer geeigneten Lösungsheuristik für das Planungsproblem stellt das in Abschnitt 3.1 beschriebene, gegenwärtig eingesetzte, manuelle Planungsverfahren dar. Es wurde folglich im Rahmen der Entwicklung eines Planungssystems versucht, „Daumenregeln“ aus dem Einplanungsverhalten des menschlichen Planers zu extrahieren und diese dann als Heuristiken zu formulieren. Hierbei konnten zwar grundlegende Handlungsstrukturen des menschlichen Planers erkannt werden (z. B. „plane zuerst Komponenten mit vielen Schnittstellen am Stück ein“ oder „erzeuge zuerst zufälligen Plan und löse anschließend Restriktionsverletzungen so auf, dass alle unbelegten Zeitslots, welche größer sind als der zu verschiebende Zeitslot, überprüft werden und die entsprechenden Tests an die Stelle verschoben werden, welche am besten die Zielfunktionen aus Abschnitt 2.1 erfüllt“), es ist aber nicht gelungen, diese so zu beschreiben und miteinander zu kombinieren, dass eine daraus resultierende Heuristik zu zufrieden stellenden Ergebnissen geführt hätte.

3.4. Metaheuristiken

Metaheuristiken (engl. „metaheuristics“) bauen auf Heuristiken auf und verwenden einfache, oft der Natur oder der Physik entlehnte Optimierungs- und Planungsprinzipien („naturinspiriert“). Oft ermöglichen sie eine zügige Modellierung und effiziente Lösung von Planungsproblemen.

Beispiele für Metaheuristiken sind lokale Suchverfahren (z. B. „Simulated Annealing“ oder „Tabu Search“), Evolutionäre Algorithmen [Gold89] oder auch kombinierte Verfahren (z. B. COSA [Wend95]). Derartige Verfahren wurden in den letzten Jahren mit Erfolg in der Produktionsfeinplanung [KuRo95, KSS95, ScMe00, WaZä00], bei der Erstellung von Zeitplänen [AaLe97] sowie anderen Anwendungsbereichen (vgl. [BiNi95, Voss+99, Alan00]) eingesetzt.

Metaheuristiken lösen ein Planungsproblem dadurch, dass sie mit Hilfe von geeigneten Suchoperatoren iterativ neue, zufällige Lösungen generieren und sich bei der Suche durch den Lösungsraum auf gute Lösungen (bezüglich einer vorgegebenen Zielfunktion) fokussieren. Für den Einsatz ist es notwendig, dass vollständige Lösungen des Planungsproblems (z. B. komplette Zeitpläne) so kodiert werden, dass geeignete Suchoperatoren angewandt und die vollständigen Lösungen bezüglich ihrer Qualität bewertet werden können.

Metaheuristiken lassen sich bezüglich der eingesetzten Suchoperatoren sowie der Anzahl der jeweils gleichzeitig betrachteten Lösungen charakterisieren. Eine grobe Charakterisierung unterscheidet zwischen lokalen und rekombinationsbasierten Verfahren. Bei den meisten lokalen Suchverfahren wird in jeder Iteration jeweils nur eine einzige Lösung betrachtet und im nächsten Suchschritt durch eine geringe Veränderung eine neue Lösung mit ähnlichen Eigenschaften erzeugt. Unterschiede zwischen den einzelnen Verfahren liegen in der Definition des entsprechenden Suchoperators und in der Steuerung der Suche. Im Gegensatz hierzu wird bei rekombinationsbasierten Metaheuristiken (z. B. Genetische Algorithmen oder „Scatter Search“) üblicherweise eine Menge von Lösungen (Population) verwendet und neue Lösungen iterativ mit Hilfe von Rekombination erzeugt. Rekombinationsoperatoren erzeugen aus einer Menge von Ausgangslösungen neue Lösungen durch die Kombination von Teilen der Ausgangslösungen.

Metaheuristiken zeichnen sich dadurch aus, dass sie universell einsetzbar sind und deswegen sehr einfach und flexibel auf unterschiedliche Problemstellungen angepasst werden können. Darüber hinaus sind die Anforderungen an die Struktur der zu lösenden Probleme sehr gering (es muss nur möglich sein, die Lösungsqualität von unterschiedlichen Lösungen zu bestimmen) und sie können auch für nichtlineare, nichtdifferenzierbare oder nichtstetige Probleme eingesetzt werden. Als Nachteile sind zu sehen, dass Standard-Metaheuristiken nur für kleine Probleme gute Ergebnisse liefern und größere Probleme nur durch speziell angepasste Verfahren sowie einer problemadäquaten Auswahl und Einstellung der jeweiligen Methodenparameter gelöst werden können. Darüber hinaus ist aufgrund des zufälligen Charakters der Suche nicht

sichergestellt, dass die optimale Lösung gefunden wird und der rechentechnische Aufwand für das Finden von guten Lösungen oft recht hoch.

4 Konzeption eines metaheuristikbasierten Planungssystems

Im Rahmen der Überlegungen zur Auswahl eines geeigneten Ansatzes für die Lösung des Schnittstellenplanungsproblems wurde ein Genetischer Algorithmus (GA) ausgewählt. Darüber hinaus wurde im Rahmen von konzeptionellen Vorstudien versucht, geeignete Heuristiken zur Lösung des Problems zu entwickeln (vgl. Abschnitt 3.3). Die praktische Anwendung zeigte jedoch, dass entweder die Qualität der damit erzeugten Testpläne unzureichend war oder nicht alle Nebenbedingungen ausreichend erfüllt wurden.

4.1. Testplankodierung und Testplanerzeugung

Wie in Abschnitt 3.4 beschrieben, muss beim Einsatz von Metaheuristiken ein gültiger und vollständiger Testplan so kodiert werden, dass Suchoperatoren darauf angewendet werden können (vgl. [Roth06]). Dies bedeutet, dass vollständige Lösungen Informationen darüber enthalten müssen, welche Schnittstellentests zu welchen Zeiten auf welchen Testsystemen durchgeführt werden. Da jeder Softwarekomponente genau ein Tester zugeordnet ist, muss keine zusätzliche Zuordnung zwischen Schnittstelle und Ansprechpartner (Tester) vorgenommen werden und es ist ausreichend, nur die jeweilige Schnittstelle bei der Planung zu betrachten. Für das vorliegende Schnittstellenplanungsproblem wurde eine gewichtete Kodierung [Bean92] zur Kodierung eines Zeitplans eingesetzt. Gewichtete Kodierungen können für Permutationsprobleme verwendet werden und wurden bisher für Maschinenbelegungs-, Fahrzeugrouting-, Kapazitätszuordnungs-, Netzwerk- und „Travelling-Salesperson“-Probleme verwendet. Für einen Überblick über den Einsatz von gewichteten Kodierungen vergleiche [Norm95, RaJu00, Roth06].

Beim Einsatz einer gewichteten Kodierung wird durch einen Vektor r aus n zufälligen reellwertigen Zahlen $r_i \in [0, k[$ mit $r_i \neq r_j \forall i \neq j$ beschrieben, in welcher Reihenfolge die n Schnittstellen auf den k Testsystemen eingeplant werden. Hierbei wird jede einzuplanende Schnittstelle mit einer eindeutigen Nummer $i \in \{0, 1, \dots, n-1\}$ und jedes Testsystem mit einer eindeutigen Nummer $j \in \{0, 1, \dots, k-1\}$ versehen. Die zeitliche Abfolge der Tests wird durch die Position und die relativen Werte der einzelnen Elemente r_i des Vektors bestimmt. Darüber hinaus wird bei

der Erzeugung des Testplans Schnittstelle i jeweils auf dem Testsystem $j = \lfloor r_i \rfloor$ eingeplant. Die Reihenfolge, in der die jeweiligen Tests durchgeführt werden, wird durch die Positionen der einzelnen Elemente des Vektors entsprechend ihres Wertes in absteigender Ordnung festgelegt.

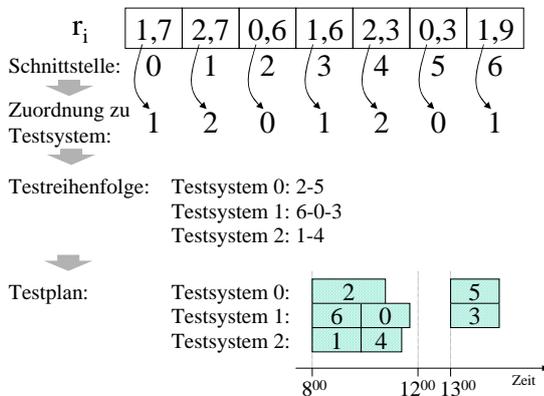


Abbildung 1: Erzeugung eines Zeitplans aus einer kodierten Lösung

Abbildung 1: Erzeugung eines Zeitplans aus einer kodierten Lösung folge 2-5 ($r_2 > r_3$) eingeplant. Auf Testsystem 1 werden die Schnittstellen 0 ($r_0 = 1,7$), 3 ($r_3 = 1,6$) und 6 ($r_6 = 1,9$) in der Reihenfolge 6-0-3 ($r_6 > r_0 > r_3$) getestet. Weiterhin werden auf Testsystem 2 die Schnittstellen 1 ($r_1 = 2,7$) und 4 ($r_4 = 2,3$) in der Reihenfolge 1-4 ($r_1 > r_4$) eingeplant.

Entsprechend dieser Vorgehensweise kann aus jedem gewichteten Vektor r der Länge n mit $r_i \in [0, k[$ eine zulässige Reihenfolge der Schnittstellentests auf den k Testsystemen gewonnen werden. Durch eine derartige Kodierung wird ein Teil der Nebenbedingungen berücksichtigt. Da bei der Verwendung einer gewichteten Kodierung jede Schnittstelle nur ein einziges Mal eingeplant werden kann, wird jeder Schnittstellentest auch genau nur ein einziges Mal durchgeführt. Darüber hinaus kann auf einer Maschine zum gleichen Zeitpunkt auch nie mehr als ein Test durchgeführt werden, sodass automatisch Nebenbedingung (2) eingehalten wird.

Nachdem durch die Kodierung festgelegt worden ist, in welcher Reihenfolge die Tests auf den jeweiligen Testsystemen eingeplant werden, muss in einem zweiten Schritt aus der Reihenfolge ein Zeitplan erstellt werden. Bei dessen Erstellung werden die Restriktionen (7) wie z. B. maximale tägliche Arbeitszeit, Mittagspause oder Wochenenden berücksichtigt (vgl. Abbildung 1). Am Ende des Konstruktionsprozesses steht ein Zeitplan, welcher Nebenbedingungen (2) und (7) berücksichtigt.

4.2. Testplanbewertung

Ein mögliches Beispiel für einen gewichteten Vektor stellt $r = (1,7; 2,7; 0,6; 1,6; 2,3; 0,3; 1,9)$ mit $n = 7$ und $k = 3$ dar (vgl. Abbildung 1). Es werden also sieben Schnittstellen (Vektor hat sieben Elemente) auf drei Testsystemen (die drei möglichen „Vorkommaziffern“ sind 0, 1 oder 2) eingeplant. Im Zeitplan werden daher die Schnittstellen mit der Nummer 2 ($r_2 = 0,6$) und 5 ($r_5 = 0,3$) auf Testsystem 0 in der Reihen-

Beim Einsatz von Metaheuristiken muss die Qualität eines Testplans mithilfe einer Bewertungsfunktion bewertet werden. Aufgrund der vorhandenen unterschiedlichen Ziele liegt ein multikriterielles Problem vor [Ehrg05], bei dem paretooptimale Lösungen bezüglich der vorgegebenen Planungsziele ermittelt werden können. Prinzipiell sind populationsbasierte Metaheuristiken gut für die Lösung von multikriteriellen Problemen geeignet [Deb01], allerdings ist ein größerer Aufwand für die Ermittlung von guten Lösungen notwendig. Da für die Lösung des vorliegenden Planungsproblems keine Menge von paretooptimalen Lösungen benötigt wird, wird daher eine additive Bewertungsfunktion verwendet, bei der ein Zeitplan einzeln bezüglich der vorgegebenen Ziele und Restriktionen bewertet wird und die sich daraus ergebenden Teilbewertungen

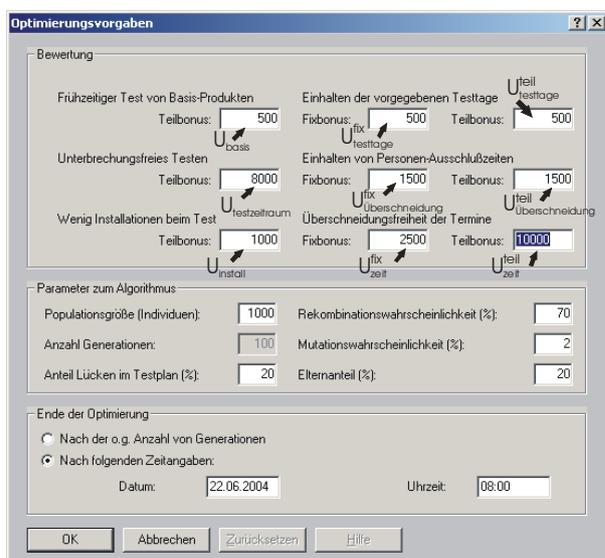


Abbildung 2: Gewichtungen für Zeitplanbewertung und Parameter für GA

gen additiv zusammengefasst werden.

Die Bewertung eines Zeitplans setzt sich somit aus Teilbewertungen zusammen, welche vom Erreichen der unterschiedlichen Planungsziele aus Abschnitt 2.1 und dem Grad der Einhaltung der Nebenbedingungen aus Abschnitt 2.2 abhängen. Die jeweiligen Gewichtungen U für die einzelnen Ziele und Restriktionen werden durch den Entscheider vorgegeben (vgl. Abbildung 2) und beeinflussen die Qualität der durch die Metaheuristik erzeugten Zeitpläne. Je größer

die jeweilige Gewichtung gewählt wird, desto stärker wird dieses Ziel bzw. die Einhaltung einer Restriktion berücksichtigt. Die Bewertung eines Zeitplans hängt von folgenden Kriterien ab:

Zusammenhängende Testzeiträume/Unterbrechungsfreies Testen: Die Bewertung eines Zeitplans wird um $B_{\text{testzeitraum}} = U_{\text{testzeitraum}} (l_{\text{testzeitraum}} / n)$ erhöht, wobei $l_{\text{testzeitraum}}$ die kumulierte Anzahl der Schnittstellen bezeichnet, welche jeweils ohne Unterbrechung durch einen Tester getestet werden können, und n die Gesamtanzahl der einzuplanenden Schnittstellen angibt.

$U_{\text{testzeitraum}}$ kann durch den Entscheider vorgegeben werden (vgl. Abbildung 2).

Frühzeitiger Test von Basiskomponenten: Die Bewertung eines Zeitplans wird um die Teilbewertung $B_{\text{basis}} = U_{\text{basis}} (1/l_{\text{basis}} \sum_{\text{Basiskomp.}} (1 - (pos_i / n)))$ erhöht, wobei l_{basis} die Anzahl der Basiskomponenten und pos_i die Position der Basiskomponente i im Zeitplan bezeichnet.

Minimierung der Installationszeiten: Die Bewertung eines Zeitplans wird um $B_{install} = U_{install} (m_{install} / (n_{install} k))$ erhöht, wobei $m_{install}$ die Anzahl der im Zeitplan geplanten Installationen, $n_{install}$ die Anzahl der notwendigen Installationen und k die Anzahl der Testsysteme bezeichnet.

Analog zu den Zielen wird der Grad der Nichtberücksichtigung von Nebenbedingungen ebenfalls bei der Planbewertung berücksichtigt:

Anzahl benötigter Testtage: Falls die Anzahl der vorgegebenen Testtage eingehalten wird, wird die Bewertung eines Zeitplans um $B_{testtage} = U_{testtage}^{fix}$ erhöht. Ansonsten gilt $B_{testtage} = U_{testtage}^{teil} (2 - (t - t_{min}))$, wobei t_{min} die minimal mögliche Anzahl an Testtagen und t die aus dem Zeitplan resultierende Anzahl an Testtagen bezeichnet.

Gleichzeitiger Test: Falls keine Überschneidungen auftreten, wird die Bewertung eines Zeitplans um $B_{Überschneidung} = U_{Überschneidung}^{fix}$ erhöht. Ansonsten gilt $B_{Überschneidung} = U_{Überschneidung}^{teil} (1 - (n_{ii} / n))$, wobei n_{ii} die Gesamtanzahl der Überschneidungen bezeichnet.

Zeitrestriktionen für Tester: Falls keine Zeitrestriktionen verletzt werden, gilt $B_{Zeit} = U_{Zeit}^{fix}$. Ansonsten ergibt sich $B_{Zeit} = U_{Zeit}^{teil} (1 - (n_z / m_z))$, wobei n_z die Anzahl der Schnittstellen im Zeitplan bezeichnet, bei denen Zeitrestriktionen verletzt werden, und m_z die Gesamtanzahl der Schnittstellen darstellt, an denen Tester mit Zeitrestriktionen beteiligt sind.

Die Gesamtbewertung für einen Zeitplan berechnet sich somit als $B_{gesamt} = B_{testzeitraum} + B_{basis} + B_{install} + B_{testtage} + B_{Überschneidung} + B_{Zeit}$. Durch die Wahl entsprechender U werden die einzelnen Ziele und Restriktionen unterschiedlich stark gewichtet und dementsprechend mögliche Lösungen (Zeitpläne) unterschiedlich gut bewertet.

4.3. Initialisierung von Startlösungen

Bei der Verwendung von Metaheuristiken werden die Startlösungen üblicherweise zufällig erzeugt. Auch im vorliegenden Fall wird eine Menge von Zeitplänen dadurch zufällig erzeugt, dass unterschiedliche Vektoren r mit jeweils zufälligen Elementen $r_i \in [0, k[$ mit $r_i \neq r_j, \forall i \neq j$ ermittelt werden. Darüber hinaus können bei einer geschickten Initialisierung sowohl vorhandenes Vorwissen über Eigenschaften von qualitativ guten Lösungen als auch bestehende Nebenbedingungen berücksichtigt werden. Von beiden Möglichkeiten wurde bei der Lösung des Schnittstellenplanungsproblems Gebrauch gemacht.

Bei der Untersuchung des Verhaltens von menschlichen Planern (vgl. Abschnitt 3.1) hat sich gezeigt, dass bei guten Lösungen Tester immer eine große Menge an Schnittstellen ohne Unterbrechungen testen und nur geringe Lücken im Zeitplan auftreten. Dieses Vorwissen über die Struktur von guten Zeitplänen wurde bei der Initialisierung der Startpopulation berücksichtigt und Tester mit vielen Schnittstellen jeweils ohne zeitliche Unterbrechungen im Zeitplan eingeplant. Dies wurde so realisiert, dass nach der zufälligen Initialisierung alle Schnittstellen ermittelt wurden, welche durch einen bestimmten Tester getestet werden. Anschließend wurden die den einzelnen Schnittstellen zugeordneten Elemente r_i im kodierten Zeitplan so gewählt, dass diese untereinander nur geringe Unterschiede aufweisen (jeweils nur um 0.001 unterschiedlich). Dadurch, dass alle Schnittstellen eines Testers sehr ähnliche Gewichte r_i im Vektor r aufweisen, werden sie bei der Testplanerzeugung nacheinander im Zeitplan eingeplant und der entsprechende Tester kann ohne Unterbrechungen testen. Entsprechend dieser Vorgehensweise wurden alle Schnittstellen der zehn Tester mit der größten Anzahl an zu testenden Schnittstellen modifiziert. Da allerdings jeder Schnittstelle zwei verschiedene Tester zugeordnet sind, kann dieses Vorgehen nur für eine kleine Anzahl an Testern (jeweils mit der höchsten Anzahl an zu testenden Schnittstellen) eingesetzt werden und darüber hinaus können bei diesem Vorgehen auch schon modifizierte Schnittstellen abermals verändert werden.

Neben der Berücksichtigung von Vorwissen über die Struktur von guten Lösungen werden bei der Initialisierung auch schon die Nebenbedingungen (5) und (6) berücksichtigt. Zur Erfüllung dieser Nebenbedingungen ist es notwendig, dass manche Schnittstellen nur auf bestimmten Testsystemen getestet werden. Daher werden nach dem zufälligen Erzeugen der Startlösungen und der stückweisen Einplanung von Schnittstellen (vgl. vorherigen Absatz) in einem nächsten Schritt alle Schnittstellen betrachtet, welche Nebenbedingungen (5) oder (6) verletzen. Bei allen derartigen Schnittstellen i wird das entsprechende Element r_i des Lösungsvektors so abgeändert (die Vorkommastelle modifiziert), dass keine Verstöße gegen die zwei Nebenbedingungen mehr auftreten.

Ursprünglich ist im Rahmen des Projektes davon ausgegangen worden, dass es möglich ist, einen lückenlosen Testplan zu finden, bei dem keine Verletzungen von Nebenbedingungen auftreten. Es hat sich allerdings bei der Realisierung des Planungssystems schnell gezeigt, dass die Anzahl der Restriktionen so hoch ist, dass ein lückenloser Testplan nicht möglich ist sondern - ähnlich wie bei manuell erstellen Testplänen - zusätzliche Lücken im Testplan eingeführt werden müssen. Daher wurde zu den schon vorhandenen Schnittstellen zusätzliche „Leerschnittstel-

len" hinzugefügt (üblicherweise ca. 15-20%), welche analog zu den real existierenden Schnittstellen behandelt werden, aber keine Restriktionen verletzen und auch keinen Ansprechpartner besitzen. Dadurch erhöht sich die Anzahl der zu planenden Schnittstellen (real existierende Schnittstellen plus Lücken im Testplan) und die Länge n der kodierten Lösungen muss entsprechend erhöht werden.

4.4. Suchoperatoren und Steuerung der Suche

Als Metaheuristik zur Steuerung der Suche wurde ein GA [Gold89] gewählt. GA wenden Suchoperatoren (Rekombination und Mutation) auf eine Menge (Population) von Problemlösungen (Individuen) über mehrere Iterationen (Generationen) an. Die Verwendung von Rekombinations- und Mutationsoperatoren führt hierbei zu einer höheren Diversität der Lösungen, wohin gehend Selektion zu einer Intensivierung der Suche führt. GA wurden schon mit Erfolg für ähnliche Planungsprobleme eingesetzt [KSS95, KuRo95, WaZä00] und sind einfach anwend-

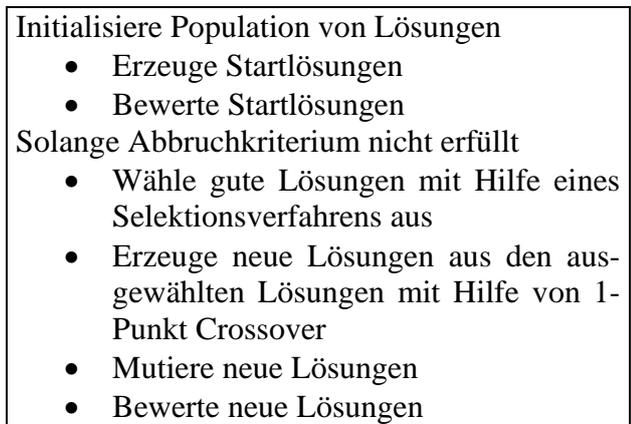


Abbildung 3: Ablaufschema eines GA

bar. Das Ablaufschema des verwendeten GA ist in Abbildung 3 dargestellt. Als Abbruchkriterium wurde eine maximale Anzahl an Suchschritten bzw. eine maximal zur Verfügung stehende Rechenzeit gewählt (vgl. hierzu Abbildung 2).

Als Rekombinationsverfahren wurde 1-Punkt Crossover mit Wahrscheinlichkeit P_{cross} gewählt. Als Mutationsoperator wird eine zufällige Veränderung von r_i durchgeführt. Das bedeutet also, dass die Einplanung (sowohl Testsystem, als auch Reihenfolge) der Schnittstelle i zufällig verändert wird. Bei der Durchführung einer Mutation wird sichergestellt, dass Restriktionen (5) und (6) berücksichtigt werden. Für den Ablauf des GA wurde als Wahrscheinlichkeit für Crossover $P_{cross} \approx 0,8$ und für Mutation $P_{mut} \approx 2/n$ gewählt. Als Selektionsverfahren wurde eine $\mu + \lambda$ -Strategie gewählt (üblicherweise $\mu = 200$ und $\lambda = 800$). Dies bedeutet, dass aus μ Ausgangslösungen jeweils λ neue Lösungen erzeugt werden. Anschließend werden aus den $\mu + \lambda$ Lösungen dann wieder die μ besten Lösungen ausgewählt.

5 Einsatz und Praxiserfahrungen

Das im vorherigen Abschnitt beschriebene Konzept eines Planungssystems wurde in enger Zusammenarbeit mit der Qualitätssicherungsabteilung konzipiert und vollständig innerhalb des Unternehmens mit Hilfe von Visual Basic implementiert. Visual Basic wurde aus Gründen der Homogenität zu den schon vorhandenen Anwendungen in der Qualitätssicherung eingesetzt. Abbildung 5 zeigt den Hauptdialog der entstandenen Anwendung.

Neben dem eigentlichen Planungssystem zur Ermittlung geeigneter Zeitpläne wurden zusätzliche Funktionalitäten im Rahmen des EUS realisiert. So werden die für die Planung jeweils benötigten Daten automatisch aus den entsprechenden Datenbanken (MS-Access) ausgelesen, dargestellt und verwaltet. Weiterhin kann der durch das metaheuristikbasierte Planungssystem ermittelte Testplan durch den Entscheider noch manuell verändert werden und basierend auf dem endgültigen Testplan sowohl automatisch Einladungen an die jeweiligen Tester verschickt als auch Testpläne für die Durchführung der Tests ausgedruckt werden.

Aufgrund der Komplexität des Planungsproblems und der großen Menge an Nebenbedingungen sind die Rechenzeiten für die Ermittlung von guten Zeitplänen hoch. Im praktischen Einsatz werden bei einer bevorstehenden Qualitätssicherungsphase üblicherweise zwei Planungsläufe auf zwei unterschiedlichen Arbeitsplatzrechnern über Nacht (ca. 8-10 Std.) durchgeführt. Die während dieser Zeit ermittelten Zeitpläne entsprechen den Anforderungen des Planers und es werden in den allermeisten Fällen nur minimale Änderungen am so ermittelten Zeitplan vorgenommen. Ursprünglich wurde geplant, Laufzeitverbesserungen durch genauere Untersuchungen der jeweils am besten geeigneten GA-Parameter zu versuchen und darüber hinaus die Implementierung (Visual Basic mit vielen Datenbankzugriffen) der Verfahren effizienter zu gestalten. Da allerdings die gegenwärtigen Rechenzeiten sowohl aus der Sicht des Unternehmens als auch aus Sicht der Qualitätssicherungsabteilung unproblematisch sind, wurde darauf verzichtet.

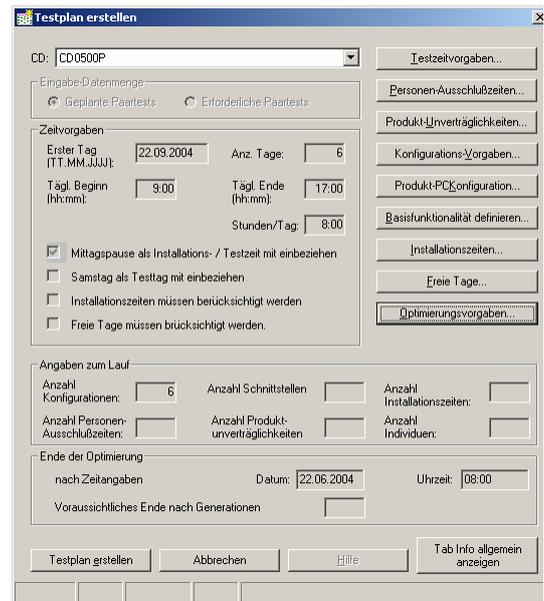


Abbildung 4: Hauptdialog Planungswerkzeug

6 Zusammenfassung

Bei der Erstellung von komponentenbasierten Informationssystemen muss im Rahmen der Qualitätssicherung nicht nur die Funktionalität der einzelnen Komponenten sondern auch das korrekte Zusammenspiel unterschiedlicher Komponenten überprüft werden. Hierzu müssen Schnittstellentests an zentraler Stelle durchgeführt und ein entsprechender Zeitplan für die Durchführung dieser Tests aufgestellt werden. Das Aufstellen eines derartigen Zeitplans ist aufgrund der Komplexität des Planungsproblems und einer Vielzahl von Zielkriterien und Restriktionen sehr aufwendig und wurde bisher durch einen menschlichen Planer durchgeführt.

Das Ziel des Beitrags ist die Beschreibung der Konzeption und Umsetzung eines Entscheidungsunterstützungssystems, welches den menschlichen Planer bei der Planung von Schnittstellentests in der Qualitätssicherung unterstützt. Wichtigstes Element des EUS ist ein automatisiertes Planungssystem, welches Zeitpläne für die Durchführung der Schnittstellentests erzeugt. Im Rahmen des Beitrags wird zuerst eine genaue Beschreibung des zu lösenden Planungsproblems vorgenommen. Es ist ein Zeitplan für die Durchführung von Schnittstellentests zu ermitteln, welcher festlegt wann und auf welchen Testsystemen die einzelnen Schnittstellen getestet werden sollen. Es wird aufgezeigt, welche Planungsziele existieren und welche Nebenbedingungen dabei berücksichtigt werden müssen. Nach der Diskussion von möglichen Lösungsansätzen für das Problem wird ein metaheuristikbasiertes Planungssystem zur Erstellung von Zeitplänen entwickelt. Dieses wurde in enger Zusammenarbeit mit dem Partnerunternehmen konzipiert und umgesetzt. Bei der Konzeption des Planungsverfahrens sind die kritischen Erfolgsfaktoren die Kodierung und Erzeugung von Zeitplänen, die Bewertung der Qualität von Testplänen und die Erzeugung von Startlösungen. Als Metaheuristik für die automatisierte Zeitplanerzeugung wird ein GA eingesetzt. Das resultierende Planungssystem wurde in ein EUS mit zusätzlichen Funktionalitäten zur Verwaltung und Durchführung von Schnittstellentests integriert und wird im Rahmen der Qualitätssicherung mit Erfolg eingesetzt.

Literaturverzeichnis

- [AaLE97] Aarts, E., Lenstra, J. K.. Local Search in Combinatorial Optimization. John Wiley & Sons; New York, 1997.

- [Alan00] Alander, J. T.: Indexed bibliography of genetic algorithms in economics. University of Vaasa, Department of Information Technology and Production Economics, 94-1-ECO, 2000.
- [Bean92] Bean, J. C.: Genetics and random keys for sequencing and optimization (Technical Report 92-43). Ann Arbor, MI: Department of Industrial and Operations Engineering, University of Michigan, 1992.
- [BiNi95] Biethahn, J.; Nissen, V.: Evolutionary Algorithms in Management Applications. Springer-Verlag, Berlin, 1995.
- [Bruc99] Brucker, P.; Drexl, A.; Möhring, R.; Neumann, K.; Pesch, E.: Resource-constrained project scheduling: Notation, classification, models, and methods, European Journal of Operational Research 112, 1999, S. 3-41.
- [Brü95] Brüggemann, W.: Ausgewählte Probleme der Produktionsplanung: Modellierung, Komplexität und neuere Lösungsmöglichkeiten. Physica, Heidelberg, 1995.
- [Deb01] Deb, K.: Multi-Objective Optimization using Evolutionary Algorithms. John Wiley & Sons, Chichester, UK, 2001.
- [Ehrg05] Ehrgott, M.: Multicriteria Optimization. Springer, Berlin, 2005.
- [Gold89] Goldberg, D. E.: Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading, 1989.
- [KäTe04] Käschel, J.; Teich, T.: Produktionswirtschaft Band 1: Grundlagen, Produktionsplanung und -steuerung. GUC Verlag, Chemnitz, 2004
- [KSS95] Kurbel, K.; Schneider, B.; Singh, K.: Parallelization of Hybrid Simulated Annealing and Genetic Algorithm for Short-term Production Scheduling. In Zhong, E. (Hrsg.): Proc. of the Int. Symp. on Intelligence, Knowledge and Integration for Manufacturing, S. 321-326, Nanjing, China, 1995.

- [KuRo95] Kurbel, K.; Rohmann, T.: Ein Vergleich von Verfahren zur Maschinenbelegungsplanung: Simulated Annealing, Genetische Algorithmen und mathematische Optimierung, *Wirtschaftsinformatik* 36(6), 1995, S. 581-593.
- [Norm95] Norman, B. A.: Scheduling using the random keys genetic algorithm. Unveröffentlichte Dissertation, University of Michigan. Ann Arbor, Michigan, 1995.
- [RaJu00] Raidl, G. R.; Julstrom, B. A.: A weighted coding in a genetic algorithm for the degree-constrained minimum spanning tree problem. In *Proceedings of the 2000 ACM Symposium on Applied Computing*. ACM Press, 2000, S. 440–445.
- [Roth06] Rothlauf, F.: Representations for genetic and evolutionary algorithms. 2. Auflage, Heidelberg, Springer, 2006.
- [SWM95] Schultz, J.; Weigelt, M.; Mertens, P.: Verfahren für die rechnergestützte Produktionsfeinplanung – ein Überblick, *Wirtschaftsinformatik* 37(6), 1995, S. 594-608.
- [ScMe00] Schultz, J.; Mertens, P.: Untersuchung wissensbasierter und weiterer ausgewählter Ansätze zur Unterstützung der Produktionsfeinplanung - ein Methodenvergleich, *Wirtschaftsinformatik* 42(1), 2000, S. 56-65.
- [Voss+99] Voß, S.; Martello, S.; Osman, I. H.; Roucairol C.: *Advances and Trends in Local Search Paradigms for Optimization*, Kluwer, Boston, 1999.
- [WaZä00] Wasner, M.; Zäpfel, G.: A heuristic solution concept for a generalized machine sequencing problem with an application to radiator manufacturing, *International Journal of Production Economics* 68, 2000, S. 199-213.
- [Wend95] Wendt, O.: Tourenplanung durch Einsatz naturanaloger Verfahren. Promotion. Wiesbaden, Gabler, 1995.
- [Zäpf94] Zäpfel, G.: Entwicklungsstand und –tendenzen in PPS-Systemen. In: *Handbuch Produktionsmanagement*. Hrsg.: Corsten, H., Gabler, Wiesbaden, 1994, S. 719-745.

Water Contamination Detection

Armin Fügenschuh

Fachbereich Mathematik
Technische Universität Darmstadt
64289 Darmstadt
fuegenschuh@mathematik.tu-darmstadt.de

Simone Göttlich, Michael Herty

Fachbereich Mathematik
Technische Universität Kaiserslautern
67653 Kaiserslautern
{goettlich,herty}@mathematik.uni-kl.de

Abstract

For the detection of contaminations in urban water supply networks we introduce a continuous optimal control model governed by partial differential equations. We derive a linear mixed-integer model by discretization of the dynamics of the partial differential equations and by approximations to the cost functional. Finally, we present numerical results for artificial and real-world networks.

1 Introduction

Since the early days of Newton and Leibnitz, phenomena involving physical laws such as growth and decay, transport of mass, or conservation of energy are described by differential equations. In our days, continuous models based on partial differential equations (PDEs) are used in various areas of applications, for example, in the simulation of production processes, the understanding of traffic flow in street networks, or the simulation and control of the energy transport from power plants to customers. PDEs provide the finest level of description for many

physical and economical processes. In particular, for simulations of large quantities, only the continuous formulations of the PDE models can provide an accurate description of the underlying physical process. Usually, these models rely on quantities such as density (parts per length) and flux (parts per time unit), where the involved parts are not considered individually, but as a continuum. Moreover, we distinguish between simulation and optimization in the following sense. The *simulation* of a physical process is given by a solution to the PDE model under some fixed parameters, e.g., workload of machines, state of compressors in gas networks or demands in water pipeline systems. Thus, simulation means just solving the PDE model, which in itself can be difficult and time-consuming. On the other hand, the *optimization* of processes which are described by a continuous PDE model tries to identify an optimal set of parameters or optimal states in the system, with respect to an objective function or functional. There are now several approaches to achieve this task. The perhaps most simple one is the so-called black-box approach, which is applicable to any simulation process. It consists in successively simulating the PDE model and afterwards gradually changing the input parameters until an “optimal” solution is found. Of course, this heuristic only leads to local optimal solutions without any solution guarantee. Additionally, this approach can be very time consuming in practice.

The new aspect of our work is to introduce a relationship between continuous models and mixed integer programming (MIP) models. The advantage of this MIP approach to PDE models is twofold. First, in many cases MIPs can be solved even for large scale instances in reasonable time by state-of-the-art numerical solvers. Second, the solutions come with a quality warranty, that is, either optimality is proven or an estimation of the optimality gap for the best-known solution is returned. Linear mixed-integer programming can be applied in many ways to continuous PDE models. Below we present a case study related to water quality management. There, we derive a simple PDE model including all major effects and achieve a linear mixed-integer reformulation by a discretization of the PDE model.

The article is organized as follows. We introduce to water quality management in Section 2. We present in Section 3 the continuous water contamination model and in Section 4 the corresponding MIP model. Numerical results for the MIP model are given in Section 5. In Section 6 other areas of application are described. We end up with conclusions in Section 7.

2 Water quality management

Water quality management has many facets. Here we consider the transport of a containment in water networks. Water networks can be understood in a broad sense ranging from municipal water networks supplying a major city up to networks of rivers. Since water is the source of life, the control of water quality has always been a topic of interest. Possible threats to water quality are intentional or accidental contaminations by industry, farms or individual persons. We discuss the problem of identifying sources of containments from given measurements. The networks under investigation are of large scale, which makes it impossible to prevent containments by physical security. We also assume that the measure stations are expensive, such that only a small number of sensors exists throughout the network, which provide information on possible containments. Based on this information we want to identify the origin of the containment. A fast and reliable determination of the sources allows to start activities to prevent further containment. To this end, we present a model for simulating the spreading of the containment within the network, and extend it to an optimization problem for identifying the sources' locations.

The literature on propagation of water and containments through pipes and networks is rich and mathematical approaches for simulation can be found, for example, in [GuLe2003], [GLS2003], [RBA1993], and the references therein. The first researchers who studied the inverse problem, i.e., finding the sources and transient inflow profiles that correspond to given measurements, were Laird, Biegler, and van Bloemen Waanders (see [LBV2006a], [LBV2006b], [LBVB2005]). They formulate the problem as a continuous optimization problem which is similar to our approach. The main difference to the approach presented in the sequel is that they solve the problem using non-linear optimization techniques. Since the resulting non-linear programming problems are of enormous size, only small instances can be solved to optimality.

3 The Continuous PDE Model

Before we introduce the source detection problem, we provide a model for transport of containment inside the network, similar to the model presented in [LBVB2005].

3.1 Simulation

We model the water network as a finite graph $(\mathcal{V}, \mathcal{A})$ with node set \mathcal{V} and arc set \mathcal{A} . We cannot consider each atom of the containment individually, since there are too many of them. Hence it is natural to analyze the flux f^j of the contamination on arc j . *Flux* means the amount of containment per unit time. The flux depends on the *density* (or *concentration*) c^j of the containment, i.e., the amount of containment per unit length. We assume that the containment is totally dissolved in water. If u^j denotes the *velocity* (or *speed*) of water in arc j (in length per time), then the flux of containment is given by $u^j \cdot c^j$. Since the whole water network system is dynamic, the density itself is not a constant, but depending on and changing with the time t . It is also depending on the space, because the density of the containment can be different in different parts of the network (it can even vary within one arc). If we model the arcs as one-dimensional (i.e., the atoms of the containment cannot overtake each other), then a single coordinate $x \in [a^j, b^j]$ is sufficient to uniquely determine every location within the arc. Hence we have $c^j = c^j(x, t)$ and $u^j = u^j(x, t)$, and thus

$$f^j(c^j(x, t)) = u^j(x, t) \cdot c^j(x, t). \quad (1)$$

$c^j(x, t)$ describes the concentration of the containment at position x at time t , and $u^j(x, t)$ describes the velocity, also depending on the position and the time. Note that the signum of $u^j(x, t)$ indicates the direction of the flow. For $u^j(x, t) > 0$ the flow is in direction of the arc, for $u^j(x, t) < 0$ the water flows in the opposite direction. We assume that for every time step t we either have $u^j(x, t) \geq 0$ or $u^j(x, t) \leq 0$, that means, the direction of the flow does not change within the arc.

At time t the amount of containment in a section of some arc j between positions x_1 and x_2 is given by

$$\int_{x_1}^{x_2} c^j(x, t) dx = \text{amount of containment between } x_1 \text{ and } x_2 \text{ at time } t. \quad (2)$$

In the same way, we can express how many containment flows through position x during a time interval from t_1 to t_2 :

$$\int_{t_1}^{t_2} f^j(c^j(x, t)) dt = \text{amount of containment passing position } x \text{ during } t_1 \text{ and } t_2. \quad (3)$$

Now we have a *mass balance*, that is, the amount of containment on the arc between x_1 and x_2 at time t_2 equals the amount of containment in this section at time t_1 plus the *inflow* at x_1 minus the *outflow* at x_2 during the time interval from t_1 to t_2 . We assume that no water is lost or produced during the transport. Expressed as an equation, the material balance can be stated as

$$\int_{x_1}^{x_2} c^j(x, t_2) dx = \int_{x_1}^{x_2} c^j(x, t_1) dx + \int_{t_1}^{t_2} f^j(c^j(x_1, t)) dt - \int_{t_1}^{t_2} f^j(c^j(x_2, t)) dt. \quad (4)$$

From the *fundamental theorem of calculus* we obtain

$$c^j(x, t_2) - c^j(x, t_1) = \int_{t_1}^{t_2} \frac{\partial}{\partial t} c^j(x, t) dt, \quad (5)$$

$$f^j(c^j(x_1, t)) - f^j(c^j(x_2, t)) = - \int_{x_1}^{x_2} \frac{\partial}{\partial x} f^j(c^j(x, t)) dx, \quad (6)$$

which implies

$$\int_{x_1}^{x_2} c^j(x, t_2) dx - \int_{x_1}^{x_2} c^j(x, t_1) dx = \int_{x_1}^{x_2} \int_{t_1}^{t_2} \frac{\partial}{\partial t} c^j(x, t) dt dx, \quad (7)$$

$$\int_{t_1}^{t_2} f^j(c^j(x_1, t)) dt - \int_{t_1}^{t_2} f^j(c^j(x_2, t)) dt = - \int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial}{\partial x} f^j(c^j(x, t)) dx dt. \quad (8)$$

Combining these results with the material balance equation (4) yields

$$\int_{x_1}^{x_2} \int_{t_1}^{t_2} \frac{\partial}{\partial t} c^j(x, t) dt dx + \int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial}{\partial x} f^j(c^j(x, t)) dx dt = 0. \quad (9)$$

If we assume that equality (9) holds for every segment x_1, x_2 in the processor and for each time interval t_1, t_2 , and if the function $c^j(x, t)$ and its partial derivatives of order one are continuous functions, then from an elementary integration property of continuous functions we obtain that

$$\frac{\partial}{\partial t} c^j(x, t) + \frac{\partial}{\partial x} f^j(c^j(x, t)) = 0. \quad (10)$$

Moreover, the containment can react with the surrounding according to some chemical laws expressed for the sake of simplicity in an additional equation as

$$\frac{\partial}{\partial t} c^j(x, t) = R^j(c^j(x, t)), \quad (11)$$

where R^j is a function depending only on the density of the containment and the arc (i.e., the size of the water pipe's cross section). In the case under consideration we use a decay of containment over time and, to be more precise, we assume that

$$\frac{\partial}{\partial t} c^j(x, t) = -r c^j(x, t), \quad (12)$$

for some constant $r \ll 1$.

It remains to prescribe the coupling of the network pipes at nodes $v \in \mathcal{V}$. Denote by A^j the cross section area of the water pipe. Denote by

$$\delta^+(v, t) := \{(w, v) \in \mathcal{A} : u^{(w,v)}(x, t) > 0\} \cup \{(v, w) \in \mathcal{A} : u^{(v,w)}(x, t) < 0\} \quad (13)$$

the set of incoming arcs to v , and by

$$\delta^-(v, t) := \{(v, w) \in \mathcal{A} : u^{(v,w)}(x, t) > 0\} \cup \{(w, v) \in \mathcal{A} : u^{(w,v)}(x, t) < 0\} \quad (14)$$

the set of outgoing arcs. By $e^j(t)$ we denote the coordinate of the arc's outflow end, that is, $e^j(t) := b^j$ for $u^j(x, t) > 0$ and $e^j(t) := a^j$ for $u^j(x, t) < 0$. The amount of water leaving the arc is proportional to the cross section area and the velocity, that is, it is proportional to $A^j |u^j(e^j(t), t)|$. We assume that at each node v the total incoming flux

$$\sum_{j \in \delta^+(v, t)} A^j |u^j(e^j(t), t)| \quad (15)$$

is distributed proportional to the ratio of the outgoing fluxes

$$\frac{A^i |u^i(e^i(t), t)|}{\sum_{j \in \delta^-(v, t)} A^j |u^j(e^j(t), t)|}, \quad \forall i \in \delta^-(v, t). \quad (16)$$

That is, in the simplest possible case of water pipes of the same diameter and having the same velocity, we would observe a distribution with the same flow on each outgoing pipe. Since the containment is dissolved in the water, it is natural to assume that the concentration traveling along with the water flow is distributed exactly as the water flow itself. At the node $v \in \mathcal{V}$ we prescribe the inflow of the containment by functions $q^v(t)$. Hence, we obtain

$$c^i(e^i(t), t) = \frac{q^v(t) + \sum_{j \in \delta^+(v, t)} A^j |u^j(e^j(t), t)| c^j(e^j(t), t)}{\sum_{j \in \delta^-(v, t)} A^j |u^j(e^j(t), t)|}, \quad \forall i \in \delta^-(v, t). \quad (17)$$

This setting guarantees conservation of the containment through nodes, because adding all incoming arcs $i \in \delta^-(v, t)$ yields the flow conservation condition

$$q^v(t) + \sum_{j \in \delta^+(v, t)} A^j |u_j(e^j(t), t)| c^j(e^j(t), t) = \sum_{j \in \delta^-(v, t)} A^j |u^j(e^j(t), t)| c^j(e^j(t), t). \quad (18)$$

Finally, we assume that initially at $t = 0$, no containment is present on the arcs of the network,

$$c^j(x, 0) = 0. \quad (19)$$

3.2 Optimization

For the presented model we formulate the source inversion problem. Given measurements of containment $\bar{c}^j(x, t)$ for $j \in \mathcal{A}_{meas}$ we try to identify sources q^v for $v \in \mathcal{V}$ such that the time evolved concentrations c^j coincide with the measurements on arcs $j \in \mathcal{A}_{meas}$. Moreover, we identify the time-evolution of these sources, i.e., we are looking for functions $t \rightarrow q^v(t)$.

We introduce an objective function measuring the distance between the predicted contamination and its measured value. There are several possible choices to define a measure for the distance between predicted and measured contamination. For example, [LBV2006a] proposed to measure the time and spaced averaged quantity

$$\sum_{j \in \mathcal{A}_{meas}} \int_0^T \int_{a^j}^{b^j} (c^j(x, t) - \bar{c}^j(x, t))^2 dx dt \quad (20)$$

Therein, T denotes the total time horizon of the measurements. We assume that the measurements \bar{c}^j are given for all points in space and time on pipe $j \in \mathcal{A}_{meas}$. If we only have information on a single point \bar{x}^j of this pipe, then a possible objective functional might also be given by

$$\sum_{j \in \mathcal{A}_{meas}} \int_0^T (c^j(\bar{x}^j, t) - \bar{c}^j(\bar{x}^j, t))^2 dt + \rho \sum_{v \in \mathcal{V}} \int_0^T q^v(t) dt \quad (21)$$

for some $\rho > 0$. This penalize the number of sources and its intensity, see again [LBVB2005] and the references therein.

In view of the latter mixed-integer approximation we propose the following objective functional also measuring the difference between simulated and measured contamination and penalizing the number of sources:

$$\sum_{j \in \mathcal{A}_{meas}} \max_{t \in (0, T)} (c^j(\bar{x}^j, t) - \bar{c}^j(\bar{x}^j, t)) + \rho \sum_{v \in \mathcal{V}} \max_{t \in (0, T)} q^v(t). \quad (22)$$

4 A Linear Mixed-Integer Model

For a numerical solution of the continuous model in the case of large scale networks and in near real-time, we reformulate this model as a linear mixed-integer program. We first transform the partial differential equations for the transport (10) and the decay (12) into a set of linear constraints. For the time derivative we use the *forward difference approximation*

$$\frac{\partial}{\partial t} c^j(x, t) \approx \frac{c^j(x, t + \Delta t) - c^j(x, t)}{\Delta t}, \quad (23)$$

whereas for the spatial derivative we take the *backward difference approximation*

$$\frac{\partial}{\partial x} c^j(x, t) \approx \frac{c^j(x, t) - c^j(x - \Delta x, t)}{\Delta x}. \quad (24)$$

We introduce a discretization for the time as $t_n := n\Delta t$ for $n \in \{0, 1, \dots, N\}$, where Δt is a constant step size (see below), and $N = \lceil T/\Delta t \rceil$. From now on, the time is not running continuously, but in discrete time steps.

For each arc j of the water network we introduce three variables $c_t^{j,in}$, $c_t^{j,mid}$ and $c_t^{j,out}$, corresponding to the containment concentration at three points inside the water pipe. The direction of the flow plays an important role, so we make a distinction between the cases $u^j(x, t) \geq 0$ and $u^j(x, t) \leq 0$. In the first case, we take $c_n^{j,in} := c^j(a^j, t_n)$, $c_n^{j,mid} := c^j(\frac{a^j+b^j}{2}, t_n)$, and $c_n^{j,out} := c^j(b^j, t_n)$, and in the second case we have $c_n^{j,in} := c^j(b^j, t_n)$, $c_n^{j,mid}$ as before, and $c_n^{j,out} := c^j(a^j, t_n)$. Accordingly, we use $u_n^{j,in}$, $u_n^{j,mid}$, $u_n^{j,out}$ as abbreviations for the function values $u^j(x, t)$ at the corresponding coordinates. Note that by this settings, we implicitly introduced a spatial discretization of $(\Delta x)^j := \frac{b^j-a^j}{2}$. Using these variables and the difference approximations (23) and (24), the so-called *upwind discretization* of (10) reads

$$c_{n+1}^{j,mid} = c_n^{j,mid} - \frac{\Delta t}{(\Delta x)^j} u_n^{j,mid} (c_n^{j,mid} - c_n^{j,in}), \quad (25a)$$

$$c_{n+1}^{j,out} = c_n^{j,out} - \frac{\Delta t}{(\Delta x)^j} u_n^{j,out} (c_n^{j,out} - c_n^{j,mid}). \quad (25b)$$

To avoid numerical problems, that is, for the *stability* of the discretization, the *CFL condition*

$$\frac{\Delta t}{(\Delta x)^j} |u^j(x, t)| \leq 1 \quad (26)$$

is required (named after Courant, Friedrichs, and Lewy). To fulfill the CFL condition we thus set

$$\Delta t := \min_j \left(\frac{(\Delta x)^j}{\max_{x,t} |u^j(x, t)|} \right). \quad (27)$$

Using (23) we obtain the following discretization for the decay equation (12):

$$\frac{c^j(x, t + \Delta t) - c^j(x, t)}{\Delta t} = -r c^j(x, t), \quad (28)$$

or

$$c^j(x, t + \Delta t) = (1 - r\Delta t) c^j(x, t). \quad (29)$$

Putting together (25) and (29), we arrive at

$$c_{n+1}^{j,mid} = (1 - r\Delta t) \left(c_n^{j,mid} - \frac{\Delta t}{(\Delta x)^j} u_n^{j,mid} (c_n^{j,mid} - c_n^{j,in}) \right), \quad (30a)$$

$$c_{n+1}^{j,out} = (1 - r\Delta t) \left(c_n^{j,out} - \frac{\Delta t}{(\Delta x)^j} u_n^{j,out} (c_n^{j,out} - c_n^{j,mid}) \right). \quad (30b)$$

We remark that the proposed upwind discretization of the transport equation is just one of many possibilities of finite-difference approaches to this partial differential equation, see e.g. [Leve1990], [SUP2002], [RBA1993], [RoBo1996]. In particular, [LBV2006a] propose a Lagrangian method as discretization of the transport equation. For this particularly simple advection equations both discretizations yield the same results. However, we used the upwind method in order to incorporate the decay of the containment more easily.

The unknown possible sources $q^j(t)$ are discretized using the variables $q_n^j := q^j(t_n)$. The discretization of the coupling (17) is straight-forward and obtained as

$$c_{n+1}^{i,in} = \frac{q_n^j + \sum_{j \in \delta^+(v, t_n)} A^j |u_n^{i,out}| c_n^{j,out}}{\sum_{j \in \delta^-(v, t_n)} A^j |u_n^{j,in}|}, \quad \forall i \in \delta^-(v, t_n). \quad (31)$$

Finally, the objective functional (22) needs to be discretized. To this end, we use the following linear objective function:

$$\sum_{j \in \mathcal{A}_{meas}} \sum_n |c_n^{j,mid} - \bar{c}_n^{j,mid}| \Delta t + \rho \sum_v \theta^v. \quad (32)$$

Here $\theta^v \in \{0, 1\}$ is a binary decision variable indicating whether there is an inflow of containment at node v . This variable is coupled to the inflow via

$$q_n^v \leq M\theta^v, \quad (33)$$

where M is a sufficiently large constant.

5 Computational Results for Water Quality Management

We present computational results on artificial test networks in Section 5.1 to computationally evaluate the complexity of the involved mixed-integer programming problems. In Section 5.2 we show how to find the containment source within a real-world network using the presented approach.

5.1 Simulation and Optimization of Test Networks

Consider the small test net shown in Figure 1. Within this network, the water is always circulating. Every arc has the same length and cross section. We assume that the velocity is a constant function in space and time. As decay rate we select r such that $1 - r\Delta t = 0.98$, that is, after each time step 2% of the containment vanished. At time $t = 0$ we inject the containment at the node marked with q of this network for the 5 next time steps. On the arc marked with m we measure the containment flow and, based on this measured informations, try to estimate the transient inflow profile. The time horizon for the simulation is $T := 100$ time steps. The results are shown in Figure 2. The left picture shows the inflow profile q , the next four picture show the $c^{j,mid}$ values on arcs $1, \dots, 4$.

The next computational tests aim at a comparison of different algorithms for linear programming. With a branch-and-bound approach for the solution of the MIP models, the integrality condition on the variables is dropped at first. The integrality is then reintroduced via iteratively selecting some node $v \in \mathcal{V}$, and creating two subproblems, one with $\theta^v = 0$ and the other with $\theta^v = 1$. For the numerical solution of the resulting linear programs, several algorithms are known. Here we test the primal and the dual simplex, the network simplex, and the barrier

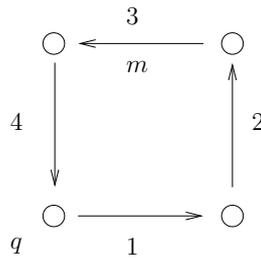


Figure 1: Test network with four arcs $i = 1, 2, 3, 4$.

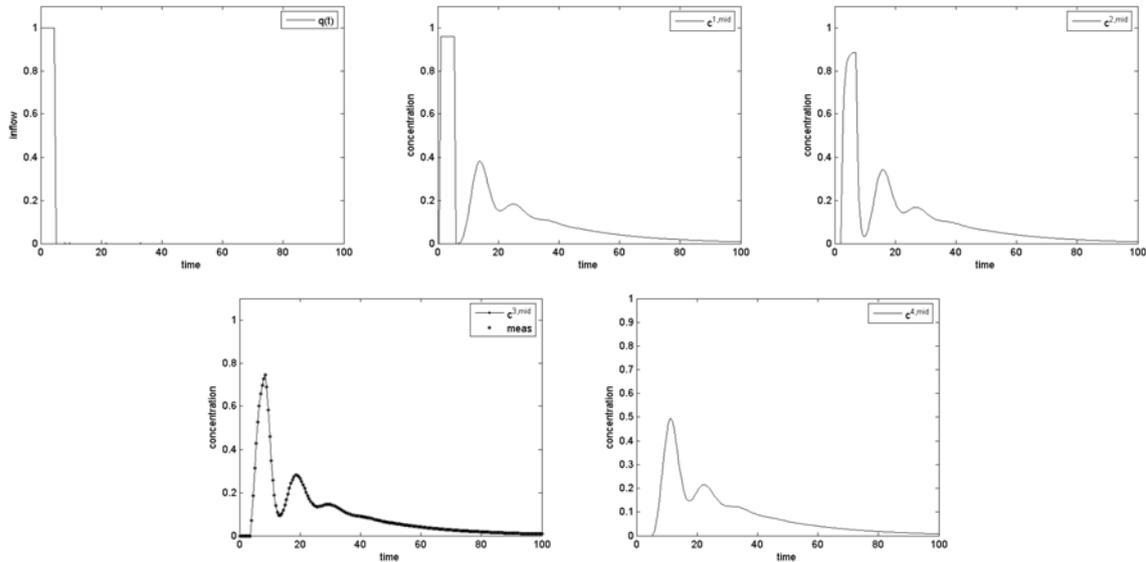


Figure 2: Optimization results for q and arcs $i = 1, 2, 3, 4$ (from left to right).

method (an interior point algorithm), which can all be found within the ILOG Cplex 10 solver suite. Moreover, we want to test the behaviour of these algorithms on instances of different sizes with respect to the number of nodes and the number of time steps. The test networks are rectangular compositions of the network shown in Figure 1 (which in this sense is a 2×2 network). The computational results (solution times in seconds on a standard 2.4 GHz AMD4800X2 personal computer) for a single linear programming relaxation can be found in Table 1.

5.2 Case Study: The Macao Water Supply Network

As a real-world case study we consider the water supply backbone network of the Chinese town Macao, which is inhabited by around half a million people. The network has today a length

size	T	primal	dual	network	barrier
5×5	50	0.1	0.1	0.1	0.1
5×5	100	0.2	0.2	0.2	0.2
5×5	250	0.4	0.4	0.4	0.4
10×10	50	0.7	0.7	0.6	3.7
10×10	100	1.0	0.9	1.0	4.0
10×10	250	2.1	1.8	1.9	5.1
15×15	50	4.1	4.4	6.7	33.5
15×15	100	6.9	6.6	9.8	57.6
15×15	250	9.2	8.9	12.1	59.8
20×20	50	4.3	5.3	8.0	24.5
20×20	100	23.3	37.9	58.4	506.2
20×20	250	27.4	41.6	62.3	401.5

Table 1: Results for various network sizes, time horizons, and LP algorithms.

of about 410 km (for further details see [MACW2006]). The entire backbone network of the three neighbouring cities Macao, Taipa, and Coloane is modeled as a graph that consists of 377 nodes and 601 arcs (see Figure 5). The subnet belonging to Macao is represented by a graph having 237 nodes and 407 arcs. The simulation of a contamination is solved on this graph. We now induce a containment in the network at a certain arc, marked with a “q” in Figure 6. On two other arcs, marked with an “m”, the containment is measured. The distribution of the containment over the time within the network is shown in Figure 6, which is a magnification of the rectangular region depicted in Figure 5. Each time step in the discrete simulation represents 5 minutes of real time.

We now demonstrate what happens within the branch-and-bound algorithm, when the decision variable θ^v is set to 1 for some arc v . If the inflow originates in this node, then the measured profiles and the resulting concentration profiles on $j \in \mathcal{A}_{meas}$ coincide (see middle and right pictures in Figure 3). The inflow profile of the containment at this node is shown on the left picture in Figure 3.

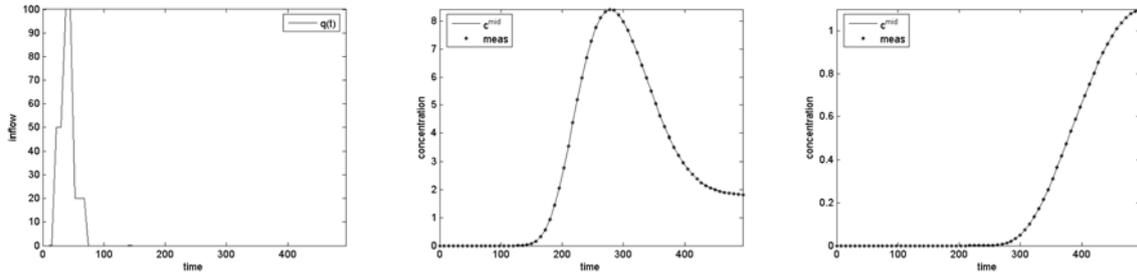


Figure 3: Optimization results for the “right” inflow node

If on the other hand the wrong variable θ^v was selected for branching, then in general we have a gap between the measured and the resulting concentrations (see middle and right picture in Figure 4). This gap leads to a high objective function value, hence the search tree will be pruned as soon as the right node (as above) was found. The inflow profile of the containment of the selected node also has a “strange” appearance (see left picture in Figure 4).

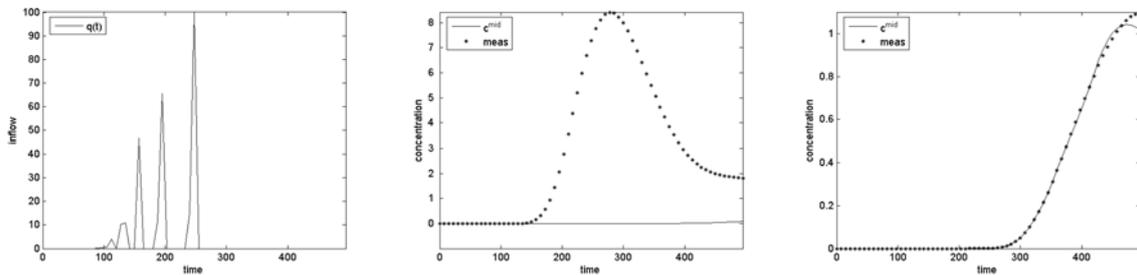


Figure 4: Optimization results for the “wrong” inflow node

6 Other Areas of Application

The applicability of linear mixed-integer programming in continuous PDE optimization is by far not limited to water quality management. In this section we briefly present two other applications where the linear mixed-integer approximation have been successfully applied in order to simulate and optimize processes on large networks.

6.1 Traffic Flow

The first example is related to traffic flow control for road networks. We briefly summarize the approach and findings of [FHKM2004]. Therein, the starting point has been a continuous model based on partial differential equations for the evolution of traffic flow in networks. The partial differential equation is nonlinear and differs from the one discussed above. Moreover, at each traffic intersection derivation suggestions to drivers are modeled as controls. Then, the detailed model is a coupled system of nonlinear partial differential equations with nonlinear constraints at the road intersections. The task is now to utilize the traffic network in such way, that all cars reach their final destination at the earliest possible time. This determines the derivation suggestions at each single intersection. The difficulty stems from the fact that, due to the partial differential equations which control the dynamics on each arc, a higher utilization of the road yield a lower average traveling speed. Furthermore, for an optimization approach the computational costs for solving the coupled system of equations are too expensive. Therefore, several model reductions have been performed and a simplified dynamics on the roads and at intersections was derived. These can be seen as a coarse-grid discretization of the partial differential equation and an additional averaging for estimating the conditions at intersections. In the next step, all nonlinearities have been approximated by linear equations and binary and real variables. A qualitative comparison showed the applicability of the simplified models and in particular in free flow traffic situations these models perform comparable with the models based on partial differential equations. Moreover, these linear mixed-integer models allowed for optimization of networks of realistic size. For further details we refer to [FHKM2004].

6.2 Production Planning

Production planning in supply chain management has been investigated in relation with continuous and discrete models. In [ADR2006] a completely new access to production planning is introduced. There, a linear partial differential equation for the conservation of goods is derived. This equation is assumed to hold for each supplier. For the modelling of a whole supply network buffering queues are introduced in front of each supplier. Thus, the supply network consists of

several suppliers which all have the possibility to store goods. The coupled system of partial and ordinary differential equations describing the supply network allows for simulating large amounts of goods over a long time. In this regard, an optimal control problem is composed of linear (PDE) and nonlinear (queues) constraints. The objective function is also linear and given by maximizing the output and minimizing the amount of goods in all queues. Possible controlling parameters are the distribution rates (used on each vertex where the flux of one supplier is splitted into two new suppliers) or the processing velocities. Similar to the traffic flow control problem we discretize the equations on a coarse grid and substitute the nonlinearity by using binary variables. This leads to a linear mixed-integer programming model. In cases of simple networks where only one supplier is linked to one supplier the continuous model and the MIP model yield same results. A nice property of the MIP is the easy extensibility to additional constraints such as the maintenance of suppliers or bounded queues. For further details we refer to [GHK2005], [GHK2006].

7 Conclusions

In this work, we pointed out the relation between continuous models, governed by partial differential equations, and linear mixed-integer programming. This relation is based on a coarse-grid discretization of the partial differential equation. This technique guarantees the conservation of the original dynamics and also allows for large scale network simulation and optimization. Moreover, the presented way is different from other approaches by Laird, Biegler et al. where non-linear optimization techniques and Lagrangian discretization are applied.

We gave several fields of application where this relation can be used to solve optimal control problems with partial differential equations as constraints. An emphasis was put on the containment source determination in water quality management. In conclusion, we presented a new approach using MIP models, linear programming, and branch-and-bound algorithms for solving continuous optimal control problems.

References

- [LBVB2005] Laird, Carl D.; Biegler, Lorenz T.; van Bloemen Waanders, Bart G.; Bartlett, Roscoe A.: Contamination source determination for water networks. In: A.S.C.E. Journal of Water Resources Planning and Management, Vol. 131, No. 2, 2005, pp. 125 – 134.
- [LBV2006a] Laird, Carl D.; Biegler, Lorenz T.; van Bloemen Waanders, Bart G.: Real-time, large scale optimization of water network systems using a subdomain approach. To appear in proceedings of the Second CSRI Conference on PDE-Constrained Optimization.
- [LBV2006b] Laird, Carl D.; Biegler, Lorenz T.; van Bloemen Waanders, Bart G.: A mixed integer approach for obtaining unique solutions in source inversion of drinking water networks. In: Special Issue on Drinking Water Distribution Systems Security, Journal of Water Resources Planning and Management, Vol. 132, No. 4, 2006, pp. 242 – 251.
- [ADR2006] Armbruster, Dieter; Degond, Pierre; Ringhofer, Christian: A Model for the Dynamics of large Queuing Networks and Supply Chains. In: SIAM J. Applied Mathematics, Vol. 66, 2006, pp. 896 – 920.
- [FHKM2004] Fügenschuh, Armin; Herty, Michael; Klar, Axel; Martin, Alexander: Combinatorial and Continuous Models for the Optimization of Traffic Flows on Networks. In: SIAM J. Optimization, Vol. 16, No. 4, 2006, pp. 1155 – 1176.
- [GHK2005] Göttlich, Simone; Herty, Michael; Klar, Axel: Network models for supply chains. In: Communications in Mathematical Sciences, Vol. 3, No. 4, 2005, pp. 545 – 559.
- [GHK2006] Göttlich, Simone; Herty, Michael; Klar, Axel: Modelling and optimization of supply chains on complex networks. In: Communications in Mathematical Sciences, Vol. 4, No. 2, 2006, pp. 315 – 330.
- [ILOG2006] ILOG CPLEX Division, 889 Alder Avenue, Suite 200, Incline Village, NV 89451, USA. Information available at URL <http://www.cplex.com>.

- [GuLe2003] Gugat, Martin; Leugering, Günter: Global boundary controllability of the St. Venant equations between steady states. In: *Annales des l'Institut Henri Poincaré, Non-linear Analysis*, Vol. 20, No. 1, 2003, pp. 1 – 11.
- [GLSS2001] Gugat, Martin; Leugering, Günter; Schittkowski, Klaus; Schmidt, E.J.P. Georg: Modelling, stabilization and control of flow in networks of open channels. In: M. Grötschel, S. Krumke, J. Rambau (Eds.): *Online optimization of large scale systems*. Springer, 2001, pp. 251 – 270.
- [GLS2003] Gugat, Martin; Leugering, Günter; Schmidt, E.J.P. Georg: Global controllability between steady supercritical flows in channel networks. In: *Mathematical Methods in the Applied Sciences*, Vol. 27, No. 7, 2004, pp. 781 – 802.
- [Leve1990] LeVeque, Randall J.: *Numerical methods for conservation laws*. Birkhäuser Verlag, Basel, Boston, Berlin 1990.
- [MACW2006] The Macao Water Company. Information available at URL <http://www.macaowater.com>.
- [RBA1993] Rossman, Lewis A.; Boulos, Paul F.; Altman, Tom: Discrete volume-element method for network water-quality models. In: *Journal for Water Resource Planing and Management*, Vol. 119, No. 5, 1993, pp. 505 – 517.
- [RoBo1996] Rossmann, Lewis A.; Boulos, Paul F.: Numerical methods for modeling water quality in distribution systems. In: *Journal for Water Resource Planing and Management*, Vol. 122, No. 2, 1996, pp. 137 – 146.
- [SUP2002] Shang, Feng; Uber James G.; Polycarpou, Marios M.: Particle backtracking algorithm for water distribution systems analysis. In: *Journal on Enviromental Engineering*, Vol. 128, No. 5, 2002, pp. 441 – 450.

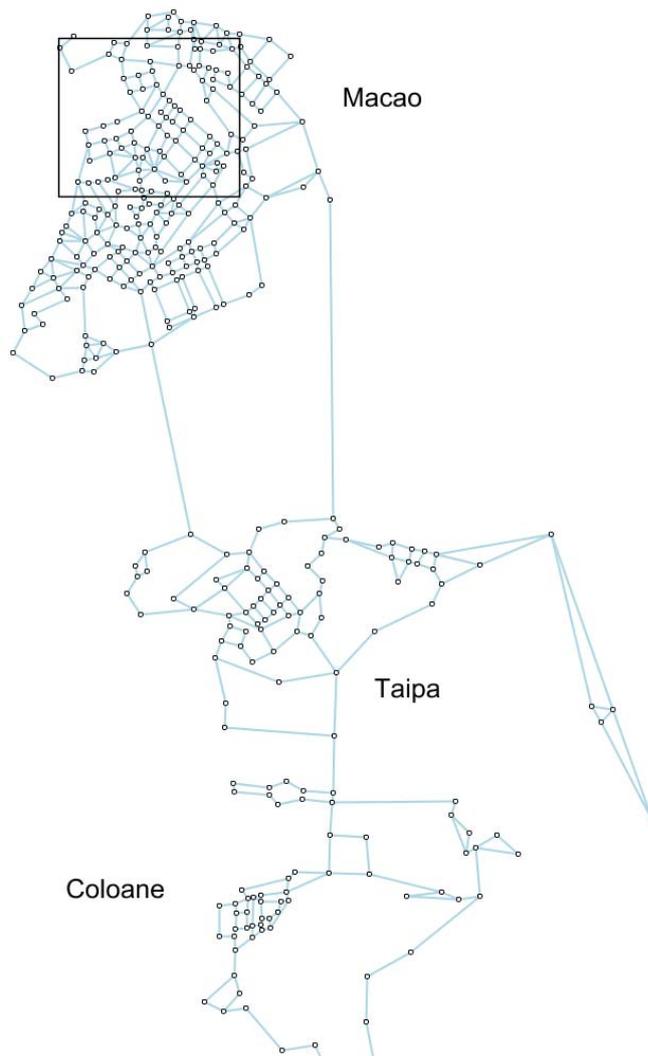


Figure 5: The water supply backbone network of Macao, Taipa, and Coloane.

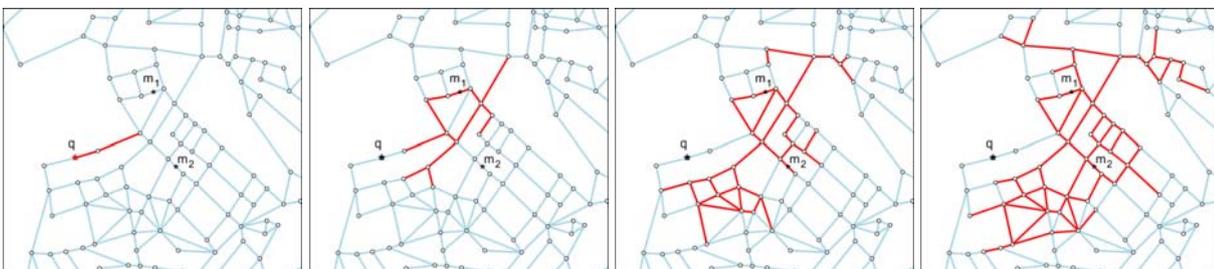


Figure 6: Containment simulation for $t = 9, 28, 47, 66$ (from left to right).

Einführung in den Track

Wissensmanagement

Prof. Dr. Norbert Gronau

Universität Potsdam

Prof. Dr. Steffen Staab

Universität Koblenz-Landau

Prof. Dr. Marcus Spieß

LMU München

Dr. Walter Waterfeld

Software AG

Der Geschäftserfolg jedes Unternehmens beruht auf dem zielgerichteten Umgang mit Wissen. Wissensmanagement zielt darauf ab, Effizienz und Effektivität im Umgang mit Wissen zu steigern, d.h. die Kommunikation von Wissen zu verbessern, das Bewahren von Wissen sicherzustellen, neues Wissen im Unternehmen zu entwickeln und Wissen besser für den Unternehmenserfolg zu nutzen.

Dies geschieht vor dem Hintergrund eines dynamischen Umfelds, in dem sich wandelnde Firmenstrategien, Geschäftsprozesse und Kundenwünsche genauso berücksichtigt werden müssen wie personelle Veränderungen und eine sich wandelnde IT-Infrastruktur. Existierende organisatorische und informationstechnische Lösungen für den Lebenszyklus von Wissen müssen dieser Dynamik gerecht werden. Hierfür werden neue Konzepte und Methoden benötigt, die Schwachstellen identifizieren und sich an die vorgegebene Dynamik anpassen.

Der Track "Wissensmanagement" möchte Wissensmanagementkonzepte, -methoden und Lösungen vor dem Hintergrund dieser Anforderungen betrachten. Wir möchten diskutieren, wie -- von der Anwendung neuartiger Funktionalitäten in Unternehmensportalen bis hin zur Analyse sozialer Netzwerke -- das Wissensmanagement im dynamischen Umfeld heutiger Unternehmen erfolgreich verankert werden kann.

Programmkomitee:

Dr. Brigitte Bartsch-Spoerl, BSR Consult

Prof. Dr. Ralph Bergmann, Universität Trier

Prof. Dr. Dimitris Karagiannis, Universität Wien

Prof. Dr. Helmut Krcmar, TU München

Prof. Dr. Klaus North, FH Wiesbaden

Prof. Dr. Ulrich Reimer, FH St. Gallen

Prof. Dr.-Ing. Bodo Rieger, Universität Osnabrück

Hans-Peter Schnurr, Ontoprise GmbH

Prof. Dr. Martin Schader, Universität Mannheim

Third Generation Knowledge Management Systems

Towards an Augmented Technology Acceptance Model

Kai Dingel, Sarah Spiekermann
Institut für Wirtschaftsinformatik
Humboldt-Universität zu Berlin
Spandauer Strasse 1, 10178 Berlin
{sspiek,dingel}@wiwi.hu-berlin.de

Abstract

The paper examines the applicability and sufficiency of the Technology Acceptance Model (TAM) in the context of social software and newer generation knowledge management systems (KMS). A reinterpretation of the two TAM constructs “Perceived Usefulness” and “Perceived Ease of Use” in light of expectancy-valence theory reveals that the TAM predominantly focuses on performance expectations on different behavioral levels and, thus, fails to account for the entire range of drivers and barriers related to KMS usage.

1 Introduction

In recent years, the prevailing conception of knowledge management (KM) in the scientific literature and its practical applications has undergone major changes and paradigmatic shifts, which are often retrospectively referred to as the “three generations” of knowledge management [Snow02; Snow03; Schü03a; Schü03b]. Unlike its predecessors, the third generation of knowledge management thinking pursues a novel holistic perspective by taking into account the embedded and multifaceted nature of knowledge and by omitting the predominant focus on knowledge sharing. A further constituting aspect of the alternation of generations or “phase shift in thinking” [Snow03, 23] is the increasing attention to the factors that drive or impede knowledge workers’ commitment to KM initiatives. Earlier generations of corporate KM activities typically overemphasized the coordinating role of IT as a key driver for the success of knowledge management programmes. In taking such a technology centric approach, they often

fell below expectations, failing to attain the necessary acceptance of the users. The shortcomings of these earlier generations convey that the participation and cooperation of knowledge workers should not be taken for granted. While IT is indeed a “hygiene factor” [Snow02], which can effectively enable and facilitate the creation and sharing of knowledge, equal weight has to be put on the motivation of knowledge workers and the mitigation of other, non-technological barriers to the participation in knowledge-related activities. In order to remedy acceptance problems, it is insufficient to merely link knowledge sharing “with bonus schemes, appraisals and targets” [Snow02]. The prevailing literature nowadays emphasizes that “sharing cannot be forced” [HuWi04, 90], or as stated by Snowden [Snow02, Snow03], “knowledge can only be volunteered, it cannot be conscripted”. In this way, “commitment” substitutes “compliance” as a driver of knowledge workers’ participation in newer generations of knowledge management thinking [Malh03].

Prominent examples like the online project Wikipedia, which is an instance of a knowledge repository, or social networking websites like LinkedIn or openBC, being representatives of the “personalization strategy” [cf. HaNT99], convey the dynamics and power of this new generation of systems. They successfully instigate intense system usage by tapping users’ intrinsic or natural motives to participate and by mitigating usage barriers. In so doing, they belong to a socially enriched type of system, which is often labeled “social software”, a term that is very much discussed in the scene of blogs and online forums while having yet only little impact in the scientific literature. Eagle [Eagl04] shortly defines “social software” as “programs that enable a group of people to accomplish common goals”, i.e. software that encourages social interaction and collaboration. According to Avram [Avra06, 1], social software involves “the use of computing tools to support, extend, or derive added value from social activities”. Quite similar, Thomas et al. [ThKE01, 872] use the term “social computing”, which comprises “digital systems that draw upon social information and context to enhance the activity and performance of people, organizations, and systems”.

One aspect of the “social” nature of social software is the way “it adapts to the user, instead of forcing the user to adapt to it; becomes part of the user’s means of representation, and augments human interaction, instead of narrowing it down” [Avra06, 7]. With social software, sharing is not imposed. Instead, social software “leaves the control of knowledge with the individuals

owning it” [Avra06, 1]: They are able to self-assign to communities based on their personal preferences and maintain their own space over which they have personal control. Consistent with this notion of a socially enriched type of software, one assumption is inherent in many of the current publications on knowledge management: Acceptance is of pivotal importance for the success of KM initiatives in general and the success of knowledge management systems in particular. To successfully tap the desires and needs of knowledge workers, knowledge management has to meet the terms of a new principle of self-organizing [Schü03b], thereby partially breaking with the traditional top-down management imperative. Processes of knowledge creation and transfer rather require a supporting, cultivating, and nurturing responsibility of management, instead of being manageable in the usual sense.

Successful systems such as Wikipedia or LinkedIn are promising examples of the usefulness of newer generation knowledge management systems. They show that social software or third generation KMS have the potential to make invaluable contributions to organizational KM initiatives, leveraging the human and social capital of an organization. However, we can only learn from these examples if we understand the underlying drivers of system acceptance and how these drivers operate in shaping usage intentions. With the rising interest in the determinants of knowledge workers’ motivation to KMS usage, an extensive body of research has compiled a long list of potentially important elements of the “knowledge management puzzle” [ThKE01, 872], i.e. factors that characterize a good KM strategy and supportive KMS. These factors, mostly gathered by means of theoretical analyses and qualitative case studies, range from concrete system characteristics to abstract phenomena such as “trust”, “intrinsic motivation”, “social obligation” or “reciprocity”. However, research is scattered into divergent perspectives and lacks a common frame of reference. In addition, few publications have so far empirically investigated possible causal models of knowledge management system acceptance and usage, which succeed to integrate the large number of qualitative findings or examine their relative importance in explaining and predicting KMS usage.

A noteworthy exception is the IS Success Model by DeLone & McLean [DeMc92; DeMc03], which is often utilized as a framework to structure the variety of success factors [MaHä01; AlLe01, 130-131] or as a fundament of quantitative studies [QiBo05]. Although the goal of our article is likewise to develop an underlying framework of critical factors in knowledge

management, we intend to rely on a slightly different theoretical grounding: the Technology Acceptance Model (TAM), which is one of the most prominent models that *explicitly* investigate the causal antecedents of the intention to use. The Technology Acceptance Model [Davi89; DaBW89] hypothesizes that information system adoption and usage can be explained and predicted by considering two focal behavioral beliefs: “Perceived Usefulness” and “Perceived Ease of Use”. It calls attention to the fact that a mere focus on the usability-oriented “Ease of Use” of a system is insufficient to explain system acceptance and should be complemented by the purpose-oriented system “Usefulness” [DaBW89, 1000]. The quantitative study by Money & Turner [MoTu04] is one of the first that empirically investigated the appropriateness of the TAM for the context of knowledge management. Even though their results should be confirmed by other studies with larger sample sizes and more advanced methodology, it seems that the original finding of the TAM, being able to predict about 40% to 50% of the variance in the behavioral intention, can be replicated in the field of knowledge management. Yet, despite these encouraging results, it is questionable whether the two constructs, “Ease of use” and “Usefulness” cover all major behavioral beliefs behind knowledge acquisition and sharing as well as behind KMS usage. In view of the large number of factors currently discussed in the knowledge management research, the two TAM constructs appear insufficient to account for the full richness of decisive motives and barriers in knowledge management. We therefore hypothesize in line with Money & Turner that “it may be necessary to add other theory-based individual beliefs to the current TAM belief constructs” to “increase the explanatory power” of the TAM [MoTu04, 8].

The first step in elaborating on this hypothesis has to be concerned with a better understanding of the content and origins of the TAM constructs and how they are embedded into the much larger landscape of motivation theory. Moreover, it may be interesting to identify other important salient behavioral beliefs that are presumably relevant to third generation KMS but yet not part of it. We approach this first step by arguing below that expectancy-valence theory – being a crucial fundament of TAM – can be used to reinterpret and extend the TAM in the face of third generation knowledge management. In essence, expectancy-valence theory argues that individuals invest and direct effort with a view to the expected outcomes of behavior [Vroo64; PoLa68]. Although most of the prominent models of system acceptance and usage rely on such “expectations-based frameworks” [cf. Sedd97, 247], they mostly just refer to expectancy-

valence theory in order to substantiate the assumed dependence of acceptance on anticipated behavioral consequences. In contrast to this rather superficial respect of expectancy-valence theory, we claim that expectancy concepts can as well be useful to systematically identify and delimit a fuller spectrum of concrete behavioral beliefs relevant for KMS usage.

In order to provide evidence to this argument, the rest of this paper is organized as follows: In the next chapter, we will give a short introduction into expectancy-valence theories of work motivation. The subsequent chapter 3 will proceed by formulating an integrated, two-tier expectancy framework that delimits two interdependent behavioral levels to which the expectancy-valence approach may apply. Based on these theoretical groundings, the two TAM belief constructs will be revisited in section 4.1 as well as extended in section 4.2. Even though not being able to provide a complete specification of possible enhancements, the paper will indicate several important behavioral beliefs still omitted in the TAM. Chapter 5 will conclude the paper by summarizing its core statements.

2 A Short Introduction into Expectancy-Valence Theories

Expectancy-valence theories hypothesize that individuals choose between different behavioral alternatives and between different levels of effort by anticipating the impact of their decisions on resultant outcomes (see Figure 1). Individuals thus determine a set of relevant, salient consequences that may arise from their actions. These consequences, however, are not of equal importance but are valued differently. The particular value or “valence” of an outcome [Vroo64, 15] can have two sources [cf. Vroo64, 16]: On the one hand, it may originate from the “instrumentality” (❶) of the outcome in allowing for further possible outcomes. An often-mentioned example of such an instrumental outcome is a monetary reward. On the other hand, valence may be due to some intrinsic value created to the individual (❷), like enjoyment or feelings of achievement.

Furthermore, individuals are theorized to form expectations in terms of anticipated, “subjective” probabilities that particular outcomes will actually be obtained [Vroo64]. The obtainment of an outcome may be impeded by a variety of internal or external factors. Most of these impediments originate from the fact that the attainment of consequences is bound to the *successful*

implementation of a behavior. Although outcomes may also be directly contingent on behavior and effort (❸), as is the case with the cost of exerted effort, outcomes are typically attained as a function of personal performance and success [PoLa68]. The expectancy component can hence be decomposed into two separate beliefs: First, expectations are formed on whether effort is likely to result in fulfillment of aspired success, which we call “performance expectations” (❹). Second, an expectation is formed on whether achieved performance is going to be followed by valued outcomes, which is called the “instrumentality” of performance (❺). Unlike performance expectations, perceived instrumentality can also be negative if actions are counterproductive in achieving certain goals and, thus, has a range from -1 to +1. According to Vroom [Vroo64], performance expectations, perceived instrumentality and outcome valence jointly shape the motivational “force” of a behavior, which in turn is the basis of a relative evaluation of different behavioral alternatives.

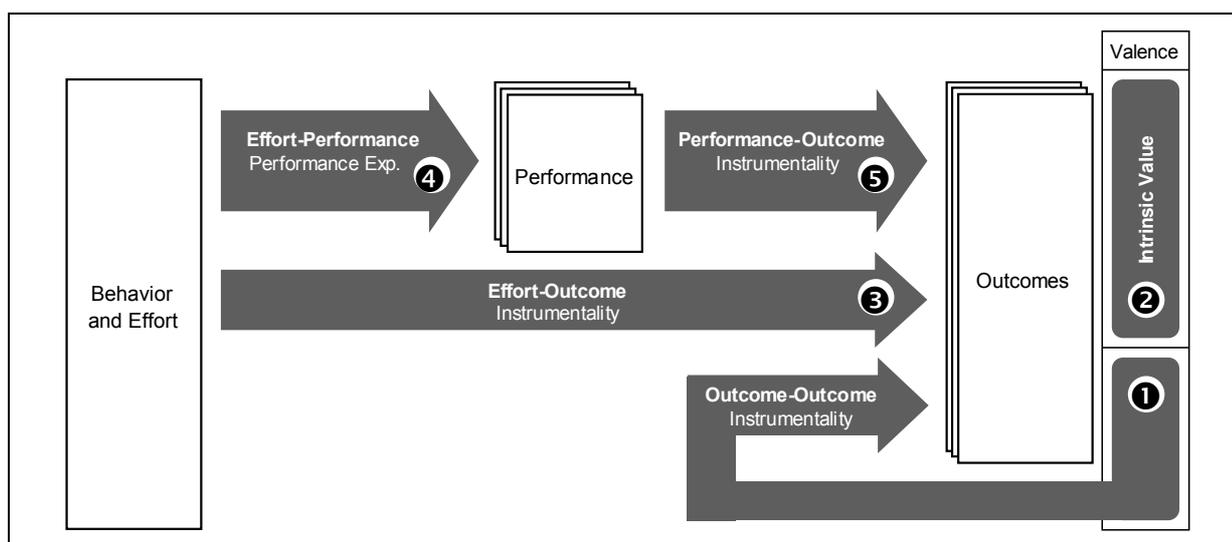


Figure 1: Core schema of expectancy-instrumentality-valence theories (own visualization)

3 Extending Expectancy-Valence Theory for Third Generation KM

3.1 Towards an Integrated Expectancy-Valence Model for Knowledge Management

In the knowledge management literature, the analysis of acceptance of *knowledge management systems* is typically intertwined with the more general investigation of *people’s commitment* to knowledge-related activities. This mingling of two dimensions of acceptance may be due to the

fact that the identification of motives for system usage requires a holistic point of view, which takes the organizational environment and organizational culture into consideration. In this vein, we argue that in formulating a holistic KMS acceptance model, it is important to recognize these two distinct tiers. They can be referred to as organization-level and tool-level acceptance (for a distinction between tool-level and organizational-level tasks see e.g. [TeCZ06, 229-231]).

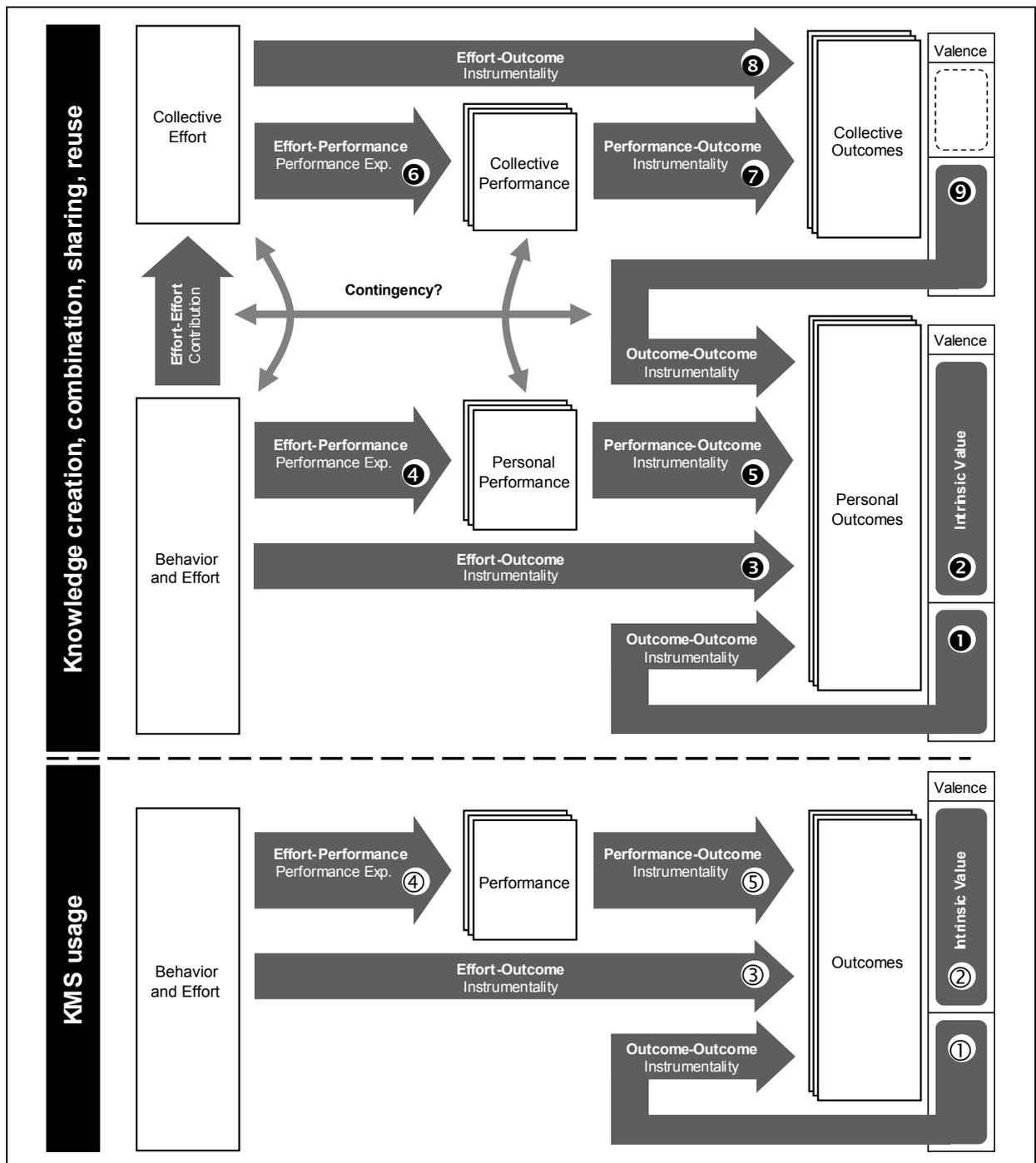


Figure 2: A two-tier expectancy model of behavioral beliefs in knowledge management

KMS usage has to be understood as being related to, but also separable from the more general organization-level tasks of the individual or collective to create, combine, share or reuse knowledge. Figure 2 gives an overview of an integrated expectancy-valence framework we propose to use for analyzing KMS usage. In particular, we claim that the expectancy-valence model introduced above can be applied to each of the two interdependent behavioral levels, because behavioral alternatives will exist on each of these tiers and will be evaluated based on idiosyncratic expectancies and anticipated outcomes.

A knowledge worker, who decides to enroll into knowledge-related activities and to KMS usage, may start out by evaluating his *expected personal performance* in sharing or reusing knowledge (④). These expectations will not only reflect personal capabilities, but also external factors such as sufficient organizational support. Moreover, expected personal performance in KMS usage (④) will play an important role within this cognitive appraisal. For instance, a user of the Wikipedia website, who wants to contribute his personal knowledge to the project, will evaluate his general capabilities of putting his knowledge into words [cf. CaCa02, 700] and compiling useful new articles or article revisions. Herein, his expected performance will be partially determined by his skills in using the Internet or the Wikipedia website. Therefore, it seems reasonable to assume a bottom-up influence of lower-level performance expectations (④) on the attainable upper-tier performance (④).

The second dimension of the decision to engage in knowledge-related activities is concerned with the *expected outcomes* of a participation. On the one hand, these outcomes directly follow from exerted effort (③), exemplified by the time-related opportunity costs of a participation [CaCa02, 688; ArPW03, 70; Kall03, 119] or a risk of jeopardizing knowledge-based status in the organization [CaCa02, 697; ArPW03, 69]. On the other hand, they may be obtained as a function of accomplishable individual performance (⑤). For instance, attaining approval by the Wikipedia community may necessitate that the contributed article meets relevant community standards to have a longer lasting impact.

The latter example also illustrates that personal outcomes, if contingent on personal success, normally necessitate that personal contributions are identifiable and separable as well as measurable through comparison with existing quality or performance standards [cf. Shep93;

KaWi01]. Consequently, the “visibility” of individual contributions may be a further aspect of the scope of behavioral beliefs that relate to the instrumental value of personal performance (⑤/⑤).

One question remained unanswered so far: Why is the upper tier of the expectancy-framework relevant to the decision to use a concrete knowledge management system? This top-down influence can be explained by the fact that upper-level outcomes (⑤) contribute to the expected outcomes of lower levels (⑤), though mediated by upper-level performance expectations (④). Even if the usage of a KMS may also directly imply personal consequences like enjoyment or intrinsically valued feelings of competence, motives to the usage of a KMS will foremost stem from higher, purpose- or task-oriented levels. However, upper-level outcomes only act as motives to the lower level of KMS usage, if they are supported by sufficient expected performance on the upper level. Stated differently, expected rewards that are contingent on sharing knowledge are only motivating factors to the use of a KMS, if the individual is actually confident in his or her personal ability to share knowledge. The same applies of course to other knowledge-related activities.

3.2 KMS Usage as a Commitment to Collective KM Activities

Aside from the delimitation of the two behavioral levels discussed above, a further source of behavioral beliefs, germane to the upper tier, can be the collective or collaborative nature of knowledge-related activities that are often embedded into the formal or informal communities of the organization. A contemporary stream of research tries to apply the expectancy-valence approach to such settings of collective actions. Herein, one is faced with contradicting phenomena: On the one hand, it is argued that the sole presence of others often positively enhances exerted effort and subsequent performance, which is called “social facilitation” [see e.g. Vroom64, 230; Shep93, 67], whereas another stream of research examines a phenomenon that Latané, Williams & Harkins [LaWH79] labeled “social loafing” and that implies quite the opposite effect: People often reduce their productivity and effort in case of working collectively.

Among others, Shepperd [Shep93] as well as Karau & Williams [KaWi93; KaWi01] have conducted an extensive review on the topic of „social loafing“, „social dilemmas” and „collective work motivation“ and presented an integrated model, based on Vroom’s expectancy-

valence theory, to analyze the interaction of various influential factors. Even if many of these group implications operate on a personal level by affecting individual motives or barriers (like enabling community support or reputation-based expert status), there are also implications related to a more general, collective level. Therefore, Karau & Williams state that individuals not only care for personal level outcomes, but also for the impact of their behavior on the collective, they identify with. This latter collective level is concerned with the success of the whole group and the attainment of collective outcomes as a function of collective effort and performance. In line with this reasoning, we argue that the original expectancy-valence model as shown in Figure 1 should be extended to a version as shown in the upper tier of Figure 2. This extended model would account for the additional motives and barriers, which are important in such settings of collective activities, and would integrate elements of the „Collective Effort Model” by Karau & Williams [KaWi93; KaWi01] that for example has already been applied to the contexts of online communities [LBLW05] or open source communities [HeNH03].

In analyzing the elements of the collective level, single individuals will firstly assess the collective performance, which the group is likely to accomplish (⑥). As Bandura [Band01] underlines by distinguishing between “self-efficacy” and “collective-efficacy”, collective performance is more than the sum of members’ individual contributions but also involves “transactional dynamics” such as coordination and concerted interaction. Moreover, collective performance expectations not only result from a mere evaluation of other group members’ capabilities, but also comprise an expectation on whether these members will actually contribute. In addition, individuals will also evaluate the collective outcomes that are obtainable in view of the likely collective performance (⑦) as well as the costs (and other outcomes), which are directly contingent on collective effort (⑧).

Mutual participation and valuable contribution by peers are vital for two reasons: On the one hand, individuals may identify with the outcomes of collective effort and these are likely to be of higher value the more effort is denoted to the common pool. For example, the contributors of the online project Wikipedia may be motivated by the projects’ intention to establish an open, freely available encyclopedia. In this case, they may value the sheer number of articles, although not being interested in each of these topics. On the other hand, participation of peers

also increases the likelihood that contributors can *personally* benefit from others' contributions, only being interested in specific subparts of the collective good.

In any case, it is crucial that collective performance and collective outcomes are perceived to be instrumental in attaining personally valued outcomes in order to act as a motivating factor for the commitment of knowledge workers (☉). For example, individuals may personally value the collective achievements of the group. However, even if the collective level is linked to the personal level, own contributions of effort and performance have to be deemed crucial for the sufficiency of collective effort and performance in turn. Otherwise, individuals would lack an incentive to *personally* contribute effort to the collective pool [KaWi01]. Even if this latter appraisal, which is visualized by the smaller arrows in Figure 2, may also be a subject of strategic considerations, knowledge workers are often simply just not able to recognize the relevance of their knowledge to the group [AlLe01, 126; ArPW03, 70; CaCa02, 700]. Researchers therefore discuss means to increase the perceived dependency of group success on the participation of all its members or to emphasize the indispensability and non-redundancy of single contributions.

4 Towards an Augmented Technology Acceptance Model

4.1 A Reconsideration of the TAM in Light of Expectancy-Valence Theory

What can be learned about the Technology Acceptance Model (TAM), if reinterpreted in light of this two-tier expectancy-valence approach? Which expectancies of the framework are already covered by TAM's two belief constructs "Perceived Ease of Use" and "Perceived Usefulness"?

4.1.1 Perceived Ease of Use

Davis [Davi89, 320] defines the construct "Perceived Ease of Use" as "the degree to which a person believes that using a particular system would be free of effort". Interestingly, the underlying items of this TAM construct all have a rather similar, narrow focus on two particular behavioral beliefs: On the one hand, nearly all items express individual expectancies concerning the likely personal performance results of system usage. This is for example evident in items

such as “I would find it easy to get THE SYSTEM to do what I want it to do.” [Davi89]. As these items refer to personal performance in the interaction with the particular system, “Perceived Ease of Use” can be regarded as relating to effort-performance-associations on the lower level of KMS usage (④).

In addition, items like “Learning to operate THE SYSTEM would be easy for me.” [Davi89] also refer to the monetary or non-monetary costs that directly result from effort spent on KMS usage. As these expenses are interpretable as a negative instrumentality of effort for obtaining positively valued outcomes, these items additionally correspond to effort-outcome-association on the level of KMS usage (③).

4.1.2 *Perceived Usefulness*

According to Davis et al. [DaBW89, 985], “Perceived Usefulness” can be defined as the “prospective user’s subjective probability that using a specific application system will increase his or her job performance within an organizational context”. Although “Usefulness” is often interpreted to cover the expected outcomes or consequences of system usage [CoHi95, 197], the focus is actually on a “use-performance relationship” [Davi89, 320]. In line with this understanding of the construct, “Perceived Usefulness” seems to be a type of performance expectation as well, similar to “Perceived Ease of Use”, though on a higher level (④). A reconsideration of its underlying items in light of expectancy-valence theory confirms this proposition. For example, the focus on performance increases is obvious in items such as “Using THE SYSTEM in my job would increase my productivity.” or “Using THE SYSTEM would improve my job performance.” [Davi89]. In addition, some of the items again tap the cost dimension and therefore correspond to direct effort-outcome associations (⑤).

4.2 **Extending TAM’s Behavioral Beliefs on the Background of Expectancy-Valence Theory**

As the discussion in the preceding section has conveyed, both TAM belief constructs each cover specific aspects of the two-tier expectancy framework, as they both relate to expected personal performance on the two different behavioral levels (④/④). Due to systems’ ability to alter expected personal performance on both behavioral levels, it seems reasonable to include

constructs such as “Perceived Ease of Use” or “Perceived Usefulness” in a model of KMS acceptance and usage. However, TAM’s *predominant* focus on these dimensions inherently implies that personal performance is assumed to be followed by some kind of positively valued outcome or avoidance of negative ones. While this latter assumption may be justified in traditional applications of the TAM, the literature on knowledge management emphasizes that the instrumental value of individual performance cannot be taken for granted.

In fact, the participation in knowledge-related activities can involve significant disincentives in terms of costs and risks, which are often not offset by associated positive outcomes. Although most studies herein refer to opportunity costs that originate from time-related constraints [see e.g. AILe01, 127; ArPW03, 70; CaCa02, 694], there is actually a wide variety of motivational barriers to the participation of knowledge workers. For example, sharing one’s ideas may imply disclosing personal secrets, loosing position-based status or expert status or personal competitiveness [AILe01, 69; CaCa02, 694; AILe01, 126]. In addition, individuals may fear criticism [ArPW03, 70] through revealing personal weaknesses or a personal lack of knowledge, or may hesitate to reveal the superiority of others. Likewise, contributors may have a “fear of abuse” [cf. Snow02] or misuse [ArPW03, 72] or may decide on a private usage of knowledge because of confidentiality considerations [ArPW03, 70].

Given these costs, it is decisive whether knowledge sharing or reuse is actually perceived to be important for one’s own personal aims (⑤). A general unawareness of potential benefits is a frequently mentioned barrier to the participation of employees in knowledge-related activities [e.g. Kall03, 121; CaCa02, 688]. Even if the rationales behind engaging in knowledge management may originate from job- or task-related motives and rewards (especially in *harvesting* or *reusing* knowledge), research has by now delimited many other potential objectives. For example, knowledge workers may strive for gaining expert status [ArPW03, 69; CaCa02, 694], reputation [CaCa02, 695], or formal or informal social recognition [ArPW03, 69; CaCa02, 696]. Here, they may perform what is called “impression management”, i.e. they may desire to deliberately shape their image as perceived by their environment, such as conveying their uniqueness, their indispensability, or their social embeddedness. Furthermore, contributing to knowledge management initiatives can encompass personal feelings of competence, proficiency, creativity, or achievement, feelings of relatedness and belonging, affiliation, or

group cohesiveness. For instance, people may enjoy working in a team as well as helping others [CaCa02, 692] and doing a kind of mentoring. Moreover, personal contributions may be elicited by desires for competition, for contesting one's ideas as well as out-performing others; or they may be the result of "moral obligations" [ArPW03, 69] or social norms of "reciprocity" [CaCa02, 692].

Neither of the above-mentioned motives and barriers is addressed explicitly by the core TAM constructs. The same is true for motives that are related to the collective level. The collective level may for example be a source of further motives, if individuals identify with collective actions, value collective outcomes or intrinsically enjoy collective activities (⑨). It raises question like whether knowledge management is seen as important for the collective or the organization as a whole (⑦/⑧) and whether the organization is supposed to be able to accomplish sufficient collective performance and the "critical mass" herein [CaCa02, 699] (⑥).

However, besides its tendency to omit the instrumentality of individual effort and performance, it is likewise also questionable whether the two TAM constructs actually cover all relevant aspects of performance expectations. For example, the construct "Perceived Usefulness", unlike "Ease of Use", does not focus on expected *absolute* personal performance, but on systems' ability in bringing about performance *increases*, compared to a not explicitly defined base case. Although this is consistent with the relative evaluation of behavioral alternatives in expectancy-valence theories, it would nevertheless be interesting to ask respondents whether they believe to generally have the necessary capabilities and resources to accomplish sufficient personal performance (④). Even if a KMS is able to assist the user in his knowledge-related activities, other factors like a lack of other resources or organizational support can severely undermine knowledge workers' motivation.

5 Summary and Outlook

In the introduction of this paper, we claimed that earlier generations of corporate knowledge management initiatives typically overemphasized the coordinating and facilitating role of IT. A similar point of view seems to be inherent in the Technology Acceptance Model with its predominant focus on performance expectations or increases on different behavioral levels.

However, the preceding chapters have indicated that this perspective fails to account for the entire set of behavioral beliefs pertinent to the field of knowledge management. The decision to use a KMS is subject to a much wider variety of behavioral beliefs. Taking a too narrow focus not only reduces the predictive validity of the acceptance model, but also underestimates the power of this new, socially enriched generation of KMS to successfully shape the motives and mitigate the barriers of potential users.

Further research should try to empirically test this conjecture by complementing the performance-oriented TAM with additional, outcome-oriented constructs. Presumably, some of these model extensions can directly be taken from the large number of re-specifications and enhancements to the Technology Acceptance Model, which researchers have already proposed and empirically validated since its initial publication. Here, the two-tier expectancy framework may serve as a unifying framework for integrating the various TAM extensions with the broad literature on barriers and motives in knowledge management. Such a stream of research would invaluablely contribute to a deeper understanding of the success factors of newer generation KMS and facilitate the transfer of their characteristics to a wider range of organizational applications.

References

- [AlLe01] *Alavi, M.; Leidner, D.E.*: Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. In: *MIS Quarterly*, 25 (2001) 1, pp. 107-136.
- [ArPW03] *Ardichvili, A.; Page, V.; Wentling, T.*: Motivation and barriers to participation in virtual knowledge-sharing communities of practice. In: *Journal of Knowledge Management*, 7 (2003) 1, pp. 64-77.
- [Avra06] *Avram, G.*: At the crossroads of knowledge management and social software. In: *The Electronic Journal of Knowledge Management*, 4 (2006) 1, pp. 1-10. <http://www.ejkm.de>, retrieved 2006-04-17.
- [Band01] *Bandura, A.*: Social Cognitive Theory: An agentic perspective. In: *Annual Review of Psychology*, 52 (2001), pp. 1-26.

- [CaCa02] *Cabrera, Á.; Cabrera, E.F.*: Knowledge-sharing dilemmas. In: *Organization Studies*, 23 (2002) 5, pp. 687-710.
- [CoHi95] *Compeau, D.R.; Higgins, C.A.*: Computer self-efficacy: Development of a measure and initial test. In: *MIS Quarterly*, 19 (1995) 2, pp. 189-211.
- [Davi89] *Davis, F.D.*: Perceived usefulness, perceived ease of use, and user acceptance of information technology. In: *MIS Quarterly*, 13 (1989) 3, pp. 319-340.
- [DaBW89] *Davis, F.D.; Bagozzi, R.P.; Warshaw, P.R.*: User acceptance of computer technology: A comparison of two theoretical models. In: *Management Science*, 35 (1989) 8, pp. 982-1003.
- [DeMc92] *DeLone, W.H.; McLean, E.R.*: Information systems success: The quest for the dependent variable. In: *Information Systems Research*, 3 (1992) 1, pp. 60-95.
- [DeMc03] *DeLone, W.H.; McLean, E.R.*: The DeLone and McLean model of information systems success: A ten-year update. In: *Journal of Management Information Systems*, 19 (2003) 4, pp. 9-30.
- [Eagl04] *Eagle, N.*: Can serendipity be planned? In: *MIT Sloan Management Review*, 46 (2004) 1, pp. 10-14.
- [HaNT99] *Hansen, M.T.; Nohria, N.; Tierney, T.*: What's your strategy for managing knowledge? In: *Harvard Business Review*, (1999) March-April, pp. 106-116.
- [HeNH03] *Hertel, G.; Niedner, S.; Herrmann, S.*: Motivation of software developers in open source projects: An Internet-based survey of contributors to the Linux kernel. In: *Research Policy*, 32 (2003), pp. 1159-1177.
- [HuWi04] *Huysman, M.; de Wit, D.*: Practices of managing knowledge sharing: Towards a second wave of Knowledge Management. In: *Knowledge and Process Management*, 11 (2004) 2, pp. 81-92.
- [Kall03] *Kalling, T.*: Organization-internal transfer of knowledge and the role of motivation: A qualitative case study. In: *Knowledge and Process Management*, 10 (2003) 2, pp. 115-126.

- [KaWi93] *Karau, S.J.; Williams, K.D.*: Social loafing: A meta-analytic review and theoretical integration. In: *Journal of Personality and Social Psychology*, 65 (1993) 4, pp. 681-706.
- [KaWi01] *Karau, S.J.; Williams, K.D.*: Understanding individual motivation in groups: The collective effort model. In: *Turner, M.E. (Ed.): Groups at Work: Theory and Research*. Erlbaum, Mahwah 2001, pp. 113–142.
- [LaWH79] *Latané, B.; Williams, K.; Harkins, S.*: Many hands make light the work: The causes and consequences of social loafing. In: *Journal of Personality and Social Psychology*, 37 (1979) 6, pp. 822-832.
- [LBLW05] *Ling, K.; Beenen, G.; Ludford, P.; Wang, X.; et al.*: Using social psychology to motivate contributions to online communities. In: *Journal of Computer-Mediated Communication*, 10 (2005) 4. <http://jcmc.indiana.edu/vol10/issue4/ling.html>, retrieved 2006-06-07.
- [MaHä01] *Maier, R.; Hädrich, T.*: Modell für die Erfolgsmessung von Wissensmanagementsystemen. *Wirtschaftsinformatik*, 43 (2001) 5, pp. 497-509.
- [Malh03] *Malhotra, Y.*: Why knowledge management systems fail? Enablers and constraints of knowledge management in human enterprises. In: *Srinkantaiah, K.; Koenig, M.E.D. (Eds.): Knowledge Management Lessons Learned*. Information Today, Medford 2003. http://liris.cnrs.fr/alain.mille/enseignements/Ecole_Centrale/projets_2004/KM_fails.pdf, retrieved 2006-04-26.
- [MoTu04] *Money, W.; Turner, A.*: Application of the Technology Acceptance Model to a knowledge management system. In: *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.
- [PoLa68] *Porter, L.W.; Lawler, E.E.*: Managerial attitudes and performance. Irwin-Dorsey, Homewood 1968.
- [QiBo05] *Qian, Z.; Bock, G.W.*: An empirical study on measuring the success of knowledge repository systems. In: *Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005.

- [Schü03a] *Schütt, P.*: Die dritte Generation des Wissensmanagements. In: KM-Journal, 1 (2003), pp. 1-7.
- [Schü03b] *Schütt, P.*: The post-Nonaka Knowledge Management. In: Journal of Universal Computer Science, 9 (2003) 6, pp. 451-462.
- [Sedd97] *Seddon, P.B.*: A respecification and extension of the DeLone and McLean Model of IS success. In: Information Systems Research, 8 (1997) 3, pp. 240-253.
- [Shep93] *Shepperd, J.A.*: Productivity loss in performance groups: A motivation analysis. In: Psychological Bulletin, 113 (1993) 1, pp. 67–81.
- [Snow02] *Snowden, D.*: Complex acts of knowing: Paradox and descriptive self-awareness. Speech at the International Conference on the Future of Knowledge Management in Berlin, March 2002. <http://amor.rz.hu-berlin.de/~h04440am/wissensmanagement.htm>, retrieved 2006-04-02.
- [Snow03] *Snowden, D.*: Complex acts of knowing: Paradox and descriptive self-awareness. In: Bulletin of the American Society for Information Science and Technology, (2003) April/May, pp. 23-28.
- [TeCZ06] *Te'eni, D.; Carey, J., Zhang, P.*: Human Computer Interaction - Developing Effective Organizational Information Systems. John Wiley & Sons, New York 2006.
- [ThKE01] *Thomas, J.C.; Kellogg, W.A.; Erickson, T.*: The knowledge management puzzle: Human and social factors in knowledge management. In: IBM Systems Journal, 40 (2001) 4, pp. 863-884.
- [Vroo64] *Vroom, V.H.*: Work and motivation. John Wiley & Sons, New York 1964.

Unterstützung wissensintensiver Prozesse im Produktlebenszyklus durch Suche in Produkt- und Prozessdaten

Thomas Barth, Christian Lütke Entrup, Walter Schäfer

Institut für Wirtschaftsinformatik
Fachbereich Wirtschaftswissenschaften
Universität Siegen
57068 Siegen
{barth, luetke-entrup, jonas}@fb5.uni-siegen.de

Abstract

Entlang des Produktlebenszyklus generierte Daten enthalten das in einer Vielzahl von Prozessen angewandte Wissen der an Planung, Entwicklung, Herstellung etc. beteiligten Experten eines Unternehmens. Ein Ansatz zur Unterstützung komplexer, wissensintensiver Prozesse ist, dieses Wissen so wieder zu verwenden, dass zu einer neuen Aufgabenstellung ähnliche Fälle in historischen Daten gesucht, identifiziert und ggf. als Vorlage verwendet werden können.

In diesem Beitrag wird ein ausgewählter Prozess – die Angebotserstellung – auf das notwendige Wissen hin analysiert und ein generischer Suchprozess auf Produkt- und Prozessdaten als Ansatz für die Wiederverwendung von Wissen in Prozessen motiviert. Der Nutzen dieses Ansatzes wird anhand eines Werkzeugs in einem industriellen Szenario eines mittelständischen, auftragsorientierten Serienfertigers der Automobilzulieferindustrie demonstriert.

1 Einleitung

Der Produktlebenszyklus – und im Idealfall ein adäquates Produktlebenszyklusmanagement (PLM) – stellt für produzierende Unternehmen mit hohem Anteil komplexer Engineering-Leistungen eine Möglichkeit der Integration von Prozessen, IT-Infrastruktur und Organisation dar [AbSL05]. Daher ist es sinnvoll, den Produktlebenszyklus als Orientierung für die Analyse von Prozessen, des darin benötigten Wissens und den entsprechenden Daten aufzufassen, um Möglichkeiten der Prozessunterstützung identifizieren und realisieren zu können.

Entlang des Produktlebenszyklus wird eine erhebliche Menge an Daten generiert (meist strukturierte Dokumente oder Datensätze mit numerischen und alphanumerischen Inhalten sowie CAD-Zeichnungen mit Geometrieinformationen), die idealer Weise in Produktdatenmanagement-(PDM-)Systemen abgelegt sind. Diese Daten enthalten das Wissen der an Planung, Entwicklung, Herstellung etc. beteiligten Experten aller Bereiche eines Unternehmens, da dieses Expertenwissen in den Prozessen des Unternehmens angewendet wird und damit in die Daten einfließt, die während der Prozesse erzeugt werden. Eine systematische Erfassung, Verwaltung und Nutzung des in den Prozessen angewendeten relevanten Wissens ist ein Ansatz, Prozesse entlang des Produktlebenszyklus sowohl zu unterstützen und zu optimieren als auch den Verlust von Expertenwissen zu vermeiden.

Einleitend sollen für die weiteren Betrachtungen Abgrenzungen der zentralen Begriffe der Wissens- (implizites ↔ explizites Wissen) und Prozesssicht (Wissensprozess ↔ wissensintensiver Prozess) vorgenommen werden. Implizites Wissen bezeichnet schwer formulier- und formalisierbares Wissen, dessen sich evtl. selbst der Träger des Wissens nicht bewusst ist und das damit nicht direkt durch IT-Werkzeuge erfassbar, abbildbar und anwendbar ist. Unter explizitem Wissen wird demgegenüber Wissen verstanden, das bspw. in Form von Regeln formalisiert und daher mit IT-Werkzeugen verwaltet und angewendet werden kann [NoTa97].

Bezüglich der Prozesse kann grundsätzlich zwischen „Wissenprozessen“ und „wissensintensiven Prozessen“ unterschieden werden: Wissensprozesse dienen dazu, die Ressource Wissen für die Durchführung der wissensintensiven Prozesse zu identifizieren, zu erfassen, aufzubereiten, zu verwalten und zur Nutzung zu verteilen. Demgegenüber zeichnen sich wissensintensive Prozesse dadurch aus, Expertenwissen im jeweiligen Anwendungsbereich zu erfordern, das beispielsweise durch Wissensprozesse zur Verfügung gestellt werden kann [Remu02].

Insbesondere in kleinen und mittelständischen Unternehmen (KMU) ist eine explizite Erfassung, Verwaltung und Nutzung des relevanten Wissens durch eigenständige und aufwändige Wissensprozesse organisatorisch kaum möglich. Daher muss für eine effiziente Unterstützung der wissensintensiven Prozesse eine transparente Integration der relevanten Wissensprozesse durch standardisierte Komponenten vorgenommen werden.

In diesem Beitrag wird eine Herangehensweise vorgestellt, mit der eine Unterstützung wissensintensiver Prozesse in einem mittelständischen Unternehmen – in diesem Szenario aus der Automobilzulieferindustrie – realisiert werden kann. In dieser Branche dominieren KMU, die ein hohes Maß an (Erfahrungs-)Wissen über ggf. lange Zeiträume aufgebaut haben und deren wirt-

schaftlicher Erfolg bzw. ihre Überlebensfähigkeit von diesem Wissen über Produkte, Herstellungsprozesse, Fertigungstechnologien, Kunden, Marktentwicklungen etc. abhängt. Dieses Wissen in Gegenwart der folgenden konkurrierenden Rahmenbedingungen

- Kostendruck (von Kunden- und Mitbewerberseite),
- steigende Produkt- und Prozesskomplexität und -qualität
- Verkürzung des Produktlebenszyklus insbesondere in den frühen Phasen

effizient zu finden und wieder verwenden zu können ist daher eine Herausforderung an Unternehmen, der durch geeignete Methoden und Werkzeuge zur Unterstützung wissensintensiver Prozesse begegnet werden kann.

Der Beitrag gliedert sich wie folgt: In Kapitel 2 werden wissensintensive Prozesse im Produktlebenszyklus charakterisiert. Zur Unterstützung dieser Prozesse wird der Ansatz einer Suche in Produkt- und Prozessdaten begründet und deren wesentliche Bestandteile erläutert. Kapitel 3 hat das Beispielszenario des wissensintensiven Prozesses der Angebotserstellung in einem mittelständischen Automobilzulieferunternehmen zum Gegenstand, untersucht dabei den Prozess sowie das dabei einfließende Wissen und stellt ein Werkzeug vor, das die Suche nach ähnlichen Produkten zur Wiederverwendung des eingeflossenen Wissens realisiert. Im vierten Kapitel werden relevante Forschungsarbeiten in benachbarten Forschungsgebieten dargestellt, bevor im abschließenden Kapitel eine Zusammenfassung sowie ein Ausblick auf weitere Forschungsarbeiten gegeben werden.

2 Unterstützung wissensintensiver Prozesse durch Wiederverwendung von Wissen

2.1 Wissensintensive Prozesse im Produktlebenszyklus

Wissensintensive Prozesse lassen sich beispielsweise durch Eigenschaften wie individuell variierende Abläufe, hohe Autonomie der Mitarbeiter, deren vergleichsweise breiter Entscheidungsspielraum bei der Erfüllung ihrer Funktion und geringe Unterstützung durch IT-Werkzeuge charakterisieren [Remu02], also durch eine prozessinhärente strukturelle Komplexität. Das in diesen Prozessen relevante Wissen lässt sich näher dadurch charakterisieren, dass es sich um sehr spezifisches Expertenwissen der jeweiligen Domäne handelt, dessen Anwendung in den

Prozessen unverzichtbar ist und daher einen erheblichen Wert darstellt [Star92]. Auch in Routinevorgängen kann Wissen erforderlich sein, das allerdings in diesem Kontext keine Klassifikation des Prozesses als wissensintensiv zulässt.

Insbesondere die Qualität und Effizienz der Durchführung wissensintensiver Prozesse sind daher für Unternehmen besonders wesentlich, da sie – wie etwa im Falle von Prozessen in der Produktentwicklung – für die Qualität der Produkte entscheidend sind oder aber – wie im Falle von Prozessen in der Planung – über die strategische Ausrichtung von Aktivitäten in der Entwicklung und im Vertrieb und damit über die Zukunft eines Unternehmens mitentscheiden. In Abschnitt 3.1 wird die Relevanz des wissensintensiven Prozesses der Angebotserstellung detailliert beschrieben. Abb. 1 zeigt eine Klassifikation von Prozessen nach deren struktureller Komplexität und Wissensintensität. Im oberen rechten Quadranten sind die wissensintensiven Prozesse aufgeführt, während im unteren linken Quadranten eher formalisierbare Routinetätigkeiten platziert sind. Hervorgehoben (Fettdruck) sind Beispiele für Prozesse, die in diesem Kontext und mit dem hier vorgestellten Lösungsansatz zu unterstützen sind. Tätigkeiten rund um die Entwicklung und Konstruktion neuer Produkte stehen dabei im Mittelpunkt. Darüber hinaus sind auch Prozesse relevant, die beispielsweise durch Analyse der bisherigen Produktpalette und der Entwicklung auf Kundenseite im Bereich des strategischen Vertriebs die Ausrichtung von Vertriebs- und Entwicklungstätigkeiten überprüfen und an die Marktentwicklung anpassen.

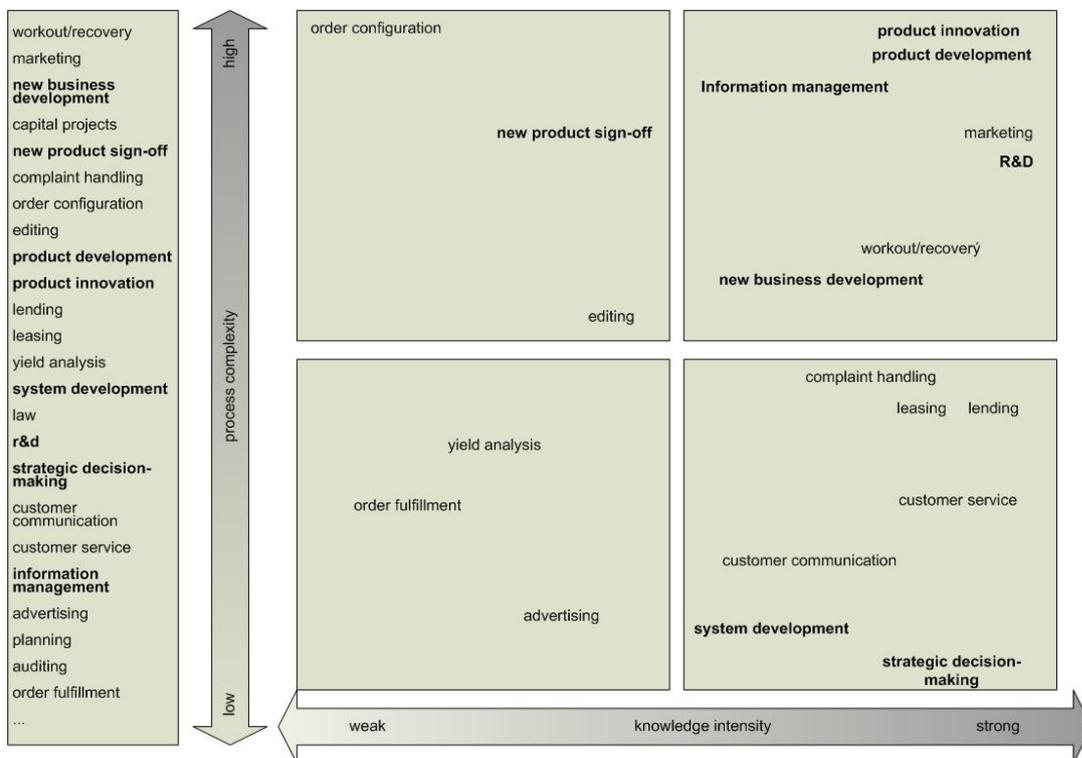


Abb. 1 Klassifikation von Prozessen nach Komplexität und Wissensintensität (vgl. [EpSR99])

2.2 Anforderungen an Werkzeuge zur Unterstützung wissensintensiver Prozesse

Wie bereits bei der Charakterisierung wissensintensiver Prozesse dargestellt, sind die inhaltliche wie strukturelle Komplexität der Abläufe sowie mangelnde Unterstützung durch IT-Werkzeuge allgemeine Merkmale.

Eine Unterstützung von Prozessen dieser Charakteristik stellt daher aus unserer Sicht an prozessübergreifende Lösungsansätze und entsprechende Werkzeuge die folgenden Anforderungen:

- **Prozess-übergreifende Einsetzbarkeit**

Verbreitet und angewendet wird Wissen in unterschiedlichen Phasen des Produktlebenszyklus, in unterschiedlichen Funktionen und mit unterschiedlichen Zielen. Daher sollte ein generischer Ansatz eine solche Nutzung unterstützen.

- **Unterstützung dynamischer Prozesse**

Entscheidungen, die auf Basis von Wissen in Prozessen getroffen werden, können auf unterschiedliche Art zustande kommen (z.B. in Abhängigkeit vom jeweils Entscheidenden). Die Unterstützung sollte daher losgelöst von a priori vorgegebenen Abläufen sein.

- **Kombinierbarkeit von Verfahren**

Wenn unterschiedliche Verfahren zur Nutzung von Wissen verfügbar sind, sollten beliebige Kombinationen von Verfahren möglich sein und unterstützt werden, um eine möglichst hohe Adaptierbarkeit an unterschiedliche Vorgehensweisen zu ermöglichen.

- **Integration mit PLM/PDM- und /oder ERP-Systemen**

Datenbasis für die Nutzung des Wissens sind verbreitet PLM/PDM- bzw. ERP-Systeme, die als Integrationsplattform eine konsistente Sicht auf betriebswirtschaftliche wie technische Produkt- und Prozessdaten ermöglichen [PBFL00]. Daher müssen Werkzeuge auf diesen Datenbestand zugreifen können, um Redundanzen und Inkonsistenzen zu vermeiden.

- **Anwendbarkeit in verteilten Szenarien**

Da der Lebenszyklus von Produkten und Herstellungsprozessen in zunehmendem

Maße geografisch verteilt und organisationsübergreifend durchlaufen wird, ist eine Unterstützung in einem verteilten Szenario anzustreben.

2.3 Generische Suchprozesse als Lösungsansatz

Hinrichs klassifiziert die Probleme, zu deren Lösung die hier betrachteten wissensintensiven Prozesse durchgeführt werden, als „open world“-Probleme, die sich durch generell unvollständiges Wissen zu ihrer Lösung auszeichnen [Hinr92]. Nach Hinrichs sind sie damit weder durch eine abgeschlossene Menge an Regeln hinreichend genau zu formulieren noch können Experten bei der Problemlösung durch regelbasierte Systeme unterstützt werden. Der Ansatz des fallbasierten Schließens („case-based reasoning“, CBR [AaPl94], s.a. Abschnitt 4.1) wird dahingegen als geeignet für die Problemlösung in solchen „offenen“ Anwendungsbereichen angesehen [AlBa96]. Der hier präsentierte Ansatz basiert daher auf einer generischen Suche nach Ähnlichkeiten in den Produkt- und Prozessdaten.

Die Suche als Werkzeug zur Wiederverwendung von Wissen bietet gegenüber bspw. regelbasierten Ansätzen den Vorteil, dass die Datenbasis und damit die Menge möglicher Ergebnisse und des verfügbaren Wissens nicht in Regeln transformiert werden muss. Da Regeln der Form *if Bedingungen then Aktionen* mit einem bestimmten Ziel (der Aktion) definiert werden müssen, ist die Verwendung eines einheitlichen Regelwerks in unterschiedlichen Prozessen problematisch. Z.B. kann kaum ein einheitliches Regelwerk sowohl in der Konstruktion als auch in der Kostenrechnung angewendet werden. Ebenso entfällt die Notwendigkeit der Pflege (Aktualisierung, Konsistenzerhaltung) einer wachsenden Regelbasis.

Bei der Analyse (Beobachtung, Interviews) der Arbeitsweise der Experten in ihren jeweiligen Funktionen entlang der Prozesskette werden Gemeinsamkeiten in der Vorgehensweise deutlich:

- Verwendung von funktionspezifischen Informationen
- Variierende Reihenfolge in der Analyse der Informationen
- Notwendigkeit variierender, fallspezifischer Informationen
- Nutzung von unvollständigen und unscharfen Daten/Informationen

Für einen generischen Suchprozess hat dies zur Folge, dass es rollenspezifische Mengen der Daten gibt, nach denen gesucht werden kann. Darüber hinaus ist sowohl die Kombination der Informationen als auch die Reihenfolge, in der gesucht wird, a priori unbekannt und kann fallweise variieren. Da eine Suche in unterschiedlichen Phasen entlang des Produktlebenszyklus

eingesetzt werden kann, ist es notwendig, auch mit unvollständigen oder unscharfen Informationen sinnvolle Ergebnisse erzielen zu können, falls überhaupt möglich. Daher ist eine unscharfe Suche (d.h. eine Suche mit vorzugebender Genauigkeit/Unschärfe, s. a. Abschnitt 2.4) nötig.

2.4 Komponenten der Suchprozesse

Voraussetzung für eine effiziente Suche innerhalb von Produkt- und Prozessdokumenten ist die Identifikation der Daten, in die das relevante Wissen eingeflossen ist. So ist beispielsweise ein Arbeitsplan, bestehend aus Maschinen, Beschreibungen von Arbeitsgängen, Parametern der Fertigung wie die Angabe „Hub pro Minute“ oder „Vorschub pro Hub“ etc., das Ergebnis eines wissensintensiven Prozesses, in den Wissen über Fertigungstechnik, Materialeigenschaften, Kosten etc. einging. Eine Suche nach diesem spezifischen Wissen sollte als Ergebnis Arbeitspläne (bzw. vollständige Produktdatensätze) liefern, die bzgl. einer Metrik ähnlich sind.

Es ergeben sich die folgenden Voraussetzungen für eine Implementierung von Suchprozessen:

- Identifikation von relevantem Wissen

Das in den Prozessen relevante Wissen hängt von den jeweiligen Funktionen und den Kompetenzen und Verantwortlichkeiten der Funktionsträger ab. Dieses Wissen muss qualitativ erfasst und in die Prozesse eingeordnet werden. Auf dieser Basis kann ein Datenmodell für die zu durchsuchenden Daten erstellt werden.

- Identifikation von Daten

Das identifizierte Wissen ist für die Suche auf die verfügbaren Daten abzubilden. Wenn bspw. das Wissen über Materialeigenschaften bestimmter Bauteile und Fertigungstechnologien relevant ist, sind die für die Suche relevanten Daten z.B. Stückliste und Arbeitsplan, während Informationen über Maschinenstundensätze hier irrelevant sind. Prinzipiell lassen sich diese Daten nach Typen klassifizieren, für die dann jeweils Ähnlichkeitsmaße spezifiziert werden können:

- Numerische Daten

Technische bzw. betriebswirtschaftliche, skalare numerische Daten wie etwa Materialstärke und -preise, Brutto-/Nettogewicht, Vorschub, Hubzahl, Instandhaltungskosten etc.

- Alphanumerische Daten

Beschreibungen von Produkten, Texte/Schlüsselworte von Beschreibungen

gen der Arbeitsgänge, frei formulierte Bemerkungen oder Identifikatoren (etwa Teilenummern gemäß einer unternehmensspezifischen Nummernsystematik)

- Geometrische Daten
Spezifikation der Geometrie eines Produkts durch CAD-Dokumente, Bitmap-Grafiken etc.

- Identifikation von Ähnlichkeitsmaßen

Einfachster Ansatz für die Definition der Ähnlichkeit numerischer Daten ist die Berechnung des Abstands zwischen Werten, der innerhalb einer vorgegebenen Schranke liegen muss. Für Vektoren kann bspw. die euklidische Distanz oder ein im Kontext sinnvolles Distanzmaß gewählt werden. Bei der Berechnung von Ähnlichkeiten zwischen alphanumerischen Daten wird in diesem Kontext die Ähnlichkeit einzelner Worte betrachtet. Dazu kann beispielsweise die Hamming- oder die Levenshtein-Distanz verwendet werden, die auf der (normierten) Anzahl von Einfüge-, Lösch- oder Änderungsoperationen basiert, die eine Zeichenkette in eine andere überführt (s. z.B. [Apos99]). Im Zusammenhang mit ingenieurtechnischen Anwendungen werden Geometrien typischerweise in Form von CAD-Dokumenten spezifiziert und ausgetauscht. Daher werden Ähnlichkeitsmaße zwischen Geometrien sinnvoller Weise ebenfalls auf CAD-Dokumenten definiert. Zur Berechnung der Ähnlichkeit zwischen Geometrien können Ansätze wie D2 [OFCD01], Lightfield Descriptor [Psra06] und Spherical Harmonics [SaVr01] eingesetzt werden. Sie unterscheiden sich in ihrem jeweiligen Ansatz, CAD-Dokumente zu analysieren. Es werden die Distanzen von zufällig ausgewählten Punkten auf der Oberfläche des Bauteils errechnet und miteinander verglichen (D2), zu vergleichende Objekte werden aus mehreren Perspektiven miteinander verglichen (Lightfield Descriptor) oder es werden Schnittmengen des Objektes mit konzentrischen Kugeln als Basis für einen Vergleich berechnet (Spherical Harmonics).

In Abb. 2 ist ein allgemeines Schema eines solchen iterativen Suchprozesses dargestellt, der in unterschiedlichen wissensintensiven Prozessen durchgeführt werden kann, aber in jedem Fall auf der gleichen Wissensbasis aufsetzt. Wesentlich ist, dass sowohl die Attribute als auch die Genauigkeit (oder „Unschärfe“), mit der gesucht werden soll, durch den Benutzer beliebig gewählt werden kann. Somit kann die Suche auf die Qualität und den Umfang der vorliegenden

Daten abgestimmt werden. Da die Genauigkeit von Daten entlang des Produktlebenszyklus (z.B. eine genauer werdende Beschreibung des Herstellungsprozesses im Arbeitsplan) typischerweise variiert, ist auch dies mit einer derartigen unscharfen Suche zu berücksichtigen. Darüber hinaus kann dadurch von Schreib- oder Übertragungsfehlern abstrahiert werden. Prozesse können dabei durch diese Suche auf das Wissen zurückgreifen, das in den produkt- und prozessbezogenen Daten enthalten ist. Da dieses Wissen nicht – etwa in Form von Entwurfsregeln o. ä. – explizit gemacht wird, ist es in einer impliziten Wissensbasis abgelegt und wird durch ein Wissensmodell, ein daraus abgeleitetes Datenmodell und einer Menge von Ähnlichkeitsmetriken zu diesem Datenmodell zugreifbar gemacht.

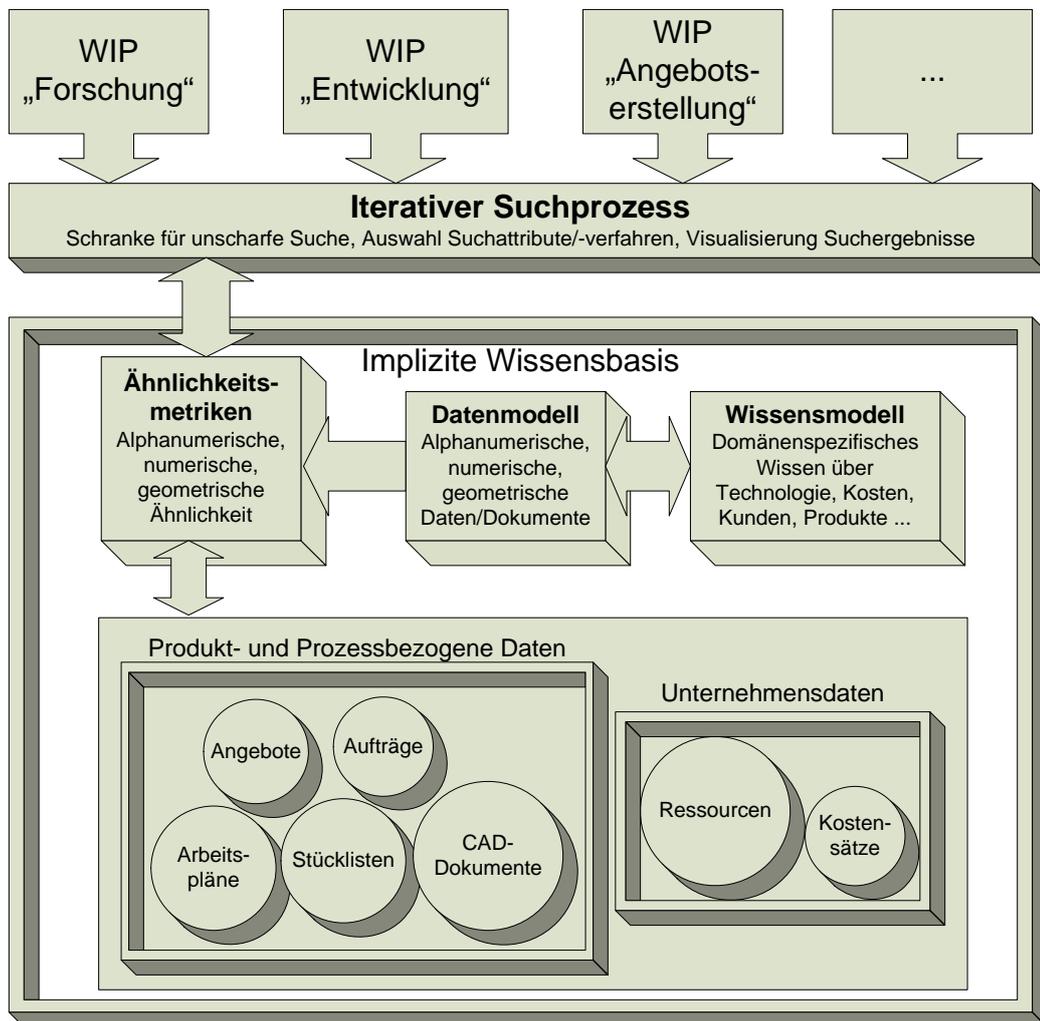


Abb. 2: Schematische Darstellung eines unscharfen Suchprozesses auf Produkt- und Prozessdaten zur Unterstützung unterschiedlicher wissensintensiver Prozesse (WIP) entlang des Produktlebenszyklus.

3 Unterstützung des Prozesses „Angebotserstellung“ in der Automobilzulieferindustrie durch Suchprozesse

3.1 Der wissensintensive Prozess „Angebotserstellung“

Der Prozess der Angebotserstellung zeichnet sich in vielen Unternehmen – in diesem Kontext mittelständische, auftragsorientierte Serienfertiger der Automobilzulieferindustrie – durch hohen Zeit- und Kostendruck aus, wobei gleichzeitig sehr hohe Anforderungen an das Wissen der Beteiligten gestellt werden. In komplexen und wissensintensiven Prozessen wie der Angebotserstellung wird bereits ein Großteil der bei der Herstellung anfallenden Kosten (ca. 70% [Bind98] bis 90% [Saku97]) festgelegt. Allerdings führen nur ca. 10% der erstellten Angebote zu Aufträgen, was eine effiziente Angebotserstellung notwendig macht [ToBT95]. Da insbesondere bei Unternehmen kleiner und mittelständischer Größe und Organisation – für die Automobilzulieferindustrie der dominierende Unternehmenstyp – nicht davon ausgegangen werden kann, dass notwendige Kompetenzen (Mitarbeiter) mehrfach vorhanden sind, kommt einer effizienten Nutzung des impliziten Wissens unabhängig von konkreten Personen erhebliche Bedeutung zu.

Das notwendige Wissen lässt sich dabei wiederum in unterschiedliche Bereiche gliedern:

- (Fertigungs-) Technisches Produkt- und Prozesswissen

In diesem Prozess muss die technische Machbarkeit eines Produkts mit (bestehenden oder neu anzuschaffenden) verfügbaren Maschinen und dem bestehenden Fertigungs-„Know how“ dargestellt werden („Herstellbarkeitsanalyse“). Ergebnis dieser Analyse ist ein (vorläufiger) Arbeitsplan und eine Stückliste, welche die Kosten des Produkts bestimmen.

- Betriebswirtschaftliches Produkt- und Prozesswissen

Aus den mit dem Ergebnis der Herstellbarkeitsanalyse verbundenen Kosten (Material-, Lohn- und Logistikkosten, Investitionen, etc.) muss ein Selbstkostenpreis ermittelt und auf dieser Basis ein Angebotspreis festgelegt werden.

In Abb. 3 ist ein Modell des Angebotsprozesses, seine Einordnung in die Prozesskette entlang des Produktlebenszyklus sowie das einfließende Wissen in einzelnen Prozessschritten informell dargestellt. In der oberen Prozesskette ist der Produktlebenszyklus eines Automobilherstellers angedeutet, der, nach abgeschlossener Planung und Konstruktion, Aufträge unter den Zuliefer-

unternehmen ausschreibt. Für auftragsorientierte Fertigungsunternehmen ist diese Anfrage der Startpunkt der Prozesskette zur Anfragebearbeitung. Zur Analyse der Anfrage ist bspw. Wissen über die aktuelle und zukünftige Produktpalette relevant, um die Wichtigkeit eines Auftrags für das eigene Unternehmen abschätzen zu können. Die Wahrscheinlichkeit einer erfolgreichen Angebotserstellung kann ebenfalls zu diesem Punkt auf Basis von Erfahrungen und Wissen über die eigenen Produktionsstätten und die des Kunden abgeschätzt werden. Auf dieser Grundlage wird entschieden, ob der ggf. kostspielige Prozess der Angebotserstellung überhaupt gestartet wird. Ist dies der Fall, wird ein technisches Konzept erstellt, in das etwa Wissen über die verfügbare und vom Kunden geforderte Fertigungstechnologie und die zu verwendenden Materialien und ihre Eigenschaften für ein bestimmtes Produkt einfließt. Wird auf dieser Basis die Machbarkeit positiv eingeschätzt, wird als Ergebnis eine vorläufige Stückliste sowie ein vorläufiger Arbeitsplan erstellt, der zur kostenrechnerischen Bewertung weitergeleitet wird. Auf Basis u. a. von Lohn- und Energiekosten an unterschiedlichen Standorten, aktuellen und prognostizierten Materialkosten über die gesamte Auftragslaufzeit (typischerweise mehrere Jahre) und anderen Kostensätzen wird ein Selbstkostenpreis ermittelt. Der daraus abgeleitete Angebotspreis beinhaltet ggf. Aspekte, die von der Wichtigkeit der Erlangung eines Auftrags abhängen. Für die eigentliche Erstellung eines Angebotsdokuments sind schließlich Kenntnisse der kundenspezifischen Anforderungen an das Format und die Inhalte eines Angebots erforderlich.

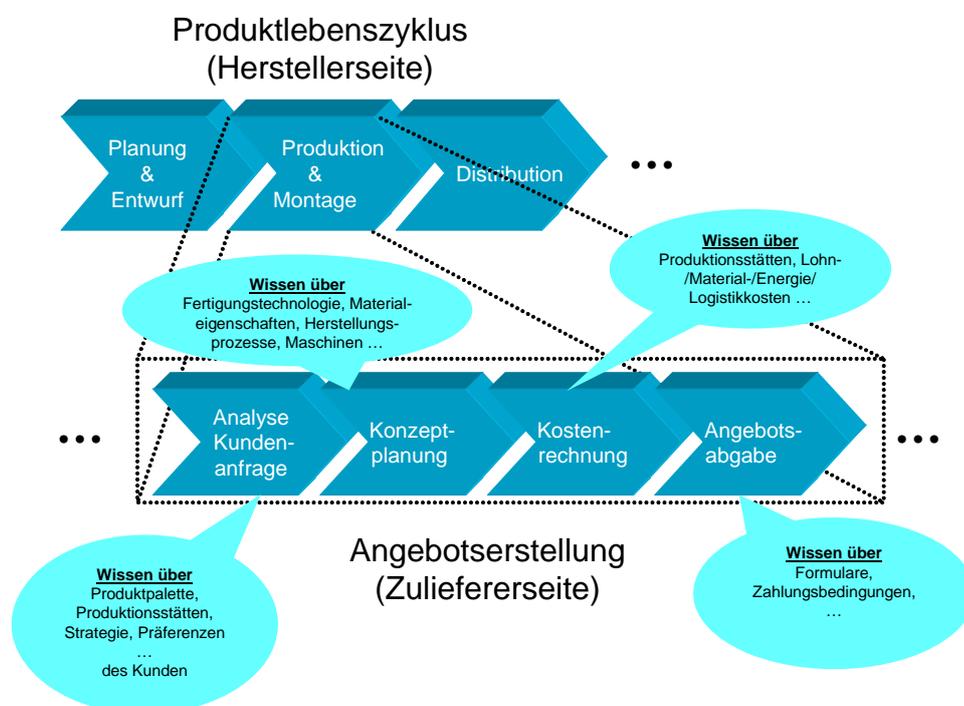


Abb. 3: Modell des Produktlebenszyklus (Ausschnitt oben) und der Angebotserstellung mit Beispielen des jeweils notwendigen Wissens

Im hier untersuchten Szenario war ein PDM-/PLM-System als gemeinsame Datenbasis noch nicht verfügbar. Daher wurde in einer Analyse des in diesem Prozess benötigten Wissens mit den in unterschiedlichen Anwendungssystemen verfügbaren Daten abgeglichen und ein Datenmodell erstellt (s. Abb. 4), welches den Suchraum des wieder verwendbaren Wissen definiert. In diesem Modell lassen sich unterschiedliche Arten von Wissen identifizieren: das notwendige Produktwissen wird repräsentiert durch die evtl. mehrstufige Stückliste. Komponenten einer Stückliste können dabei wiederum Stücklisten sein, die als Einzelteile (Baugruppen) in ein komplexeres Produkt einfließen. Neben selbst hergestellten Stanz- oder Walzteilen können Verpackungen, fremdbezogene Zukaufteile und Handelswaren Bestandteil einer Stückliste sein. Das Prozesswissen wird in diesem Modell abgebildet durch den Arbeitsplan, der eine Folge von Arbeitsgängen enthält. Jeder Arbeitsgang kann mit Werkzeugen ausgeführt werden und auf einer bestimmten Walz- oder Stanzmaschine (gegeben durch eindeutige Arbeitsplatznummern), von Hand oder fremd vergeben durchgeführt werden. Jenseits dieser beiden Arten von Wissen ist Wissen über die unternehmensweiten Ressourcen (Standorte, standortspezifische Kostensätze, Maschinen) und die jeweiligen kundenspezifischen Projekte relevant.

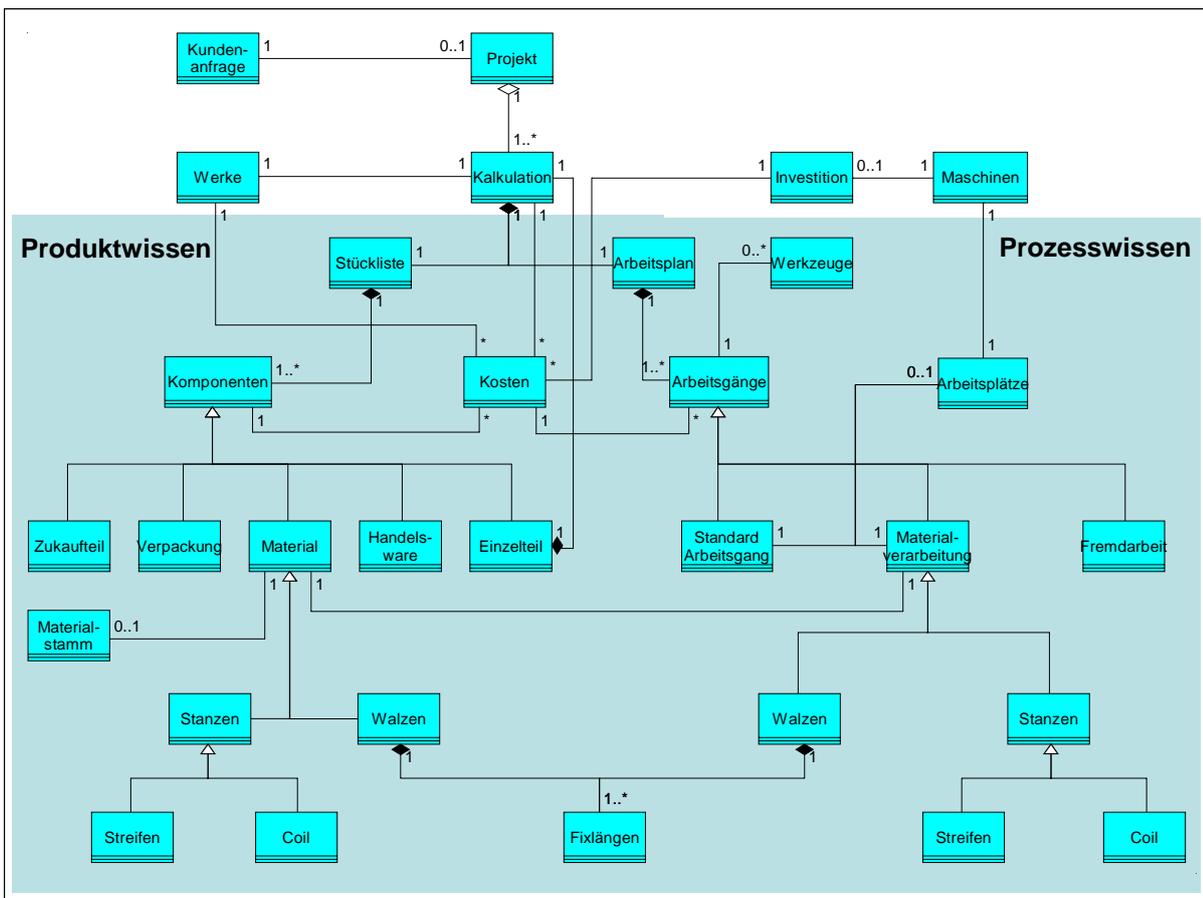


Abb. 4 Ausschnitt des Datenmodells des identifizierten Wissens

3.2 Ein Suchwerkzeug für Produkt- und Prozessdaten

Für beliebige Suchprozesse auf den identifizierten Daten (s. Abb. 4) und mit einer variablen, vorzugebenden Genauigkeit (oder Unschärfe) wurde ein datenbankbasiertes Werkzeug in Java prototypisch implementiert. Zur Aufbereitung der Suchergebnisse wurde eine Visualisierung integriert, die CAD-Dokumente im verbreiteten Austauschformat STL (Standard Transformation Language) dreidimensional rotierbar und mit der Möglichkeit zur Vergrößerung darstellt. Da ein visueller Abgleich durch Experten nur unvollständig formalisier- und durch Ähnlichkeitsmetriken abbildbar, jedoch nach Aussage der Anwender unverzichtbar ist, kommt dieser Funktionalität in der Praxis hohe Bedeutung zu.

In Abb. 5 ist das Ergebnis eines iterativen Suchprozesses nach alphanumerischen Daten (Bezeichnung des Produkts) sowie nach aus den CAD-Dokumenten extrahierten Informationen zur Lage des Produkts im Raum dargestellt. Da sich die Bezeichnungen der Teile voneinander unterscheiden und Suchergebnisse ausschließlich auf Basis einer unscharfen Suche in diesen Bezeichnungen keine hinreichend kleine Trefferliste ähnlicher Produkte ergibt, kann durch die Verkettung mit der geometrischen Suche nach Lageinformationen eine reduzierte Ergebnisliste erreicht werden, die durch visuellen Vergleich der verbleibenden Suchtreffer dem Experten die Auswahl des ähnlichsten Produkts und damit die Wiederverwendung von Stückliste, Arbeitsplan etc. und dem darin enthaltenen Wissen erlaubt.

Nr.	SM-Ähnlichkeit	Primärschlüssel	Kundenname	Teilebezeichnung	Einzelteilbezeichnung	Teilenummer	Einzelteilenummer	Bearbeitungsart	Gesamtgewicht	Bruttogewicht	Nettogewicht
1.	92 %	anonymisierte Projektnummer	anonymisiert	DACHRAHMEN VO	DACHRAHMEN VO	anonymisiert	anonymisiert	Stanzen vom Coil	156,4	214,1	156,4
2.	89 %	anonymisierte Projektnummer	anonymisiert	DACHRAHMEN VO	DACHRAHMEN VO	anonymisiert	anonymisiert	Walzen	204,2	271,5	204,2
3.	86 %	anonymisierte Projektnummer	anonymisiert	DACHRAHMEN HI	DACHRAHMEN HI	anonymisiert	anonymisiert	Stanzen vom Coil	220,9	339,1	220,9
4.	74 %	anonymisierte Projektnummer	anonymisiert	DACHRAHMEN AU LJ/RE	DACHRAHMEN AU LJ/RE	anonymisiert	anonymisiert	Stanzen vom Coil	77,3	142,6	77,3

Abb. 5 Maske des Werkzeugs zur Visualisierung von Suchergebnissen anhand der CAD-Dokumente

Eine objektive und systematische Bewertung der Suchergebnisse zur Evaluierung der Genauigkeit der Suche anhand einer Metrik ist aufgrund der Vielzahl der Kriterien, die ein Experte an Ergebnisse anlegt, schwer zu realisieren. Bisher liegen qualitative Beurteilungen der Suchergebnisse durch eine Anzahl von Experten aus unterschiedlichen Bereichen der Automobilzulieferindustrie (Kostenrechnung, technische Konzeptplanung, Produktentwicklung) vor, die einheitlich diese Herangehensweise bestätigen und den Prototypen als einsetzbar in unterschiedlichen Prozessen der jeweiligen Bereiche einstufen. Insbesondere als integraler Bestandteil von PLM wird ein solches Suchwerkzeug als unverzichtbar eingeschätzt.

4 Related work

Zur Unterstützung wissensintensiver Prozesse – insbesondere in Anwendungsbereichen mit ingenieurtechnischem Hintergrund, wie beispielsweise in der Automobilindustrie – lassen sich drei relevante Bereiche identifizieren: Fallbasiertes Schließen („case-based reasoning“, CBR) wissensbasierte Konstruktion („knowledge-based engineering“, KBE) sowie das organisationale Wissensmanagement.

4.1 Fallbasiertes Schließen („case-based reasoning“, CBR)

Wie bereits in Abschnitt 2.3 dargestellt, kann die Ähnlichkeitssuche in Produkt- und Prozessdaten als Teil des CBR-Zyklus [AaPI94] aufgefasst werden. Der vollständige Zyklus enthält allerdings weitere Phasen („reuse“, „revise“ und „retain“), die explizit die Art der Wiederverwendung gefundener Fälle sowie die Pflege der Fallbasis behandeln. Diese Funktionalitäten werden im hier präsentierten Ansatz nicht weiter betrachtet, da zum einen die Wiederverwendung selbst als Teil des zu unterstützenden Prozesses eingeordnet wird und sich damit einer allgemeinen Betrachtung entzieht. Eine Versionierung/Revisionierung von in den Produkt- und Prozessdaten gefundenen Fällen (in diesem Szenario zum Beispiel die Erstellung einer Kopie eines Angebots als Vorlage und die nachfolgende Modifikation für eine neue Anfrage) ist Funktion eines PLM/PDM-Systems und nicht unbedingt in allen Prozessen erforderlich. Eine explizite Pflege der Fallbasis (Phasen „revise“ und „retain“) ist ebenfalls kein unabdingbarer Bestandteil, da die gefundenen Fälle/Dokumente durch die jeweiligen Benutzer gemäß des unterstützten Prozesses ausgewertet und beurteilt werden. Weiterentwicklungen des Ansatzes sind in Richtung einer automatisierten Analyse von Versionen der Dokumente möglich, um Fragen nach der Entwick-

lung eines Produkts bzw. eines Herstellungsprozesses über den gesamten Produktlebenszyklus hinweg – etwa bzgl. der verwendeten Technologie oder des Materials und die Konsequenzen auf die Kosten – beantworten zu können.

Zur Realisierung der Suche in Produkt- und Prozessdaten sind Verfahren zur Bestimmung von Ähnlichkeiten zwischen Datensätzen notwendig, die aus dem Information Retrieval bzw. Text Mining stammen (s. z.B. [BaRi99]); da sich im hier dargestellten Ansatz die verwendeten Verfahren im Bereich der Ähnlichkeitsanalyse von numerischen, alphanumerischen und geometrischen Daten auf Standardverfahren beschränken und derzeit nicht im Fokus der Betrachtung stehen, wird auf eine eingehende Diskussion verzichtet.

4.2 Wissensbasierte Konstruktion („knowledge-based engineering“, KBE)

Wissensbasierte Konstruktion hat vorrangig zum Ziel, Konstrukteure von Routinearbeiten bei der Erstellung von Produktentwürfen zu entlasten [Sand03]. Dies kann beispielsweise durch die Vorgabe und automatische Prüfung von Regelwerken und das entsprechende Erzeugen von CAD-Dokumenten realisiert sein [PBFL00]. Ebenso können standardisierte Entwurfsprozesse damit forciert werden. Mit Hilfe dieser Werkzeuge kann sich der Konstrukteur auf die komplexen Aspekte des Produkt- und Prozessentwurfs konzentrieren.

Eine Integration mit PLM/PDM-Systemen als unternehmensweiter Basis für Produkt- und Prozessdaten ist für Systeme des wissensbasierten Entwurfs sinnvoll, um eine konsistente Datenbasis für die Wiederverwendung von Wissen nutzen zu können [Vajn01, PBFL00]. Allerdings ist der generelle Ansatz losgelöst von der Integration mit PLM/PDM-Systemen, um eine möglichst breite Einsetzbarkeit zu gewährleisten.

Insgesamt ist der Ansatz der wissensbasierten Konstruktion jedoch auf die Unterstützung von Konstrukteuren und dabei schwerpunktmäßig auf die Erstellung von CAD-Dokumenten ausgerichtet und hat damit die ingenieurtechnischen Aspekte der Produktentwicklung und Konstruktion im Zentrum. Eine Abstraktion des hierbei verwendeten Wissens und dessen Nutzung in anderen Prozessen wird nicht thematisiert. Der prinzipielle Ansatz der Extraktion von Wissen in Form von Entwurfsregeln und die darauf basierende Unterstützung von Prozessen ist allerdings als viel versprechend einzustufen, falls die Regelbasis automatisch erzeugt und immer aktuell gehalten werden kann. Die manuelle Pflege (und Erhaltung der Konsistenz) einer Regelbasis ist für komplexere Fälle als in der industriellen Praxis als kaum umsetzbar einzuordnen. Im hier präsentierten Ansatz wird auf eine Extraktion von Wissen in Form von Regeln verzichtet (s. a. Abschnitt 2.3).

4.3 Organisationales Wissensmanagement

Wissensmanagement stellt eine Vielzahl von Modellen und Methoden zur Verfügung, um Wissen (der Mitarbeiter) innerhalb eines Unternehmens systematisch zu erhalten und zu verteilen [MeHV03]. Dazu wurden Ansätze wie Wissenslandkarten, Content Management, Gelbe Seiten, Communities, „lessons learned“, „best practices“ und Suchmaschinen [Remu02] entwickelt, die den Austausch von Wissen – manuell oder halbautomatisch – unterstützen können. Daher sind sie auf die Kompetenzen der Mitarbeiter angewiesen, Wissen formalisieren und speichern zu können und darauf, dass dieses persönliche Wissen aktiv ausgetauscht und dieser Austausch gewollt wird. Daraus resultieren zwei Nachteile: Zur Unterstützung von Prozessen nutzbares Wissen muss explizit verfügbar und dazu durch den Wissensträger expliziert worden sein. Zum anderen wird durch diesen Ansatz nur Wissen erfasst, das auch dem Wissensträger bewusst ist.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Ansatz zur Unterstützung wissensintensiver Prozesse durch Suche in Produkt- und Prozessdaten vorgestellt. Die Relevanz wissensintensiver Prozesse wurde erläutert und diese von Wissensprozessen abgegrenzt. Nach der Anforderungsanalyse und Motivation der Suche als Lösungsansatz wurden die Komponenten des Suchprozesses eingeführt und erläutert. Der Prozess der Angebotserstellung wurde als Rahmen zur Bestätigung der Anwendbarkeit dieser Systematik ausgewählt und detailliert vorgestellt. Über ein Wissensmodell wurde ein Datenmodell erstellt, auf dessen Basis die Suche vorgenommen werden konnte. Zur Überprüfung des Ansatzes und der Komponenten wurde ein Suchwerkzeug prototypisch implementiert.

Die Weiterentwicklung dieses Ansatz beinhaltet zwei Aspekte: Zum einen den Entwurf und die Implementierung eines service-orientierten Frameworks, das die Integration in verteilte Workflows auf Basis von Web bzw. Grid Services erleichtert. Ein weiterer Ansatzpunkt für zukünftige Arbeiten ist die Entwicklung einer allgemeinen Vorgehensweise für den Einsatz von Suchprozessen zur Unterstützung wissensintensiver Prozesse. Dazu ist geplant, einen Kriterienkatalog für die Identifikation der wissensintensiven Prozesse, des darin enthaltenen relevanten Wissens sowie der zu untersuchenden Daten aufzustellen. Auf dieser Basis lassen sich zu Prozessen und Daten „passende“, effiziente Suchprozesse generieren, die sich dann in entsprechende Workflows der Prozesse im Produktlebenszyklus integrieren lassen. Eine Optimierung dieser

Suchprozesse ist etwa durch Analyse der Vorgehensweise bei der Suche (z.B. Auswahl und Reihenfolge der Attribute, Wahl der Genauigkeiten, Länge der Suchprozesse bis zum Erreichen einer hinreichend kleinen Treffermenge oder Abbruch) und der entsprechenden Anpassung der Suchmasken und -attribute geplant. Durch Prozess- und Benutzerspezifische spezifische „Muster“ für (Erfolg versprechende) Suchprozesse könnte die Effizienz der Suche gesteigert werden.

Danksagung

Teile der hier präsentierten Arbeiten wurden durch das Bundesministerium für Bildung und Forschung im Rahmen der D-Grid-Initiative im Projekt In-Grid (<http://www.ingrid-info.de>) unter #01AK806C gefördert.

Literaturverzeichnis

- [AaP194] Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches, AI Comm., IOS Press, 7 (1), S. 39-59.
- [AlBa96] Althoff, K.-D., Bartsch-Spörl, B.: Decision support for case-based applications, Wirtschaftsinformatik 38 (1996) 1, S. 8-16.
- [Apos99] Apostolico, A.: General Pattern Matching. In: Atallah, A. (eds.): Algorithms and Theory of Computation Handbook. CRC Press 1999, S. 13-5 - 13-10.
- [AbSL05] Abramovici, M., Schulte, S., Leszinski, S.: Best Practice Strategien für die Einführung von PLM. In: industrie management 2/2005, GITO-Verlag.
- [Bind98] Binder, M.: Produktkostenmanagement in Entwicklung und Konstruktion. Gabler 1998.
- [BaRi99] Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval, Addison-Wesley 1999.
- [EpSR99] Eppler, M., Seifried, P., Röpnack, A.: Improving Knowledge Intensive Processes through an Enterprise Knowledge Medium. SIGCPR Conf., New Orleans 1999.

- [Hintr92] Hinrichs, T.: Problem Solving in Open Worlds – A Case Study in Design, Lawrence Erlbaum Associates 1992.
- [MeHV03] Mertins, K.; Heisig, P.; Vorbeck, J.: Knowledge Management. Springer 2003.
- [NoTa97] Nonaka, I., Takeuchi, H., Die Organisation des Wissens, Campus-Verlag 1997.
- [OFCD01] Osada, R.; Funkhouser, T.; Chazelle, B. und Dobkin, D.: Matching 3D Models with Shape Distributions. In: Shape Modeling International (2001), S. 154-166.
- [PBFL00] Penoyer, J., Burnett, G., Fawcett, D., Liou, S.-Y.: Knowledge based product life cycle systems: principles of integration of KBE and C3P, Computer-Aided Design 32 (2000), S.311-320.
- [Psra06] Princeton Shape Retrieval and Analysis Group: Shape Representations. <http://www.cs.princeton.edu/gfx/proj/shape/>, (Webabruf 2006/03/23).
- [Remu02] Remus, U.: Prozessorientiertes Wissensmanagement – Konzepte und Modellierung, Dissertation, Universität Regensburg 2002.
- [Saku97] Sakurai, M.: Integratives Kostenmanagement. Vahlen, München 1997
- [Sand03] Sandberg, M.: Knowledge-based engineering – in product development, Tech. Bericht, LTU-TR-2003/05-SE, Technische Universität Lulea 2003.
- [SaVr01] Saupe, D., Vranic, D. V.: 3D Model Retrieval with Spherical Harmonics and Moments. DAGM-Symposium 2001, S. 392-397.
- [Star92] Starbuck, W.: Learning by Knowledge-intensive Firms, Journal of Management Studies 29 (6), 1992, S. 713-740.
- [ToBT95] Tönshoff, H. K.; Brunkhorst, U.; Tracht, K.: Angebotsplanung in der Einzelfertigung. In: CIM-Management (1995) 5, S. 42-45.
- [Vajn01] Vajna, S.: Approaches of Knowledge-based Design. In: Proc. Int. Design Conf. 2002, Dubrovnik, S. 93-100.

Mobile Knowledge Management Support in Fire Service

Organisations

Andrea Freßmann, Ralph Bergmann

Lehrstuhl für Wirtschaftsinformatik II
Universität Trier
54286 Trier, Germany
{fressmann,bergmann}@uni-trier.de

Brian Taylor, Ben Diamond, Gary Carr-Smith

Fire Service College
Moreton in Marsh, Gloucestershire, GL56 ORH, Great Britain
btaylor@fireservicecollege.ac.uk, Ben.Diamond@wmfs.net, Gary.Carr-Smith@avonfire.gov.uk

Abstract

In emergency domains like fire services, the need for support systems enabling instant mobile access to mission-critical knowledge from different heterogeneous knowledge sources has increased significantly during the past years. A comprehensive socio-technical study that focuses on the current state of working and on the needs of the people working in fire services was conducted within the European Project AMIRA (Advanced Multimodal Intelligence for Remote Assistance). Based on the results of this study, the AMIRA system was designed, which supports knowledge management by a system that integrates various knowledge sources through an intelligent mediator component called *CAKE (Collaborative Agent-based Knowledge Engine)*. The knowledge can be assessed through multi-modal mobile and stationary clients. The developed system has been implemented as prototype and successfully evaluated in a comprehensive evaluation based on user trials.

1 Introduction

In recent years, the demand for support systems for emergency services, such as fire services, has increased significantly in order to improve methods for all types of protection. Today, a

focus is put on training, qualifying and supporting the members of emergency services, but there is also an increased awareness of the potential of systematic knowledge management. In particular in fire service organisations, there is a demand for supporting inexperienced *Incident Commanders (ICs)*. As first attendance at incident grounds, ICs are in charge of the fire crew and are required to make dynamic decisions in a safety and risk critical environment, utilising a range of skills in a time critical manner. This decision making process can be improved by providing knowledge of different kinds, for different purposes, and from different sources, according to the particular emergency situation. This knowledge provision is embedded in a collaborative process and involves several people in the fire service working with different roles in different locations. To address this challenge, the European AMIRA (Advanced Multimodal Intelligence for Remote Assistance) R&D project was initiated. AMIRA has particularly the objective of developing advanced knowledge-based, mobile support techniques integrated with recent speech and search technology. The resulting technology aims at supporting users by obtaining and providing knowledge for making decisions in time and business critical situations.

To establish a solid ground for the development of such a knowledge management technology, a comprehensive analysis of the requirements of fire service organisations has been performed. It was necessary to analyse in advance what the actual needs of users in fire service organisations are, what they use today, and what they could realistically need to use tomorrow. Particularly, the flow of information within fire service organisations was analysed with the goal to improve information exchange among the collaboratively working fire service staff in the future. This socio-technical study has been conducted in close collaboration with the *Fire Services College (FSC)*, which is the UK governmental training and research centre establishment for UK fire services. Based on this requirements analysis, the AMIRA system has been designed and implemented. It supports knowledge management through a support system that integrates various knowledge sources through an intelligent mediator component called *CAKE (Collaborative Agent-based Knowledge Engine)*. The people involved in a fire service organisation can access all the available knowledge through multi-modal mobile and stationary clients in their particular working environment.

The remainder of the paper is organised as follows. Section 2 presents the main results of the socio-technical study. Section 3 sketches the AMIRA system, which is evaluated in detail in Section 4. Finally, a short conclusion closes this paper.

2 Socio-technical Study

The goal of the socio-technical study is to elicit the user needs of fire service members with respect to optimisations of their current state of working. In accordance with human behaviour, user opinions and working conditions three points were figured out that had to be taken into consideration to reach the analysis goal: analysis of the current state of working practices, elicitation of user requirements, and development of use cases. Based on these points a methodology consisting of four different analysis levels was designed:

- **Information source interrogation and analyses:** analysis of existing documentations, studies, and data sources;
- **Interviews and questionnaires:** knowledge acquisition from users;
- **Requirement assessment:** elicitation of user requirements and use cases;
- **Requirement evaluation:** verification of the requirements and use cases.

This methodology was also aligned to socio-technical research in information society [Huws02, Huws04, Soci02] for ensuring a comprehensive view on user needs. In the following the main achievements of the conducted socio-technical study [FrBe05] are presented.

2.1 Current State of Working Practices

For elaborating the current state of working practices, related projects, available documents and existing studies (e.g. a study from the “Chief and Assistant Chief Fire Officers Association” (CACFOA) [CACF01]) were analysed. Furthermore, questionnaires were developed that were specially tailored to young ICs in order to acquire knowledge about user needs and to uncover deficiencies in terms of the required support for ICs. These questionnaires and interviews were conducted at FSC at which representatives of different UK fire services took part. Altogether 78 fire fighters and ICs were involved in the interviews and debriefings. As a result, several typical important skills of people involved in fire service organisations were identified [FrBe05]:

- Keen perception for capturing situations;
- Flexible acting in a fast changing environment;
- Retaining facts in memory;
- Collaborative working.

Both, the first and the second skill is highly time critical in emergency situations and is needed at both the incident ground and the supporting level in terms of fire control personnel and human experts. Keen perception for capturing situations is the precondition for identifying the nature of incident and further for recognising signs for special dangers, i.e. hazardous materials, leaking vessels and poisons. By ignoring such signs, fire fighters take a risk of misconceiving incidents that could lead to a crucial issue. Above all, it depends on the reports of the fire fighters how ICs make their decisions. Beyond keen perception, fire fighters have to cope with occurring external events having impact on their decisions and incident procedures.

The third skill concerns long-term processes with which fire fighters have to cope. Because of complexity and diversity of incidents, fire fighters have to retain much knowledge about their operations. It is hard for fire fighters to keep such a lot of facts in their minds and to undertake further study, particularly for novices or fire fighters who work in the auxiliary fire brigade.

The last skill addresses all kinds of collaborative working during operations, e.g. information exchange between colleagues and the collaboration with experts. When ICs are not sure in making a decision they request experienced colleagues or fire control for support; communication is a big issue in fire services.

Another result in terms of current state of working practices was that over 30 knowledge sources are in use during work, e.g. operational notes, fire facts, incident information, location information [FrBe05]. Some knowledge sources are only available on papers, other knowledge sources can be accessed via electronic devices, and colleagues or human experts can be seen as knowledge sources as well. Usually, for fire fighters who wear their working clothes it is very cumbersome to access these knowledge sources and, additionally, they have to be aware what kind of information is provided by what knowledge source.

2.2 User Requirements

User requirements were deduced from information source interrogations and from the conducted interviews. Because different fire services have different requirements, only those requirements that are common between UK fire services are presented. Analysis of the interview transcripts and questionnaires highlights several specific issues [FrBe05]. One main issue is the surfeit of knowledge sources that are in use in fire service organisations. Some fire fighters talk about the danger of information overload or overflow. Due to this overload of information, a large number of text-based documents have to be read in order to determine the specific information wanted. In summary, the following requirements were identified:

Information support (R1): ICs should be supported with information and working instruction for decision making. Only ICs demand such support because fire services are hierarchically structured, so that other fire fighters involved in the incident should get information only from the responsible IC.

Mobility (R2): For supporting mobile workers, the envisaged information support must be able to provide information at any location the mobile worker can be: Mobile access is needed. While working in time and business critical processes, mobile workers demand short response time regarding information support.

Ease of use (R3): The usability of the envisaged computer support has to be aligned to the requirements during work, e.g. that fire fighters wear operational kits, such as helmets, goggles or breathing apparatus and thick gloves. The information support enables users to smoothly access information by a hands-free access. One solution is the integration of a speech dialogue system via which the ICs are able to access the information support system.

Core information (R4): The IC may not be suitably supported by highly complex information because there is no time to analyse the information in action. An overload of information has to be avoided; the emphasis must be on the concise, easily understandable core information.

Reduction of the number of systems in use (R5): Because of a large number of information sources and different access methods there is a need for reducing the number of systems in use for facilitating improved working on location.

Search on structured and unstructured data (R6): For integrating the existing information sources, the envisaged system requires sufficient integration methods due to disparate formats of information, e.g. structured or unstructured.

Knowledge exchange and collaborative working (R7): It is crucial to consider information paths within organisations: The control centre which coordinates mobile workers in action is lacking information from the field. Therefore, an enhanced information support is demanded that allows notifications and collaboration among users.

Pro-active information support (R8): In relation to the current work, the fire control lacks information about what is going on in the field. Hence, they should be pro-actively supported in getting information about their operatives.

Integration of best practices (R9): Novices and inexperienced users often lack sophisticated knowledge about *how* to assess the required knowledge in a time-efficient manner and how to avoid irrelevant results. Therefore, they need support to perform an adequate search process.

Information capturing (R10): Based on the interviews conducted, a need for procedures to update existing information sources becomes obvious. Up to now, update processes are not executed very frequently, so that knowledge sources may become less reliable. Structuring of update and review procedures would be a huge benefit for fire service domains.

User profiles (R11): User profiles are required that ensure the users' authenticity and authorisation.

This list of user requirements is not exhaustive but establishes an understanding of user needs within the fire service organisations. In the scope of this paper, the requirement R9 is of particular importance: The reuse of search processes demands process support in order to provide means of efficient combining result items from different information sources.

2.3 Use Cases

The main goal of the socio-technical study is the research of appropriate ways to integrate and enhance knowledge management in order to support mobile workers during collaborative work. The development of a prototypical AMIRA system as demonstrator of the results is one of the main outcomes. The system implements three use cases tailored to the user requirements.

2.3.1 Use Case 1: IC requests information via a speech dialogue system

Within the fire services the IC is the only person who is in charge and responsible for the decisions made at the incident. Therefore, all fire fighters involved provide her/him with precise details of the incident. The IC is then able to make decisions on how to proceed. Consequently, there is a need to support the IC when he lacks information necessary for decision-making and in estimating current resources. Instead of going back to the fire appliance, the IC should be supported in requesting information and receiving answers by using a mobile system which can be requested by speech. Figure 1 shows the first use case.

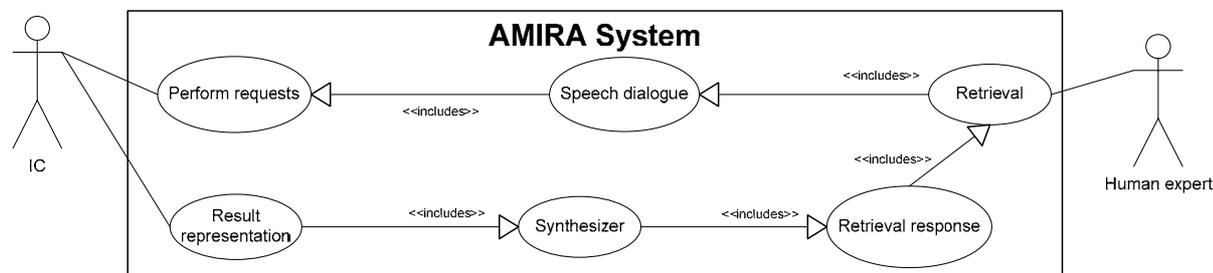


Figure 1: Use Case 1 represented as UML Use Case Diagram (please note that the arrows do not represent the flow of information but the inclusion of separately described sub- use cases)

One possible scenario of this use case is when the IC arrives at the incident, fire fighters tell him many facts about the incident but he does not know how to interpret the facts or how to act. For example, an inexperienced IC and his crew come to a road traffic accident where a truck is involved labelled with an *emergency action code (EAC)* indicating dangerous goods. The IC is not sure of the encoding, so that he uses the AMIRA system to perform a query which includes the EAC and a description of the accident like ‘vehicle accident’ and ‘fire affecting or likely to affect load’. To articulate the information request the IC uses a speech dialogue system as illustrated in Figure 1. After that the AMIRA system initiates a retrieval based on the information request. In the retrieval process human experts can also be involved. Finally, the retrieval response is synthesized into speech and presented to the IC. Thus, the IC receives instructions how to act and what has to be taken into consideration for further decisions.

2.3.2 Use Case 2: pro-active information support

By monitoring user-system interactions, the AMIRA system can elicit information directly related to the operation in which the operatives are currently working. By analysing this information, the AMIRA system is able to pro-actively support the fire control (working in the fire service headquarter) with information about the situation in the incident ground as shown in Figure 2. During or after the operation this information can be used as part of the operation debrief protocol or for reproducing actions or decision made during operation.

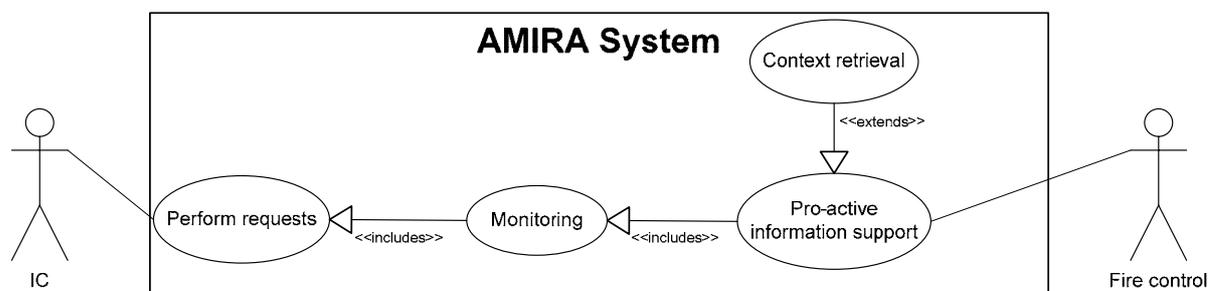


Figure 2: Use Case 2

For instance, in time critical situations it is imperative that the fire control is fully aware of the activities of fire fighters involved in the incident. Therefore, the information requests performed by the IC are of particular importance to the fire control: All requests are monitored and pro-actively delivered to the fire control. Beyond that, the system is able to retrieve additional context information by retrieving more information based on the monitored facts.

2.3.3 Use Case 3: post-incident analysis

Collaborative post-incident analysis of operations is managed by the AMIRA system. This encompasses pro-actively asking involved persons (control, IC and fire fighters) for information about their last actions concerning possible modifications to guidelines or work instructions.

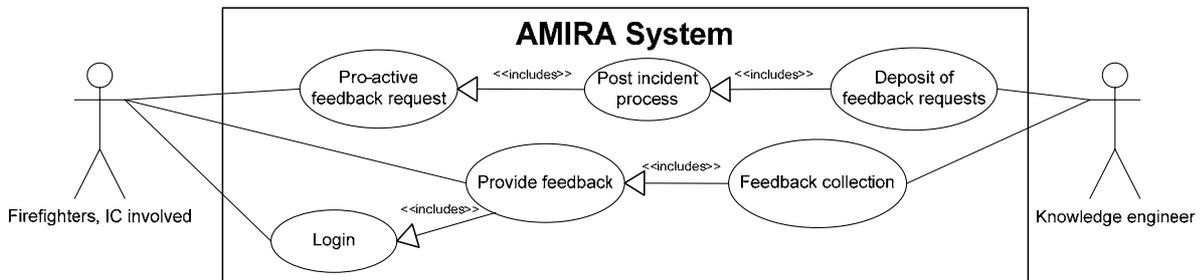


Figure 3: Use Case 3

After an incident a knowledge engineer initiates a post-incident process. Regarding the incident he/she requests feedback from the persons involved. The post-incident process sends the feedback request to all persons involved in the particular incident. Then the fire fighter or IC logs in the AMIRA system and formulates his/her feedback, which is collected in a repository. Later on, a knowledge engineer is able to access this repository, to analyse the reports, and – if necessary – to modify the information sources available within the organisation.

3 The AMIRA System

According to the socio-technical study, a knowledge management support for collaborative and mobile working was designed and realised through the AMIRA prototype system. Due to the space limitations of this paper, only short technical overview on the system is given here, focussing on the means for implementing the user requirements.

3.1 Overview

The fundamental idea of the system is to combine knowledge management with speech dialogue and search technology as depicted in Figure 4. For this purpose the *Collaborative Agent-based Knowledge Engine (CAKE)*¹ [FrMS05] has been developed. Beyond the ability to integrate various technology components (R5), CAKE interfaces with speech dialogue and

¹ For further details see also: <http://www.wi2.uni-trier.de/de/cms/projects/Cake/>

search components and acts as a mediator between them in order to guide searching on multiple information sources (R1). To support a seamless integration into working procedures, a speech dialogue component (R3) provides a mobile interface (R2) which can be used by ICs to articulate information demands. To make existing information sources accessible, appropriate search components can be connected to CAKE, so that heterogeneous information sources are integrated (R6) (e.g. search engines, knowledge bases, human experts) [Freß05].

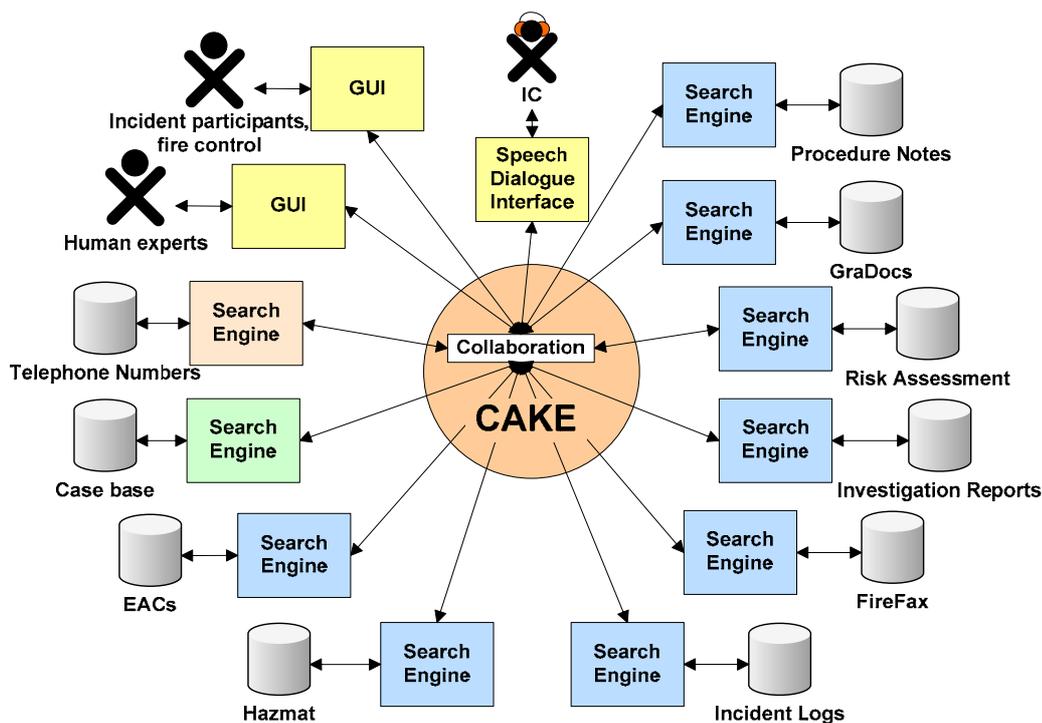


Figure 4: Architecture of the AMIRA System

In the AMIRA context, a sophisticated search engine from Fast DataSearch² provides indexing and searching of unstructured documents or data archives. Instead of indexing the whole documents only, this search engine supports the search of document fragments, called ‘scope search’. This avoids the retrieving of long and verbose documents; focusing instead on the most important ‘scope’ of the document (R4). Secondly, Kaidara³ Software provides a search engine based on a structural case-based reasoning approach [BABG04] and enables a search on past experiences that are represented by highly structured data in form of cases.

To access the system, *graphical user interfaces (GUIs)* are provided for human experts, fire control staff, and incident participants. This GUI which can be used on different stationary and

² www.fast.no
³ www.kaidara.com

mobile devices such as tablet PCs and PDAs, allows these experts to answer ICs' requests (R7). The fire control staff is pro-actively supported via a GUI (R8) and the incident participants can use a GUI in order to log in (R10) and to formulate feedback (R11).

Besides the user interfaces, Figure 4 also shows an excerpt of the information sources available in UK fire service organisations: chemical database ChemData, UN numbers (hazardous substances), procedure notes, EACs, human experts who answer information requests while doing their daily office work, etc. These information sources are related to special purposes and normally do not overlap each other by content, so that the competence area of each information source can be characterised differently. Consequently, the need for a meta-search process implementing efficient search strategies becomes obvious (R9). For example, clever search processes only request a relevant selection of available information sources and efficiently combine result responses retrieved from different information sources. The usage of such search processes enables novices to reuse search strategies of experienced users.

3.2 Unified Data Model

Mediating between search components and user interface components, CAKE addresses the challenge of dealing with differently structured search queries and search results of multiple information sources. In order to transport various queries and results through CAKE, it includes a flexible and unified data model which is constructed in an application specific manner according to the structure of the available knowledge sources and the supported types of queries. This *CAKE Data Model* describes all kinds of data that occur in the system. It is based on object-oriented modelling techniques and makes use of specialisation and aggregation. Available built-in data classes for model construction are atomic classes like Boolean, integer, double, date, or time as well as compound data classes such as aggregates, collections, and intervals. These so called *system classes* are used to define the application specific data model. We call this model a *unified data model* because it unifies all data formats and structures which occur in the AMIRA system in a consolidated structure. Queries and search results from the various GUIs and the speech dialog system are mapped to this data model and the various search engines are connected to the AMIRA system by means of wrapper agents that map the specific representation of the respective search engine to the data model. As a consequence, the unified data model allows to capture disparate search results and to transport these results within the AMIRA system.

3.3 Collaboration Patterns

The idea of integrating former search experiences in the form of processes is of particular importance in order to fulfil the requirements R1, R3–R6 and R9. Therefore, CAKE incorporates workflow technology [Freß05] in order to support real-world processes in general and search processes in particular. The process support allows different information sources to be requested in several steps and to stop when the search results have satisfied the request. A search process can be viewed as a collaborative process conducted by several search components in order to achieve a common result set. Within CAKE such collaborative processes are realised as so called *collaboration patterns* which are a specialisation of workflow definitions describing collaboration between workflow participants (humans and/or machines) based on best practice. Users are supported by collaboration patterns in automatically selecting appropriate information sources with respect to the current information need. Among other details, those search processes include information about which sources are relevant, which distribution techniques for the current query should be used, and how the result should be merged by use of well known methods from distributed information retrieval [Call00]. Figure 5 illustrates an example of a collaboration pattern.

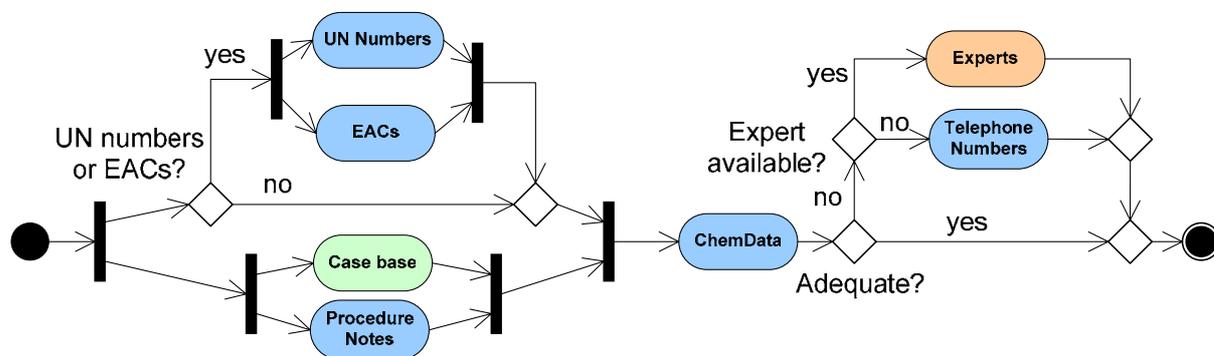


Figure 5: Example of Collaboration Patterns represented as UML Activity Diagram

According to the scenario described in the first use case, this collaboration pattern comprises information sources relevant to road traffic accidents involving hazardous materials. The processing of the pattern is as follows. Firstly, the case base and the procedure note collection are requested. Concurrently, the query is checked for EACs and UN numbers. If the query contains such encodings, information about EACs and/or UN numbers is retrieved. After that, the ChemData collection is requested. The retrieval does not need to be based on ICs query directly but on data retrieved from the first collections, so that additional information can be received in order to provide inexperienced ICs with more context information. Finally, in order

to ensure useful results, either human experts can be requested or experts' telephone numbers can be retrieved if no adequate result items have been achieved so far.

This is only one example of a collaboration pattern. According to the socio-technical study many best practices were encountered and hence several specific collaboration patterns were collected. Each pattern describes a search process for a particular kind of information request and is characterised by a meta-data description of the information requests for which it is appropriate (*collaboration pattern characterisation*). Consequently, for each individual request it is necessary to select the best possible collaboration pattern. CAKE incorporates special case-based retrieval component which automatically select an appropriate collaboration pattern with respect to the users' information request [Freß05]. In a nutshell, this retrieval process works as follows: After the IC performed an information request, the incoming request is analysed for keywords pre-defined in the unified data model. The keywords are used as query input for a similarity-based retrieval [FMMS05] allowing the request and available collaboration pattern characterisations to be compared at the semantic level. The most suitable collaboration pattern is then initiated for processing the current request. To translate the query for each underlying knowledge source, CAKE uses the unified data model.

4 Evaluation

In order to assess the AMIRA system against the socio-technical study, an evaluation based on 19 field trials were conducted between May and June 2006. Besides user requirements and use cases, conceptual aspects like collaboration patterns were also objects of evaluation. The use cases acted as starting points during the design of the trials and they were used as guidance to verify the expectations suggested by the socio-technical study. A further aim of the evaluation was to assess to what degree the user requirements were met. Therefore, hypotheses were formed and assessed in conjunction with domain experts.

4.1 Evaluation Design

In literature many evaluation methodologies are available. The presented evaluation design was developed in accordance with existing evaluation approaches and methods [Nati06, LaLS06] in order to evaluate the AMIRA system in the particular field of collaborative and mobile working. The evaluation design includes a methodology which specifies the conduction of field trials.

These trials are appropriate for testing the system in remote locations and under real-world conditions. Field trials are field experiments and build the basis for testing the ability, quality and performance of the AMIRA system. Based on these trials, conclusions established by observation and user opinions can be drawn in order to obtain an overall evaluation of the AMIRA system. The underlying evaluation methodology is divided into four steps:

1. Definition of concrete scenarios
2. Elaboration of questionnaires focused on scenarios, use cases and user requirements
3. Trial procedure consisting of three phases
 - (a) Conduction of trials with the AMIRA system at work
 - (b) Conduction of debriefs, interviews and questionnaires
 - (c) Intermediate trial analysis
4. Analysis of debriefs, interviews and questionnaires.

Concrete scenarios have been precisely delineated because they have direct impact on user instruction and field trials. Elaborating a questionnaire ensures the particular focus on the objects of interests and the consideration of the scenarios, use cases and user requirements. Questionnaires had to be well organised and carefully prepared because this step determines the data which is collected about the trials and which build the basis for analysis.

4.2 Evaluation Results

Field trials were conducted at the Avon Fire & Rescue Service, at Westmidland Fire Service and at the Swedish Fire College. At the end of each trial statistical data was collected during the debriefs and interviews which was the foundation of the final analysis. Therefore, several hypotheses to be met by the AMIRA system were formulated in advance and with respect to the socio-technical study. The most important hypotheses were:

- **H1:** The AMIRA system is useful.
- **H2:** The AMIRA system supports useful information.
- **H3:** The AMIRA system is well suited for a decision support system.
- **H4:** The AMIRA system means an added value to incident procedures.

We now present the evaluation results with respect to these hypotheses.

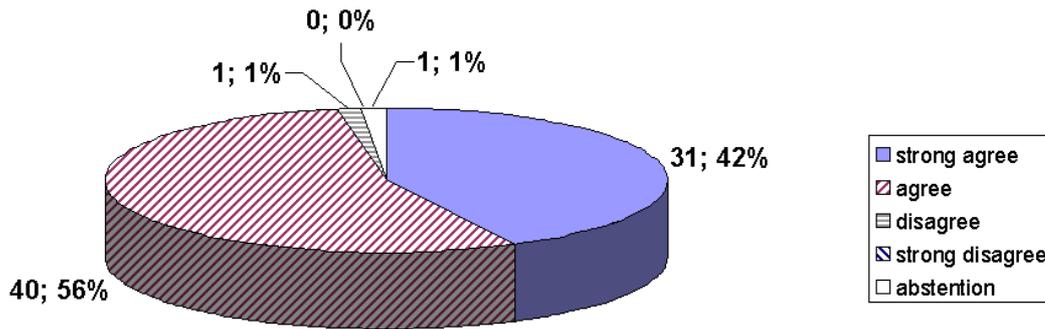


Figure 6: AMIRA system is useful

Altogether 73 fire fighters, ICs and fire service member playing representative roles within fire services are involved in the evaluation process. As depicted in Figure 6 nearly all of them (71 interviewees) agreed or strongly agreed that the AMIRA system was useful for fire service organisations. This leads to subjective evidence that this hypothesis holds.

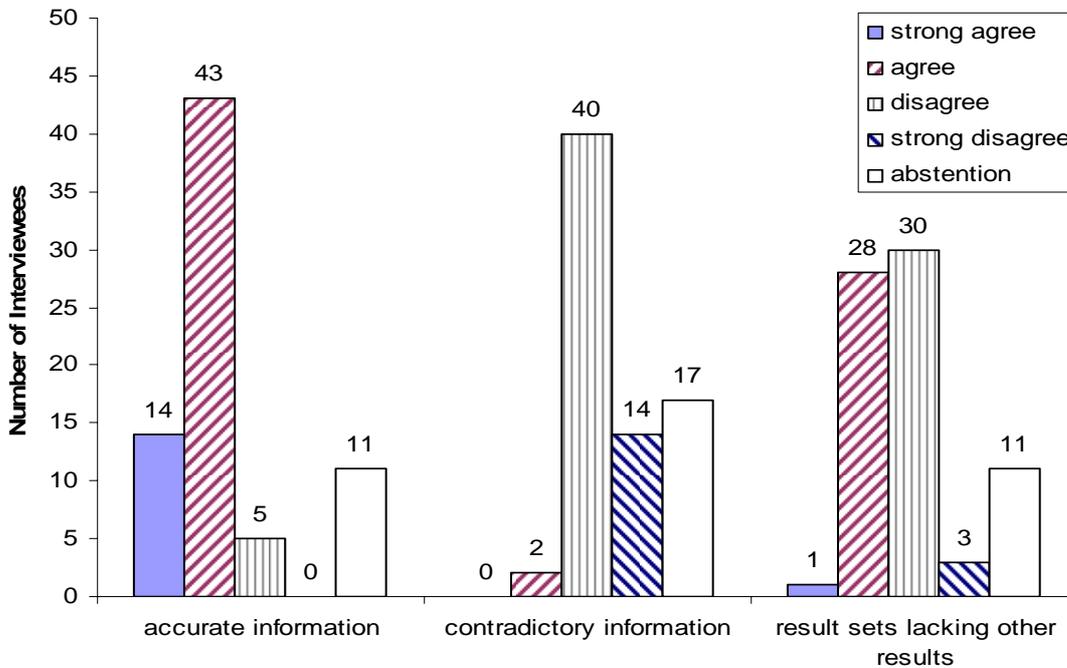


Figure 7: Quality Criteria

The second hypothesis states that the system provides ICs with information of high quality. For collecting opinions about the quality, questions concerning accuracy of information, contradictory information and incomplete result sets were included in the questionnaire. Figure 7 presents the questionnaire results of the quality criteria. Based on the corresponding answers, conclusions about precision and recall could be drawn. A majority of interviewees confirmed that the AMIRA system supported accurate information: 57 interviewees gave positive

assessments whereas only 5 persons disagreed. This fact stresses the precision achieved by the retrieved information. Additionally, the interviewees were asked to estimate the percentage of the accurate information: 69% of all interviewees estimated the percentage between 95% - 100%. The next question related to whether contradictory information was retrieved by the AMIRA system. The main trend was to disagree that contradictory information had been retrieved, however 17 abstained from voting. At last, half of the interviewees agreed that the result sets lack other results which would be of particular importance with respect to the current information request. This indicates a deficit in terms of recall but it can be traced back to the prototypical implementation of the AMIRA system, which does not include all organisational information sources.

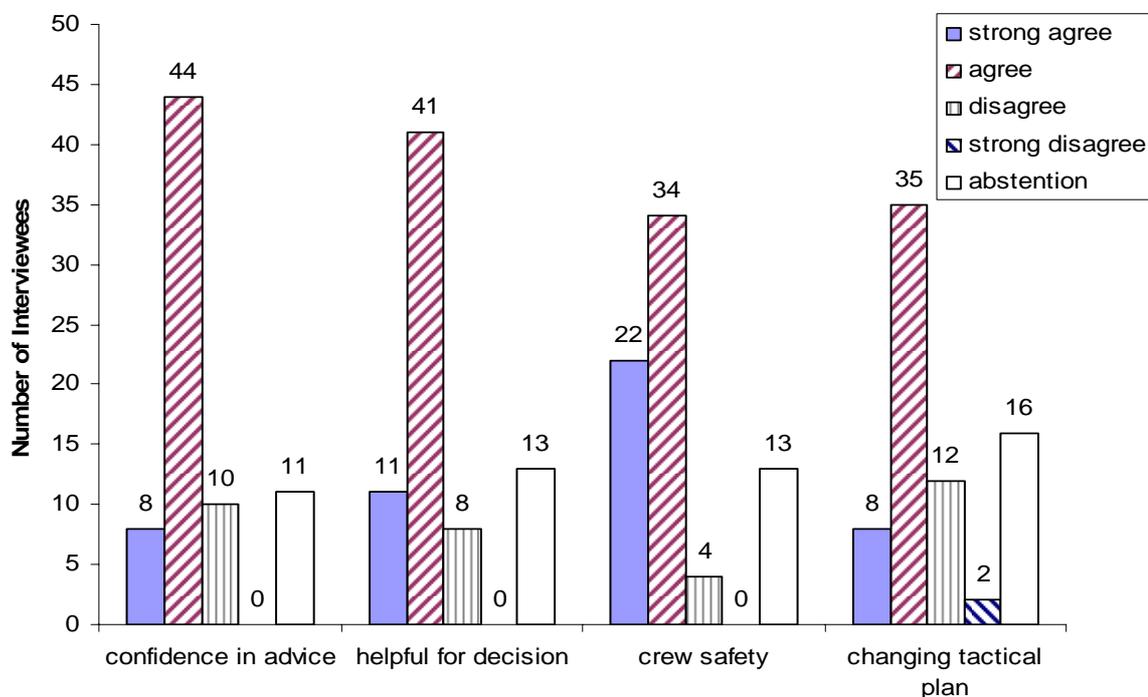


Figure 8: Decision-Making

H3 hypothesises that the AMIRA system is well suited as a decision support system. Therefore, the questionnaire included questions relating to confidence in the AMIRA advice, helpfulness for decision, crew safety and consequences triggered by the advice. All questions were positively assessed as illustrated in Figure 8. One of the most important results was that the interviewees were confident in the advice given by the AMIRA system. This confidence is fundamental for decision making. Beyond that, the advice was assessed as helpful for decision, particularly helpful for resolving the incident. Furthermore, the given advice supported crew safety which is of crucial importance for all fire services. Finally, the interviewees classified the

AMIRA system as decision support system by agreeing that they had changed their tactical plans based on the advice provided. As a result, H3 is agreed with respect to the questionnaires.

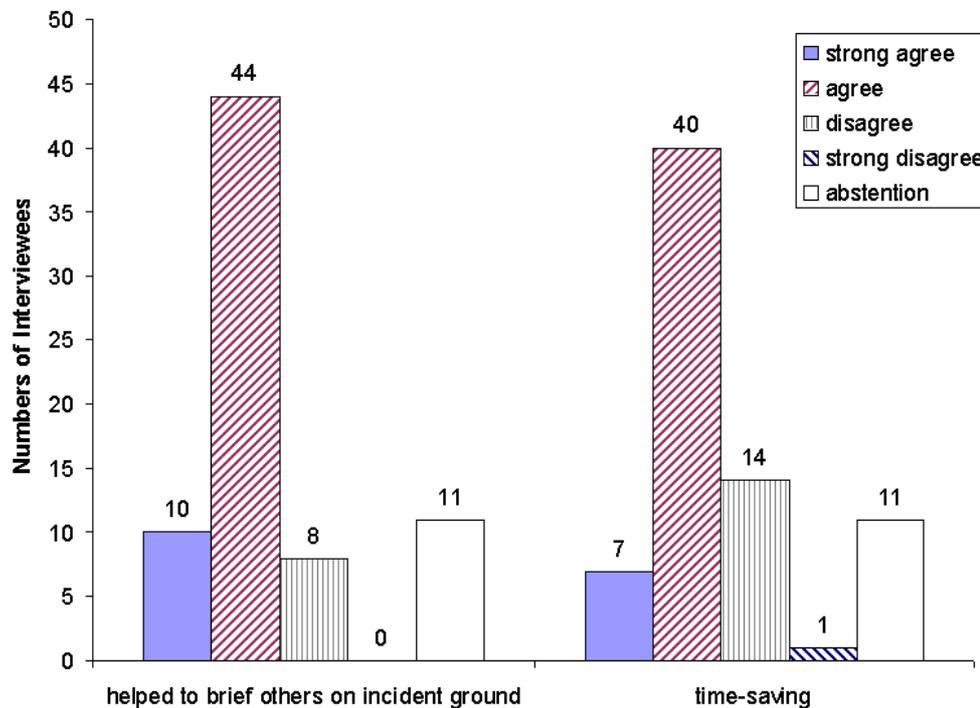


Figure 9: Consequences on Incident Processes

The last hypothesis concerns the consequences triggered by the AMIRA system. The results of the questionnaire are illustrated in Figure 9: On the one hand AMIRA advice helped to brief others on the incident ground and on the other hand time-saving was achieved by this system. Both perspectives are positively assessed by over half of the interviewees, so that H4 holds.

5 Conclusion

In this paper, a case study of developing a knowledge management support in fire service organisations is presented. Based on interviews and questionnaires, a socio-technical study was conducted in order to elaborate user requirements and use cases addressing current user needs and demands in practice. To meet these requirements, a concept of knowledge management support for collaborative and mobile working was developed that allows users and knowledge sources to collaborate in an efficient and user-oriented way. Within the AMIRA project this concept was put into practice and resulted in the AMIRA system. This system realises

collaboration of knowledge sources by incorporating collaboration patterns which are based on best practices and formally represent search processes on multiple data sources.

Finally, an evaluation was conducted for assessing the AMIRA system against the socio-technical study. The evaluation process was mainly based on field trials, conducted several times, and questionnaires. Since questionnaires collect user opinion by nature, the evidence of evaluation results was achieved from a subjective point of view. This emphasises the user-oriented approach of the AMIRA project. In summary, the usability of the system is acceptable, but to make it more easily applicable to a busy hands and busy eyes environment, the hardware used need to be developed further to provide a more compact package which is better integrated. Furthermore, the AMIRA system plays an important role with respect to innovation because current information systems applied in fire service organisation do not allow mobile access to ICs. Further results indicate that in the majority of cases, users think that information provided by the AMIRA system would bring incidents to a more effective conclusion and provide for improved crew safety. This will increase efficiency and provide a more cost-effective service to the communities which fire service organisations serve. Finally, strategic managers see this from a different perspective and are intrigued by the possibility of providing experiential knowledge and supporting information to the ICs. They see the emergency response environment as one where organisations are exposed to corporate risk.

Acknowledgement

The authors acknowledge the European Commission for funding AMIRA (www.amira.no) under grant number FP6, project IST-2003-511740.

Literature

- [BABG04] *Bergmann, R.; Althoff, K.-D.; Breen, S.; Göker, M.; Manago, M.; Traphöner, R.; Wess, S.: Developing Industrial Case-Based Reasoning Applications. The INRECA Methodology. 2. Aufl., Springer, LNAI 1612, 2004.*
- [CACF01] *Chief and Assistant Chief Fire Officers Association (CACFOA): Mobile Data Study. <http://www.fire-uk.org/mobdata/study/default.htm> verified on 2005-08-05, 2001.*

- [Call00] *Callan, J.*: Distributed Information Retrieval. In: *Croft, W.B. (Hrsg.): Advances in Information Retrieval*. Kluwer Academic Publishers, 2000, S. 127-150.
- [FMMS05] *Freyßmann, A.; Maximini, K.; Maximini, R.; Sauer, T.*: CBR-based Execution and Planning Support for Collaborative Workflows. In: Workshop "Similarities - Processes - Workflows" on the Sixth International Conference on Case-Based Reasoning (ICCBR 2005), Workshop Proceedings, Chicago, Illinois (USA), 2005, p. 271-280.
- [FrBe05] *Freyßmann, A.; Bendeck, F.*: Socio-economic Study on User Needs. Technical Report, http://www.wi2.uni-trier.de/publications/2005_technicalReport_D210.pdf, 2005.
- [Freß05] *Freyßmann, A.*: Adaptive Workflow Support for Search Processes within Fire Service Organisations. In: Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. IEEE Computer Society, 2006, p. 291-296.
- [FrMS05] *Freyßmann, A.; Maximini, R.; Sauer, T.*: Towards collaborative agent-based knowledge support for agile projects. In: *Althoff, K.D.; Dengel, A.; Bergmann, R.; Roth-Berghofer, T. (Hrsg.): WM2005: Professional Knowledge Management Experiences and Visions*. DFKI, Kaiserslautern 2005, S. 383-388.
- [Huws02] *Huws, U.*: Towards a Definition of Socio-Economic Research for the RESPECT Project. <http://www.respectproject.org/genpapers/definition.pdf> verified on 2005-08-05, 2002.
- [Huws04] *Huws, U.*: Socio-Economic Research in the Information Society: a user's guide from the RESPECT project. Sussex: Institute for Employment Studies, 2004.
- [LaLS06] *Laudon, K. C.; Laudon, J. P.; Schoder, D.*: Wirtschaftsinformatik – Eine Einführung. Pearson Education Deutschland GmbH, München 2006.
- [Nati06] *National Science Foundation*. <http://www.nsf.gov/ehr/rec/evaldesign.jsp>, verified on 2006-07-06.
- [Soci02] *Social Research Association*: Commissioning Social Research: a Good Practice Guide. 2. Aufl., London 2002.

Kontextbasiertes Information Retrieval

zur Unterstützung von Entwicklungsprozessen

Andreas Henrich, Karlheinz Morgenroth

Lehrstuhl für Medieninformatik
Fakultät Wirtschaftsinformatik und Angewandte Informatik
Otto-Friedrich-Universität Bamberg
96045 Bamberg
{andreas.henrich|karlheinz.morgenroth}@wiai.uni-bamberg.de

Abstract

Mitarbeitern die richtige Information zum richtigen Zeitpunkt zugänglich zu machen, stellt heute einen entscheidenden Wettbewerbsfaktor dar. Auch wenn in den vergangenen Jahren durch Maßnahmen und Systeme aus dem Bereich des Wissensmanagements in Organisationen umfangreiche Wissenssammlungen angelegt worden sind oder auch nur durch den Einsatz entsprechender IT-Systeme mehr Information in digitaler Form vorliegt, so wird die Identifikation von relevanter Information für die einzelne Person immer schwieriger und zeitaufwändiger.

Der von uns präsentierte Ansatz erfasst, beschreibt und nutzt historische und aktuelle Kontextinformation sowohl über Mitarbeiter als auch über die Entstehung von Dokumenten, um zum einen manuelle Suchanfragen von Mitarbeitern zu präzisieren und zum anderen das Entstehen von Informationsbedürfnissen bei Mitarbeitern automatisch erkennen und mit geeigneten Suchanfragen selbstständig beantworten zu können. Unser Ansatz und dessen Implementierungen decken dabei insbesondere ingenieurmäßige Produktentwicklungsprozesse ab.

1 Motivation

Mit der zunehmenden Menge an verfügbarer Information wird die Identifizierung relevanter Information für die einzelne Person immer schwieriger. Für die einzelnen Mitarbeiter reicht die in einer Organisation technisch abrufbare Information von eher allgemeinen Inhalten über neue Themen und Erkenntnisse im eigenen oder in angrenzenden Fachgebieten, die zur Weiter- und Fortbildung dienen können, über die Anwendung von und Erfahrungen mit Vorgehensmodel-

len, Methoden und Werkzeugen, bis zu konkreten Teilergebnissen oder Komponenten, die direkt in einem aktuell bearbeiteten Projekt Verwendung finden können. Die Anwendung bewährter Vorgehen, aber auch die Wiederverwendung von Komponenten stellt einen Faktor dar, der zur Steigerung der Effizienz und Produktivität innerhalb eines Entwicklungsprozesses – bspw. im Software Engineering oder Maschinenbau – und zu einer höheren Qualität bei den entwickelten Produkten führen kann. Dies bedeutet, dass der Identifikation von relevanten Inhalten, die zu einer Wiederverwendung herangezogen werden können, ein erheblicher Stellenwert zukommt.

Trotz des Einsatzes von Suchmaschinen sind die gefundenen Ergebnisse für eine anfragende Person oft nicht zielführend und erschöpfend im Hinblick auf das bestehende Informationsbedürfnis. Gründe hierfür können in einem zu unscharf artikulierten Informationsbedürfnis oder schlichtweg in der Wahl falscher Suchbegriffe liegen. In der Konsequenz unterlassen Anwender häufig die Durchführung einer Suche wegen des damit verbundenen Aufwands und der ungewissen bzw. in der Vergangenheit unbefriedigenden Qualität der Ergebnisse (vgl. z. B. [Wils99]).

Einen Ansatz die Ergebnisqualität zu erhöhen, bildet die Einbeziehung des Kontextes des Anfragestellers. Dabei werden neben seinen aktuellen und zurückliegenden Tätigkeiten und deren Inhalten unter anderem auch persönliche Kenntnisse erfasst und zur Verfeinerung von manuell gestellten Anfragen verwendet. Darüber hinaus kann der Kontext auch dazu verwendet werden, auf ein latent vorliegendes Informationsbedürfnis zu schließen, automatisch Suchanfragen zu stellen und die Suchergebnisse dem Anwender proaktiv zu präsentieren. Auf der anderen Seite kann u. a. der Aufgabenkontext, in dem bestimmte Dokumente erstellt worden sind, sowie deren Bearbeiter erfasst werden, um neue Beschreibungs- und Suchkriterien für diese Dokumente zu erhalten.

In Abbildung 1 ist ein Szenario dargestellt, in dem ein Entwickler innerhalb eines laufenden Entwicklungsprojektes mit der Arbeit an einem Gesamtentwurf beginnt. Die Grundlage für den Gesamtentwurf sind Anforderungslisten, prinzipielle Lösungen sowie ein Vorentwurf. Um nun für den aktuellen Gesamtentwurf Elemente existierender Vorentwürfe oder ganze Teilkomponenten wieder verwenden zu können, muss der Entwickler Kenntnis über entsprechende Artefakte entweder in früheren Iterationen des eigenen Projektes oder anderen Projekten haben. Diese Kenntnis könnte bei einer Person durch eine ständige Einbeziehung in andere Projekte geschaffen werden. Diesem Vorgehen sind aber offensichtlich zeitliche und organisatorische

Grenzen gesetzt. Eine Alternative bildet der Einsatz eines kontextbasierten Information Retrieval Systems, das eine Person – in diesem Fall einen Entwicklungsingenieur – mit der für ihre aktuell vorliegende Aufgabe benötigten Information versorgt.

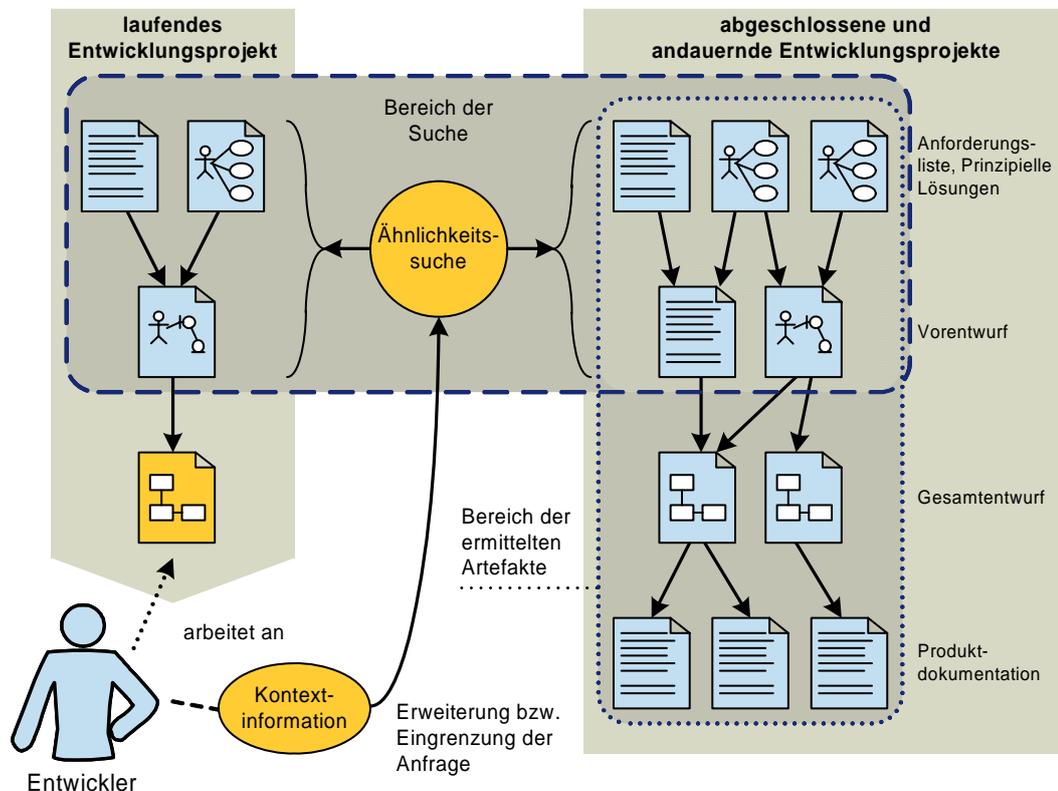


Abb. 1: Szenario eines laufenden Entwicklungsprojektes (in Anlehnung an VDI 2221 [VDI93]) mit einem Entwickler, der aktuell einen Gesamtentwurf beginnt

Auch wenn Entwicklungsprozesse, wie z. B. VDI 2221 [VDI93] im Maschinenbau oder ein Rational Unified Process [Kruc04] im Software Engineering, zunächst eine starke Strukturierung aufweisen mögen, so bieten sie in den einzelnen Teilaufgaben den beteiligten Personen eine sehr hohe Entscheidungsvielfalt in der Art und Weise wie die gestellten Aufgaben im Detail auszuführen sind und welche Werkzeuge oder Materialien dazu zum Einsatz kommen sollen. Gleichzeitig lässt sich ein Erfolg oder Misserfolg dieser Entscheidungen erst in folgenden Aufgaben und Phasen beurteilen.

2 Verwandte Ansätze

In der Vergangenheit wurde eine Reihe von Forschungsansätzen präsentiert, die sich sowohl mit der Erfassung und Modellierung des Kontextes von Individuen als auch mit der automatischen Belieferung von Personen mit zu deren aktueller Tätigkeit relevanter Information beschäftigten.

Über den folgenden kurzen Abriss hinausgehend ist eine ausführliche Würdigung von themenverwandten Forschungsansätzen bei [Morg06] zu finden.

Eine Gruppe verwandter Ansätze bilden Systeme, die den Text eines aktuell von einer Person bearbeiteten Dokumentes zur Formulierung von Anfragen verwenden (vgl. [RhSt96], [BuKr99], [DCSH04]). Gesucht werden dabei relevante Dokumente, wobei der betrachtete Kontext sich aber im Wesentlichen über den aktuell bearbeiteten Text definiert. Andere Ansätze beobachten einen Benutzer beim „Surfen“ im Web und geben Empfehlungen zu thematisch als ähnlich erachteten Webseiten (vgl. [Libe95], [JoFr97], [PaMB96]). Darüber hinaus kommen Methoden des Collaborative Filtering sowohl in Empfehlungssystemen für Texte als auch für Produkte im E-Commerce zum Einsatz (vgl. [RIS+94], [ScKR01], [ShMa95]). Eine Integration entsprechender kontextbasierter Suchdienste zeigen bspw. [Meie00] für Managementinformationssysteme und [PPK03] für Portalsysteme.

Im Forschungsgebiet des Workflow-Management verfolgte das Forschungsprojekt KnowMore sowie das sich daran anschließende DECOR-Projekt einen partiell ähnlichen Ansatz [ABH+00, ABN+01]. Dabei wurden allerdings existierende Workflow-Schemata zusätzlich um weitere Metadaten ergänzt, die u. a. für jede Aufgabe eine generische Anfrage enthielten [Maus01]. Im Unterschied zu unserem im Folgenden skizzierten Ansatz wurden die Anfragen für ausgewählte und vordefinierte Aufgaben also a priori ausformuliert. Zum Ausführungszeitpunkt einer Aufgabe wurden diese generischen Anfragen mit Instanzdaten gefüllt und zur Ausführung gegenüber einer Reihe von so genannten Suchagenten gebracht. Ebenfalls an einer starren Prozessmodellierung orientiert sich das PreBIS-Projekt [BöHä05].

Unser Ansatz hebt sich von den genannten Ansätzen durch eine umfassendere Betrachtung des Kontextes und durch ein wesentlich mächtigeres Konzept zur Relevanzbewertung sowie durch den Bezug auf den Produktentwicklungsprozess ab. Darüber hinaus findet sowohl eine Unterstützung von stark strukturierten Prozessrahmen als auch von eher unstrukturierten Einzeltätigkeiten statt.

3 Der COBAIR-Ansatz

Mit dem COBAIR-Ansatz [Morg06] wird ein grundlegender methodischer Ansatz zur Nutzung des Kontextes von Personen und deren Tätigkeiten für die Suche nach relevanter Information vorgestellt. Darüber hinaus wird ebenso der Kontext von Dokumenten, bzw. Artefakten i. A.

betrachtet, in dem diese entstanden sind. Der COBAIR-Ansatz umfasst dabei die Erfassung und Beschreibung von Kontextinformation, die Strukturierung und gleichzeitige Vernetzung der erfassten Kontextinformation, die Ableitung von Informationsbedürfnissen als auch die Präzisierung und Durchführung von Ähnlichkeitssuchen, teilweise oder gänzlich auf Basis der Kontextinformation.

3.1 Erfassung und Beschreibung des Kontextes

Die Ausgangsbasis von Suchvorgängen von Personen ist ein mehr oder minder stark ausgeprägtes Informationsbedürfnis. Dieses lässt sich jedoch nicht direkt beobachten, vielmehr können nur die aus einem Informationsbedürfnis motivierten Handlungen einer Person beobachtet werden. Gleichzeitig ist ein Informationsbedürfnis durch Eigenschaften einer Person selbst, ihrer Umgebung sowie die aktuellen Aufgaben und ausgeführten Tätigkeiten – allgemein den Kontext einer Person – motiviert. Zu diesen Eigenschaften zählen unter anderem psychologische, kognitive und affektive Faktoren, Erfahrungen und Kenntnisse einer Person, aktuelle wie auch vergangene Aufgaben und Tätigkeiten einer Person, die Einordnung einer Person in eine Organisation, das Arbeitsumfeld sowie die politische, ökonomische und technische Umgebung. Wie bei anderen Ansätzen und Systemen (vgl. Abschnitt 2) wird auch hier implizit von der Annahme ausgegangen, dass eine entsprechende Person zur Verrichtung ihrer Tätigkeit, in deren Rahmen bestimmte Informationsbedürfnisse auftreten können, Softwaresysteme (z. B. CAD-, Office-, Groupware-Systeme) einsetzt, die zur Ableitung von Informationen über die aktuell verrichtete Tätigkeit in der Lage sind bzw. diesbezüglich angepasst werden können.

Der Kontext eines Anwenders während der Ausübung seiner Tätigkeit mit Hilfe eines Softwaresystems lässt sich daher zum einen aus der von ihm vollführten Benutzerinteraktion mit dem Softwaresystem ableiten. Zum anderen umfasst der Kontext auch die zu einer Aktion gehörenden Vor- und Ergebnisdokumente.

Für die Erfassung von Informationen über aktuelle und zurückliegende Kontexte von Personen dienen verschiedenste Softwaresysteme als Quelle. Im Folgenden werden diese nach ihren technischen Architekturmerkmalen unterschieden und Beispiele für zu erfassende Kontextinformation gegeben.

- *Klassische Client-Anwendungen.* In diese Kategorie fallen neben Office-Programmen (inkl. E-Mail und Terminverwaltung) auch CAD-Anwendungen, die auf dem Arbeitsplatzrechner eines Anwenders laufen. Derartige Client-Anwendungen können heute durchweg mittels so

genannter Plug-Ins um Funktionalitäten erweitert werden. Auf diesem Weg ist sowohl ein Zugriff auf die aktuell von einem Anwender bearbeiteten Daten möglich (z. B. vorliegende prinzipielle Lösungen oder einen betrachteten Vorentwurf) als auch die Erfassung der Aktionen, die eine Person ausführt.

- *Client-Server-Anwendungen.* Unter Client-Server-Anwendungen fallen bspw. Anwendungen wie Groupware-Systeme (z. B. Lotus Notes), Werkzeuge zur Projektplanung (z. B. Microsoft Project) oder PDM-Systeme, teilweise in direkter Verbindung mit CAD-Systemen. Die eigentliche Anwendung setzt sich aus mindestens zwei Teilen zusammen, die auf dem einzelnen Arbeitsplatzrechner (Client) und auf einem zentralen Server laufen und miteinander kommunizieren. Der Erweiterbarkeit dieses Typus von Softwaresystemen kommt die Tatsache entgegen, dass zunehmend so genannte offene Architekturen zum Einsatz kommen. Je nach eingesetzter Technologie und Architektur der Implementierung können derartige Anwendungen sowohl auf der Client- als auch der Server-Seite durch Plug-Ins ergänzt werden.
- *Server-Anwendungen.* In diese Kategorie fallen z. B. unternehmensweite Verzeichnisdienste (X.500, Active Directory, LDAP), denen sich Kontextinformation über die organisatorische Eingliederung einer Person in einem Unternehmen sowie deren Rollen und Aufgaben entnehmen lässt. Weitere Quellen stellen abteilungs- oder unternehmensweite Dokumenten-Management-Systeme (DMS), Content-Management-Systeme (CMS) oder Repositories und PDM-Systeme dar. Diese Systeme erfassen neben den eigentlichen Nutzdaten automatisch auch eine Reihe von Metadaten, wie z. B. Autor und Zeitpunkt einer Änderung an einem Dokument. Für die Erfassung von Kenntnissen, Erfahrungen, Interessen und auch von persönlichen Präferenzen lassen sich entweder spezialisierte Skill-Management-Werkzeuge oder erweiterte Human-Resource-Systeme einsetzen.

Sowohl die für eine Ableitung von Kontextinformation in Betracht kommenden Systeme als auch deren Inhalte sind dabei sehr heterogen. Für eine einheitliche und eindeutige Erfassung, Benennung und Kennzeichnung der Inhalte kommen daher in unserem Ansatz verschiedene Technologien des Semantic Web [Bern98] zum Einsatz.

Für die Beschreibung der erfassten Kontextinformation wird das Resource Description Framework (RDF, [KICa04]) verbunden mit Ontologien auf Basis der Web Ontology Language (OWL, [DeSc04]) genutzt. Der Kontext einer Person wird somit in Form einer automatisch ge-

wonnenen Annotation mit Metadaten beschrieben, während zur eindeutigen Bezeichnung der Inhalte der Kontextbeschreibung verschiedene in Ontologien definierte Konzepte Verwendung finden.

Um die resultierende Fülle an verfügbarer Information sinnvoll für die Erschließung eines Informationsbedürfnisses verwenden zu können, gilt es diese zunächst anhand eines Modells zu gliedern. Abbildung 2 zeigt einen Ansatz für ein einheitliches Nutzermodell, das die verschiedenen Kontextinhalte in einzelne Dimensionen einteilt und darüber hinaus auch deren Veränderung berücksichtigt. Gleichzeitig findet aber eine Vernetzung der einzelnen Kontextinhalte statt [HeMo01, HeMo03].

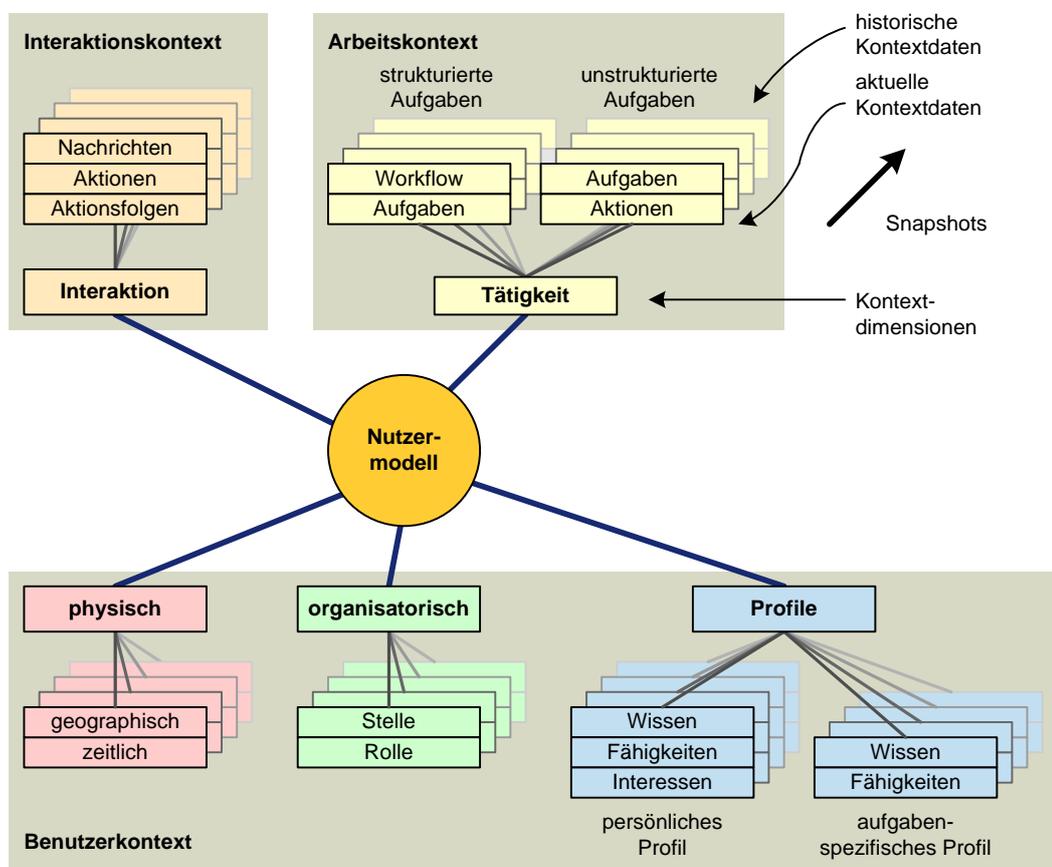


Abb. 2: Einheitliches Nutzermodell

Dieses einheitliche Nutzermodell differenziert grundsätzlich zwischen Benutzer-, Arbeits- und Interaktionskontext:

- Der *Benutzerkontext* umfasst sowohl den physikalischen und organisatorischen Kontext als auch das Profil des Benutzers. Diese einzelnen Aspekte werden im Folgenden als Kontextdimensionen bezeichnet.

Die physikalische Kontextdimension beinhaltet beispielsweise die geographischen Daten

über den Arbeitsplatz eines Mitarbeiters.

Die organisatorische Kontextdimension beinhaltet sowohl die Stelle, die eine Person innerhalb der Organisation einnimmt, als auch die einzelnen Rollen, die sie ausfüllt. Dies schließt eine Stellenbeschreibung sowie die erwarteten Fähigkeiten und das Wissen eines entsprechenden Stelleninhabers mit ein. Darüber hinaus enthält diese Kontextdimension Informationen über die einzelnen Organisationseinheiten selbst, also beispielsweise über die Abteilung, in der ein Mitarbeiter tätig ist.

Die Profildimension enthält Informationen über die Wissensgebiete und Fähigkeiten einer Person. Im Modell wird zwischen dem persönlichen Profil, das von einer Person selbst vorgegeben werden kann und beispielsweise persönliche Präferenzen und Interessen enthält, und einem aufgabenspezifischen Profil unterschieden, das sich aus den Aufgaben und Tätigkeiten bestimmt, die eine Person in der Vergangenheit ausgeführt hat.

Diese verschiedenen Teildimensionen des Benutzerkontexts haben gemein, dass sie sich über die Zeit nur sehr langsam verändern.

- Der *Arbeitskontext* charakterisiert die aktuelle Tätigkeit, die eine Person ausführt. Dabei lässt sich diese Tätigkeit durch die wohl definierten Aufgaben in einem Workflow oder durch die ausgeführten Aufgaben und Aktionen beschreiben, welche eine unstrukturierte Tätigkeit mit sich bringt. Im ersten Fall wird eine konkrete Beschreibung der einzelnen Aufgaben oder der gesamten Tätigkeit vorliegen, während im zweiten Fall typischerweise nur vage Informationen über die Tätigkeit verfügbar sind.

Die technische Sicht auf den Arbeitskontext schließt die aktuell verwendeten Module eines Anwendungssystems mit ein, die eine Person zur Ausübung ihrer Tätigkeit verwendet. Der Arbeitskontext einer betrachteten Person wird sich deutlich häufiger ändern als dies vom Benutzerkontext zu erwarten ist.

- Die aktuellen Interaktionen eines Benutzers mit den von ihm eingesetzten Anwendungssystemen spiegelt der *Interaktionskontext* wider. Diese Interaktionen werden durch Nachrichten von Menübefehlen oder Dialogen repräsentiert, die ein Benutzer während der Arbeit mit einem Anwendungssystem ausführt bzw. verwendet, beispielsweise das Öffnen eines Vorentwurfs in einem CAD-System. Der Interaktionskontext ändert sich damit dynamisch innerhalb von Minuten oder Sekunden.

Neben den aktuellen Werten der einzelnen Kontextdimensionen ist auch deren Veränderung über die Zeit für die Bestimmung eines Kontextes von Interesse. So werden in bestimmten Zeitabständen sog. Snapshots aufgezeichnet, deren Inhalt zudem über verschiedene Mechanismen aggregiert und verdichtet wird (siehe [Morg06]).

3.2 Ableitung eines Informationsbedürfnisses aus der erfassten Kontextinformation

Für die Ableitung eines Informationsbedürfnisses aus der erfassten Kontextinformation erfolgt zunächst die Bildung von verschiedenen Profilen. Dabei besteht die Hauptaufgabe in der Filterung und Aggregation der im einheitlichen Nutzermodell enthaltenen Kontextinformation. Die Filterung erfolgt in Bezug auf die Veränderungshäufigkeit der verschiedenen Kontextdimensionen typischerweise mittels Fensterfunktionen entlang des zeitlichen Verlaufs. Technisch werden dabei sowohl Histogramme, welche die Häufigkeit einzelner Konzepte der Ontologie abbilden, als auch gewichtete RDF-Daten erzeugt und weiterverarbeitet.

- So wird das *Profil der aktuellen Tätigkeit* aus der Kontextinformation des Interaktions- und Arbeitskontextes für einen vom aktuellen Zeitpunkt ausgehenden Zeitraum von z. B. 10 Minuten in die Vergangenheit ermittelt. Ziel ist dabei die Abbildung der aktuellen Tätigkeit einer Person.
- Das *Profil der zurückliegenden Tätigkeit* wird aus der Dimension Tätigkeit des Arbeitskontextes im Nutzermodell gebildet. Ziel dieses Profils ist es, die zurückliegende Tätigkeit einer Person und die von ihr bearbeiteten Inhalte abzubilden. Im Gegensatz zum Profil der aktuellen Tätigkeit soll mit dem Profil der zurückliegenden Tätigkeit nicht eine bestimmte ausgeführte Tätigkeit beschrieben werden, sondern vielmehr das Tätigkeitsgebiet des Entwicklers und dessen Inhalte in einer allgemeineren Form.
- Das *persönliche Profil* einer Person umfasst deren Wissen, Fähigkeiten und auch Interessen. Die Bildung eines persönlichen Profils kann entweder nur aus den manuell vorgegebenen Eigenschaften erfolgen oder aus den in zurückliegenden Tätigkeiten gezeigten Eigenschaften.
- Ein *Profil des organisatorischen Benutzerkontextes* enthält Kontextinformation aus der gleichlautenden Kontextdimension des Nutzermodells über die Stelle, die eine Person innerhalb einer Organisation innehat, bzw. die Rollen, die sie ausübt.

- Das *Profil des physischen Benutzerkontextes* enthält physische Kontextinformation aus der entsprechenden Dimension des Nutzermodells wie die geographische Lage eines Benutzers sowie auch zeitliche Angaben.

Die eigentliche Ableitung eines Informationsbedürfnisses erfolgt nun über mehrere Verarbeitungsstufen, die in Abbildung 3 dargestellt sind. Dieses einfache Modell motiviert sich u. a. durch verschiedene Modelle des menschlichen Informationsverhaltens (vgl. z. B. [Wils99]), die davon ausgehen, dass ein Profil einen mehr oder minder starken Einfluss auf das Entstehen eines zunächst latenten Informationsbedürfnisses hat. Im Gegenzug wirken die einzelnen Profile gleichzeitig hemmend auf das Entstehen eines tatsächlich wahrgenommenen Informationsbedürfnisses, das im Modell als erwartetes tatsächliches Informationsbedürfnis bezeichnet wird.

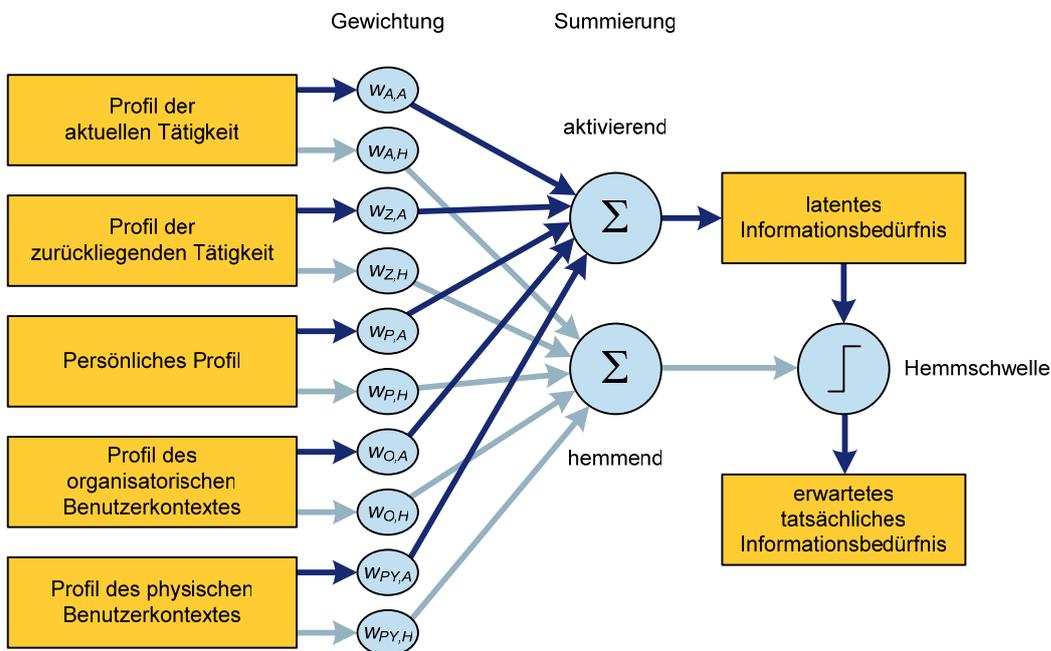


Abb. 3: Ableitung eines erwarteten tatsächlichen Informationsbedürfnisses aus den einzelnen Profilen des einheitlichen Nutzermodells (einfaches Modell)

Sehr grob gesprochen ergibt sich das Informationsbedürfnis als Abweichung des aktuellen Kontextes vom langfristigen Mittel. Einem Entwickler, der sich im Rahmen einer Analysetätigkeit z. B. mit für ihn bisher unbekanntem Legacy-Systemen beschäftigt, werden von unserem System entsprechende Dokumente über die Anbindung dieser Systeme geliefert.

Neben diesem einfachen Gewichtungmodell sind ebenfalls Regelsysteme und Systeme zur Erkennung bestimmter Handlungsweisen (z. B. mittels Hidden-Markov-Modellen) verfügbar und können alternativ eingesetzt werden (vgl. [Morg06]).

Der Inhalt des erwarteten tatsächlichen Informationsbedürfnisses kann nun zum einen zur Verfeinerung einer manuell gestellten Suchanfrage genutzt werden. Zum anderen kann bei einer systembedingten Auslösung der Hemmschwelle der Inhalt des erwarteten tatsächlichen Informationsbedürfnisses zu einer Suchanfrage umformuliert und an einen oder mehrere Suchdienste weitergeleitet werden.

3.3 Ähnlichkeitssuche mittels RDF Query by Example

Der COBAIR-Ansatz sieht sowohl die Anbindung externer Suchdienste, die z. B. im Rahmen eines existierenden Wissensmanagementsystems bereits vorhanden sind, als auch eine eigene Suchfunktionalität vor. Während typische Suchdienste auf Basis des Vektorraummodells [SWY75] arbeiten, setzt der eigene Suchdienst auf dem RDF Query by Example Ansatz [HeMo05] auf. Basis für die Präzisierung einer manuell gestellten Anfrage oder die automatische Formulierung einer Anfrage ist ein erwartetes tatsächliches Informationsbedürfnis, das in Form von gewichteten RDF-Daten vorliegt, mithin bestimmte Vernetzungen mit Gewichtungen zwischen Konzepten und externen Informationen, z. B. Dokumenten, Aufgaben oder Personen, aufzeigt. Wenngleich eine Umwandlung in eine für das Vektorraummodell adäquate Form möglich ist, so gehen dabei zahlreiche, explizite Beziehungen zwischen Konzepten verloren. Basis der in den RDF-Daten enthaltenen Konzepte sind in OWL [DeSc04] vorliegende Ontologien, die gleichzeitig den Ausgangspunkt für die Bestimmung eines Ähnlichkeitsmaßes mittels der sog. ontologischen Distanz zwischen einzelnen Konzepten darstellt.

Der RDF Query by Example Ansatz nutzt als Anfrage einen RDF-Beispieldatensatz, der z. B. im einfachsten Fall ein erwartetes tatsächliches Informationsbedürfnis repräsentiert. Gleichzeitig kommen vage Anfragebearbeitung, die Unterstützung von Volltextanfragen sowie die Ermittlung sortierter Ergebnislisten zum Einsatz. Über die Identifikation ähnlicher, aber zeitlich zurückliegender Informationsbedürfnisse anderer Personen können nicht nur deren Arbeitsergebnisse gefunden werden, sondern in der Relevanzbewertung die Bewertungen dieser Arbeitsergebnisse z.B. in Reviews mit einfließen.

RDF-Daten repräsentieren fast immer einen allgemeinen Graphen mit komplexen Strukturen. Eine Suche entspricht damit dem bekannten Problem der Subgraphisomorphie und liegt zunächst in der Komplexitätsklasse NP. Beschränkt sich die Form des Anfragegraphen auf Bäume, so erreicht die Komplexität zwar die Klasse P, bleibt aber für praktische Zwecke ungeeignet. Durch den Einsatz geeigneter Datenstrukturen, hier zweistufig verketteter invertierter Lis-

ten, und einer strombasierten Anfragebearbeitung ist dennoch eine effiziente Anfragebearbeitung möglich [HeMo05].

4 Das COBAIR-Framework

Zur Realisierung entsprechender kontextbasierter Information Retrieval Systeme geben wir neben einer Reihe von Methoden und Ansätzen auch einen Architekturrahmen an, der neben der Anbindung verschiedenster Quellen für Kontextinformation insbesondere auch die Handhabung hoher Volumina an Kontextinformation ermöglicht [HeRo00, HeRo03]. Für die effiziente Bearbeitung vager Anfragen wurde der RDF Query by Example Ansatz integriert.

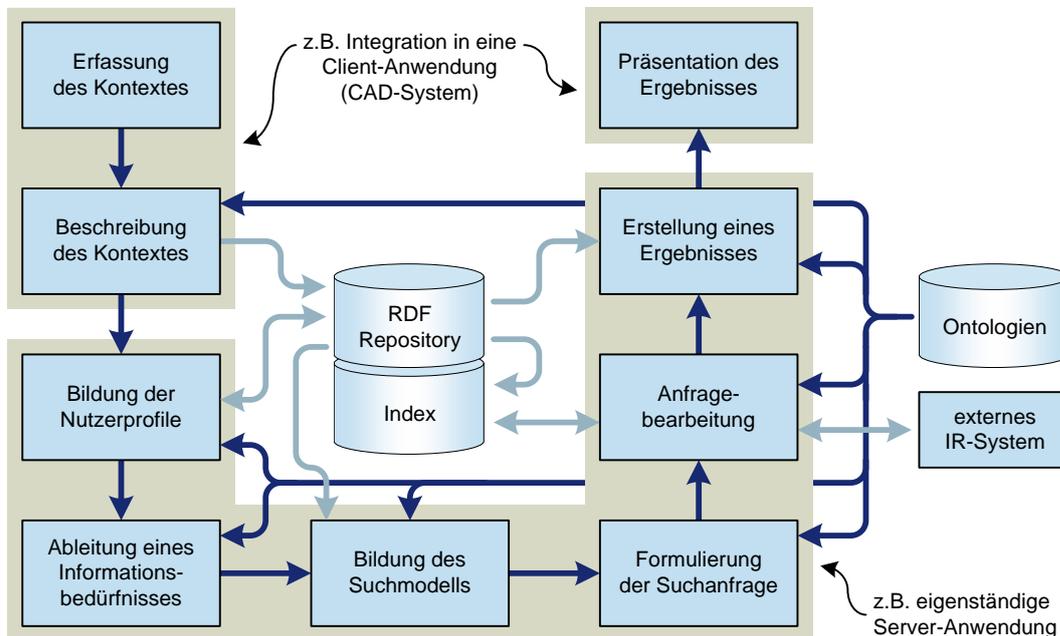


Abb. 4: Komponenten des COBAIR-Frameworks

Dieses Framework unterstützt dabei im Einzelnen die:

- Anbindung von verschiedenen Client-, Client-Server- sowie Server-Anwendungen zur Erfassung von Informationen über aktuelle und zurückliegende Kontexte von Personen,
- Beschreibung der Kontextinformation mittels RDF und Einsatz von Ontologien zur eindeutigen Benennung der einzelnen Elemente,
- Verwaltung und Abfrage von Kontextinformation sowie Gliederung der Kontextinformation in einem einheitlichen Nutzermodell, technische Speicherung in einem RDF-Repository,

- Bildung von Nutzerprofilen und Ableitung von Informationsbedürfnissen einzelner Benutzer aus deren Kontextinformation,
- Formulierung eines erwarteten tatsächlichen Informationsbedürfnisses für einen Anwender,
- Formulierung von Suchanfragen auf einem Suchindex,
- Ausführung der Suchanfragen gegenüber dem eigenen RDF-Repository oder externen IR-Systemen sowie
- Präsentation der Suchergebnisse und deren weitere Verwendung durch einen Anwender.

Das obige Framework ermöglicht mit verschiedenen Architekturmerkmalen vor allem die Verarbeitung sehr großer Mengen an Kontextinformation und gleichzeitig die Anfragebearbeitung in für die Anwender akzeptabler Zeitdauer. Die Anfragebearbeitung unterstützt dabei sowohl vage Anfragen als auch die Lieferung der nach ihrer Relevanz gegenüber der Anfrage sortierten Ergebnisse.

5 Einsatz und Evaluierung im Gebiet des Software Engineerings

Für die Anwendungsdomäne des Software Engineering erfolgte exemplarisch die Realisierung eines entsprechenden kontextbasierten Information Retrieval Systems. Mittels dieser Implementierung wurden die im Rahmen des Forschungsprojektes aufgezeigten Methoden und Modelle auf ihre Nützlichkeit und Performanz beim Einsatz in Softwareentwicklungsprozessen hin evaluiert [Morg06]. Konkrete Zielsetzung war hier, wie bereits in der Motivation und in Abbildung 1 abstrakt aufgezeigt, die Unterstützung von Softwarearchitekten und -entwicklern mit zu ihrer aktuellen Tätigkeit relevanten Artefakten aus anderen Softwareprojekten. Neben einer direkten Verwendung von Komponenten stand dabei auch das Aufzeigen von möglichen Lösungswegen, bspw. durch die Wahl geeigneter Softwarearchitekturen oder -designs im Mittelpunkt.

Wenngleich der zugrunde liegende Rational Unified Process [Kruc04] einen strengen Prozessrahmen vorgibt, so verbleiben auch hier in den einzelnen Teilaufgaben zahlreiche Entscheidungsfreiheiten für einen Entwickler, die teilweise erst in späteren Phasen negative wie positive Auswirkungen zeigen können. Dies sind z. B. Fragen der Softwarearchitektur, des Softwaredesigns, der Einsatz bestimmter Integrationswerkzeuge und Middleware-Systeme, aber auch der Einsatz von geeigneten Algorithmen.

Die prototypische Realisierung eines kontextbasierten Information Retrieval Systems basierend auf dem COBAIR-Framework integriert sich in die typische Anwendungslandschaft mit Werkzeugen zur Prozess- und Projektplanung, Verwaltung von Anforderungen, Änderungen und Fehlern sowie Dokumentenmanagementsysteme, Repositories und Verzeichnisdienste. Auf Client-Seite der Entwickler wurde die Integration in Form eines Plug-Ins in eine integrierte Entwicklungsumgebung realisiert (siehe Abbildung 5). Das zur Anwendung kommende Nutzermodell berücksichtigt neben den Fähigkeiten der einzelnen Personen auch deren in zurückliegenden Projekten durchgeführte Tätigkeiten. Auf Basis der aktuellen Tätigkeit wird so ein latentes Informationsbedürfnis antizipiert und automatisch in entsprechende Suchanfragen übersetzt. Für die Evaluierung der Suchqualität und -performanz kam eine Testkollektion aus 72 Open Source Softwareprojekten zum Einsatz.

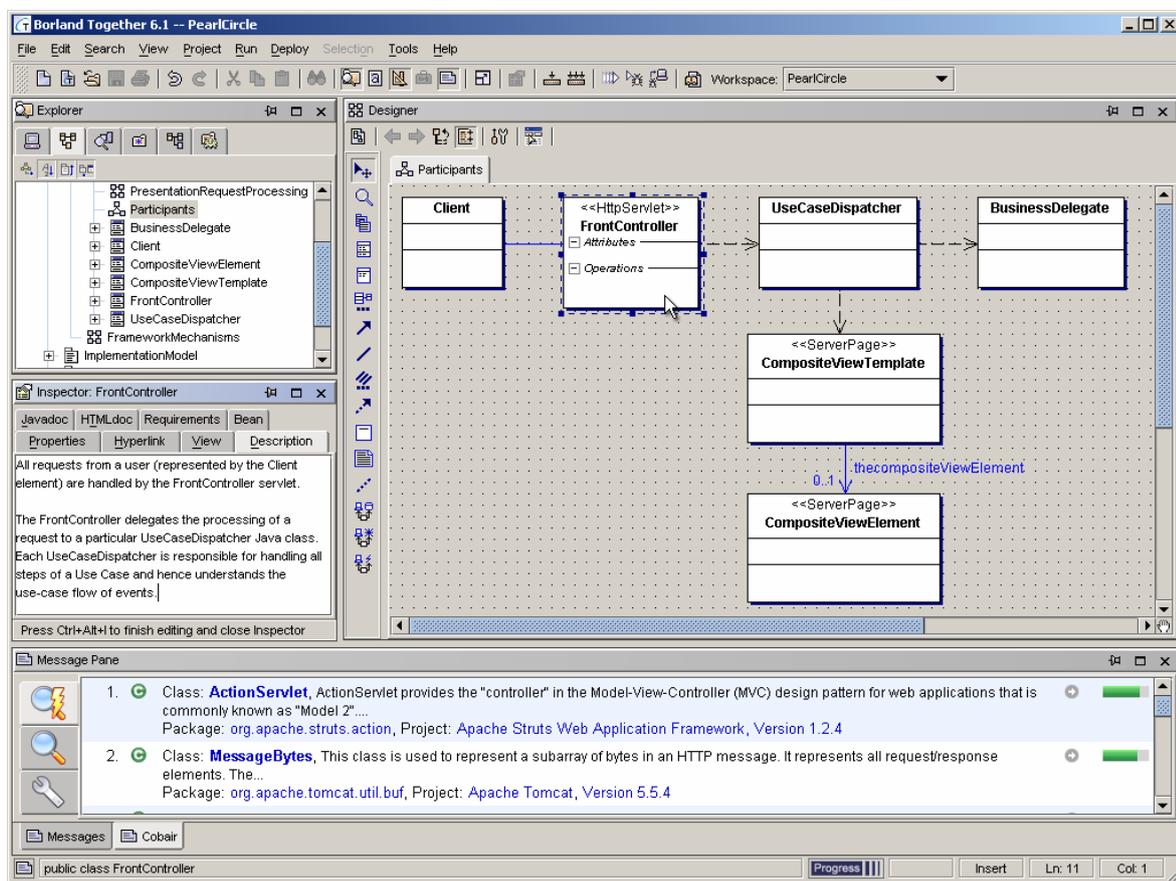


Abb. 5: Einsatz eines Plug-Ins in Borland Together zur Erfassung von Kontextinformation über einen Mitarbeiter sowie zur Anzeige von Suchergebnissen und deren weiteren Verwendung.

Das der Evaluierung zugrunde liegende Szenario umfasst die Entwicklung eines einfachen, webbasierten Auktionssystems in Form einer Mehrschichtenarchitektur, das technisch auf Basis der Java 2 Platform, Enterprise Edition (J2EE) realisiert werden soll. Dieses Beispielszenario, der Verlauf des Softwareentwicklungsprozesses sowie alle dabei entstehenden Artefakte sind

Eeles et al. [EHK02] entnommen. Damit war auch für wiederholte Evaluierungen ein stabiler Verlauf eines Softwareentwicklungsprozesses gewährleistet.

Die Ergebnisse (ausführlich in [Morg06] zu finden) zeigen neben der Tragfähigkeit des Ansatzes auch die Qualität der mit einem entsprechenden kontextbasierten Information Retrieval Ansatz erzielbaren Suchergebnisse. So waren relevante Suchergebnisse im Rahmen der Evaluierung häufig an erster oder zweiter Stelle zu finden (siehe auch Abbildung 5). Die Bearbeitung auch komplexer Suchanfragen, die automatisch auf Basis der vorliegenden Kontextinformation erzeugt wurden, dauerte dabei selten länger als eine Sekunde.

6 Zusammenfassung und Ausblick

Eine abschließende Bewertung der gesamten prototypischen Implementierung eines kontextbasierten Information Retrieval Systems für das Anwendungsgebiet des Software Engineerings zeigt, dass ein derartiges System in der Lage ist, einen Anwender automatisch mit hilfreicher und relevanter Information während seiner Tätigkeit zu unterstützen. Da diese Bewertung in einem allerdings beschränkten Testszenario stattgefunden hat, ist eine weitere Evaluierung in einem größeren Projektumfeld, die dann auch verschiedene Nutzerprofile und -modelle mit einbeziehen könnte, anzustreben.

Dies ist für die Zukunft in einem entsprechenden Rahmen geplant. Dabei sollen Produktentwicklungsprozesse in der Domäne der Konstruktionstechnik und des Maschinenbaus in einem interdisziplinären Forschungsumfeld untersucht werden.

Literaturverzeichnis

- [ABH+00] Abecker, Andreas; Bernardi, Ansgar; Hinkelmann, Knut; Kühn, Otto; Sintek, Michael: Context-Aware, Proactive Delivery of Task-Specific Information: The KnowMore Project. *Information Systems Frontiers*, 2(3/4):253–276, 2000.
- [ABN+01] Abecker, Andreas; Bernardi, Ansgar; Ntioudis, Spyridon; Herterich, Rudi; Houy, Christian; Legal, Maria; Mentzas, Gregory, Müller, Stephan: The DECOR Toolbox for Workflow-Embedded Organizational Memory Access. In: ICEIS 2001, 3rd International Conference on Enterprise Information Systems, Setúbal, Portugal, 7–10 July, 2001.

- [Bern98] Berners-Lee, Tim: A roadmap to the Semantic Web. <http://www.w3.org/DesignIssues/Semantic.html>, September 1998, Zugriff am 1.7.2006.
- [BöHä05] Böhm, Karsten; Härtwig, Jörg: Prozessorientiertes Wissensmanagement durch kontextualisierte Informationsversorgung aus Geschäftsprozessen. In: Ferstl, O. K.; Sinz, E. J.; Eckert, S.; Isselhorst, T. (Hrsg.): Wirtschaftsinformatik 2005. Heidelberg: Physica-Verlag, 2005 (eEconomy eGovernment eSociety), S. 943–962.
- [BuKr99] Budzik, Jay; Hammond, Kristian: Watson: Anticipating and Contextualizing Information Needs. In 62nd Annual Meeting of the American Society for Information Science, Medford, New Jersey, USA, 1999.
- [DCSH04] Dumais, Susan; Cutrell, Edward; Sarin, Raman; Horvitz, Eric: Implicit Queries (IQ) for Contextualized Search. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2004), Seite 594, Sheffield, UK, 25.–29. Juli 2004. ACM Press.
- [DeSc04] Dean, Mike; Schreiber, Guus: OWL Web Ontology Language Reference. <http://www.w3.org/TR/owl-ref/>, Februar 2004, Zugriff am 1.7.2006.
- [EHK02] Eeles, Peter; Houston, Kelli; Kozaczynski, Wojtek. Building J2EE Applications with the Rational Unified Process. Addison Wesley, August 2002.
- [HeMo02] Henrich, Andreas; Morgenroth, Karlheinz: Integration von kontextunterstütztem Information Retrieval in Portalsysteme. In: Tagungsband zur Teilkonferenz Management der Mitarbeiter-Expertise in IT-Beratungsunternehmen im Rahmen der Multikonferenz Wirtschaftsinformatik (MKWI 2002), 9.–11. September 2002, Nürnberg.
- [HeMo03] Henrich, Andreas; Morgenroth, Karlheinz: Supporting Collaborative Software Development by Context-Aware Information Retrieval Facilities. In: Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA) (3rd International Workshop on Web Based Collaboration Prague, Czech Republic, September 1–5, 2003). USA: IEEE Computer Society, 2003, Seite 249–253.

- [HeMo05] Henrich, Andreas; Morgenroth, Karlheinz: A Query Engine for RDF based Similarity Queries on Software Artifacts. In: Witold Abramowicz (Ed.): BIS 2005. Proceedings of 8th International Conference on Business Information Systems. Poznan, Poland, 20.–22. April 2005, Seite 169–184.
- [HeRo00] Henrich, Andreas; Robbert, Günter: Combining Multimedia Retrieval and Text Retrieval to Search Structured Documents in Digital Libraries. In Proceedings 1st DELOS Workshop on Information Seeking, Searching and Querying in Digital Libraries, Zürich, Schweiz, Dezember 2000.
- [HeRo03] Henrich, Andreas; Robbert, Günter: Ein Ansatz zur Übertragung von Rangordnungen bei der Suche auf strukturierten Daten. In Tagungsband der 10. GI-Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW'03), Leipzig, 2003.
- [JoFr97] Joachims, Thorsten; Freitag, Dayne; Mitchell, Tom M.: Web Watcher: A Tour Guide for the World Wide Web. In Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI), Seite 770–777, 1997.
- [KlCa04] Klyne, Graham; Carroll, Jeremy J.: Resource Description Framework (RDF): Concepts and Abstract Syntax. <http://www.w3.org/TR/rdf-concepts/>, Februar 2004, Zugriff am 1.7.2006.
- [Kruc04] Philippe Kruchten. The Rational Unified Process: An Introduction. Addison Wesley, 3. Auflage, 2004.
- [Libe95] Lieberman, Henry: Letizia: An Agent That Assists Web Browsing. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95), Seite 924–929, Montreal, Quebec, Kanada, 1995. Morgan Kaufmann.
- [Maus01] Maus, Heiko: Workflow Context as a Means for Intelligent Information Support. In Proceedings of 3rd International Conference on Modeling and Using Context (CONTEXT'01), Seite 261–274, 2001.
- [Meie00] Meier, Marco: Integration externer Daten in Planungs- und Kontrollsysteme – Ein Redaktions-Leitstand für Informationen aus dem Internet. Gabler, Wiesbaden, 2000.

- [Morg06] Morgenroth, Karlheinz: Kontextbasiertes Information Retrieval. Modell, Konzeption und Realisierung kontextbasierter Information Retrieval Systeme. Dissertation, Otto-Friedrich-Universität Bamberg, 2006.
- [PaMB96] Pazzani, Michael J.; Muramatsu, Jack; Billsus, Daniel; Syskill & Weibert: Identifying Interesting Web Sites. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, Seite 54–61, Portland, Oregon, USA, 1996.
- [PPK03] Priebe, Torsten; Pernul, Günther; Krause, Peter: Ein integrativer Ansatz für unternehmensweite Wissensportale. In Tagungsband 6. Internationale Tagung Wirtschaftsinformatik (WI 2003), Dresden, September 2003.
- [RhSt96] Rhodes, Bradley J.; Starner, Thad: Remembrance Agent: A continuously running automated information retrieval system. In The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology (PAAM '96), Seite 487–495, London, UK, April 1996.
- [RIS+94] Resnick, Paul; Iacovou, Neophytos; Suchak, Mitesh, Bergstrom, Peter; Riedl, John: GroupLens: an open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, Seite 175–186, Chapel Hill, North Carolina, USA, 1994. ACM Press.
- [ScKR01] Schafer, J. Ben; Konstan, Joseph A.; Riedl, John: E-Commerce Recommendation Applications. *Data Min. Knowl. Discov.*, 5(1-2):115–153, 2001.
- [ShMa95] Shardanand, Upendra; Maes, Pattie: Social information filtering: algorithms for automating word of mouth. In Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, Seite 210–217, Denver, Colorado, USA, 1995. ACM Press.
- [SWY75] Salton, Gerard; Wong, Anita; Yang, Chung shu. A vector space model for automatic index. *Communications of the ACM*, 18(11):613–620, November 1975.
- [VDI93] VDI-Richtlinie 2221: Methodik zum Entwickeln und Konstruieren technischer Systeme, 1993.
- [Wils99] Wilson, Tom D.: Models in information behaviour research. *Journal of Documentation*, 55(3):249–270, 1999.

Towards Understanding the Dynamics of Digital Communication Networks

Matthias Trier, Annette Bobrik, Tilmann Bartels

Institut für Wirtschaftsinformatik
Technische Universität Berlin
10587 Berlin
{trier, bobrik, bartels}@sysedv.tu-berlin.de

Abstract

Understanding structures and processes of virtual communication networks can help to improve knowledge sharing and collaboration in a corporate setting. One current research method for that objective is Social Network Analysis. Although this method can generate a static additive picture of the final structures resulting of virtual networks, it misses to uncover the underlying processes of structural evolvement. This article suggests and demonstrates a novel methodology which is based on longitudinal visualization of dynamic networks. It allows visualizing structural evolvement of communication networks over time and thus leads to improved understanding of networking processes. Using the corporate e-mail network of Enron, the approach is demonstrated by studying temporal activity bursts, organizational positions that integrate the virtual network, and longitudinal structural change.

1 Introduction

During the past years, electronic communication like e-mail, instant messaging, or videoconferencing have quickly climbed the ranking of preferred means to communicate and subsequently collaborate in an organization. This results in changes in organizational work towards more decentralization [LuBa94] where "electronic communications fuel the growth and effectiveness of an organization and its parts. Information, rather than being limited, controlled, and a source of power, appears to be instrumental for greater effectiveness when widely disseminated and freely available in so-called virtual electronic organizations." [DaLe93:iv].

The resulting complex communication network delivers fix points for the combination of order and chaos in the organization. Every person must be enabled to purposefully switch between the processing of expected and unexpected information. The necessary orientation for that is provided by a network, which is offering active connections (for coordination and communication), but even more importantly connections, which potentially can be activated on demand in special situations [Baec99:26].

With the above properties, novel means of communication can be conceived as a domain within traditional systems theory. It regards organizations as social systems, which are solely consisting of communication acts related to each other [Luhm84]. Following this approach, sociologically oriented systems scientists imply that a major issue for understanding organizations (or groups of people) is to understand their communication and their ability to communicate [Baec99:22]. These insights coincide with another recent approach: Communities of Practice (CoP). Here, ongoing informal communication is developing networks of employees, which eventually constitute CoPs, which are defined as a group of people bound by informal relationships who share common practices [LaWe91]. The role of extending or even replacing offline by online communication and the consequences for sociability in the enterprise is further discussed in the research domain of Computer-mediated Communication (CMC). Traditional work in CMC is focusing on the bandwidth of media (Media Richness Theory, Social Presence Theory) and determines its role for conveying the meanings or intentions of communication partners [DaLe86]. The studies found that due to low bandwidth, text-based online communication has been considered ineffective for developing interpersonal relationships, since there are greater chances of misunderstanding, and incorrect impressions are often generated [Pree00]. On the contrary, other researchers argue that such traditional theories may not apply to digital communication. For example, Parks and Floyd [PaFl96] report that an online communicator can develop an interpersonal relationship with someone met via an Internet newsgroup despite a lack of synchronous or spontaneous interactions. Walther [Walt92] also suggests that even though it may take longer than in face-to-face settings, online communicators can develop interpersonal relationships by devising their own relationship strategies in CMC to deliver social and emotional cues (also cf. [CTK05]). These relationship forming aspects are further substantiated by a recent study: Despite the virtuality and the large size of electronic groups, a survey of Berge and Collins [BeCo05] substantiates that there is a perception of community.

The increased interest of the above research domains implies the vital role of understanding the complex networks resulting from the abundance in lateral communication. It is fundamental to be able to create, foster, and manage a modern, competitive, and intelligent organization.

2 SNA – the current approach for understanding network structure

Substantially triggered by increased corporate focus on networking between employees is the desire to observe, analyze, and even ‘measure’ the actual state, an employee network is in. Next to approaches based on simple activity logging (e.g. [Coth00]), the rapid and regular advance in social network research provides a vast body of related measurements and methodologies [WaFa94]. The according field of Social Network Analysis (SNA) is defined as a framework for the analysis of structured social relationships [WaFa94], which in the organizational context can reflect role-based authority relationships of formal organizational structures, informal structures based on communication, information exchange, or affection [TTF79].

The main hypothesis of SNA is that human behavior is influenced by structural properties (e.g. restrictions). This shifts the focus towards observing relationships between actors and the actor’s embeddedness in a complex relationship network. This relationship network and the individual relationships have influential structural conditions. According to Wellman [Well97], such social networks are virtually present whenever a group of people interacts electronically. This enables the systematic examination of such computer-networked communities.

According measurements within the domain of SNA can either include composition variables, i.e. the number and properties of actors, or structural variables, i.e. the properties of relationships. Among the most important factors for evaluating actor networks are currently network size, relationship strength, network roles (broker, gatekeeper, pulsetaker, hub, isolate, transmitter, receiver, carrier), degree (activity, prominence, symmetry or reciprocity), betweenness and centrality, density, and diameter.

For example, relationships in a network can differ in their strength, although there is no specific definition of what is weak and what is strong. It has been found, that pairs who maintain strong ties are more likely to share the resources they have [WeWo90]. Further, a network’s diameter, i.e. the longest distance between two nodes in a network, is a proxy for the likelihood that information passes through the net. Network density is the ratio between existing ties and the maximum possible ties. It indicates if a network is tightly or very sparsely connected. The

density can be applied together with relationship strength to discover clusters or groups in a network. They usually are highly interconnected sets of actors (thus having a high density). The evaluation of subgroups or epicenters of communication can help to identify structural patterns of the overall network. The nodal degree is a measure of the activity and connectedness, assuming that connections need activities to maintain strength. A network can be characterized by its average nodal degree, which shows the average number of connections of each actor. If an actor has a degree of zero, i.e. no direct relationships, he is called an isolate. Related to the nodal degree is the measure of author centrality. Larger social networks tend to have more heterogeneity in their social characteristics and more complexity in their structure [WePo97], which can be measured by degree distribution.

3 Research Objectives - Towards Dynamic Network Visualization and Analysis

The existing indicators, mainly provided by the field of Social Network Analysis, help to evaluate a variety of properties of a network which emerges from ongoing virtual communication. Although relevant properties like dense areas, general (high versus low) activity levels of the network, balance of relationships, likelihood of transporting information from one end to another, and essential actors are identified, the method provides an additive picture of the final state of the network to derive its insights. For example, it can not be observed, if the actor reacted to external events, or if he had a long steady or a short but quick growth in his degree. Alternatively, an actor with a low degree might have experienced restrictions for his connectedness induced by his specialized field of expertise. Finally, the strength of a relationship does not inform the analyst, if this relationship is stable or eventually already decaying again.

Such shortcomings of static visualizations and measurements motivated the research project Commetrix, which started in 2004 (cf. [Trie04; Trie05]). Next to providing the maximum possible visual transparency of otherwise invisible communication networks, the research seeks to extend the insights from Social Network Analysis by enabling observation of dynamic network processes. Periods and areas of decay, stagnation, growth, acceleration, or deceleration can lead to improved analysis of network stability (or change). The emergence of central nodes and clusters can be visually traced and quantitatively measured, which helps with the issue that

“understanding of the betweenness of these centre nodes changes once the temporal nature of the network is revealed” [MMB05:1218]. This can include observation, visualization, and measurement of changes in density, the integration of dispersed substructures, the stability of components, the growth of relationships or their maintenance, equilibrating (balancing) processes, or the sensitivity and reactions to events. By working on such novel means of observing and analyzing real networking processes, the actual underlying lifecycle of a community can become the focus of analysis. The research is based on the fact, that the development of Social Network Analysis has always been strongly related to visualizations in the form of Sociograms [More34]. We hence suggest that extensions towards dynamic SNA should consequently also be founded on sophisticated means of longitudinal visualization to cope with the resulting complexity. This is also substantiated by McGrath and Blythe [McBl04] who find, that visual motion has a positive effect on the accuracy of viewers' perceptions of change in status in a social network.

In summary, our research objective is hence to develop a methodology based on longitudinal visualization of evolving network structures, which is able to extend the insights of Social Network Analysis by detecting and showing the dynamic behavior behind the static measures utilized by SNA. The visualization method builds on dynamic network graphs and needs to be able to transparently capture time periods and structural areas of decay, stagnation, growth, acceleration, or deceleration. It needs to visually trace and quantitatively measure the emergence of central nodes. Another focus is the observation of overall network stability (density, integration of substructures, stable components, growing relationships, relationship maintenance, balancing processes), change, sensitivity, and reactions to events.

Following the Design Research methodology [e.g. cf. VaKu04], this research objective will now comprise the framework for the design of a novel visualization approach for analyzing dynamic network evolution. Afterwards, its novel insights are demonstrated and evaluated by analyzing the dynamic evolution of Enron's e-mail corpus.

4 Designing a Method for Dynamic Visualization and Measurement

Advancing from current pictures of social network graphs (Sociograms) to animations of graph evolution over time raises methodological and technical questions. How can the graph be

defined to cover for longitudinal data streams and how can such a graph best be visualized to create an organic view on network evolution.

The methodological aspect should be based on graph theoretic foundations, but needs to consider time-related data. The according sociomatrix [WaFa94] is formally described with a set of actors and their relations. Both constitute a Graph $G = (N, E)$, which consists of a finite set of nodes N and a finite set of edges E that are constituted by pairs (u,v) of nodes [Batt99:3]. To accommodate for the usually sparse networks found in the computer-mediated communication domain, we store such sociomatrices in list form to be able to manage large datasets efficiently. Capturing time related data into such a model now requires a list of events, i.e. communication acts. In other words, the individual message (and its reply) constitutes an event on a time scale, which impacts the relationship strength. This can either be switching from null (the relationship is absent) to one (the relationship exists) or it is growing stronger by a certain value (weighted relationships). This approach can be extended to represent more information, as other changes in data can in principle also constitute events. For example, change in author properties, decay of a relationship, the appearance of an author without acting, etc.

The above model is extended to represent additional information. This allows capturing as much information from individual electronic discourses as possible in a systematic and standardized way. It is allowing for subsequent analysis in one model, which we termed 'communigraph'. Such a communigraph consists of two elements. First, a data model has to be defined to capture the individually specified properties of a digital communication network. Primary elements are authors, their messages, and the relational information, i.e. how the authors and their messages relate to other authors' comments. Additional data stores information about the authors, including their names, e-mails, or indexes for anonymous analysis. Further, messages are saved together with their subjects, body texts and time stamp. Finally, dynamic properties can be defined to store individual data about each discourse (i.e. peer evaluations or organizational departments (cf. application in [CTK05]). We further include keywords in the data model, as we regard the technological challenge of integrating keyword analysis and Social Network Analysis as a major potential in understanding large decentralized network structures.

Next to the data model of the communigraph, there need to be visual model specifications, which utilize the underlying data structure to enable insights into the complex structures,

activities, and contents of electronic collaboration. The two main elements of the visual model are author and relation. The relation is the sum of interpersonal contacts between authors in the form of replying messages. Authors are represented by a node (sphere). Means like labels, node colors, node size, or a number of rings around the node are available to represent author core attributes like name, e-mail, and index number and their properties (defined as conveying a qualitative or quantitative meaning) like for example the author's location, his number of contacts, the most important keywords, or his organizational hierarchy. Relations are graphically represented by an edge with properties like the number of messages it represents, the average evaluation of its contents, the time stamps of its messages, or its main keywords.

The technical aspect of visualizing and animating this data structure over time is very challenging. The essential element of a longitudinal visualization of the above event-based graph is the handling of transitions between different network states. This aspect can be termed 'transition problem'.

In general, our experiences are consistent with related work of Moody et al. [MMB05] and Gloor et al. [GLZD04], two research groups which are working in the field of longitudinal network visualization. Both suggest technical approaches to visualize temporal social networks by creating movies of animated graph structures. Generally, both existing approaches handle the transition problem between two successive network layouts via a linear transition between different states of a network, represented by a sliding window. For example, Gloor et al. suggest a time frame of $[d, d+n]$, where d represents the current day and n is the size of the time frame (e.g. 30 days). All communication within this frame will be used to calculate a network's state at the time point d . Subsequently, "the animation of the changing layout is interpolated between [...] keyframes" [GLZD04].

By looking at the resulting visualization, it can be found, that such rendering of transition frames disturbs the impression of organic movement. Nodes move in changing speeds and moreover in changing directions across the screen, eventually also crossing other nodes on their way. The main problem arises from inconsistent repositioning methods. For example, force-directed layouting could first move a node away from the center. Subsequent interpolation to bring a node to the next force-directed layout position of the subsequent time frame often results in a translation where nodes move sideward but equidistant to the center. The user has to mentally switch between the organic impression of attraction and repulsion of particles and the linear translation between two subsequent network positions. In visual outputs using

interpolation frames, centers often move without being attracted by other nodes (in the transition phase), intuitively resulting in an unstable impression of an actually stable network transition. Being trained to evaluate stable parts by their inertia, the user is distracted from observing how new nodes find their position, because established centers also shift positions and all adjacent nodes in their clusters with them. We conclude that though being a technical option, interpolative transition between sequential network states (also known as ‘tweening’) results in suboptimal impression of node movement and network evolution and thus seems to be a suboptimal approach to the transition problem. The approach suggested in this paper is omitting the idea of ‘morphing’ between two layout frames. Rather, we suggest an adapted version of the layout algorithm, which uses iterative layout steps to animate the changes. New nodes are added to the layouting algorithm at the time, when the resulting events occur. To keep a conceivable behavior of existing components through time, new layout calculations are initialized with the node positions from the previous time frame. We propose a relative temperature approach which is able to produce an ergonomic and conceivable animation (for visual results see next section). It builds on personal node temperatures to control the graph’s overall temperature function (i.e. the energy function where low remaining tension in the graph corresponds with a high fit to optimal position). Personal node temperatures correspond with the nodes’ degrees. It indicates how inert they are in their position (i.e. how far they can move). Nodes with high degrees automatically would get a lower temperature and accordingly are less movable. These layout modifications reduce the movement of large structures while new nodes, entering the network, efficiently move towards their equilibrium position.

The adapted force-directed layout approach further includes longitudinal temperature control: Keeping the temperature changes between the layout iterations low and decreasing temperature slowly over time will force the nodes to slowly move around each other. Since the proceeding state is well layouted and the changes between states of a network are usually small, new animation frames will most likely be well layouted themselves. Mainly the parts of the network will move which are changing since the rest will - in terms of temperature (which relates to inertia) - be close to its equilibrium state. The advantage is now, that the force directed layout is the only layouting principle (instead of interchanging between force-directed and transitional linear layout. This results in a much better and intuitive understanding of the network’s evolution as it keeps established parts as stable as they should appear, while drawing the

user's full attention to the areas where the actual change happens. Movement is thus not a general behavior of the network graph but is directly related to structural changes.

This eventually gives the impression of a real living system of interactive elements in a network, which we found to be of fundamental importance for optimal dynamic network visualization.

The resulting iterative layout approach introduced above is able to produce interim layouts for every event in a network's evolution. It is not influenced by the final structure but only shows the network as it would have been if the sample ended on the current time t . Optionally, old nodes can be excluded to show only the recent activity in the network's evolution.

To summarize this section, we are suggesting a novel method for longitudinal network visualization, which is based on an animation of a series of events (i.e. communication acts) which finally constitute the communication network. On the methodological level, we utilize data of sociomatrices stored in list form and extend them to create the communigraph. This consists of a data model and visual model elements, which together model information about author and relationship properties, like author affiliation, relationship strength, important keywords (topics), etc. Technically we propose an improved approach to longitudinal layout, based on a concept for node-dependent and dynamic temperature control.

In the next section, we will now apply and evaluate this method for visualizing streaming data of communication networks using data about the e-mail communication network of Enron. It allows for visual observation of some novel longitudinal effects.

5 Experiments –Dynamically exploring an e-mail communication network

For demonstration and evaluation of the above method we will take a sample of Enron's e-mail network. It consists of a set of 19808 messages exchanged between 151 authors during a 959 day period ranging from 11/05/1999 to 06/21/2002. Using the analytical methods of SNA, we arrive at a preliminary static picture of the network: It contains 1525 relationships with the average relationship strength (replies on messages) of 26.55. The network's density is 13.47 percent. The diameter only amounts to 4. The core group of active people, which together accumulate 80 percent of overall message volume, has a share of 27 percent. The maximum degree is 74 contacts.

To compare the insights generated by dynamic network visualization with conventional network statistics, we discuss the following four research questions, which all relate to dynamic evolution of virtual communication networks:

- 1) How can the network behavior be described in terms of separation of clusters?
- 2) How can longitudinal analysis show artifacts of additive pictures provided by static SNA, using the most central node and the establishment of its position as an example?
- 3) Which organizational roles are responsible for spanning large distances and for integrating separate parts of the corporate network? What role does organizational hierarchy play in such processes of network integration?
- 4) How can dynamic visualization help to observe and evaluate the current activity and thus also the current level of change. Is there a relation to external events noticeable?

To approach the first research question (1), Figure 1d shows the final static picture as it would have been produced by conventional Social Network Analysis. A network of only one component can be seen. The most central node (in terms of degree centrality) is at the center. The layout shows a very dense area at the center (named section 1). Further, there is a strongly interconnected section at the top right area (section 2) and some more peripheral nodes at the bottom (section 3). The other parts of Figure 1 (a-c) are snapshots taken from an animation generated with the novel method. They show three intermediate communigraphs¹. This now allows to visually trace the various activities which finally constitute figure 1d. In the first period, no separation can be observed, a central network exists. However, in period 2 (Figure 1b), section 1 is increasing its density and section 3 starts to exist. It is connected to the main section 1 by only a few nodes. First traces of section 2 are also visible. A further small component (section 4) is visible on the right hand side but it does not survive period 3 (Figure 2c). Section 2 is much more separate now, whereas section 3 has established a very broad connection (many connecting nodes) to section 1. Many authors of section 1 are growing in their degree (denoted by node size), thus improving the density of that area. In the final period 4, the bridge between section 3 and 1 gets narrower again, section 3 thus achieves more separation, section 2 is also remaining separate. However, all areas are connected with each other. The animation further highlights, that there are very important events, which largely influence the resulting structures. Two examples for very important network changes are an

¹ As this print medium does not allow for movies showing the change in networks, all animations of network evolution have been made available for viewing online at <http://www.commetrix.de/enron/>.

action caused by the CEO (K. Lay) on 08/24/2001 or the most central employee's action on 02/07/2002. Their relevance is mainly due to spreading a message to a larger group of people and receiving replies to it. However, it remains to be explored, if they are also that important from a contents perspective or only in terms of structural network activity.

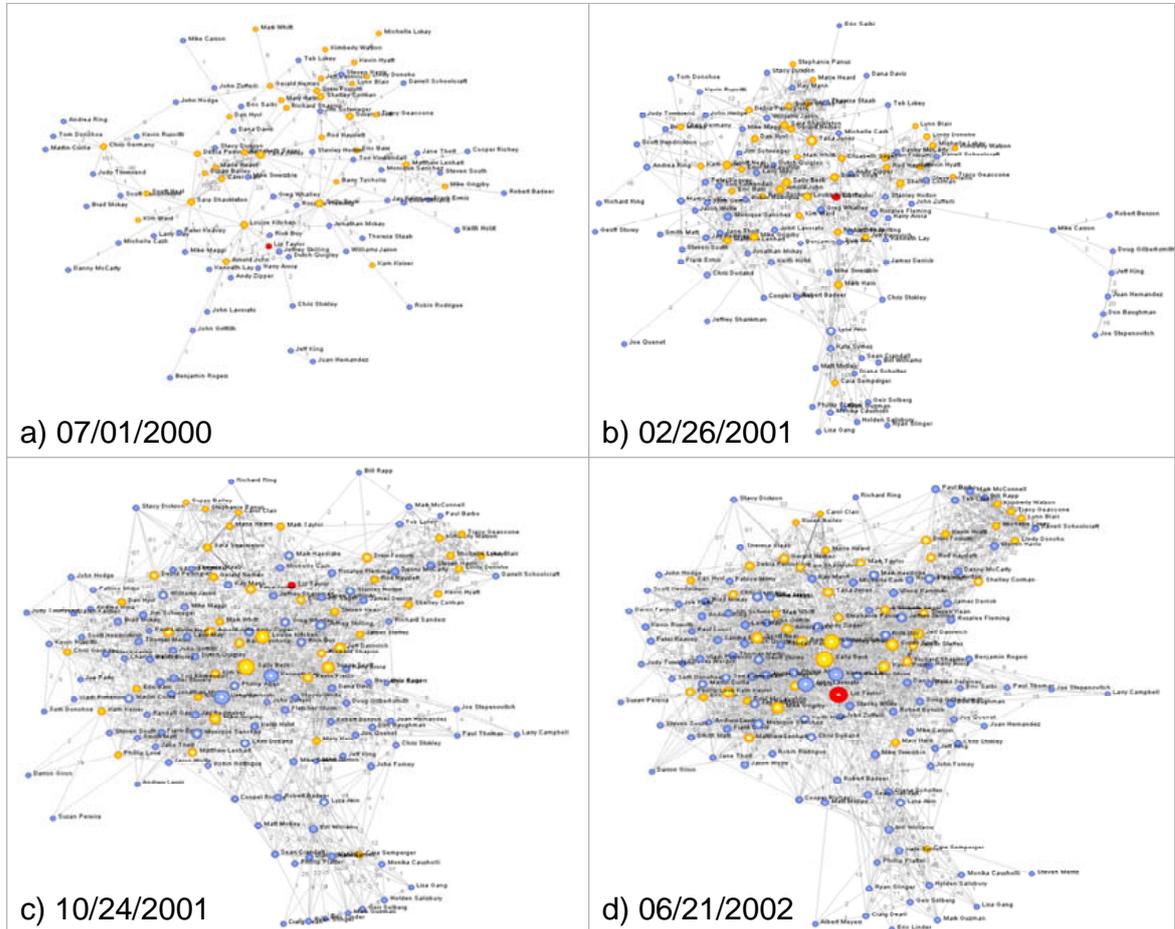


Fig. 1: The evolution of the most central author's position in the network. The observed (most central) author is marked red, the size is representing the nodes' degree. Orange nodes represent members of the 'core group' of active people. The final static picture of the ENRON e-mail corpus as would be produced by an SNA Sociomatrix is shown on the bottom right. The according animation is available at <http://www.commetrix.de/enron/>.

Figure 1 also yields a first insight into research question 2 concerning artifacts of additive pictures of networks. An example showing this wrong impression of a network structure can be demonstrated, when the most central actor gets observed to see how she established her position: Although the author was active in the network during the complete period, almost all of her centrality has been achieved in the last quarter (note the difference in node size between Figure 1c and 1d). This comparison of different states of the network can now be analyzed in more detail by looking at the dynamic evolvement in the last quarter of the sample period¹. To produce a transparent picture, only the direct ego-network around the observed node is shown.

Further, only relationships with a weight of more than one are involved. Figure 2 highlights the animation's main result: after a long period without significant change, between February 5th and 7th of 2002, the centrality of our observed node increases by 50 percent in just three days. Afterwards it remains stable again until the end. This implies that the most central node established its position in a very fast burst of activity instead via continuous growth. Color has been employed to code function. It can be seen, that many new vice presidents and common employees entered the central node's egonet during this process. The observed non-linear behavior in node activity is a first insight, which can not be derived from the final additive network picture. Rather, it shows that the eventually most connected node seems to have a short period of untypical behavior, which brought it into its position.

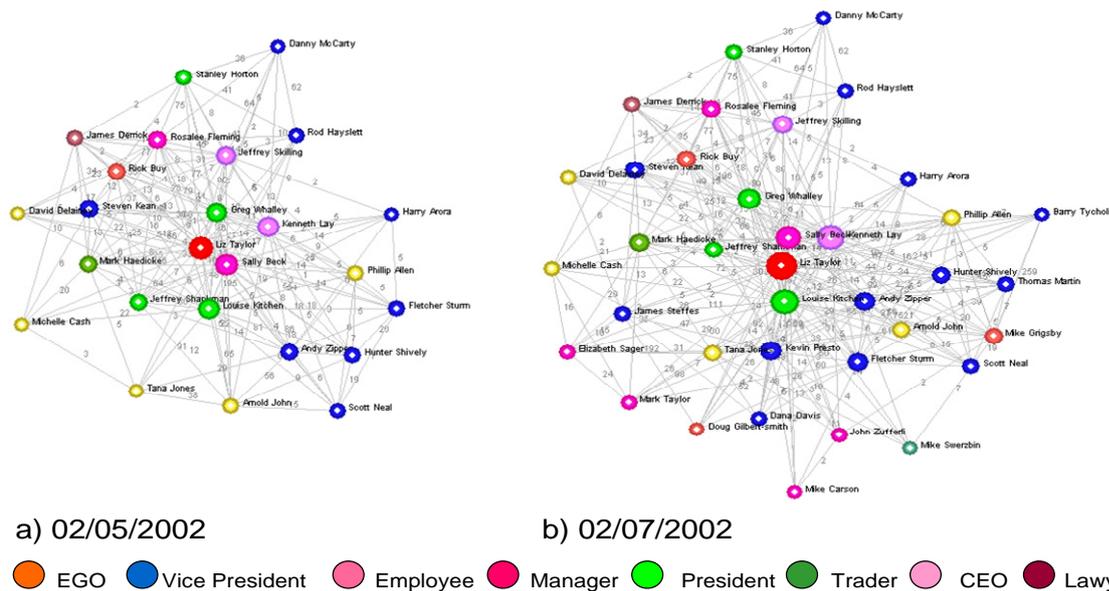


Fig. 2: The evolution of the most central author's position in his ego-network between 02/05/2002 and 02/07/2002 (constant scale). The observed author is central and marked red. Yellow nodes have no known job position. Only relationships with a weighting of 2 and more are shown. Their increase in degree is from 24 to 36 contacts. The according animation is available at <http://www.commetrix.de/enron/>.

By defining a sliding window, we can now eliminate messages which are falling outside the time window, i.e. to delete message activity from the visualization which is too old and has decayed. The result of such an animation is that the observer can follow the current activities of the network's evolution without being distracted by past structures. All visible activities directly represent change in the network. Nodes get active and can initialize or join clusters. These clusters are either dissolving or growing over time depending on the ongoing activity level. We found, that this visualization mode helps to understand, who contacts whom to actually establish (or change) a communication network. Periods of high activity can be implied, when many

nodes and relations remain active in the time frame. Brokering actions are becoming visible, even if the nodes do not establish broker positions in the long run.

This mode of analysis is now being applied to investigate the third research question: We want to analyze the brokering actions in the evolving network overtime to find out whether brokers which are higher up in the organizational hierarchy have a different role than employees. For that, we code the job positions of the e-mail network's authors by node color and label. This allows observing, which nodes were active and which connected sections to establish the final network structure.

Figure 3 depicts four frames of a part of the overall network evolution, which includes an important brokering situation. A node with job position President connects three otherwise disconnected segments of the network. Interestingly, the other segment's contacts are also above management level: one director and two vice presidents. This visualization raises questions regarding the properties of important acts of brokering between dispersed parts of the network have and regarding the role different job positions have in this process. It could be assumed, that positions higher up in the organizational network show more such brokering activity (connecting otherwise distant parts) than normal employees, as managers should be better connected across the enterprise.

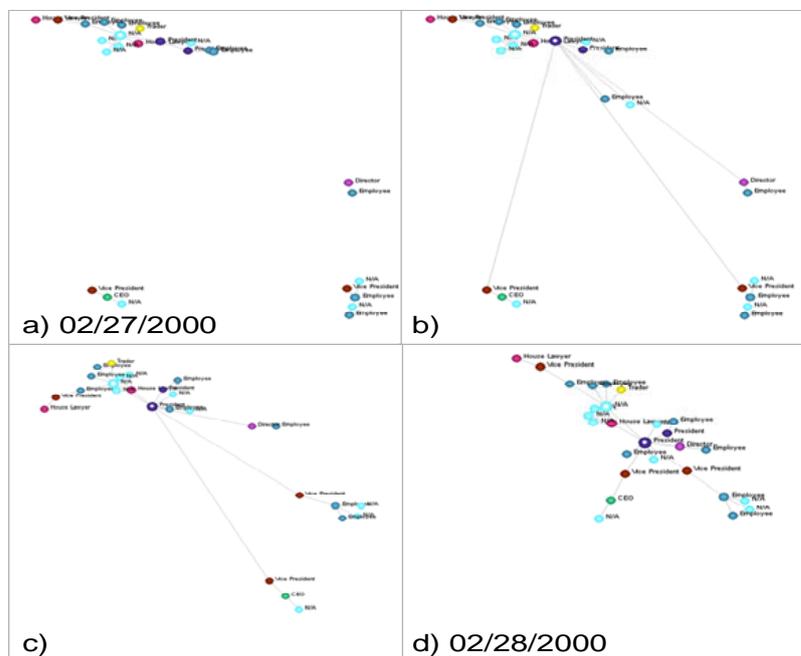


Fig. 3. A node with job position President connects three otherwise separate clusters (via two vice presidents and one director) on 02/28/2000. Four separate frames of the according animation. Time filter is set to show only one past month of mail activity.

To further investigate this question, for the year 2000 all brokering activities have been detected by monitoring the according animation¹. They include the creation of relevant shortcuts (shortcuts which are not only a result of triadic closure, i.e. bypassing the intermediary node) and the integration of clusters. Together, 74 individual actions have been classified as brokering actions in the one year sampled. 25 of those connected previously separated network segments and further 49 actions created fundamental shortcuts within one large component. For 36 of these actions, the organizational function of both adjacent participants is known (as 52 of the 151 job positions of the sample's actors have not been disclosed). With that information, the role of cross-hierarchical relationships for establishing the final interconnected network can be analyzed. Out of 36 brokering actions in the year 2000, 9 actions (25 percent) have involved only top management positions, further 9 actions (25 percent) only employees. The majority of 18 connecting actions (50 percent) have been cross-hierarchical. This quantitative pattern is supported by the visual impression, which implies that managers connect with distant employees in a brokering action to join separate parts of the communication network. Overtime this behavior results in the establishment of a single integrated component.

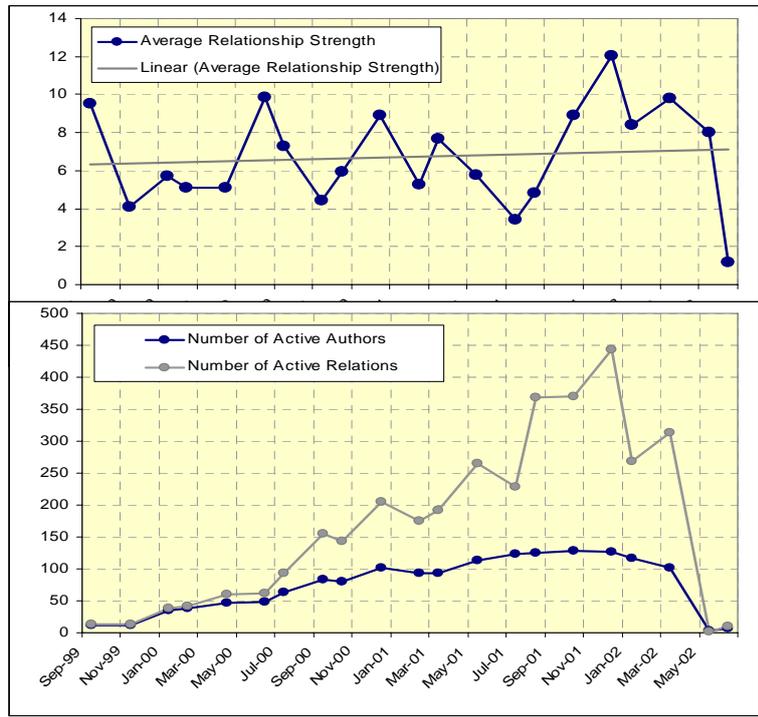


Fig. 4. Level of network change. Time window, showing only the last month of network activity to indicate the change in active authors, active relationships and current average relationship strength across the time period. A clear peak in change and in relationship strength is visible around the time when the Enron scandal was published.

The final research question (4) requires an analysis of the network's level of change. This can again be indicated by observing a time window animation (e.g. of one month). While moving forward in time measurements of active nodes, active relationships between them, and the current average relationship strength were measured. The active nodes and relationships can be interpreted as the added amount of network activity and thus the change in the network's overall structure. Figure 4 shows the quantitative results. The measurements show that a clear peak of change is visible around the time when the Security and Exchange Commission started their investigations into the Enron fraud scandal on 10/31/01 and Enron filed for bankruptcy in December 2001. The Enron e-mail corpus hence seems to react to a fundamental external event by adjusting its level of change.

In summary, these four analyses showed that the developed research method based on longitudinal visualization and measurement is able to provide new insights: It helps to visually observe and understand the dynamic behavior of virtual communication networks.

6 Conclusion and Outlook

Although the field of Social Network Analysis provides a series of useful measures for the analysis of communication networks [WaFa94], there are various shortcomings due to the static nature of the underlying model. This can lead to wrong impressions about nodes' status or even a lack of understanding, especially as far as evaluation of a network's dynamic properties is concerned.

By conducting an analysis of the dynamic behavior of a corporate e-mail network, a novel research method that is being introduced in this paper, proved its ability to help with the detection and description of periods, areas, or components of high activity. It further supports the analysis of structural integration (brokering) actions or the establishment of special network positions over time. It has been found, that the most central node is actually likely to be only a statistical artifact, as it acquired a large part of its high degree in only three days, by some obviously untypical behavior. Animating the network evolution further showed, that clusters form and decay overtime, which is largely affecting the final picture of the network. By eliciting the network's current zone of activity, change could be visually captured. This technique has been applied to visually study brokering actions, which finally resulted in a network with only one component. We found that the majority of such brokering actions included cross-

hierarchical relationships. Managers were very actively involved in connecting otherwise distant parts of the overall network. Here, findings are consistent with a recent publication by Kossinets and Watts [KoWa06:89]. They observe different network measures and their varying levels of stability. One of the main findings regarding network evolution is that bridges in their sample show an unstable nature [KoWa06:90] and seem to provide only temporary advantage for their owner. The method of only showing change also allowed to identify a peak in the speed of network evolution (activity) in fall of 2001. This coincided with an important external event.

It can thus be concluded, that the introduced research approach allows for novel insights, which are able to improve the understanding of network dynamics, both in a visual and in a quantitative way. This can help to better observe and understand corporate virtual networks and their hierarchical, organizational, or locational integration behavior or artifacts in the future.

Based on these initial experiments with longitudinal network behavior and first approaches to actually visually observe, what happens in a network's evolution, future research will now need to systematize, augment, and extend the exploratory experiments discussed in this paper to arrive at a methodology together with according measures, which is providing novel visual but also robust scientific insights into network dynamics.

References

- [Baec99] Baecker, D. (1999): Organisation als System (in German). Frankfurt am Main
- [Batt99] Battista Di, G. (1999): Graph Drawing: Algorithms for the Visualization of graphs. Upper Saddle River, New Jersey
- [BeCo00] Berge, Z. L.; Collins, M. P. (2000): Perceptions of e-moderators about their roles and functions in moderating electronic mailing lists. Distance Education: An International Journal, 21(2000)1, p. 81-100.
- [Coth00] Cothrel, J.P. (2000): Measuring the success of an online community. Strategy & Leadership 28(2000)2, p.17-21, <http://www.participate.com/research/StrategyLeadership.pdf>, Accessed 20.11.2005

- [CTK05] Cho H.-K.; Trier, M.; Kim, E. (2005): The Use of Instant Messaging in Working Relationship Development: A Case Study. *Journal of Computer-Mediated Communication*, Volume 10, Issue 4, July 2005.
- [DaLe86] Daft, R. L.; Lengel, R. H. (1986): Organizational information requirements, media richness and structural design. *Management Science*, 32 (5), p. 554-572.
- [DaLe93] Daft, R. L.; Lewin, A.Y. (1993). Where are the theories for the new organization forms? An editorial essay. *Organization Science*, 4, i-vi.
- [GLZD04] Gloor, P.; Laubacher R.; Zhao, Y., Dynes, S. (2004): Temporal Visualization and Analysis of Social Networks, NAACSOS Conference, June 27 - 29, Pittsburgh PA, North American Association for Computational Social and Organizational Science, 2004.
- [KoWa06] Kossinets,G.; Watts D.J. (2006): Empirical Analysis of an Evolving Social Network, *SCIENCE VOL 3116 JANUARY 2006*, p. 88-90
- [LaWe91] Lave, J.; Wenger, E. (1991): *Situated Learning. Legitimate Peripheral Participation*. Cambridge University Press, Cambridge
- [LuBa94] Lucas, H.C.; Baroudi, J. (1994). The role of information technology in organization design. *Journal of Management Information Systems*, 10, 9-23.
- [Luhm84] Luhmann, N. (1984): *Soziale Systeme – Grundriss einer allgemeinen Theorie* (in German). Frankfurt am Main
- [McBl04] McGrath, C.; Blythe J. (2004) : Do You See What I Want You to See? The Effects of Motion and Spatial Layout on Viewers' Perceptions of Graph Structure. In: *Journal of Social Structure*, 5 (2004)2.
- [MMB05] Moody J.; McFarland D.; Bender-DeMoll S. (2005): Dynamic Network Visualization. *American Journal of Sociology*. *AJS Volume 110 Number 4* (January 2005), p.1206–41
- [More34] Moreno, J. L. (1934): *Who Shall Survive?* Washington, DC: Nervous and Mental Disease Publishing Company

- [PaFl96] Parks, M. R.; Floyd, K. (1996): Making friends in cyberspace. *Journal of Communication*, 46 (1), p. 80-97.
- [Pree00] Preece, J. (2000): *Online Communities: Designing Usability, Supporting Sociability*. Chichester and New York: John Wiley & Sons.
- [Trie04] Trier, M. (2004): IT-Supported Monitoring and Analysis of Social Networks in Virtual Knowledge Communities. In: *Proceedings Sunbelt XXIV Conference*, Portoroz, Slovene 2004.
- [Trie05] Trier, M. (2005): IT-supported Visualization of Knowledge Community Structures., *Proceedings of 38th IEEE Hawaii International Conference of Systems Sciences*, Hawaii, USA, 2005.
- [TTF79] Tichy, N. M.; Tushman, M. L., Fombrun, C. (1979): Social Network Analysis for Organizations. *Academy of Management Review*, 4(1979)4, p. 507-519.
- [VaKu04] Vaishnavi, V.; Kuechler, W. (2004). "Design Research in Information Systems" January 20, 2004, last updated January 18, 2006. URL: <http://www.isworld.org/Researchdesign/drisISworld.htm>. Accessed: 20.06.2006
- [Walt92] Walther, J. B. (1992): Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19 (1), p. 52-90.
- [WaFa94] Wasserman, S.; Faust, K. (1994): *Social Network Analysis: Methods and Applications*. Cambridge University Press: Cambridge, 1994.
- [Well97] Wellman, B. (1997): An Electric Group is Virtually a Social Network. *Culture of the Internet*, p. 179-205.
- [WeWo90] Wellman, B., Wortley, S. (1990): Different Strokes from Different Folks: Community Ties and Social Support. *American Journal of Sociology*. 96(1990), p. 558-88.
- [WePo97] Wellman, B., Potter, S. (1997): The elements of personal community. In: Wellman, B. (ed.): *Networks in the global village*. Norwood, NJ, 1997.

Multiperspektivische Wissensvisualisierung für Wissensaustausch in heterogenen Community-Netzwerken

Jasminko Novak

Institut für Informatik, FG Informationsmanagement
Universität Zürich
8050 Zürich
novak@ifi.unizh.ch

Kurzfassung

Wissensnetzwerke, die unterschiedliche Communities of Practice einbeziehen, spielen eine wichtige Rolle für die Förderung von Innovation in Organisationen. Der gemeinschaftsübergreifende Wissensaustausch wird aber durch unterschiedliche Wissensdomänen, den fehlenden gemeinsamen Kontexts und die geringe Kommunikation erschwert. Dieser Beitrag diskutiert einen Ansatz zur Überwindung solcher Barrieren mittels Wissensvisualisierung. Die entwickelte Methode ermöglicht es, dynamische Wissenskarten zu erstellen, die persönliche und geteilte Perspektiven aus unterschiedlichen Communities abbilden, in Beziehung setzen und für die Informationssuche nutzbar machen. Eine Laborstudie zeigt, wie ein solcher multiperspektivischer Zugang zu Community-Informationsräumen den gemeinschaftsübergreifenden Wissensaustausch unterstützen kann.

1 Einführung

Die Bedeutung von Communities, als informelle, soziale Netzwerke zur Wissensgenerierung und Wissensverteilung in Organisationen, die auf gemeinsamen Interessen oder Aktivitäten basieren, wurde ausgiebig untersucht [BrDu91, Weng98]. Bisherige Arbeiten konzentrieren sich hauptsächlich auf die Unterstützung des Wissensaustausches innerhalb relativ homogener Gemeinschaften, die Teilnehmer mit ähnlichem Hintergrundwissen verbinden. Gleichzeitig ist der Bedarf gestiegen, den Wissensaustausch zwischen unterschiedlichen Gemeinschaften, mit sehr spezialisierten Fachkenntnissen zu unterstützen [BoTe95, BrDu91, Doug92]. Ein üblicher Ansatz ist die Errichtung von gemeinsamen Wissensportalen, die einen zentralen Zugang und

Treffpunkt für verschiedene organisatorische Einheiten und Gemeinschaften zur Verfügung stellen. Entsprechende Unterstützung für solche Plattformen ist jedoch noch nicht verfügbar. Sie basieren meistens auf standardisierten Wissenstaxonomien, die das Wissen unterschiedlicher Communities in eine gemeinsame, festgelegte Struktur verbinden [BoBT02]. Auch werden Kommunikations- und Koordinationswerkzeuge angeboten, die aber für den Wissensaustausch in Teams und *innerhalb* von Gemeinschaften entwickelt wurden (gemeinsame Workspaces, Chat, Awareness).

Verschiedene Studien identifizieren aber besondere Anforderungen für den gemeinschaftsübergreifenden Wissensaustausch, wie z.B. die Koexistenz unterschiedlicher, lokaler Perspektiven inklusive ihrer Visualisierung und Koordination [Dough92, BoTe95, Swan01]. Auch HCI Arbeiten zur Wissenskonstruktion bei der Informationssuche (eng. Sensemaking) betonen die kritische Rolle der Wissensstrukturierung und des kontextualisierten Zugangs zu Informationen zur Lösung von schlecht strukturierten Problemen [RSPC93, QuFu05]. Die Entwicklung von Werkzeugen, welche diese Bedürfnisse erfüllen, ist jedoch nicht sehr weit fortgeschritten.

Dieser Beitrag stellt eine Methode und ein Werkzeug zur Wissensvisualisierung vor, welche diese Aspekte berücksichtigen und speziell zur Förderung des gemeinschaftsübergreifenden Wissensaustauschs entwickelt wurden. In bisherigen Veröffentlichungen wurden die technische Realisierung der Wissenskartenmethode [NWFS03] und ein frühes Model ihrer Anwendung zur Unterstützung des Wissensaustausches [NoWu04] bzw. die formative Usability Evaluierung des ersten System-Prototyps [NoWu05] vorgestellt. Davon ausgehend wurde im iterativen Design-Prozess ein neues, theoretisch besser begründetes Wissensaustausch-Modell entwickelt und in Form eines interaktiven Werkzeugs realisiert (Knowledge Explorer II) und evaluiert.

Diese Ergebnisse werden hier vorgestellt. Zuerst werden die Barrieren für den gemeinschaftsübergreifenden Wissensaustausch geschildert. Daraufhin wird ein Modell des gemeinschaftsübergreifenden Wissensaustausches eingeführt, welches das Framework des „Perspective Making – Perspective Taking“ [BoTe95] mit der Wissenskonstruktion in der Informationssuche verbindet [RSPC93]. Anschließend wird gezeigt, wie die entwickelte Wissensvisualisierungsmethode diese Anforderungen berücksichtigt und wie ihre Anwendung zum multiperspektivischen Zugriff auf Community-Informationsräume den gemeinschaftsübergreifenden Wissensaustausch unterstützt. Zum Abschluss werden die Ergebnisse der Evaluierung in einer vergleichenden Laborstudie dargestellt.

2 Barrieren des gemeinschaftsübergreifenden Wissensaustausches

Eine typische Konstellation, in welcher der gemeinschaftsübergreifende Wissensaustausch stattfindet, sind heterogene Wissensnetzwerke. Sie verbinden Experten aus unterschiedlichen Disziplinen und Mitglieder unterschiedlicher Communities, die keine kontinuierlich geteilten Interessen haben [MeDa01]. Ihr primärer Zweck ist es, Informationen über die Grenzen einzelner Disziplinen und organisatorischen Strukturen hinaus zu sammeln und zu verteilen [Alle00]. Sie entstehen, wenn Mitglieder verschiedener Gemeinschaften Informationen, Wissen oder Leistungen erwerben müssen, die in ihrem Umfeld nicht verfügbar sind. Die fehlende Kontinuität gemeinsamer Ziele und Interessen resultiert in schwachen Beziehungen (weak ties) und geringer Kommunikation – die Teilnehmer bleiben in ihren Gemeinschaften verankert.

Die Verwendung von Community-Informationsräumen (z.B. gemeinsame Dokumentablagen oder Diskussionsforen) stellt daher nicht nur eine wichtige Modalität des Wissensaustausches *innerhalb* verteilter Gemeinschaften dar, sondern auch für den gemeinschaftsübergreifenden Wissensaustausch. Der Zugang zu Informationsräumen unbekannter Communities wird jedoch erheblich durch die Abwesenheit eines gemeinsamen kognitiven und sozialen Kontexts erschwert, auf Grund dessen die Bedeutung von Informationen bestimmt wird [LaWe91].

2.1 Mentale Modelle und Interpretationsschemata

Unterschiedliche Gemeinschaften besitzen unterschiedliche mentale Modelle [Doug92] oder Interpretationsschemata [BoTe95], welche bestimmen, wie ihre Mitglieder die Bedeutung von Informationen und Erfahrungen interpretieren. Ihr Wissen wird in unterschiedlichen Terminologien ausgedrückt, wobei die Bedeutung und Verwendung von Begriffen stark kontextabhängig ist. Gleiche Begriffe werden oft auf unterschiedliche Art und Weise verwendet [BoTe95] und ähnliche Themen im Kontext unterschiedlicher Probleme gesehen. Das erschwert es, Wissen durch Weitergabe von Informationen auszutauschen. Wissensintegration erfordert ein Verständnis der zugrunde liegenden Interpretationsschemata [Swan01], da andererseits standardisierte Konzepte den Wissensaustausch verhindern, anstatt ihn zu fördern [Dough92].

2.2 Geringe, direkte Kommunikation und mangelnde Externalisierung

Solch implizites Wissen wird entweder durch Sozialisierung oder Internalisierung gewonnen [NoTa95]. Die Sozialisierung bedarf häufiger sozialer Interaktion, die in verteilten Communities auf intensiver, informeller Onlinekommunikation basiert (Chat, Messaging). Die Internalisierung geschieht durch Verfolgung des gemeinschaftlichen Diskurses und häufige

Nutzung des gemeinsamen Informationsraumes (z.B. Diskussionsforen). Die Kommunikation zwischen Gemeinschaften ist aber selten und verfolgt das Ziel der Beschaffung von sonst nicht verfügbaren Informationen [Alle02]. Eine Sozialisierung kann hier nicht stattfinden. Implizites Wissen kann auch nicht durch herkömmliche Internalisierung geteilt werden, da Mitglieder einer Gemeinschaft weder die Motivation noch die Ressourcen haben, den Informationsaustausch in anderen Gemeinschaften zu verfolgen. Der Mangel expliziter Darstellungen gemeinschaftlicher Wissensstrukturen erschwert dieses Problem. Communities befassen sich selten mit der Erzeugung von Wissensrepräsentationen bzw. leiden an der Notwendigkeit einen breiten Konsens zu erzielen. Dies resultiert in übermäßig verallgemeinerten, wenig nutzbaren Strukturen [Caro99]. Folglich sind Wissensstrukturen von Communities größtenteils implizit und unsichtbar für Dritte.

2.3 Schlecht-definierte Probleme und unklarer Informationsbedarf

Die Notwendigkeit, auf Wissen von unbekanntem Gemeinschaften zuzugreifen, tritt bei schlecht definierten Problemen auf. Dies ist bei interdisziplinärer Arbeit, wissenschaftlicher Forschung und Innovationsprozessen in Unternehmen der Fall [Swan01]. Solche Kontexte erfordern es, die Struktur des Problems und verwandte Wissensdomänen zu erschließen, um die Wissenslücken und das notwendige Wissen zur Lösungsfindung zu identifizieren. Dabei ist der Informationsbedarf sehr mehrdeutig und schwierig mit einer zielgerichteten Suche abzudecken. Beispiele sind Strategie- und Produktentwicklung, Business Intelligence [RSPC93] und selbstgesteuertes Lernen [QuFu95].

3 Ein Modell für den Wissensaustausch in Inter-Community Netzwerken

Das Modell des "Perspective Making – Perspective Taking" [BoTe95] beschreibt, anhand des Wechselspiels zwischen der Entwicklung der gemeinschaftlichen Begriffswelten und ihrem Wissen, wie Wissen zwischen unterschiedlichen Gemeinschaften geteilt werden kann.

3.1 Perspective Making – Perspective Taking

Perspective Making (PM) umfasst Prozesse, durch welche die Mitglieder einer Gemeinschaft ihr Wissen und eine gemeinsame Perspektive entwickeln. Dies schließt Aktivitäten vom gemeinsamen Problemlösen bis hin zur kollaborativen Strukturierung von Informationen und narrativer Kommunikation ein. Durch die Erläuterung der Bedeutung existierender Begriffe, die

Entwicklung neuer Begriffe und dem Aufbau von Beziehungen zwischen ihnen, entwickeln die Community-Mitglieder einen gemeinsamen Wissensraum. Perspective Taking (PT) beschreibt die Erlangung des Verständnisses der Wissensperspektive einer unbekannt Community. Dies umfasst das Verstehen der impliziten Wissensstrukturen, in welchen die Bedeutung von Informationen begründet ist und ihrer Beziehungen zum eigenen Wissen. Das fremde Wissen wird internalisiert, indem es mit eigenen Begriffen ausgedrückt wird (PM).

Der gemeinschaftsübergreifende Wissensaustausch findet also in einem Wechselspiel von Perspective Making und Perspective Taking statt: die durch das Verstehen von „was und wie die anderen wissen“ [BoTe95] und durch die Entdeckung der Beziehungen zum eigenen Wissen gewonnenen Erkenntnisse, werden durch Aneignung von neuen Begriffen ausgedrückt und somit als neues Wissen in die eigene Community eingebracht (Abb. 2).

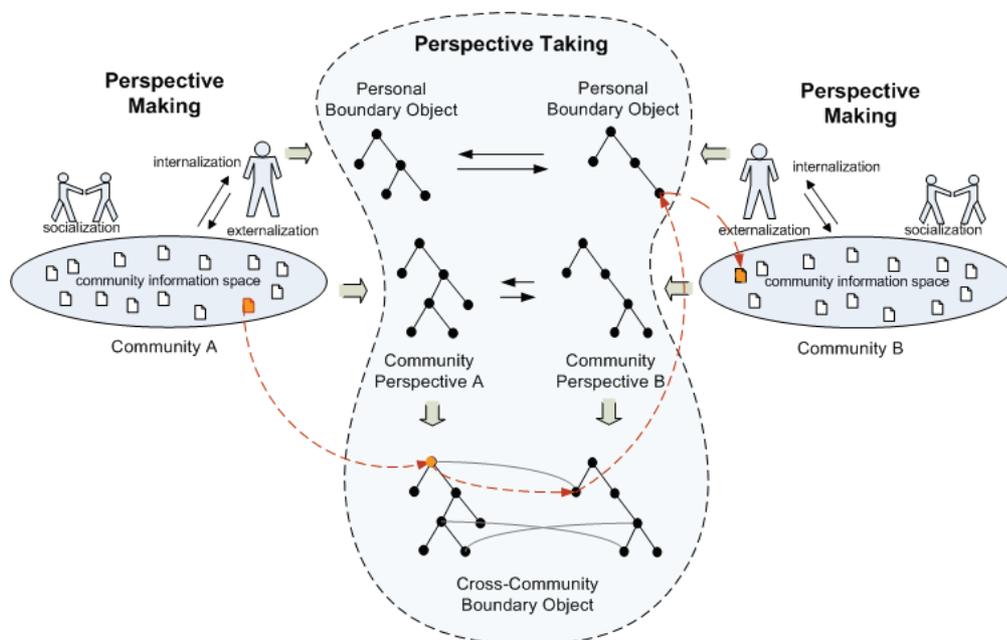


Abb. 1: Modell des gemeinschaftsübergreifenden Wissensaustausches

Die Hauptanforderung zur Unterstützung des gemeinschaftsübergreifenden Wissensaustausches ist also die Visualisierung von impliziten Wissensstrukturen individueller Communities. Diese Wissensvisualisierung muss es Mitgliedern unterschiedlicher Communities ermöglichen,

- Einsichten in implizite Wissensstrukturen fremder Communities zu gewinnen (PT),
- die Beziehungen zwischen dem Wissen anderer Communities und ihrem eigenen Wissen zu entdecken bzw. unbekannte Informationen zu kontextualisieren (PT→PM),
- neu erworbenes Wissen in ihren eigenen Begriffen bzw. durch Aneignung fremder Begriffe in ihren gemeinschaftlichen Perspektiven auszudrücken (PM).

3.2 Boundary Objects

Perspective Taking kann insbesondere durch sog. Boundary Objects („Grenzobjekte“) unterstützt werden: diese verbinden unterschiedliche Perspektiven von heterogenen Akteuren auf ein Problem oder eine Wissensdomäne und erlauben ihnen, Informationen zu interpretieren, ohne die eigene, spezielle Sichtweise aufgeben zu müssen [BoSt99]. Relevante Beispiele für Perspective Taking sind Klassifizierungssysteme und kognitive Karten [BoTe95]. Nach [BoTe95] sind auch Visualisierungen von persönlichen Wissensstrukturen der Community-Mitglieder wichtige Boundary Objects, weil oft schon innerhalb von Gemeinschaften sehr verschiedene Sichtweisen über die Bedeutung von Begriffen und Informationen bestehen.

3.3 Wissenserzeugung beim Informationszugriff

Das Wechselspiel von Perspective Taking und Perspective Making kann in den Prozess der Informationssuche in einer unbekannt Community mittels der sog. „Sensemaking-Schleife“ eingebettet werden [RSPC03]. Sie beschreibt wie Wissenskonstruktion in einem Wechselspiel zwischen der Suche nach relevanten Informationen und der Suche nach Strukturen, die Informationen kontextualisieren, stattfindet. Um eine komplexe Aufgabe zu erschließen, erstellt der Benutzer Informationsstrukturen (Schemata), welche die wesentlichen Merkmale des Informationsbedarfs der Aufgabe umfassen. Die Schemata werden benutzt, um die Identifikation relevanter Begriffe, die den Informationsbedarf charakterisieren, zu leiten und die wieder gewonnenen Informationen zu kategorisieren. Schließlich werden die mit Informationen angereicherten Schemata zur Erfüllung der Aufgabe verwendet. Die Bereitstellung persönlicher Wissensstrukturen anderer Nutzer bietet dafür eine wichtige Unterstützung [QuFu05].

3.4 Bestehende Unterstützung

Perspective Making wird durch eine große Palette von Community-Systemen und Werkzeugen unterstützt: von Diskussionsforen und gemeinsamen Arbeitsbereichen über kollaboratives Filtern und Recommender-Systeme bis hin zum kooperativen Kommentieren und Annotationen [Koch03]. Die Unterstützung von Perspective Taking ist dagegen bisher sehr beschränkt. Bestehende Ansätze setzen typischerweise standardisierte Taxonomien oder organisationsweite Ontologien ein, um Wissen über die Grenzen unterschiedlicher Organisationsstrukturen und informellen Communities hinaus zugänglich zu machen [Boni02]. Diese Ausrichtung auf eine unifizierte Struktur wird den Anforderungen von simultaner Koexistenz und Koordination zwischen unterschiedlichen lokalen Perspektiven nicht gerecht [Boni02, Swan01].

Auch die kollaborative Erstellung von Ontologien ist durch die Notwendigkeit der expliziten Verhandlung und des gemeinschaftlichen Konsens eingeschränkt. Ontologie Mapping Methoden ermöglichen es, Beziehungen zwischen unterschiedlichen Schemata zu entdecken, erfordern aber explizit modellierte Ontologien [DiFo02]. Sie vernachlässigen die implizite Natur und soziale Erzeugung von gemeinschaftlichem Wissen bzw. das Verstehen der gegebenen Beziehungen, das für Perspective Taking notwendig ist.

Wissensvisualisierungsmethoden wie Concept Mapping ermöglichen es, konzeptuelle Strukturen von Individuen und Gruppen zu erfassen, darzustellen und für den Wissensaustausch anzuwenden [TeKe05]. Im Wissensmanagement werden wiederum Wissensstrukturkarten benutzt, um einen Überblick über vorhandene Wissensressourcen darzustellen (z.B. Wissensträgerkarten, Kompetenzkarten) [Epl01]. Während die Concept Mapping Methoden dem Nutzer einen hohen Aufwand bei der manuellen Erzeugung abverlangen, bieten die Wissensressourcenkarten wenig Einsicht in semantische Strukturen von Nutzern und Gruppen.

4 Visualisierung persönlicher und gemeinschaftlicher Wissensstrukturen

Die Hauptidee des vorliegenden Ansatzes ist es, den Zugang zu Community-Informationsräumen mittels interaktiven Wissenskarten zu unterstützen, welche implizite, persönliche und geteilte Wissensstrukturen von Mitgliedern unterschiedlicher Communities darstellen. Dabei wird die Informationssuche als ein Prozess betrachtet, in dem die Nutzer durch Interaktion mit Informationen bestehendes Wissen ausdrücken und neues Wissen entwickeln.

4.1 Das Wissenskarten-Modell

Dazu wird ein Wissenskarten-Modell eingeführt, das Wissensstrukturen von Nutzern visualisiert und einen Kontext zur Interpretation der Bedeutung ihrer Informationsaktionen darstellt.

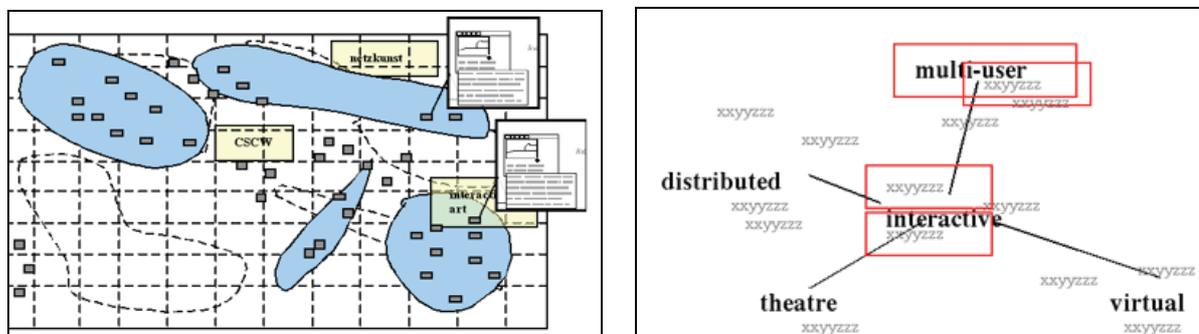


Abb. 2. Das Wissenskarten-Modell

Eine solche Wissenskarte besteht aus zwei eng miteinander verbundenen Elementen: einer Document Map und einer Concept Map. Die Document Map (Abb. 2, links) stellt eine persönliche oder Community-Informationssammlung dar, strukturiert in semantische Cluster, die thematisch verwandte Dokumente enthalten. Die Concept Map (Abb. 2, rechts) visualisiert Gruppen und Beziehungen von Begriffen, die in ähnlichen Kontexten verwendet werden. So werden Sprachmuster eines Nutzers oder einer Community sichtbar.

4.2 Erhebung und Visualisierung von persönlichen Sichtweisen

Um solche Karten basierend auf den persönlichen Sichtweisen der Benutzer zu erzeugen, wird die statistische Textanalyse mit einem selbst-organisierendem, neuronalen Netzwerk und mit Methoden des überwachten Lernens von Nutzerpräferenzen kombiniert [NWFS03]. Dem Benutzer wird zuerst eine systemgenerierte Struktur angeboten, welche er zur Informationssuche benutzen kann. Die Ergebnisse seiner Suche kann der Nutzer in einer neuen, persönlichen Document Map anlegen (Dokumente auswählen, Cluster erstellen und benennen). So entsteht eine Vorlage, die seine persönliche Interpretation der Bedeutung von Dokumenten widerspiegelt. Diese wird vom System gelernt und kann, als semantischer Filter, neue Dokumente automatisch in nutzerdefinierte Themen klassifizieren. Dadurch sind persönliche Karten unmittelbar nützlich für den Kartenautor selbst, was die notwendige Motivation für ihre Erstellung sicherstellt. Die system-generierten Karten lösen dabei das Kaltstart-Problem, da sie benutzt werden können, bevor genügend persönliche Karten verfügbar sind.

Anhand der Analyse der benutzten Begriffe (Konzepte) in seinen persönlichen Document Maps, ermittelt das System für jeden Benutzer auch eine persönliche Concept Map, die ein Netzwerk der aus Benutzersicht relevantesten Konzepte und ihren Verbindungen darstellt. Die Konzeptrelevanz wird anhand der Anzahl der zugeordneten Dokumente bestimmt. Durch die Benennung von Dokumenten-Clustern erstellen die Nutzer auch implizite Beziehungen zwischen eigenen Konzepten und anderen Begriffen, die in den Dokumenten vorkommen. So werden Beziehungen zu eigenen und fremden Community-Begriffen hergestellt.

4.3 Zusammenfassung persönlicher Ansichten in Community-Wissensstrukturen

Wird diese Methode auf Dokumentenkarten aller Benutzer einer Community angewandt, entsteht die gemeinsame Community Concept Map. Die Clusterbezeichnungen von Mitgliedern einer Community beschreiben die wichtigsten Konzepte, die in dieser Community auftreten. Die Verbindungen von Konzepten zu Dokumenten veranschaulichen dabei die Bedeutung der

benutzten Begriffe, indem sie konkrete Beispiele ihrer Verwendung aufzeigen. Beziehungen zwischen unterschiedlichen Konzepten, die verschiedene Nutzer für ähnliche Dokumentencluster (Themen) verwendet haben, werden aus der Ähnlichkeit der Dokumentencluster abgeleitet. Diese wird text-basiert anhand einer Stichwort-Analyse und kontext-basiert anhand Dokumenten-Kollokation gemessen [NWFS03]. So werden unterschiedliche persönliche Ansichten aus einer Community in einer gemeinsamen Struktur verbunden.

5 Multiperspektivischer Zugang zu Community-Informationsräumen

Auf Basis der vorgestellten Methode wurde ein Wissensvisualisierungsmodell entwickelt, das es erlaubt, verschiedene persönliche und gemeinschaftliche Wissensstrukturen gleichzeitig darzustellen, die Beziehungen zwischen ihnen zu entdecken und Informationen thematisch zu kontextualisieren. Dieses wurde im Werkzeug Knowledge Explorer II implementiert, das die Exploration von Community-Informationsräumen aus verschiedenen Sichtweisen ermöglicht.

5.1 Der Knowledge Explorer (II)

Das Werkzeug Knowledge Explorer II ist in Abb. 3 dargestellt (siehe [NoWu05] für Version I). Die Document Maps werden als ein zweidimensionales Gitter dargestellt, auf dem die Dokumente nach Ähnlichkeitsbeziehungen und Clusterzugehörigkeit verteilt werden.

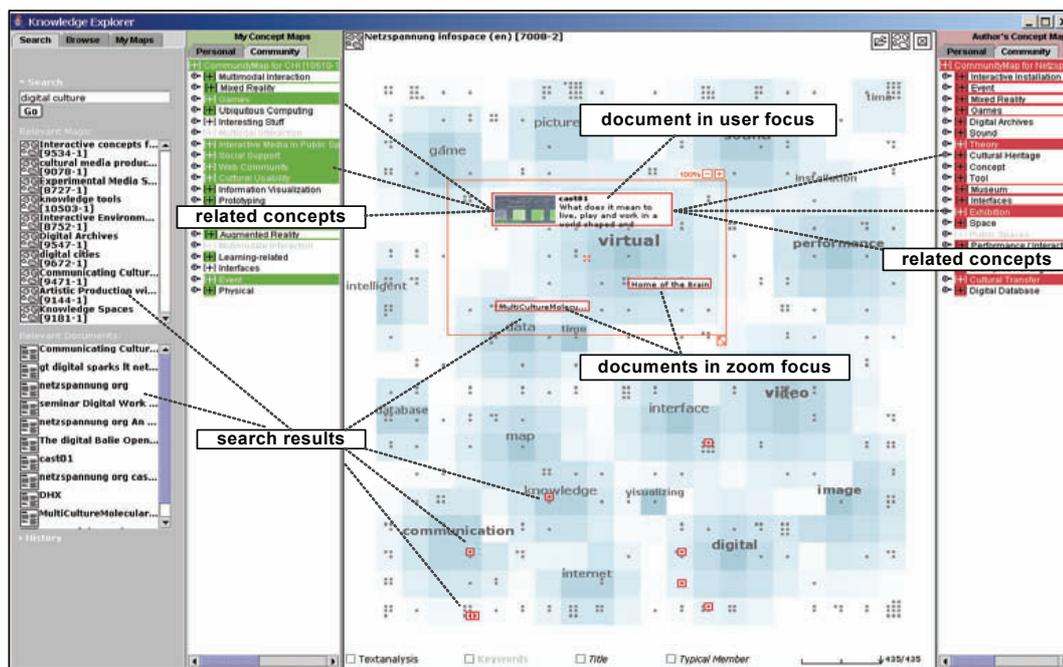


Abb. 3. Das Knowledge Explorer II Werkzeug

Die vom Clusterzentrum abfallende Farbtintensität, stellt den Zugehörigkeitsgrad einer Gitterzelle zu einem Cluster dar. Die Concept Maps werden in einem Verzeichnisbaummodell abgebildet, das die visuelle Komplexität reduziert und intuitive Navigation ermöglicht.

Ein typischer Nutzungsablauf beginnt mit einer Suchanfrage, die eine Liste relevanter Dokumente und Karten liefert (erzeugt von einem Empfehlungsdienst [NWFS03]). Die relevanteste Document Map wird im zentralen Panel angezeigt (Abb. 3). Zusätzlich werden die dazugehörige Concept Map (rechts) und die persönliche Concept Map des Nutzers (links) geöffnet. Die Suchergebnisse werden durch die farbige Kennzeichnung der Dokumente auf der Document Map und der verwandten Begriffe in den beiden Concept Maps in Zusammenhang gebracht. So kann der Nutzer die Suchergebnisse auf ihre thematischen Zusammenhänge in der Document Map untersuchen, über hervorgehobene Begriffe in den Concept Maps navigieren, oder eine neue persönliche Karte erstellen, um relevante Dokumente und Themen zu speichern. Mittels Tab-Register kann er leicht zwischen den gewünschten Kombinationen von Perspektiven umschalten. So kann er die Sichtweisen einer fremden Community erforschen oder durch dessen Informationsraum aus der eigenen persönlichen oder vertrauten Community-Sicht navigieren.

5.2 Semantische Exploration und kontextualisierter Zugang zu unbekanntem Domänen

Die Document Map und die Concept Map einer unbekanntem Community bieten einen schnellen Überblick über die Community-Wissensstrukturen. Dies umfasst die wichtigsten Themen, Gruppen von thematisch verwandten Dokumenten sowie Begriffe und ihre Beziehungen. Mittels Begriffsnavigation in der Concept Map können Benutzer relevante Dokumente identifizieren. Durch das Auswählen von Dokumenten in der Document Map werden wiederum verwandte Begriffe angezeigt. Dies unterstützt Perspective Taking, da die Nutzer ein Verständnis für relevante Themen und die Terminologie der fremden Community entwickeln können. Die Visualisierung der zu einem Suchergebnis verwandten Begriffe erleichtert ihnen zu lernen, wie sie ihren Informationsbedarf in der unbekanntem Terminologie ausdrücken können.

5.3 Multiperspektivische Navigation und gemeinschaftsübergreifende Beziehungen

Gemeinschaftsübergreifende Beziehungen, die aus der Analyse persönlicher Karten von Mitgliedern unterschiedlicher Communities hervorgehen, ermöglichen es, das Wissen einer Community im Kontext einer anderen zu lokalisieren. Solche Beziehungen können in Community Maps visualisiert und erforscht werden oder die persönlichen Karten, denen sie

entstammen, können direkt gegenüber gestellt werden. So kann die persönliche Begriffskarte eines Nutzers oder seiner Community genutzt werden, um durch die Dokumentenkarte einer unbekannt Community anhand vertrauter Begriffe zu navigieren. Die Selektion unbekannter Dokumente zeigt wiederum verwandte Begriffe in vertrauten Wissensstrukturen auf. Die Selektion eines Begriffs visualisiert sowohl die verwandten Dokumente als auch die zugehörigen Begriffe in der unbekannt Begriffskarte (Abb. 3). Dadurch wird sichtbar, wie Begriffe unterschiedlicher Communities miteinander in Beziehung gebracht werden können.

5.4 Kontextualisierung von unbekanntem Wissen in persönlichen Perspektiven

Anstatt Community Maps anzuzeigen, kann der Benutzer auch eine persönliche Karte eines anderen Benutzers öffnen. Dadurch wird nur eine Teilmenge von Dokumenten angezeigt, die das persönliche Wissen des Kartenauteurs widerspiegelt und für einen spezifischen Informationsbedarf relevant ist. Persönliche Karten können auch benutzt werden, um einen Communityraum in thematische Cluster, aus der Sichtweise des Kartenauteurs zu strukturieren. Die Klassifikation der unbekannt Dokumente beruht auf ihrer Ähnlichkeit zu Dokumenten, die bereits in der persönlichen Karte des Nutzers enthalten sind. Solche Kategorisierung bildet die nutzerspezifischen Kriterien der Dokumentenzugehörigkeit zu einem Thema implizit ab. Die Zuordnung von unbekannt Dokumenten zu persönlichen Wissensstrukturen erleichtert es dem Nutzer zu verstehen, ob und wie sich unbekanntes Wissen auf sein vorhandenes Wissen bezieht (PT → PM). Die persönlichen Karten unterstützen auch den Sensemaking-Prozess: sie bieten Vorschläge zu richtigen Suchbegriffen, relevanten Dokumenten für einen bestimmten Informationsbedarf und erlauben die Kontextualisierung unbekannt Informationen.

5.5 Externalisierung von persönlichem Wissen und Beziehungen zu anderen Nutzern

Mit persönlichen Karten drücken die Nutzer das bei der Informationssuche entwickelte Wissen aus. Durch das Hinzufügen von Dokumenten aus fremden in eigene Karten erstellt ein Nutzer implizite Beziehungen zwischen seinen Begriffen und den Begriffen anderer Nutzer. Für Karten aus der gleichen Community entstehen dadurch Beziehungen zwischen persönlichen Sichtweisen innerhalb der Community. Durch das Hinzufügen von Dokumenten aus Karten anderer Communities drücken die Nutzer die entdeckten Beziehungen zwischen den Begriffen und dem Wissen unterschiedlicher Communities aus (PT → PM). Die Motivation, eigene Karten anderen zur Verfügung zu stellen, basiert sowohl auf dem Reputationsmechanismus (Karten beinhalten den Namen ihres Autors) als auch auf einem unmittelbaren Gewinn für den

Kartenautor: durch die entstandenen Beziehungen zwischen seinen Konzepten und Themen und Dokumenten von anderen Nutzern erhält er Zugang zum ihren Wissen.

6 Empirische Evaluierung in einer vergleichenden Laborstudie

Die Anwendung der entwickelten Wissensvisualisierungsmethode zur Unterstützung des gemeinschaftsübergreifenden Wissensaustausches wurde in einer Laborstudie evaluiert. Dazu wurde der Knowledge Explorer II mit einem herkömmlichen Werkzeug für die Informationssuche verglichen, welches keine Wissensvisualisierung einsetzt.

6.1 Experiment Design

Basierend auf der experimentellen Simulationsmethode wurden Testnutzer mit Szenarien und Aufgaben konfrontiert, die einer realen Situation ähneln. Es wurden zwei Communities sowie dessen Informationsräume ausgewählt, welche grundlegende Unterschiede aufweisen, jedoch Schnittmengen in ihren Interessengebieten haben (Tabelle 1). Zuerst lösten Mitglieder der ersten Community (C1) einen Aufgabensatz zu vertrauten Themen in der eigenen Community, woraus ihre persönlichen und gemeinschaftlichen Wissensstrukturen extrahiert wurden. Im zweiten Schritt wurden diese Wissenskarten visualisiert und von Mitgliedern einer anderen Community (C2) zum Zugriff auf den Informationsraum der ersten Community verwendet.

Die Teilnehmer des zweiten Versuchs wurden in drei Gruppen aufgeteilt. Die erste Gruppe benutzte den Knowledge Explorer mit der gemeinsamen Community Map der Community C1 (Gruppe A). Die zweite Gruppe benutzte den Knowledge Explorer mit persönlichen Karten der Community C1 (Gruppe B). Die dritte Gruppe benutzte ein herkömmliches System für die Informationssuche ohne Wissensvisualisierung (Google Search + Mozilla Bookmarks, Gruppe C). Die Haupthypothese bestand darin, dass das Knowledge Explorer Werkzeug eine bessere Unterstützung für die Identifikation und das Verstehen von relevantem Wissen in Informationsräumen unbekannter Communities bietet, als das Referenzsystem. Zur Bewertung wurde die Qualität des Wissenszugriffs zwischen den Versuchsgruppen verglichen.

Community	Wissensdomäne	Teilnehmer	Informationsraum
C1: Medienkunst (netzspannung.org)	Medienkunst, Design & Technologie	Künstler, Kuratoren, Interface-Designer (7)	Teilbereich des Community-Portals (435 Dokumente)
C2: HCI (konstruiert)	Entwicklung interaktiver Systeme	Forscher, Doktoranden, Studenten (3x6, 1 drop-out)	Teile des ACM CHI Konf.-Archiv (604 Dokumente)

Tabelle 1. Zusammensetzung der Test-Communities, Teilnehmer und Informationssammlungen

6.2 Die Messung der Qualität des Wissenszugriffs

Nach dem vorgestellten Modell des gemeinschaftsübergreifenden Wissensaustausches kann die Qualität des Wissenszugriffs anhand von drei Hauptkriterien bewertet werden: 1) dem Ausmaß der gewonnenen Einsicht in die Wissensstrukturen der unbekannt Community, 2) der Qualität der entdeckten Beziehungen zum eigenen Wissen und 3) der Qualität der Erfüllung typischer Sensemaking-Aufgaben während der Informationssuche.

Um diese objektiv bewerten zu können, wurden die Aufgabenlösungen der Nutzer nach folgenden Indikatoren gemessen: dem Ergebnis der Informationswiedergewinnung, der Qualität der thematischen Strukturierung und dem Lerneffekt.

Das Ergebnis der Informationswiedergewinnung wird durch die Standardmaße Precision (Genauigkeit) und Recall (Vollständigkeit) gemessen. Die Genauigkeit beschreibt das Verhältnis zwischen der Anzahl der relevanten Dokumente und der gesamten Anzahl der Dokumente, welche der Nutzer gefunden hat. Die Vollständigkeit misst das Verhältnis zwischen der Anzahl der vom Nutzer gefundenen relevanten Dokumente und der gesamten Anzahl relevanter Dokumente in der Testsammlung (Referenzlösung).

Die Qualität der thematischen Strukturierung wird anhand der Anzahl und der Relevanz der Kategorien bewertet, welche der Nutzer zur Dokumenten-Strukturierung verwendet hat. Thematische Genauigkeit misst das Verhältnis zwischen der Anzahl relevanter Begriffe und der gesamten Anzahl der in der Aufgabenlösung benutzten Begriffe. Thematische Vollständigkeit misst die relevanten Begriffe im Vergleich zu allen Themen der Referenzlösung. Das ist ein relatives Maß, da eine gute Lösung nicht alle Begriffe der Referenzlösung beinhalten muss.

Ferner wird zwischen gemeinschaftsübergreifender thematischer Genauigkeit und gemeinschaftsinterner thematischer Genauigkeit unterschieden. Die erste misst die Nutzung der Begriffe aus der unbekannt Community, während die zweite die Nutzung der Begriffe aus der eigenen Community wiedergibt. Die Referenzlösung bestimmt deren relative Bedeutung. In unserer Aufgabenstellung war die gemeinschaftsübergreifende, thematische Precision wichtiger. Sie reflektiert den Umfang, in dem die Nutzer die Begriffe aus der unbekannt Community verstanden haben.

Der Lerneffekt beschreibt in welchem Maße die Nutzer dieses Wissen verinnerlichen konnten. Er wird durch einen Fragebogen ermittelt, der die Benutzer auffordert Begriffe und Themen zu nennen, die mit bestimmten Aspekten ihrer Aufgabe in Beziehung stehen. Die Antworten werden auf thematische Genauigkeit und Vollständigkeit überprüft.

Die Ergebnisse zeigen, ob die Nutzer die Begriffe aus der unbekannt Community, welche sie in ihrer thematischen Strukturierung verwendeten, auch wirklich verstanden haben.

6.3 Aufgabenstellung und Ablauf des Versuchs

In der Einführungsphase lösten die Nutzer Aufgaben zu vertrauten Themen im Informationsraum der eigenen Community. So wurden persönliche und gemeinschaftliche Wissensstrukturen erhoben. Die Testphase des Versuchs II beinhaltete Aufgaben mit „schlecht-definiertem“ Fokus und unklaren Informationsbedürfnissen, welche die Identifizierung von Dokumenten und Themen aus der unbekannt Community erforderten. Hier berichten wir über die Ergebnisse der wichtigsten Aufgabe, für welche die Teilnehmer die meiste Zeit (40min) zur Verfügung hatten: „Finden Sie relevante Dokumente zum Thema ‚Interactive Systems for Digital Culture‘ und ordnen Sie diese in Unterthemen an“ (Ergebnisse der restlichen Aufgaben stimmen überein). Die Referenzlösung wurde von Experten der Community C1 im Versuch I erstellt und um die Sichtweise der Experten aus Community C2 erweitert. Vor den Aufgaben wurde das Testsystem vorgeführt und die Teilnehmer hatten die Gelegenheit es auszuprobieren.

6.4 Testergebnisse hinsichtlich der Qualität des Wissenszugriffs

Angesichts der kleinen Stichproben wurden zur Interpretation der Ergebnisse keine statistischen Signifikanztests, sondern deskriptive Statistik und explorative Datenanalyse angewandt. Die Ergebnisse zeigen, dass Nutzer, die das entwickelte Wissensvisualisierungssystem (Knowledge Explorer) benutzt haben, eine konsistent höhere Qualität des Wissenszugriffs nach allen drei Indikatoren erreicht haben, als Nutzer des herkömmlichen Referenzsystems (Google + Mozilla). Die Unterstützung durch persönliche Karten erwies sich auch geeigneter, als die der Community Maps. Schon bei der Informationsgewinnung erzielten die Nutzer des Knowledge Explorers eine höhere Informationsgenauigkeit bei vergleichbarer Vollständigkeit (Abb. 4, links). Außerdem erzielte die Gruppe B, welche die persönlichen Karten verwendete, eine höhere Genauigkeit, als Gruppe A, welche die Community Map benutzte.

Noch größere Unterschiede sind in der Qualität der thematischen Strukturierung und beim Lerneffekt zu verzeichnen. Beide Knowledge Explorer Gruppen erzielten eine bessere Qualität bei der thematischen Strukturierung im Vergleich zu den Nutzern des Referenzsystems. Dies zeigt sowohl die höhere Anzahl der Themen als auch die höhere thematische Genauigkeit und Vollständigkeit (Abb. 4). Die Anzahl und Qualität der Themen, welche von den Nutzern des Google-Referenzsystems zur Strukturierung von relevanten Dokumenten verwendet wurden, ist

eher gering (Median=2 Themen vgl. zu 4 Themen für beide Knowledge Explorer Gruppen). Die Google-Gruppe erzielte auch die geringste gemeinschaftsübergreifende thematische Genauigkeit. Alle Nutzer in dieser Gruppe beschwerten sich über Schwierigkeiten, die unbekannte Community-Terminologie zu verstehen und relevante Themen zu identifizieren.

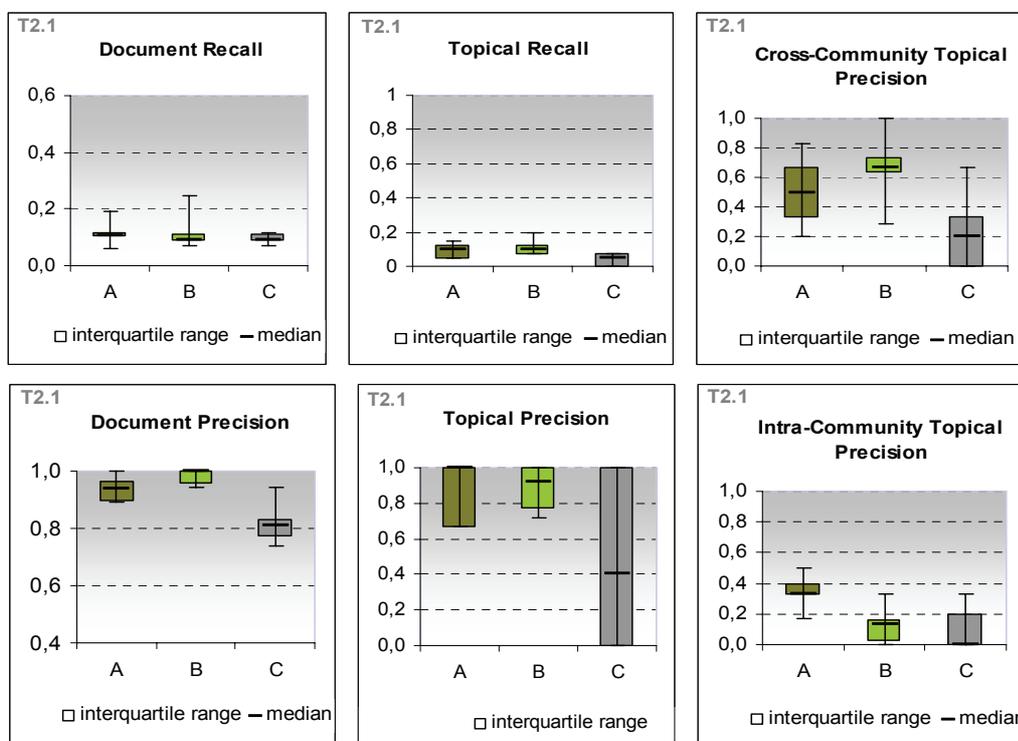


Abb. 4. Ergebnisse der Informationswiedergewinnung und der Qualität der thematischen Strukturierung

Ein wichtiger Unterschied zwischen den zwei Knowledge Explorer Gruppen zeichnet sich durch die stärkere Nutzung der Konzepte aus der unbekannt Community (höhere gemeinschaftsübergreifende Genauigkeit) seitens der Nutzer von persönlichen Karten aus (Gruppe B, Abb. 7. Mitte). Die Nutzer der Community Maps (Gruppe A) nutzten zum größten Teil vertraute HCI Konzepte, um die gefundenen Dokumente zu organisieren (Abb. 7, Mitte unten). Dieses deutet darauf hin, dass spezifische Ausschnitte, die für einen bestimmten Informationsbedarf relevant sind, die Wissenskonstruktion in unbekannt Community-Domänen besser unterstützen, als Ansichten der vollständigen Community-Wissensstruktur. Die Ergebnisse des Lerneffekt-Fragebogens lassen ähnliche Schlussfolgerungen zu (Abb. 5). Die thematische Vollständigkeit und Genauigkeit der Nutzerantworten sind bei den Fragen, die das erworbene Verständnis der Begriffe aus unbekannter Community (Q1) bzw. das Ausmaß der gemeinschaftsübergreifenden Wissensintegration (Q3) bewerten, für beide Knowledge Explorer Gruppen deutlich höher, als für die Google Gruppe. Bei Fragen bezüglich der

Wiedererkennung und Verwendung von Begriffen aus der eigenen Domäne war ihre Leistung mit den Knowledge Explorer Gruppen vergleichbar (Q2, Q4). Das Nutzerfeedback belegt: Google-Nutzer konnten sich die Begriffe der unbekannt Community nicht aneignen und konzentrierten sich darauf, bereits vertraute Begriffe zu identifizieren. Daher waren die Knowledge Explorer Gruppen auch im Lerneffekt effektiver als Google-Nutzer. Auch hier waren die Nutzer von persönlichen Karten besser als die der Community Maps (Q1, Q3).

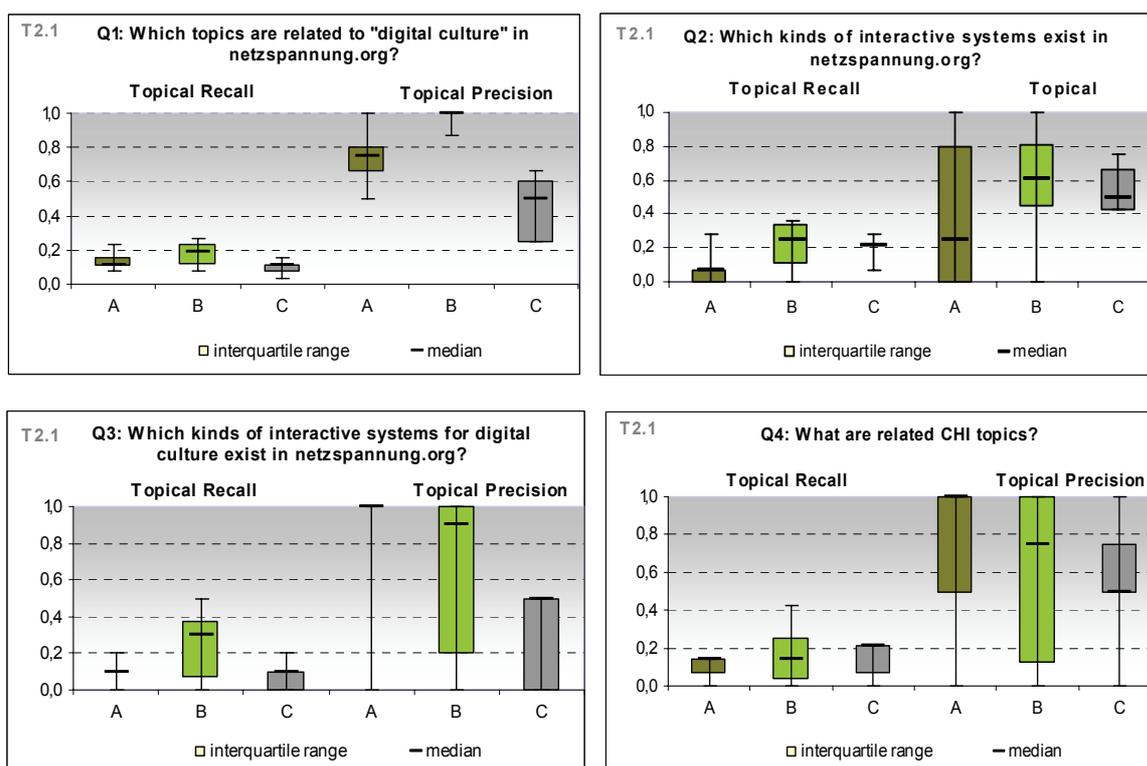


Abb. 5. Ergebnisse des Lerneffekt-Fragebogens

7 Fazit

Die vorgestellte Methode ermöglicht es, dynamische Wissenskarten zu erstellen, die persönliche und geteilte Perspektiven aus unterschiedlichen Communities abbilden, sie in Beziehung bringen und für die Informationssuche nutzbar machen. Die Evaluierungsergebnisse zeigen, dass ein multiperspektivischer Zugang zu Informationsräumen unterschiedlicher Communities, eine wertvolle Unterstützung für den gemeinschaftsübergreifenden Wissensaustausch leisten kann. Der limitierende Aspekt der Evaluierung ist ihr explorativer Charakter (kleine Stichproben), der keine Rückschlüsse auf die statistische Signifikanz erlaubt. Zur weiteren Untersuchung mit einer höheren externen Validität würden auch longitudinale Studien beitragen.

8 Literaturverzeichnis

- [Alle00] Allee, V.: Knowledge Networks and Communities of Practice. In OD Practitioner, Journal of the Organization Development Network, 32 (4), 2000
- [Boni02] Bonifacio, M. et al.: Enabling distributed knowledge management. Managerial and technological implications. Novatica and Informatik III(1), 2002
- [BoSt99] Bowker, G.C., Star, S. L.: Sorting Things Out: Classification and its Consequences. MIT, 1999
- [BoTe95] Boland J.R., Tenkasi R.V.: "Perspective Making and Perspective Taking in Communities of Knowing", Organization Science, 6,4 (July-August) 1995
- [BrDu91] Brown, J.S., Duguid, P.: Organizational Learning and Communities of Practice, Organization Science, Vol.2, No. 1, pp. 40-57, Feb., 1991
- [Caro99] Carotenuto, L. et al.: Communityspace: Towards flexible support for voluntary knowledge communities. Proc. of Workshop on Workspace Models for Collaboration, London, April, 1999
- [DiFo02] Ding, Y., Foo, S.: Ontology research and development. Part 2 – a review of ontology mapping and evolving, Journal of Information Science, 28(5), 2002
- [Doug92] Dougherty, D.: Interpretative barriers to successful product innovation in large firms. Organization Science, 3(2), 1992
- [Epp101] Eppler, M.J.: Making Knowledge Visible Through Intranet Knowledge Maps: Concepts, Elements, Cases, Proc. of HICSS-34, Maui, Hawaii, January 2001
- [GrLe02] Gruninger, M., Lee, J.: Ontology applications and design. Communications of the ACM, 45(2), p. 39 – 41, 2002
- [Koch03] Koch, M: Community-Unterstützungssysteme - Architektur und Interoperabilität, Habilitationsschrift, Technische Universität München, 2003
- [LaWe91] Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation. NY: Cambridge Univ. Press, 1991.

- [MeDa01] Merali, Y., Davies, J.: Knowledge Capture and Utilization in Virtual Communities, In Proc. of K-CAP'01, Victoria, Canada, October 2001
- [NoTa95] Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company, Oxford University Press, 1995.
- [NoWu04] Novak, J., Wurst, M.: Collaborative Knowledge Visualisation for Cross-Community Learning, in Tergan, S. & Keller, T. (eds), Knowledge Visualisation and Information Visualisation – Searching for Synergies, Springer, 2005
- [NoWu05] Novak, J., Wurst, M.: Knowledge Explorer: An Interactive Tool for Supporting Cross-Community Discovery and Sharing of Knowledge, Proc. of HCI International 2005, Lawrence Erlbaum Publishers, 2005
- [NWFS03] Novak, J., Wurst, M., Fleischmann, M., Strauss, W.: Discovering, visualizing and sharing knowledge through personalized learning knowledge maps, Proc. AAAI-AMKM 03, Springer, 2003
- [QuFu05] Qu, Y., Furnas, G.: Sources of Structure in Sensemaking. Proc. of ACM CHI 2005, Extended Abstracts and Applications, Portland, Oregon, USA, April 2005
- [RSPC93] Russell, D. M. et al.: The Cost Structure of Sensemaking. Proc. ACM INTERCHI'93, 1993
- [Swan01] Swan, J., Knowledge Management in Action: Integrating Knowledge Across Communities, Proc. Of HICSS-34, Maui, Hawaii, January 2001
- [TeKe05] Tergan, S.O., Keller, T. (Eds.): Knowledge and Information Visualisation – Searching for Synergies, Springer LNCS 3426, Springer, 2005
- [Weng98] Wenger, E.: Communities of Practice: Learning, Meaning and Identity. Cambridge Univ. Press, 1998

Einführung in den Track

Agenten- und Multiagenten-Technologien für betriebliche Anwendungen

Prof. Dr. Jörg P. Müller

TU Clausthal

Prof. Dr. Bernhard Bauer

Universität Augsburg

Softwareagenten und Multiagentensysteme beschreiben ein Entwurfsparadigma für dezentrale Systeme, das sich konsequent an lokalen Aufgabenstellungen, lokalem Problemverständnis, lokalen Situationen und Strukturen sowie lokaler Dynamik orientiert. Erfolgreiche Einsatzgebiete dieser Technologien liegen unter anderem in den Gebieten Optimierung, Verteiltes Informations- und Ressourcenmanagement und Personalisierung. Daraus ergeben sich einerseits neue Möglichkeiten für die Entwicklung betrieblicher Anwendungssysteme, andererseits entstehen daraus aber auch neue Herausforderungen an die Gestaltung soziotechnischer Informations- und Kommunikationssysteme.

Im Mittelpunkt des Tracks stehen demzufolge die Beschreibung und Analyse betrieblicher Anwendungen der Agententechnologie, das Engineering agentenbasierter Applikationen vom (Early) Requirements Engineering bis hin zu Test und Qualitätssicherung, Integrations- und Konfigurationsmanagement sowie der Wartungsproblematik. Technologische Beiträge im Anwendungskontext sind dabei ebenso willkommen wie aktuelle Projekt- und Erfahrungsberichte aus der betrieblichen Praxis.

Programmkomitee:

Prof. Dr. Stefan Kirn, Universität Hohenheim

Dr. Klaus Fischer, DFKI

Prof. Dr. Winfried Lamersdorf , Universität Hamburg

Dr. Michael Berger, Siemens

Prof. Dr. Torsten Eymann, Universität Bayreuth

Prof. Dr. Otthein Herzog, Universität Bremen

Prof. Dr. Rainer Unland, Universität Essen

Prof. Dr. Martin Bichler, Technische Universität München

Christian Danegger, Whitestein, CH

James Odell, James Odell Associates, USA

Michal Pechoucek, University Prag, CZ

Alexander Smirnov, Russian Academy of Sciences, RU

Multi-agent Simulation for the transshipment problem with a non-negligible transfer lead times and a limited transportation mean capacity

Mohamed Himden, Lamjed Ben Said, Khaled Ghédira

Unité de Recherche en Stratégie d'Optimisation des Informations et de la connaissance (SOIE)
, ISG de Tunis 41 avenue de la liberté cité Bouchoucha 2000 Tunis, Tunisia

mohamed.hmiden@essec.rnu.tn
{lamjed.bensaid, khaled.ghedira}@isg.rnu.tn

Abstract

We consider a supply chain consisting of n locations replenished at the beginning of each period by a supplier. These locations may coordinate in order to balance their inventory level through *transshipment*. Transshipment is the items transfer from location having an inventory excess to another in need. The transshipment problem consists to determine the initial inventory level where a transshipment policy is practiced. In this work, we consider the transshipment problem characterized by a non-negligible transshipment lead times and a limited transportation mean capacity. Our aim is to find a transshipment policy that reduces the inventory costs and improve the customer fill-rates. To realize this aim, we proposed a new formal transshipment model in which the period is divided into a set of sub-periods and the transshipment decision is made at the end of one of them. We also introduced a multi-agent model allowing to simulate the cooperated behavior of the inventory locations.

1 Introduction

The inventory management is a crucial activity in the supply chain. It allows the regulation of the inventory level to face the unexpected fluctuation of customer demands. In fact, a reduced inventory level causes the non satisfaction of customer requirements witch presents a shortage

cost. An inventory excess generates a holding cost. The challenge facing the inventory managers is to minimize the total inventory cost and to improve the customer fill-rates. This task becomes more complex when it concerns several locations. The transshipment is an inventory collaboration method consisting to transfer items between locations. It has been widely used in practice to reduce inventory costs and to improve the customer fill-rates. It provides an effective mechanism for correcting discrepancies between the locations observed customer demands and their available inventories. The transshipment problem is defined as the determination of two target parameters. The first is the replenishment quantity for each location and the second is the transshipment policy [Kris65]. The transshipment problem is extensively studied, where several parameters are considered. We identify three parameter types. The first is the *replenishment parameters* such as the replenishment lead times [Taga89] and the replenishment fixed costs [Here99]. The second is the *transfer parameters* such as the transshipment mean capacity [Ozde03] and the transshipment lead times [Taga02]. The third type of parameter is the *environment parameters* such as the number of locations [Robi90], [Taga92], [Week05] and the number of periods [Taga99]. We can classify the studies on the transshipment problem into two categories: (1) in the first category, exact methods are adopted to resolve the problem. This kind of research is interested in inventory system restricted to two non-identical locations or multi-identical locations [Week05, Kris65, Taga89]. (2) In the second category, meta-heuristics or simulations are adopted to find an approximate solution or to choose between several strategies. This kind of research is interested in inventory system composed by multi non-identical locations [Robi90, Ozde03, Taga99].

In this work, we investigate the two following parameter configurations: (1) a non-negligible transshipment lead times and (2) a limited transportation mean capacity. Our aim is to propose a transshipment policy that reduces the total inventory cost and improves the customer fill-rates under these considerations. The contribution of this paper is twofold. First, it introduces a formal model that takes into account the two parameters mentioned above. In this model, we propose to divide the period into a set of sub-periods at the end of one of them the transshipment decision is made. Second, it defines a multi-agent model that simulates the cooperative behavior of the locations.

The rest of the paper is organized as follows. The second section presents a description of the transshipment problem. The third section introduces our formal transshipment model. The fourth section presents the proposed multi-agent model and describes the global dynamics of the

system based on this model. The fifth section describes the realized experimentation and comments the obtained results. Finally, we discuss the future works.

2 Problem description

In our work, we consider the transshipment problem characterized by a set of locations having non-identical cost structures¹. These locations are replenished by a supplier at the beginning of each period. We assume that the inventory review is periodic for each location and the replenishment quantity is fixed. We consider also that the transshipment lead times between locations are non-negligible. In addition, we consider a limited capacity of the transportation mean. In the next sections, we adopt the following notations [Taga99] :

- U_i : excess quantity at the location L_i
- Z_i : in need quantity at the location L_i
- D_i : customer demands at the location L_i
- C_i : replenishment unit cost at the location L_i
- CS_i : shortage unit cost at the location L_i
- CH_i : holding unit cost at the location L_i
- C_{ij} : transshipment unit cost from the locations L_i to the location L_j . It is supported by L_i
- Q_i : initial inventory quantity at location L_i
- X_{ij} : transshipped quantity from the location L_i to the location L_j

The objective function is the minimization of the total inventory cost noted $C(Q)$:

$$C(Q) = \sum_{i=1}^n \left[C_i Q_i + \sum_{j=1}^n C_{ij} * X_{ij} + CH_i * U_i + CS_i * Z_i \right].$$

n is the number of locations and U_i is the excess quantity at the location L_i . This quantity is calculated after the satisfaction of the customer demands and the achievement of the transshipment actions to the other locations ($\sum X_{ij}$). Z_i is the needed quantity at the location L_i . This quantity is calculated after the partial satisfaction of the customer demands and the realization of the transshipment actions from the other locations to L_i ($\sum X_{ji}$).

¹ Cost structures designs the holding, shortage and transshipment cost

$$U_i = \max \left(0, Q_i - (D_i + \sum_{\substack{j=1 \\ j \neq i}} X_{ij}) \right)$$

$$Z_i = \max \left(0, D_i - (Q_i + \sum_{\substack{j=1 \\ j \neq i}} X_{ji}) \right)$$

3 Transshipment policy

In this work, we are interested to find a transshipment policy that takes into account a non-negligible transshipment lead times and a limited transportation mean capacity. It should contribute to minimize the inventory costs and it should also insure good customer fill-rates. To define such policy, the two following questions should be discussed: (1) what's the transshipment decision moment ? (2) What's the transshipment quantity that should be transferred from a location L_i in excess to a location L_j in need ?

3.1 Decision transshipment moment

Let L_1, L_2, L_3 and L_4 four locations buying the same item. At the beginning of each period, these locations are replenished by the same supplier. We consider in this example that a period corresponds to seven days. We assume that the transshipment from a location L_i in excess to L_j in need takes a non-negligible lead time noted t_{ij} . We assume also that $t_{ij} = t_{ji}$. Figure 1 presents the different possible transshipment actions and their respective lead times.

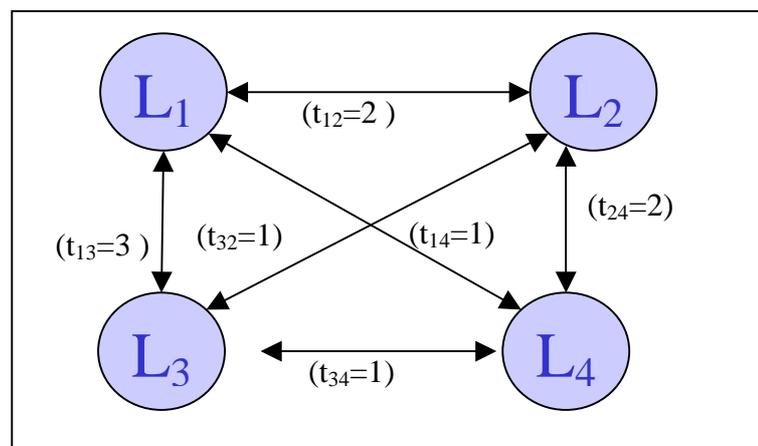


Figure 1. : transshipment actions and their Lead times.

Suppose that the location L_1 is in need at the end of the period. Its shortage could be satisfied by transshipment actions from the other locations. The transshipment action must be launched at a precise moment called the transshipment decision moment. This moment must consider the different transshipment lead times that is t_{12} , t_{13} and t_{14} . In fact, in order to satisfy the cumulated observed demand during the period, the transshipped quantity must arrive to the location L_1 before the end of the period. Thus, L_1 must evaluate its inventory at an appropriate moment within the period in order to take the transshipment decision. Let T_1 be this particular moment. The figure 2 illustrates the different transshipment possibilities and their lead times from locations L_2 , L_3 and L_4 to the location L_1 .

Three cases could be identified :

- If the supplier location is eventually L_4 then the transshipment decision must be taken by the location L_1 before the end of the period by a lead time equals to 1 day (t_{14}).
- If the supplier location is eventually L_2 then the transshipment decision must be taken by the location L_1 before the end of the period by a lead time equals to 2 days (t_{12}).
- If the supplier location is eventually L_3 then the transshipment decision must be taken by the location L_1 before the end of the period by a lead time equals to 3 days (t_{13}).

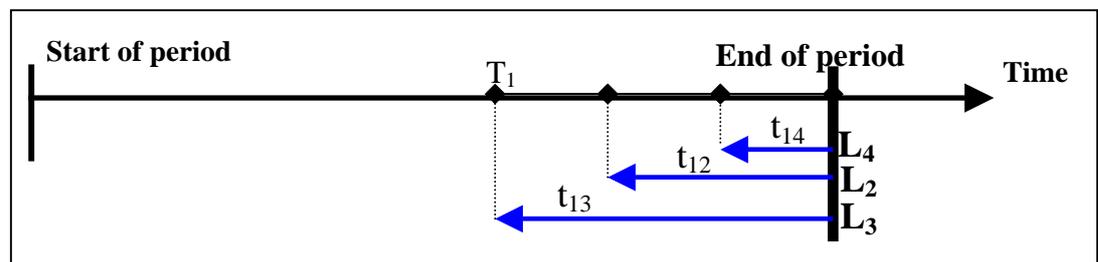


Figure 2 : Transshipment possible actions and their lead times.

In the worst case, the location L_1 is supplied by the location L_3 having the maximum lead time. In fact, the location L_1 must insure that its transshipment request could be processed by all the other locations and eventually served by at least one of them. Consequently, $T_1 = L_p - \max (t_{1j}), j \in \{2,3,4\}$ and L_p is the period length (7 days). In the general case for a location L_i : $T_i = L_p - \max (t_{ij}), j \in \{1,2,\dots,n\}, j \neq i$ and n is the number of locations.

In the majority of the research works concerning the transshipment problem, the inventory review is assumed to be periodic [Kris65, Taga99, Ozde03, Here01]. This means that the inventory situation is known only at the end of the period. However, as we proposed to evaluate the inventory situation at T_i within the period, it is necessary to divide the period into several

equal sub-periods. Thus, the transshipment decision is made at the end of a sub-period and the transfer lead time t_{ij} between locations L_i and L_j is expressed as a number of sub-periods.

3.2 Transshipment quantity

The transshipment quantity depends on the transshipment lead times which determines the transshipment decision moment. Consequently, this quantity depends on the inventory evaluation moment. Intuitively, it depends also on the transportation mean capacity of the sending locations. To determine the transshipment quantity to the location L_i in need, it is necessary to know the L_i inventory situation at the transshipment decision moment T_i .

Figure 3 illustrates the transshipment decision moment T_1 (identified for the example of the previous section) and both realized and provisional customer demands. We note $D_{obs}(T_1)$ as the observed customer demands during the four first sub-periods and $D_{prov}(T_1)$ the provisional customer demands for the rest of a period (the last three sub-periods). L_1 inventory level at T_1 noted $IL_1(T_1)$ depends on $D_{obs}(T_1)$, $D_{prov}(T_1)$ and the initial inventory level Q_1 . We obtain : $IL_1(T_1) = Q_1 - (D_{obs}(T_1) + D_{prov}(T_1))$.

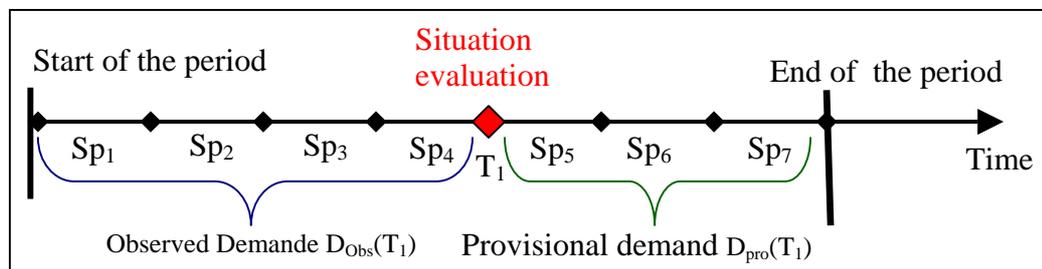


Figure 3: Inventory situation at T_1 moment.

In the general case, we obtain : $IL_i(T_i) = Q_i - (D_{obs}(T_i) + D_{prov}(T_i))$. If $IL_i(T_i)$ is positive then the location L_i is in excess else it is in need. Now, we define the quantity that must be transferred from L_i in excess to L_j in need noted X_{ij} at moment T_j . This quantity is equal to the minimum between : (1) the location sender L_i transportation mean capacity noted TC_i , (2) the excess quantity in the location L_i and (3) the needed quantity in the location L_j . We obtain : $X_{ij} = \min (TC_i, IL_i(T_j), |IL_j(T_j)|)$.

We note that this quantity depends on the provisional customer demands. In this research, we distinguish two kinds of inventory locations: *cooperative* and *egoist*. We suppose that a sub-

period customer demands follows a normal distribution $N(\mu, \sigma)$. A cooperative location favors the global benefit by forecasting a mean sub-period customer demands evaluated to μ . An egoist location favors its own benefit by forecasting a maximum sub-periods customer demands evaluated to $\mu + 3\sigma^2$.

4 Multi-agent model

Our model contains two types of agents: the *Interface agent* (IA) and the *Location agent* (LA) described in the next sections.

4.1 Interface agent

The Interface agent is defined by the following static knowledge:

- IAident : IA identifier.
- NbrLA : Number of Location agents.
- NbrPeriod : Number of periods.
- NbrSubPeriod : Number of sub-periods.
- PcentageEgoist : The percentage of the egoist LA.

The dynamic knowledge of the Interface agent is represented by a list called *LEvaluationParamters* containing total inventory cost and fill-rates.

The interface agent permits to :

- (1) Create the different Location agents.
- (2) Construct the initial Location agents coalitions.
- (3) Trigger the resolution process.
- (4) Recuperate the values of the evaluation parameters at the end of each period.
- (5) Detect the end of the simulation process and display the results.

4.2 Location agent

The Location agent represents an inventory location and it communicates with the other location agents and the Interface agent. It is defined by the following static knowledge :

² The probability that the customer demands is between $\mu - 3\sigma$ and $\mu + 3\sigma$ is equal to 0.997 [Will71]. This is valid if the customer demands is normally distributed.

- S_i : location agent i identifier
- $TypeS_i$: location agent S_i type: cooperative or egoist.
- CH_i : holding unit cost for the Location agent L_i .
- CS_i : shortage unit cost for the Location agent L_i .
- Q_i : replenishment quantity for the Location agent L_i .
- μ_i : sub-period's demand mean for Location agent L_i .
- σ_i : sub-period's demand standard deviation for the Location agent L_i .
- TC_i : the capacity of the transportation mean used by Location agent L_i .
- T_i : transshipment decision sub-period for the Location agent L_i .
- $LCoal_i$: a list containing the Location agents that L_i can communicate with them.
- $LCostLead_i$: a list that contains the transshipment costs and the lead times for the Location agent L_i .

The dynamic knowledge is:

- $Status_i$: the location L_i status (in need or in excess)
- $LReceivedDemand_i$: a list containing the transshipment requests received by L_i from the other Location agents.
- $LAcceptedDemand_i$: a list containing the transshipment requests that can be served by the Location agent L_i .
- $LOffers$: a list containing offers sent by locations as response to the L_i transshipment request.
- $LRestrainedOffer_i$: a list that contains the offers restrained by L_i .
- $LTransshipment_i$: a list containing the L_i realized transshipment.

The Location agent behavior depends on two criteria: the location inventory level and the current sub-period. During each sub-period, the Location agent executes some operations. The figure 4 illustrates the inventory level variation of the Location agent L_i during one period divided into seven sub-periods (Sp_1, Sp_2, \dots, Sp_7). It enumerates the different operations executed by this Location agent during the period.

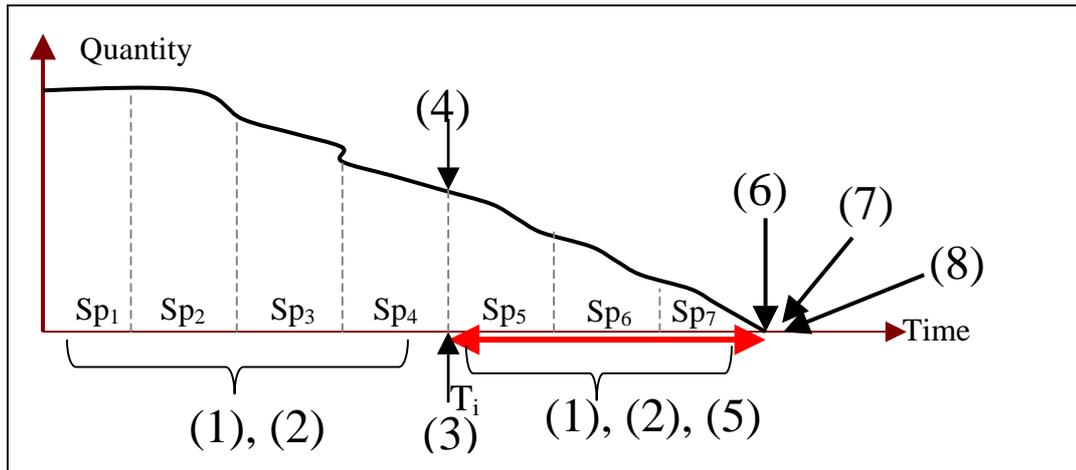


Figure 4. : Operations executed by a Location agent during one period.

The Location agent L_i executes two operations before the transshipment decision moment T_i :

- (1) *Demand observation* : this operation is executed at each sub-period.
- (2) *Transshipment request processing* : it is possible during these sub-periods that an other Location agent L_j in need had sent a transshipment request to the Location agent L_i .

At the T_i moment two other operations are executed by the Location agent L_i :

- (3) *Inventory status evaluation* : this operation consists to calculate the inventory level $IL_i(T_i)$. If this level is positive then the Location agent is in excess otherwise it is in need.
- (4) *Transshipment request launching* : in the case where the Location agent L_i is in need, it launches a transshipment request to the other Location agents.

During the rest of sub-periods (after T_i) the location L_i executes operations (1) and eventually (2). Besides these operations, it executes the operation (5) if the operation (4) was executed.

- (5) *Transshipment quantity updating* : this operation consists to execute one of the following actions :
 - a- To cancel the transshipment request : this operation is executed if the Location agent in the current sub-period is in excess and the surplus quantity can cover the shortage ones realized at the previous sub-period.
 - b- To modify the quantity to transship : this operation is executed in the case where the Location agent notes after the inventory level evaluation that the quantity to transship must be updated (increased or decreased).

Finally, at the end of the period, the Location agent L_i executes the three following operations:

- (6) *Customer demands satisfaction* : this operation consists to serve the customer demands observed during the period.
- (7) *Situation evaluation* : calculate the values of the total inventory cost and the customer fill- rates.
- (8) *Backlogging the unsatisfied customer demands* : this operation consists to backlog the unsatisfied customer demands (if ever exists) to the next period.

4.3 Global dynamics

The simulation process for identifying the best transshipment strategy is composed by the following three steps :

- Simulation initialization
- Negotiation and transshipment
- Simulation stop

4.3.1 Step 1 : Simulation initialization

This step consists to execute the following actions:

- *Creation of the Location agents* : this operation is realized by the interface agent. It consists to create and initialize the knowledge for each Location agent.
- *Creation of Location agents list accountancies* : each Location agent forms his list accountancy containing the other Location agents identifiers
- *Sorting of the Location agents list accountancies* : each Location agent sorts its list of accountancies according to their lead time.
- *Calculating the transshipment decision moment for each Location agent.*
- *Starting the simulation.*

4.3.2 Step 2 : Negotiations and transshipment

During this step, the different Location agents cooperate to determine the necessary transshipments in order to reduce the total inventory cost and to improve the customer fill-rates. The negotiation protocol between the Location agents adopted is the contract net [Davi83]. We assume that the Location agents in need are the *managers* and the Location agents in excess are the *contractors*.

The actions executed during this step by the Location agents are the following :

- *Status evaluation* : it consists to evaluate the Location agent status.

- *Transshipment request* : the Location agent in need send a message to the other Location agents belonging to its coalition.
- *Processing of received requests* : it consists to response to the transshipment requests sent by the other Location agents in need.
- *Selection of offers* : it consists to select the best offers sent by the Location agents in excess.
- *Updating of the offers* : it consists to modify the requested transshipment quantity or cancel the offers.
- *Realization of transshipment* : it consists to update the inventory level of the Location agents in need and the others in excess that participate in the transshipment operation.

In order to achieve the negotiation described above through the executed actions the following messages are exchanged between Location agents (L_i designs Location agent in need and L_j designs Location agent in excess):

- *TransshipmentRequest* ($L_i, L_j, id_{Req}, Q_{req}$): this transshipment request message identified by id_{Req} is sent by the Location agent L_i to the Location agent L_j belonging to its coalition to ask for Q_{req} items.
- *TransshipmentOffer* ($L_j, L_i, id_{Req}, Q_{off}$): this transshipment offer message is sent by the Location agent L_j (contractor agent) proposing Q_{off} to the Location agent L_i as a response to the transshipment request launched identified by id_{Req} .
- *Apology* (L_j, L_i, id_{Req}): this apology message is sent by Location agent L_j to the Location agent L_i for the latter request identified by id_{Req} .
- *AcceptedOffer*($L_i, L_j, id_{Req}, Q_{acc}$): this message is sent by the Location agent L_i to inform the Location agent L_j that its request identified by id_{Req} is accepted and the accepted quantity is Q_{acc} .
- *RefusedOffer*(L_i, L_j, id_{Req}): this message is sent by a Location agent L_i to inform the Location agent L_j that its offer corresponding to the request identified by id_{Req} is refused.
- *CanceledOffer*(L_i, L_j, id_{Req}): this message is sent by Location agent L_i to inform the Location agent L_j that its transshipment offer corresponding to the request identified by id_{Req} is cancelled.

- *ModifiedOffer*($L_i, L_j, id_{Req}, Q_{mod}$) : this message is sent by the Location agent L_i to inform the Location agent L_j that its offer is modified and the new requested quantity is Q_{mod} .
- *RealizedTransshipment* (L_j, L_i, id_{Req}) : this message is sent by the Location agent L_j to the Location agent L_i to inform it that the transshipment is realized.

4.3.3 Step 3: Simulation stop

This step is executed at the end the period. The Interface agent recuperates the total inventory cost and the customer fill-rates for each Location agent. Then, it displays the global results relative to these evaluation parameters.

5 Experimentations and results

We realized Multi-Agent Simulation tool for the TRANshipment problem (MASTRA) based on the above presented model. MASTRA have been realized with swarm multi-agent platform. It is a simulation environment realized in objective-C [Bene02]. In this section, we are interested to show the effects of differents parameters on our inventory model.

The evaluation parameters considered during our experimentation study are : the average total inventory costs, noted C_{avg} and the average customer fill-rate, noted F_{avg} . These evaluation parameters are calculated through the following formula using the notations presented in the section 2:

$$C_{avg} = (1/n) \left[\sum_{i=1}^n \left(CS_i * Z_i + CH_i * U_i + \sum_{j=1}^n C_{ij} * X_{ij} \right) \right]$$

$$F_{avg} = (1/n) \left[1 - \left(\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n D_i} \right) \right].$$

To construct the following experimentation configurations, data relative to the cost structure are randomly generated. This choice is justified by the absence of the transshipment benchmark.

Our first experimentation study is designed to compare between a cooperative inventory management adopting our transshipment model and inventory management without transshipment. 10 inventory locations are considered. The results described in the figures 5 and 6 show that our transshipment policy contributes simultaneously to reduce consequently the inventory costs and to improve the customer fill-rates.

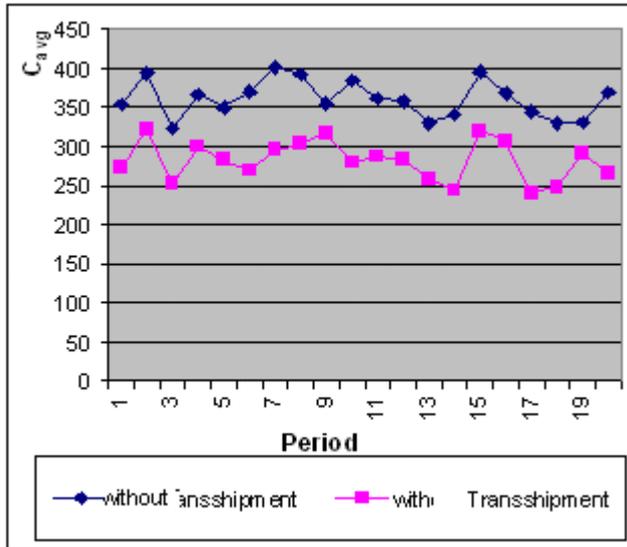


Figure 5 : Variation of C_{avg} .

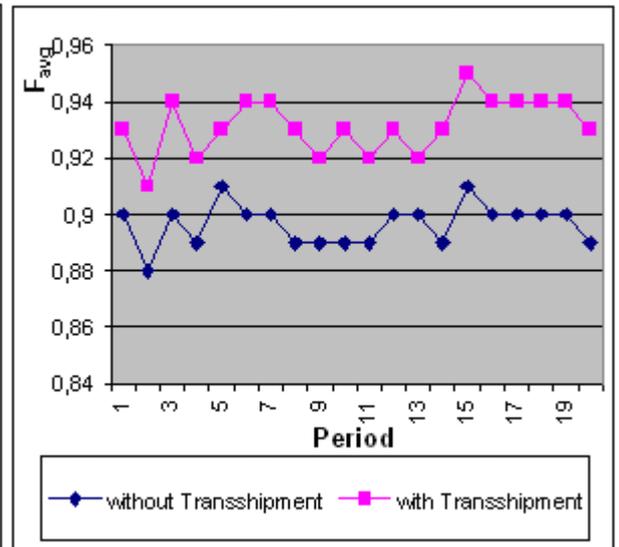


Figure 6 : Variation of F_{avg} .

The next sections describe the experimental results relative to the following parameters :

- Number of Location agents and their types
- Transshipment lead times
- Transshipment mean capacity

5.1 Number of Location agents and their types

We have compared between three inventory systems. The first includes 5 cooperatives Location agents. The second contains 10 cooperative Location agents. The third contains 20 cooperatives Locations agents. The results of this experimentation illustrated in figures 7 and 8 show that we obtain good results with the third system. We conclude that we obtain good results if the number of the Location agents participating in the transshipment actions is important. This is explained of a higher probability of cooperative interaction

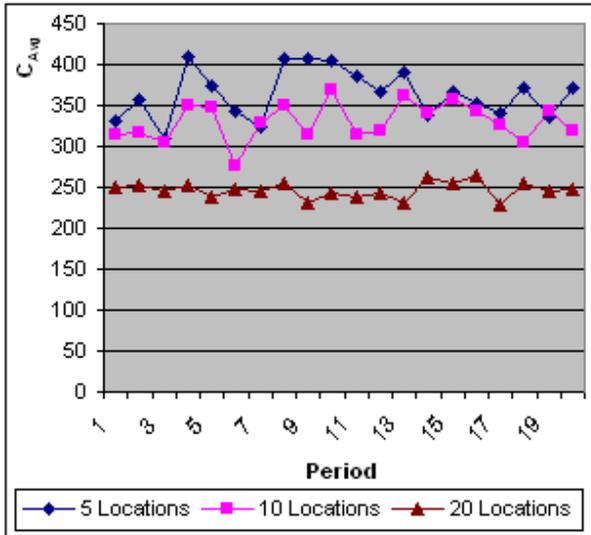


Figure 7 : Variation of C_{avg}

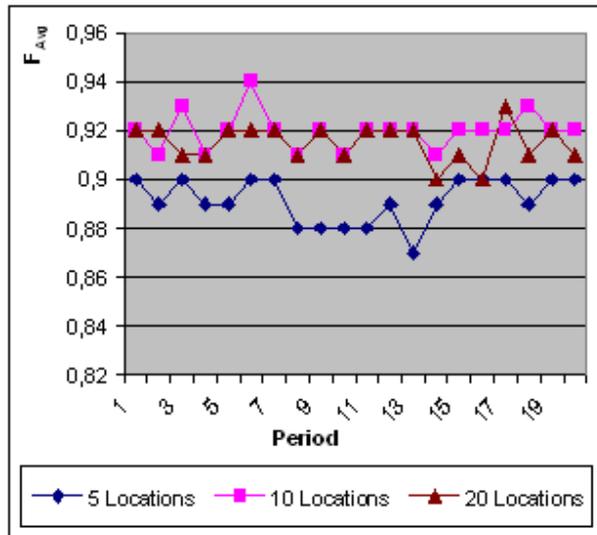


Figure 8 : Variation of F_{avg}

A second experimentation is realized to show the influence of the different Location agents types participating in the transshipment actions on the evaluation parameters. We compared between three inventory systems. The first includes only a cooperatives Location agents. The second contains only egoists Location agents. The third contains cooperatives and egoists Locations agents. The results of this experimentation illustrated in figures 9 and 10 show that we obtain good results with the third system. We conclude that we obtain good results if the population of Location agents participating in the transshipment actions is mixed.

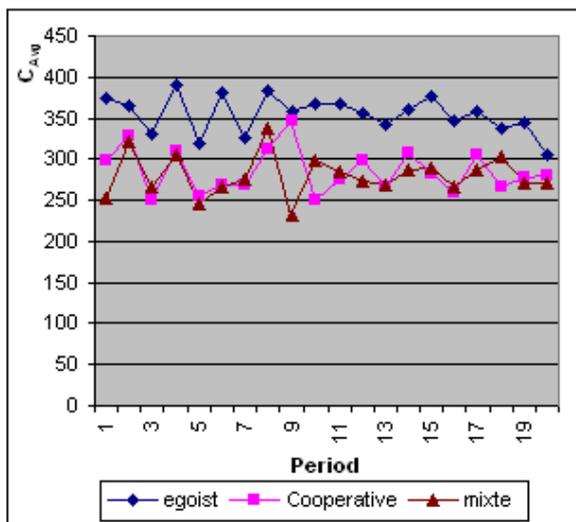


Figure 9 : Variation of C_{Avg}

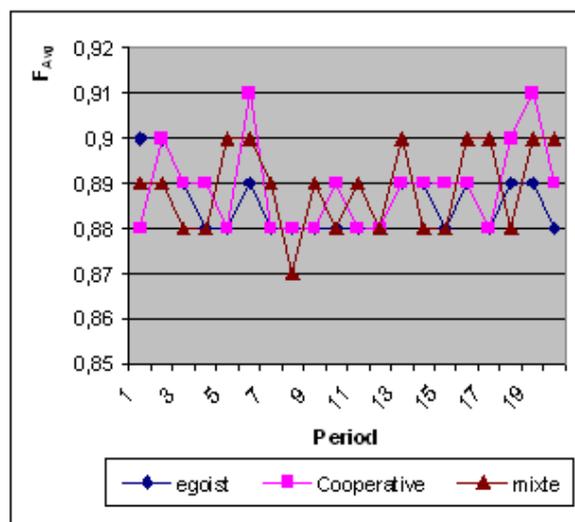


Figure 10 : Variation of F_{Avg}

5.2 Transshipment leads time

This experimentation is realized to show the influence of the transfer leads time on the evaluation parameters. We compared between two inventory systems. In the first system the transfer lead time t_{ij} is belonging to the set of days $\{1, 2\}$. However, in the second inventory system t_{ij} is belonging to the set of days $\{3, 4\}$. The results of this experimentation illustrated in figures 11 and 12 show that we obtain good results with the first system. We conclude that we obtain good results if the transfer lead time is reduced.

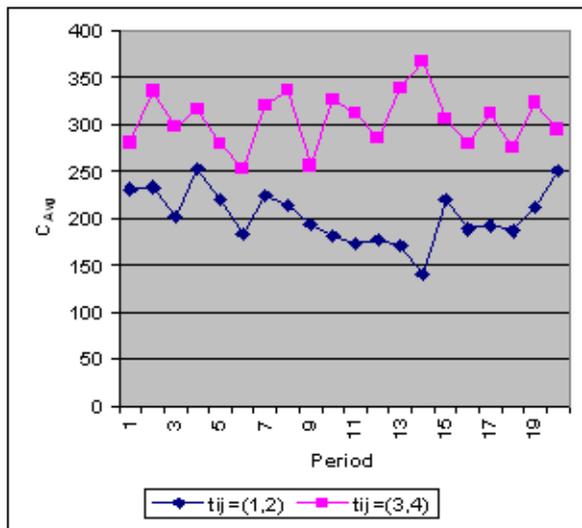


Figure 11: Variation of C_{avg}

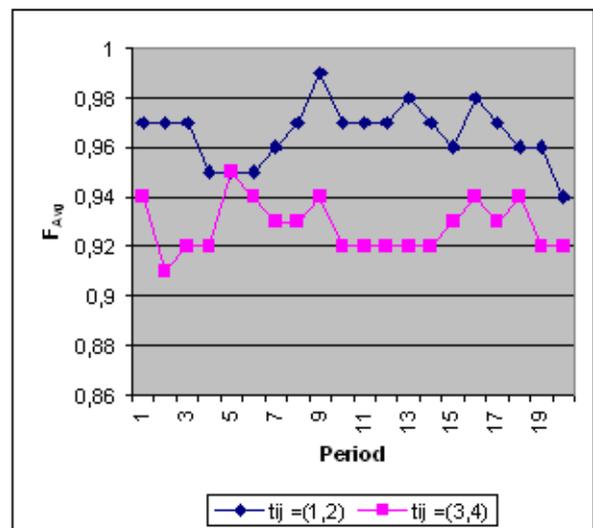


Figure 12 : Variation of F_{avg}

5.3 Transportation mean capacity

This experimentation is realized to show the influence of the transportation mean capacity on the evaluation' parameters. We are compared between two inventory systems. The first system uses a transportation mean having a capacity equal to 20 item units. However, the capacity of the transportation mean used by the second system is 50 item units. The results of this experimentation illustrated in figures 13 and 14 show that we obtain good results with the second system. We conclude that we obtain good results if the capacity of the transportation mean used is large.

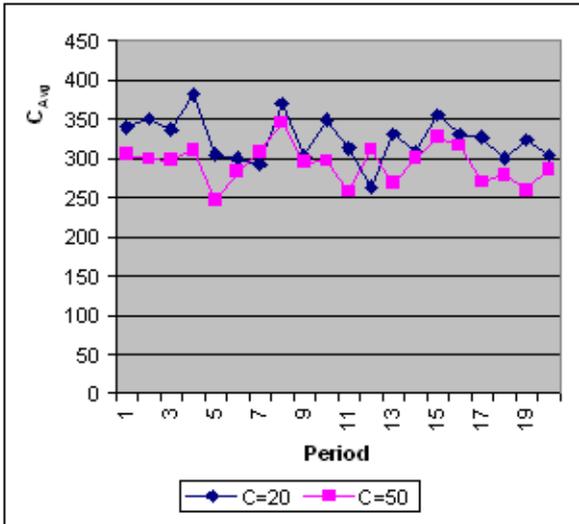


Figure 13 : Variation of C_{Avg}

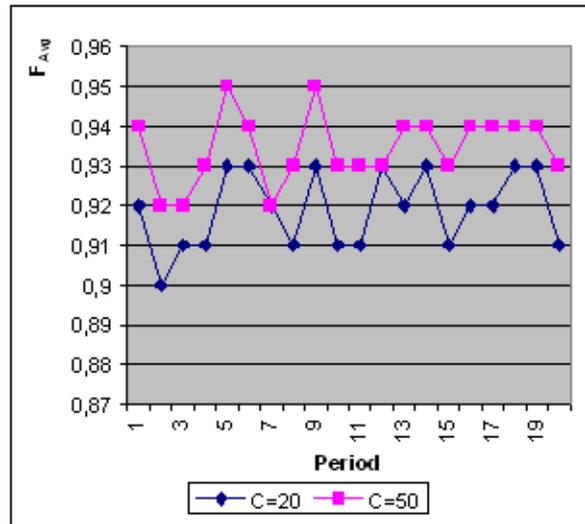


Figure 14 : Variation of F_{Avg}

6 Conclusion

In this paper we have proposed a new transshipment policy that takes in account a non-negligible transshipment lead times and a limited transportation mean capacity. In order to satisfy the maximum of the customer demands, we have required that the transshipped items participate to satisfy the demands of the current period. So, we have proposed to divide the period into several sub-periods and at the end of one of them the transshipment decision is made. We have introduced a multi-agent model that simulates the behavior of the collaborative network locations. Our experimental results demonstrate that: (1) the number of locations influences the total inventory cost. In fact, we obtain a good result if the number of locations is important, (2) We obtain a good result if the population of the Location agents is mixed, composite of egoist Location agents and cooperatives ones, (3) The transshipment lead times affect the total inventory cost. With a reduced transshipment lead times we obtain better results, (4) the transportation capacity influence the total inventory cost. In fact, with a large capacity we obtain a good result.

Our future works are to determine the initial inventory level, where the transshipment strategy described above is practiced and to use the learning techniques to evaluate the risk caused by the participating in the transshipment operations.

References

- [Axsä90] Axsäter S. : Modelling emergency lateral transshipments in inventory systems, *Management Science*, 1990, 36, p. 1329-1338.
- [Bene02] Benedikt S. And Francesco L : *Economic Simulations in Swarm: Agent-Based Modelling and Object Oriented Programming*, Springer , 2002
- [Davi83] Davis, R. and R. G. Smith : Negotiation as metaphor for distributed problem solving, *Artificial intelligence*, 1983, 20, p. 63-109.
- [Ferb95] Ferber J. : *Les systèmes multi-agents vers une intelligence collective*. Ed InterEditions Paris, 1995.
- [Here99] Herer, Y. T. and A. Rashit. : Lateral Stock Transshipments in a Two-Location Inventory System with Fixed and Joint Replenishment Costs, *Naval Research Logistics*, 1999, 46 p. 525-547.
- [Kris65] Krishnan, K. S. and V. R. K Rao: Inventory control in N warehouses, *The journal of industrial engineering*, 1965, 16, p. 212-215.
- [Ozde03] Ozdemir, D., M. Tzur and Y.T. Herer : A Monte Carlo simulation approach to the capacitated multi-location transshipment problem. *Proceedings of the winter simulation conference*, 2003, p. 1729-1736.
- [Robi90] Robinson, L.W.: Optimal and Approximate Policies in Multiperiod, Multilocation Inventory Models with Transshipments. *Operations Research*, 1990, 38, p. 278-295.
- [Rudi02] Rudi , N. and V. Miegheem : Newsvendor Networks: Inventory Management and Capacity Investment with Discretionary Activities. *Manufacturing and service operations management*, 2002, 4 , p. 313-335.
- [Sher92] Sherbrooke, C.C.: Multi-echelon inventory systems with lateral supply. *Naval Research Logistics* ,1992, 39, p. 29-40.
- [Taga89] Tagaras, G. : Effects of Pooling on the Optimization and Service Levels of Two-Location Inventory Systems. *IIE Transactions*, 1989, 21, p. 250-257.
- [Taga99] Tagaras, G.: Pooling in Multi-Location Periodic Inventory Distribution Systems. *Omega*, 1999, p. 39-59.
- [Taga92] Tagaras, G. and M. Cohen. : Pooling in Two-Location Inventory Systems with Non-

Negligible Replenishment Lead Times. *Management Science*, 1992, 38, p. 1067-1083.

[Taga02] Tagaras, G. and D. Vlachos : Effectiveness of stock transshipment under various demand distribution and non-negligible transshipment times production and operation management, 2002 ,11(1), p. 183-198.

[Week05] Week, K. E., M. Dada : Optimal Policies for transshipment inventory in Retail Network. *Management Science*, 2005, 51(10), p. 1515-1533.

[Will71] William, Feller : An introduction to the probability theory and its applications . Wiley, New York 1971.

[Zhan05] Zhang, J : Transshipment and its Impacts on Supply chain Members' Performances, *Management Science*, 2005, 51(10), p. 1534-1539.

Einführung in den Track

eFinance

Prof. Dr. Hans Ulrich Buhl

Universität Augsburg

Prof. Dr. Wolfgang König

Universität Frankfurt a.M.

Prof. Dr. Peter Gomber

Universität Frankfurt a.M.

Wolfgang Gaertner

Deutsche Bank

Nach einem Fokus auf Kostenreduktion und Prozessoptimierung in der Ertragskrise öffnet sich die Finanzdienstleistungsbranche nun sukzessive in Richtung einer Verbesserung der Erträge sowie einer Optimierung der Ertrags-/Risiko-Positionen. Hierzu gehören etwa die IT-gestützte Steigerung des Kundenverständnisses und darauf aufbauend der wirtschaftlichen Beratungsqualität; ebenso Produkt- und Prozessinnovationen (z. B. im Bereich Wertpapierhandel) sowie strategische Allianzen innerhalb der Branche und über deren Grenzen hinaus im Verbund mit dem hierfür notwendigen effizienten Informationsmanagement.

Trotzdem ist der Trend zur Konsolidierung, zur Industrialisierung der Prozesskette und zur Weiterentwicklung von Sourcingstrategien ungebrochen. Darüber hinaus sehen sich die Häuser einer hohen Dichte regulatorischer Anforderungen (z. B. Basel II, Sepa oder MiFID) gegenüber, die direkten Einfluss auf Geschäftsmodelle, Prozesse und IT-Infrastruktur haben. Auf Basis der regulatorisch und Reporting-getriebenen IT-Investments nun auch die Ertragsziele besser zu erreichen, steht bei vielen Finanzdienstleistern im Vordergrund des Interesses.

Ziel des Tracks ist, eine Plattform für den Austausch von Forschungsergebnissen zwischen Praxis und Wissenschaft zu aktuellen und IT-bezogenen Fragestellungen der Finanzindustrie, der Finanzmärkte und der betrieblichen Finanzwirtschaft bereitzustellen.

Programmkomitee:

Dr. Mario Daberkow, Postbank AG

Prof. Dr. Jürgen Moormann, Hochschule für Bankwirtschaft

Frank Annuschein, Commerzbank AG

Prof. Dr. Günther Müller, Universität Freiburg

Prof. Dr. Robert Winter, Universität St. Gallen

Klaus Lintelmann, IBM Deutschland GmbH

Rainer Riess, Deutsche Börse AG

Prof. Dr. Andreas Hackethal, EUROPEAN BUSINESS SCHOOL (ebs)

Prof. Dr. Christof Weinhardt, Universität Karlsruhe (TH)

Ralf Schmid, MLP Finanzdienstleistungen AG

Außerbörslicher Emittentenhandel: Handelsplatzwahl von Online-Investoren

Dennis Kundisch, Tobias Stuber

Lehrstuhl WI-IF
Universität Augsburg
86135 Augsburg
dennis.kundisch@wiwi.uni-augsburg.de

Abstract

Der außerbörsliche Emittentenhandel stellt für Privatanleger eine attraktive Alternative zur Nutzung von börslichen Handelsplätzen dar. Bislang existieren weder theoretische noch empirische Ergebnisse zur Handelsplatzwahl von Privatinvestoren. Auf Basis einer Online-Umfrage unter Online-Investoren wird im vorliegenden Beitrag die Neigung außerbörslich zu handeln untersucht. Es zeigt sich, dass vor allem die Bekanntheit, die expliziten Transaktionskosten, Sicherheits- und Transparenzfragestellungen sowie die Handelszeiten eine Rolle spielen.

1 Einleitung

Der außerbörsliche Handel war lange Zeit institutionellen Investoren vorbehalten. [Gomb00, S. 35] Getrieben durch die rasante Entwicklung des Internets, den Boom so genannter Retail-Derivate und den Auf- bzw. Ausbau von außerbörslichen Handelsplattformen (bspw. CATS-OS, Tradelink oder T.I.Q.S.), hat sich diese Situation in den letzten Jahren jedoch verändert. Inzwischen bieten beinahe alle großen Online-Broker in Deutschland Privatanlegern die Möglichkeit außerbörslich zu handeln. Dabei wird neben dem außerbörslichen Aktien- insbesondere der außerbörsliche Retail-Derivatehandel offeriert. Deutschland ist im Bereich des Retail-Derivatehandels Weltmarktführer - sowohl bezogen auf den Umsatz als auch auf die Anzahl der gehandelten Produkte.¹

¹ Der Orderbuchumsatz des Handelssegments EUWAX der Börse Stuttgart - der börsliche Marktführer im Retail-Derivatehandel - betrug 41 Mrd. Euro in 2005. Ende 2005 waren dort über 80.000 Produkte gelistet. Der

Privatanleger können zwischen verschiedenen Handelsplätzen wählen. Speziell für den Handel von Retail-Derivaten stehen bspw. das Handelssegment EUWAX der Börse Stuttgart, das Handelssegment Smart Trading der Börse Frankfurt und der außerbörsliche Emittentenhandel zur Auswahl. Es ist davon auszugehen, dass die Handelsplatzentscheidung durch die Vielzahl der zur Verfügung stehenden Handelsplätze mit teilweise unterschiedlichen Marktmodellen [Holt04] für Privatanleger einen komplexen Entscheidungsprozess darstellt. Da das Angebot eines jeden Handelsplatzes Kosten verursacht, ist es für Anbieter von Online-Brokerage von Interesse, warum Online-Investoren bestimmte Handelsplätze in Anspruch nehmen. Offensichtlich haben auch die konkurrierenden Handelsplatzbetreiber ein Interesse daran zu erklären, welche Faktoren Privatanleger dazu veranlassen börslich oder außerbörslich zu handeln. Während bereits einige wenige Studien zur Handelsplatzwahl institutioneller Investoren vorliegen, [Schi95; Aver98] existieren bislang keine Arbeiten, welche die Handelsplatzwahl von Privatanlegern untersuchen. Allerdings wurde bereits darauf hingewiesen, dass die Handelsplatzwahl von Privatinvestoren möglicherweise von anderen Faktoren beeinflusst wird als bei institutionellen Investoren. [MeWi94] Aus diesen Überlegungen leitet sich die Forschungsfrage für den vorliegenden Beitrag ab: *Welche Faktoren beeinflussen die Neigung von (privaten) Online-Investoren außerbörslich zu handeln?*

Der Beitrag ist wie folgt gegliedert: Abschnitt 2 beinhaltet zentrale Begriffsabgrenzungen. In Abschnitt 3 werden die Einflussfaktoren bei der Handelsplatzwahl von Online-Investoren theoretisch abgeleitet und zum außerbörslichen Handel in Beziehung gesetzt sowie die zu testenden Hypothesen formuliert. Im Abschnitt 4 werden das empirische Design sowie die Durchführung der Untersuchung beschrieben. Abschnitt 5 stellt die empirischen Ergebnisse vor und liefert eine Interpretation der Daten. Schließlich werden in Abschnitt 6 die Ergebnisse zusammengefasst.

2 Begriffsabgrenzung und Definitionen

Da Privatanleger nicht selbst Mitglied an einem Handelsplatz werden können, benötigen sie zu deren Zugang einen Intermediär, den so genannten *Broker*. Nach § 2 (3) Wertpapierhandelsgesetz führen Broker Geschäfte in eigenem Namen und für die Rechnung ihrer Kunden aus.

geschätzte außerbörsliche Anteil am Gesamtmarktumsatz in Retail-Derivaten liegt bei ca. 70%. [Euwa05] Insgesamt kann daher von einem Gesamtmarktumsatz von weit über 100 Mrd. Euro in 2005 ausgegangen werden.

Hier wird der Begriff des *Online-Brokers* als umfassende Bezeichnung für solche Anbieter verwendet, deren Leistungsspektrum die Verschaffung von Marktzugang sowie das Orderrouting umfasst und die ihre Leistungen (auch) über elektronische Vertriebskanäle erbringen.² [KuKr06]

Die Dienstleistungen von Online-Brokern werden hauptsächlich von Privatanlegern und externen Vermögensverwaltern in Anspruch genommen. Ein Privatanleger, welcher ein Depot bei mindestens einem Online-Broker unterhält und über dieses Wertpapiertransaktionen abwickelt, wird im Folgenden als *Online-Investor* bezeichnet.

Viele Wertpapiere können - nach ihrer Emission - auf dem Sekundärmarkt an unterschiedlichen Handelsplätzen gehandelt werden. Diese Handelsplätze stehen teilweise in Konkurrenz zueinander und ihr Ziel ist es, ein möglichst großes Ordervolumen auf sich zu vereinen. [Heil02, S. 168] Auf oberster Ebene konkurrieren außerbörsliche mit börslichen Handelsmöglichkeiten um den Orderflow von Investoren.

Die Börsenregulierung erfolgt in jedem Land separat. Es existiert im deutschen Börsengesetz keine Legaldefinition einer *Börse* im materiellen Sinne. Folgende Aspekte kennzeichnen jedoch das Begriffsverständnis: Örtliche und zeitliche Handelszentralisierung, exklusive Teilnahme von professionellen Händlern, Zusammenschluss der professionellen Händler zu einem Verbandbetrieb, der Verbandbetrieb verfasst die Handelsregeln und ist Eigentümer über die Infrastruktur der Marktveranstaltung. [PiBR96, S. 9] Demgegenüber setzt sich der außerbörsliche Handel im ursprünglichen Verständnis aus einer Vielzahl von heterogenen Teilmärkten zusammen. Diese Teilmärkte lassen sich nach den gehandelten Produkten und nach den Marktteilnehmern voneinander abgrenzen. Charakteristisch für den außerbörslichen Handel ist das Fehlen von festgelegten Handelszeiten und einer räumlichen Lokalisierung. [Gomb00, S. 35] Im Gegensatz zum börslichen Handel gibt es auch keine standardisierten Kontrakte. Handelsgeschäfte können individuell festgelegt und ausgestaltet werden. Daraus resultiert einerseits eine größere Flexibilität, andererseits fehlt eine regulierende Instanz, die für Preistransparenz sorgt. Gomber definiert den *außerbörslichen Handel* als: „Gesamtheit von unter professionellen Marktteilnehmern organisierten, nicht-börslichen Märkten für Finanzprodukte auf der Basis heterogener Kommunikationsnetzwerke“. [Gomb00, S. 35] Für Online-Investoren ist der Begriff des außerbörslichen Handels allerdings einzugrenzen. Sie haben nicht wie institutionelle Investoren die Möglichkeit Kontrakte individuell auszuhandeln. Der außerbörsliche Handel, der

² Dies umfasst sowohl klassische Online-Broker (bspw. Comdirect, Cortal Conors) als auch die Online-Brokerage-Funktionalitäten von traditionellen Geschäftsbanken (bspw. Dresdner Bank, Commerzbank).

den Online-Investoren derzeit angeboten wird, kann als *standardisierter außerbörslicher Handel für Online-Investoren* bezeichnet werden. Er stellt nur eine Teilmenge des außerbörslichen Handels im traditionellen Sinne dar. Im vorliegenden Beitrag wird der *außerbörsliche Handel* im Hinblick auf das Angebot für Online-Investoren folgendermaßen definiert: Gesamtheit der von Online-Brokern angebotenen, standardisierten, nicht-börslichen Märkte für Wertpapiere, auf Basis heterogener Kommunikationsnetzwerke.

Man unterscheidet zwei Arten des außerbörslichen Handels: Von *Internalisierung* spricht man, wenn der Online-Investor direkt mit seinem Online-Broker als Kontrahent in eine Handelsbeziehung tritt. Der Broker handelt dabei gegen seinen eigenen Bestand. Beim *außerbörslichen Emittentenhandel* handelt der Online-Investor direkt mit dem Emittenten oder Market Maker³ eines Wertpapiers. Die Emittenten bzw. Market Maker stellen dabei für Wertpapiere unverbindliche Quotes, die vom Online-Investor über die Handelsapplikation des Online-Brokers angefragt werden können. Auf solche Quotes kann ein Online-Investor reagieren und per Mausklick ein Angebot zum gerade veröffentlichten Preis abgeben.⁴

Im außerbörslichen Handel für Online-Investoren nimmt der außerbörsliche Emittentenhandel die bedeutendere Stellung ein. Zum einen ist das Wertpapierangebot im Emittentenhandel wesentlich größer, zum anderen wird der Emittentenhandel von deutlich mehr Online-Brokern angeboten. Der Fokus der Untersuchung liegt daher in diesem Beitrag auf dem außerbörslichen Emittentenhandel.

3 Theorie zur Handelsplatzwahl

Die Frage, wo eine Order platziert werden soll, stellt sich bei jeder Wertpapiertransaktion aufs Neue. Es ist anzunehmen, dass die Vielzahl der wählbaren Alternativen die Entscheidung des Online-Investors erschwert. Zudem sorgt die Fülle der Kriterien, nach denen ein Handelsplatz beurteilt werden kann, für Komplexität. Aus der Theorie lässt sich eine Reihe von Einflussfak-

³ Ein Market Maker ist eine Bank oder eine Händlerfirma, welche verbindliche Kauf- und Verkaufskurse für ein Wertpapier stellt und damit als Marktpartner während der Handelszeit ständig zur Verfügung steht. Insofern ist der Oberbegriff Emittentenhandel etwas unscharf, da ein Market Maker Emittent eines von ihm betreuten Wertpapiers sein kann, aber nicht sein muss.

⁴ Rechtlich gesehen ist erst ein Handelsgeschäft zustande gekommen, wenn der Emittent seinerseits dieses Angebot annimmt. Durch die Ausgestaltung des Emittentenhandels bei den Online-Brokern ist aber anzunehmen, dass viele Online-Investoren davon ausgehen, sie reagierten auf ein Angebot seitens des Emittenten bzw. Market Makers. Zu Marktmodellen im außerbörslichen Emittentenhandel (Angebote für den Online-Investor überwiegend „quote driven“) und im börslichen Handel (Angebote für den Online-Investor überwiegend „order driven“) vgl. [Thei98, S.8 ff., Holt04].

toren auf die Handelsplatzwahl ableiten und begründen. Dabei werden die Entscheidungssituation und die zu beachtenden Entscheidungskriterien aus der Sicht institutioneller Investoren untersucht. [Aver98, Schi95] Die Ergebnisse und Überlegungen aus diesen Untersuchungen sind daher nur bedingt auf die vorliegende Fragestellung übertragbar, auch wenn es zwischen Online-Investoren und institutionellen Investoren Parallelen gibt. [MeWi94, S. 538] Des Weiteren wurden die Faktoren meistens im Hinblick auf Börsenplatzentscheidungen analysiert. Berücksichtigt man auch den Fall des außerbörslichen Handels bzw. speziell des außerbörslichen Emittentenhandels, erscheinen weitere Faktoren relevant.

3.1 Im Modell berücksichtigte Einflussfaktoren

3.1.1 Explizite Transaktionskosten

Auch wenn eine einheitliche Definition von Transaktionskosten in der Literatur nicht existiert, [Aver98, S. 68] so werden diese häufig in explizite und implizite Transaktionskosten unterteilt. Explizite Transaktionskosten bezeichnen die Kosten, welche für den Anleger unmittelbar ersichtlich sind. Sie werden üblicherweise auf der Handelsabrechnung des Online-Brokers ausgewiesen. Explizite Transaktionskosten sind für den Privatanleger einfach nachzuvollziehen, lassen sich schon im Vorfeld einer Transaktion sehr genau abschätzen [Schi95, S. 21] und spielen bei seiner Entscheidung für den Handelsplatz eine bedeutende Rolle. [Gerk93, S. 733; MeWi94, S. 538; HiSt96, S. 527ff.] Implizite Transaktionskosten bezeichnen die für den Anleger nicht direkt ersichtlichen Kosten, welche vor allem aus dem erzielten Preis am jeweiligen Handelsplatz resultieren. Implizite Transaktionskosten sind - wenn überhaupt - erst nach Ausführung einer Order zu ermitteln. Bezüglich expliziter Transaktionskosten stellt der außerbörsliche Handel im Vergleich zu den Börsenplätzen in der Regel eine günstigere Alternative dar: Börsegebühren und Makler Courtage fallen hierbei zum Beispiel nicht an.

Es wird außerdem ein moderierender Einfluss der Tradeanzahl auf die Bedeutung der expliziten Transaktionskosten vermutet, da [KuKr06] Indizien für unterschiedliche Einstellungen bei Online-Investoren in Abhängigkeit der Tradeanzahl finden. Es wird davon ausgegangen, dass die Transaktionskosten mit steigender Tradeanzahl zunehmend Beachtung finden, da potenziell von über die Handelsangebote besser informierten Investoren ausgegangen werden kann.

3.1.2 Handelszeit

Viele Online-Investoren schätzen die Möglichkeit ohne Zeitbeschränkung handeln zu können. [StRa04, S.74] Vor allem für Investoren, die dem Online-Brokerage nicht als Hauptbeschäftigung nachgehen, dürften lange Handelszeiten interessant sein, vor allem weil auch nach Börsenschluss in Deutschland noch auf Entwicklungen in US-Märkten reagiert werden kann.⁵ Online-Investoren, die der Handelszeit eine hohe Bedeutung beimessen, dürften daher im außerbörslichen Emittentenhandel eine attraktive Alternative zum Börsenhandel sehen.

3.1.3 Vertrauen

Vertrauen spielt bei Käufer-Verkäufer-Beziehungen grundsätzlich eine wichtige Rolle. In einigen Bereichen unterscheidet sich die (Anleger-)Sicherheit, welche für Vertrauen sorgen kann, zwischen börslichen und außerbörslichen Angeboten nicht (bspw. Anwendung des Wertpapierhandelsgesetzes). So gelten sowohl für börsliche als auch für außerbörsliche Handelsplätze sogenannte Mistrade-Regelungen, bei denen die Handelspartner dazu berechtigt sind, eine Transaktion unter bestimmten Bedingungen rückgängig zu machen. Börsen werden in Deutschland zusätzlich durch das Börsengesetz (BörsG) reguliert. Die Vorschriften bezüglich Fairness, Effizienz und Transparenz im BörsG gewährleisten dabei einen höheren Anlegerschutz. Dies sollte es für Börsen daher vergleichsweise einfacher machen, das Vertrauen des Privatanlegers zu gewinnen. Online-Investoren, denen die Vertrauenswürdigkeit eines Handelsplatzes besonders wichtig ist, werden demnach eher dazu geneigt sein börslich zu handeln.

3.1.4 Involviertheit in den Handelsprozess

Durch die Möglichkeiten des Online-Investierens können Privatanleger wesentlich schneller und unmittelbarer am Geschehen der Wertpapiermärkte teilhaben. [StRa04, S. 75 ff.] Während früher kein direkter Zugang zu Marktdaten und weitergehenden Anlageinformationen vorhanden war, ist es heute möglich in „real-time“ am Marktgeschehen teilzunehmen. Die durch das Internet bedingte stärkere Involviertheit der Privatanleger wird in der Literatur auch als Grund für den erhöhten Wertpapierhandel genannt. [BaOd02, S. 476] Das Konstrukt der „Involviertheit in den Handelsprozess“ beschreibt die Bedeutung, die Online-Investoren der Markt-

⁵ *Außerbörslich*: werktags meist von 8 bis 22 Uhr, vereinzelt auch am Wochenende, *börslich*: werktags von 9 bis 20 Uhr, XETRA nur bis 17.30 Uhr. kein Handel am Wochenende. Vgl. [Schi95, S. 40] für theoretische Gründe für die Bedeutung der Handelszeit.

und Handelsnähe beim Wertpapierhandel beimessen. Der außerbörsliche Emittentenhandel ist durch Schnelligkeit, Einfachheit und Direktheit gekennzeichnet. Online-Investoren können auf Quotes entsprechend reagieren und wissen, dass dies auch aller Voraussicht nach die Ausführungspreise sind, während sich Preise im börslichen Handel über ein Limit nur nach oben (Kauf) bzw. nach unten (Verkauf) begrenzen lassen und der Anleger passiv auf Ausführung seiner Order warten muss. Es ist deswegen davon auszugehen, dass Online-Investoren, für welche die Involviertheit in den Handelsprozess wichtig ist, eher dazu tendieren außerbörslich zu handeln.

3.1.5 Individualisierung des Handelsprozesses

Durch Orderzusätze (bspw. Ausführungsbestimmungen, Gültigkeitsbeschränkungen) können Online-Investoren ihre Orders individuell gestalten. Die Nutzung von Orderzusätzen hängt in der Regel vom zugrunde liegenden Ordersystem und Marktmodell sowie vom Funktionsumfang der Handelsapplikation des Online-Brokers ab. Zudem sind manche Orderzusätze auch auf bestimmte Wertpapierkategorien begrenzt. Im außerbörslichen Emittentenhandel ist die Verwendung von Orderzusätzen nur wenig verbreitet. Inzwischen gibt es zwar vereinzelt Emittenten, die die Verwendung von Orderzusätzen anbieten,⁶ der börsliche Handel hat jedoch in Bezug auf eine Individualisierung von Orders aber derzeit noch deutliche Vorteile. Es ist daher anzunehmen, dass Online-Investoren, welche die Bedeutung der Individualisierung des Handelsprozesses hoch einschätzen, weniger geneigt sein werden außerbörslich zu handeln.

3.1.6 Bekanntheit

Die Bekanntheit eines Produktes bzw. einer Marke wird als wichtiger Einflussfaktor auf die Auswahl der Produktalternativen und auf die Auswahl eines Produktes angesehen. [MaSh00, S. 5] Die Bekanntheit eines Angebots hat gleichzeitig einen Einfluss auf die Wahrscheinlichkeit vom Kunden gekauft zu werden und auf die dem Produkt zugeschriebene Qualität. [HoBr90, S. 145 ff.] Handelsplätze bieten spezifische Leistungen an, mit welchen sie um die Gunst der (Privat-)Anleger werben. [Gomb00, S. 168] Ähnlich wie substitutive Produkte eines Marktes, stehen sie in Konkurrenz zueinander. Investoren wählen dabei Handelsplätze nicht nur nach objektiven sondern auch nach subjektiven Kriterien - bspw. dem Image - aus. [Schi95, S. 43] Bevor ein Handelsplatz allerdings ein Image aufbauen kann, muss er bekannt werden. Es liegt daher

⁶ Zu nennen sind hier die Emittenten BNP Paribas und SG Warrants.

nahe, dass die Markenbekanntheit bzw. die Bekanntheit eines Handelsplatzes einen Einfluss auf die Kundenwahrnehmung hat. Beim relativ jungen Phänomen des außerbörslichen Emittentenhandels ist davon auszugehen, dass dessen Bekanntheit im Vergleich zu den großen, seit Jahren etablierten Börsenplätzen wie Frankfurt oder Stuttgart noch gering ist. Daher wird angenommen, dass Personen, welche den außerbörslichen Emittentenhandel gut kennen, eher geneigt sein werden außerbörslich zu handeln.

3.2 Im Modell nicht berücksichtigte Einflussfaktoren

Theoretische Überlegungen bezüglich der Einflussfaktoren auf die Handelsplatzwahl von Investoren legen eine Reihe weiterer Einflussfaktoren nahe. Vor allem die Arbeit von Schiereck [Schi95] nennt einige weitere Konstrukte. Diese werden im Folgenden kurz vorgestellt und es wird knapp begründet, warum sie nicht in das Modell aufgenommen wurden.

- *Liquidität*: Die Ordergrößen von Online-Investoren haben in der Regel keinen Einfluss auf die Markttiefe. Die Marktbreite ist zwar auch für den Privatanleger interessant, im Vergleich zu institutionellen Investoren wird der Bid-Ask-Spread allerdings bei vielen Online-Investoren eine geringere Rolle spielen.
- *Image*: Das Image wurde im zugrunde gelegten Modell nicht explizit berücksichtigt, da diverse Aspekte des von Schiereck [Schi95, S. 43] definierten Konstruktes separat modelliert wurden („Vertrauen“, „Involviertheit in den Handelsprozess“, „Bekanntheit“).
- *Abwicklungseffizienz*: Die Effizienz mit der das Clearing und Settlement [Schi95, S. 27] durchgeführt werden ist vor allem für institutionelle Anleger eine der wichtigsten Leistungskomponenten im Wertpapierhandel. [Aver98, S. 75] Für Online-Investoren spielt die Abwicklungseffizienz eine deutlich geringere Rolle. [MeWi94, S. 538]
- *Produktpalette*: Es ist davon auszugehen, dass mit einer Verbreiterung der Produktpalette auch die Attraktivität eines Handelsplatzes steigt [MeWi94, S. 538]. Im Hinblick auf Retail-Derivate liegt ein fast identisches Produktangebot an den drei oben genannten Handelsplätzen vor.
- *Markteffizienz* [Thei98, S. 184]: Der Zusammenhang zwischen Markteffizienz bzw. Informationseffizienz und der Wahl eines Handelsplatzes ist bspw. bei der Implementierung von Arbitragestrategien nicht eindeutig.
- *Markttransparenz* [Schi95, S. 30]: Durch die Börsenregulierung wird im Vergleich zum außerbörslichen Handel für eine deutlich höhere Pre- und Post-Trade-Transparenz gesorgt.

Dieser Aspekt ist jedoch bereits im Konstrukt „Vertrauen“ enthalten und wurde nicht gesondert modelliert.

3.3 Hypothesen

Folgende Hypothesen werden aufbauend auf die Überlegungen in Abschnitt 3.1 überprüft:

Hypothese H1: Je höher die Bedeutung der expliziten Transaktionskosten, desto höher ist die Neigung außerbörslich zu handeln.

Hypothese H2: Je höher die Tradeanzahl, desto höher ist die Bedeutung der expliziten Transaktionskosten.

Hypothese H3: Je höher die Bedeutung der Handelszeit, desto höher ist die Neigung außerbörslich zu handeln.

Hypothese H4: Je höher die Bedeutung von Vertrauen, desto geringer ist die Neigung außerbörslich zu handeln.

Hypothese H5: Je höher die Involviertheit in den Handelsprozess, desto höher ist die Neigung außerbörslich zu handeln.

Hypothese H6: Je höher die Bedeutung der Individualisierung des Handelsprozesses, desto geringer ist die Neigung außerbörslich zu handeln.

Hypothese H7: Je höher die Bekanntheit des Angebots außerbörslich zu handeln, desto höher ist die Neigung außerbörslich zu handeln.

4 Empirische Studie

4.1 Design und Durchführung der Befragung

Aufgrund der vorliegenden Fragestellung war davon auszugehen, dass die zu befragenden Personen alle über einen Internetanschluss verfügen und mit dem Internet vertraut sind. Die Befragung wurde deswegen online durchgeführt. Die einschlägigen Empfehlungen von [ADM01] bei der Gestaltung von Online-Umfragen wurden beachtet. Im Mittelpunkt des Interesses der Umfrage stehen Online-Investoren, welche entweder bereits erste Erfahrungen mit dem außerbörslichen Emittentenhandel gemacht haben oder zumindest schon über erste Kenntnisse darüber verfügen. Repräsentative Aussagen könnten theoretisch also nur für den Kreis der Online-Investoren gemacht werden. Für den Personenkreis der Online-Investoren sind jedoch keine

derartigen Informationen vorhanden.⁷ Daher kann für die vorliegende Online-Umfrage keine Repräsentativität gewährleistet werden. Die Rekrutierung der Teilnehmer erfolgte über Texthinweise, Banner und Newslitereinträge von Emittenten, Finanzwebseiten, Börsen und Online-Brokern.⁸ Als Incentive für eine Teilnahme wurden drei Mal 7.500 Miles&More Prämienmeilen verlost.⁹

Zur Überprüfung der vorgestellten Hypothesen sind die Konstrukte „Bedeutung der expliziten Transaktionskosten“, „Bedeutung der Handelszeit“, „Bedeutung von Vertrauen“, „Involviertheit in den Handelsprozess“, „Bedeutung der Individualisierung des Handelsprozesses“ und „Bekanntheit“ zu messen. Jedes Konstrukt wurde mittels Multi-Item-Messungen abgefragt, denen auf einer 5-stufigen Skalen mehr oder minder zuzustimmen war (1 = Trifft überhaupt nicht zu, 5 = Trifft voll und ganz zu). Bei Vorliegen von interner Konsistenz wurden die einzelnen Messvariablen pro Konstrukt und pro Auskunftsperson arithmetisch gemittelt. Die Statements wurden teilweise abwechselnd negativ und positiv formuliert. Zur Überprüfung, ob die Werte der Items zu einer gemeinsamen Skala zusammengefasst werden können, wurde mittels einer Reliabilitätsanalyse der Wert für Cronbach's Alpha bestimmt. Für die Dummyvariable „Tradeanzahl“ wurde der Median der Tradeanzahl der Auskunftspersonen ermittelt.

Der eingesetzte Fragebogen¹⁰ wurde in einem iterativen Prozess unter anderem mit der Unterstützung des Zentrums für Umfrage- und Marktforschung (ZUMA) in Mannheim erstellt. Nachdem der Fragebogen in einer ersten Version fertig gestellt war, wurde er einem Pretest unterzogen. Den Auskunftspersonen stand ein Online-Glossar zur Verfügung. Der Fragebogen selbst bestand aus vier logischen Teilen. Teil A beinhaltete Fragen zum allgemeinen Handelsverhalten der Auskunftspersonen. Am Ende von Teil A entschied sich, ob die Auskunftspersonen im weiteren Verlauf zu den außerbörslich unerfahrenen oder den außerbörslich erfahrenen Nutzern gezählt werden und die dementsprechenden Fragen gestellt bekommen. Personen, welche weniger als 20% ihrer Trades außerbörslich handeln (außerbörslich unerfahrene Nutzer), bekamen die Fragen zu den Konstrukten gestellt (Teil B).¹¹ Zu jedem Konstrukt wurden drei bis

⁷ [NFO03] liefert Hinweise auf Verteilungen gewisser Merkmale von Online-Investoren. Von repräsentativen Merkmalsverteilungen wird aber auch dort nicht gesprochen.

⁸ Eine Präselektion der Umfrageteilnehmer kann gleichwohl nicht ausgeschlossen werden.

⁹ Um die Motivation zur Teilnahme an einer Online-Umfrage zu erhöhen, können Incentives zum Einsatz kommen. Allerdings sollten diese nicht das zentrale Motiv sein, sondern eine Aufwandsentschädigung darstellen. Der sinnvolle Einsatz von Incentives knüpft diese an die formal richtige Beantwortung des Fragebogens. Die richtige Anwendung von Incentives führt zu keiner Verzerrung im Sinne eines Volunteer Bias. [Theo00, S. 73]

¹⁰ Der Fragebogen kann unter www.wi-if.de > Wissenschaft > Publikationen heruntergeladen werden.

¹¹ Teil D enthielt die Fragen für außerbörslich erfahrene Nutzer - so genannte außerbörsliche Heavytrader. Diese haben für den vorliegenden Beitrag keine Relevanz.

fünf Fragen gestellt. Diese waren als Pflichtfragen gestellt. Teil C enthielt Fragen zur Demografie, welche von allen Auskunftspersonen beantwortet wurden. Hierbei wurde größtenteils auf Fragen und Skalen zurückgegriffen, welche bereits in anderen Studien [Glas03, NFO03] Anwendung gefunden haben. Die Feldzeit des Fragebogens betrug 4 Wochen im Sommer 2005.

4.2 Multiple Regressionsanalyse mit Interaktionseffekt

Das Modell wird simultan, mittels einer multiplen Regressionsanalyse mit Interaktionseffekt getestet. Die Konstrukte werden als direkte Einflussgrößen auf das Handelsverhalten angenommen. Die Variable „Tradeanzahl“ wird als moderierende Einflussgröße auf die „Bedeutung der expliziten Transaktionskosten“ aufgenommen. Für das in Abschnitt 3 vorgestellte Modell wird folgende Regressionsgleichung modelliert:

$$Y_{NAH} = \beta_0 + \beta_1 X_{BTAK} + \delta_1 X_{BTAK} X_{TA} + \beta_2 X_{BH} + \beta_3 X_{BVT} + \beta_4 X_{IHP} + \beta_5 X_{BIHP} + \beta_6 X_{BT} \quad (1)$$

Tabelle 1 gibt einen Überblick über die verwendeten Variablen.

Y_{NAH}	Neigung außerbörslich zu handeln	Abhängige Variable
X_{BTAK}	Bedeutung der expliziten Transaktionskosten	Unabhängige Variable, direkter Einfluss
X_{TA}	Tradeanzahl	Unabhängige Variable, moderierender Einfluss auf X_{BTAK}
X_{BH}	Bedeutung der Handelszeit	Unabhängige Variable, direkter Einfluss
X_{BVT}	Bedeutung von Vertrauen	Unabhängige Variable, direkter Einfluss
X_{IHP}	Involviertheit in den Handelsprozess	Unabhängige Variable, direkter Einfluss
X_{BIHP}	Bedeutung der Individualisierung des Handelsprozesses	Unabhängige Variable, direkter Einfluss
X_{BT}	Bekanntheit	Unabhängige Variable, direkter Einfluss

Tab. 1: Verwendete Variablen

Das Variable „Tradeanzahl“ wird in eine Dummyvariable umkodiert. Als Trennwert dient der Median der getätigten Trades pro Jahr.

5 Darstellung und Analyse der Ergebnisse

5.1 Stichprobe und Rahmendaten

Insgesamt kamen 5.123 Personen auf die Startseite der Online-Umfrage. Davon begannen 2.660 Personen den Fragebogen. Die Ausschöpfungsquote liegt damit bei 52 Prozent. 1.804 Personen bearbeiteten den Fragebogen vollständig, 1.190 von diesen bearbeiteten den Fragenteil B der sich spezifisch an außerbörslich unerfahrene Online-Investoren richtete. Diese stellen die Stichprobe der Untersuchung für den vorliegenden Beitrag dar. Das vorliegende Sample wurde mit bereits durchgeführten Studien [Glas03, NFO03, Onvi04] verglichen, um zumindest Hinweise auf die Repräsentativität zu erhalten. Es zeigt sich, dass das Sample im Durchschnitt etwas mehr Männer enthält (92% vs. 85%) sowie etwas risikofreudiger und älter (44 vs. 37 Jahre) als der Online-Investor in der Studie von NFO Infratest ist. [NFO03] Dies lässt eine vergleichsweise etwas höhere Trading-Orientierung des Samples vermuten. Dies könnte durch den Umstand erklärt werden, dass auf Grund des Untersuchungsdesigns natürlich nur zum Befragungszeitpunkt in irgendeiner Form finanzwirtschaftlich aktive Online-Investoren an der Umfrage teilgenommen haben, während in [NFO03] die Rekrutierung aus einem vorhandenen Panel erfolgt.

5.2 Ergebnisse der Regressionsanalyse

Um zu überprüfen, ob die Items pro Konstrukt zu einem gemeinsamen Wert zusammengefasst werden können, wurde mittels Reliabilitätsanalyse jeweils der Wert für Cronbach's Alpha bestimmt. Zudem wurden einige Items rekodiert, um Vergleichbarkeit mit den anderen Items herzustellen. Die Anpassungen der Items und die jeweiligen Werte für Cronbach's Alpha pro Konstrukt sind in folgender Tabelle dargestellt.

Konstrukt	Anpassungen	Cronbach's Alpha (vor Eliminierung)
Y_{NAH}	Item 3 (B8c) rekodiert; Eliminierung Item 3 (B8c)	0,817 (0,758)
X_{BTAK}	Item 4 (B1d) rekodiert	0,669
X_{BH}	Item 3 (B3c) rekodiert	0,692
X_{BVT}	Item 2 (B2b) rekodiert	0,655
X_{IHP}	Eliminierung Item 2 (B6b)	0,711 (0,688)
X_{BIHP}	Item 3 (B5c) rekodiert; Eliminierung Item 4 (B5d)	0,850 (0,755)
X_{BT}	Eliminierung Item 3 (B4b)	0,817 (0,758)

Tab. 2: Cronbach's Alpha für die sieben Konstrukte

Da alle Werte für Cronbach's Alpha (zum Teil deutlich) größer 0,6 sind und es sich um neu entwickelte Konstruktfragen handelt, sind die Mindestanforderungen erfüllt. [Nunn67, S. 226]

Die Ergebnisse der Regression sind in Tabelle 3 zusammengefasst.

Hypothese	Parameter	Zugehörige Variable	Wert	Standardabweichung	t-Wert	Signifikanz
	β_0	Konstante	1,153	0,181	6,370	< 1%
H1	β_1	X _{BTAK}	0,124	0,024	5,082	< 1%
H2	δ_1	X _{TA}	-0,020	0,012	-1,599	nicht signifikant
H3	β_2	X _{BH}	0,199	0,199	0,023	< 1%
H4	β_3	X _{BVT}	-0,180	0,030	-5,955	< 1%
H5	β_4	X _{IHP}	0,063	0,025	2,476	< 5%
H6	β_5	X _{BIHP}	0,022	0,024	0,925	nicht signifikant
H7	β_6	X _{BT}	0,323	0,019	17,426	< 1%

Tab. 3: Ergebnisse für die Parameter der Regressionsanalyse mit Interaktionseffekt

Der Parameterwert β_0 für die Konstante beträgt 1,153 und ist zum 1%-Niveau signifikant. Dies deutet auf Einflussfaktoren hin, die im zugrunde gelegten Modell nicht berücksichtigt wurden. Der Parameterwert β_1 , der den Einfluss der expliziten Transaktionskosten auf die Neigung außerbörslich zu handeln beschreibt und wegen des positiven Vorzeichens einen positiven Zusammenhang zwischen den beiden Größen postuliert, ist zum 1%-Niveau signifikant. Die Hypothese H1 kann demnach nicht verworfen werden. Demgegenüber muss Hypothese H2 wegen des Parameterwertes δ_1 nahe 0 sowie wegen fehlender Signifikanz verworfen werden. Der Hypothese, dass eine hohe Tradeanzahl einen positiven Einfluss auf die Bedeutung der expliziten Transaktionskosten hat, kann also nicht zugestimmt werden. Der Parameterwert β_2 für den Einfluss der Bedeutung der Handelzeit auf die Neigung außerbörslich zu handeln, ist mit 0,199 deutlich positiv und kann zum 1%-Niveau bestätigt werden. Somit kann Hypothese H3 nicht verworfen werden. Der Parameterwert für β_3 von -0,180 ist deutlich negativ. Daher kann Hypothese H4 ebenfalls zum 1%-Niveau bestätigt werden, welche einen negativen Einfluss der Bedeutung des Vertrauens auf die Neigung außerbörslich zu handeln postuliert. Die Hypothesen H5 und H6 müssen beide verworfen werden. Im Fall von H5 liegt zwar ein signifikanter Wert zum 5%-Niveau vor, dieser liegt mit 0,063 aber nahe 0 und drückt damit keinen positiven Zusammenhang zwischen der Involviertheit in den Handelsprozesses und der Neigung außerbörslich zu handeln aus. Im Fall von H6 ist der - ebenfalls nahe an Null liegende - ermittelte Wert nicht signifikant. H7 schließlich kann wieder zum 1%-Niveau bestätigt werden. Für r^2 ergibt

sich ein Wert von 0,348. Das getestete Modell erklärt damit knapp 35 Prozent der Varianz und ist somit als mittelmäßig einzustufen.

Hypothese		Ergebnis
H1	<i>Je höher die Bedeutung der expliziten Transaktionskosten, desto höher ist die Neigung außerbörslich zu handeln.</i>	☑
H2	<i>Je höher die Tradeanzahl, desto höher ist die Bedeutung der expliziten Transaktionskosten.</i>	✗
H3	<i>Je höher die Bedeutung der Handelszeit, desto höher ist die Neigung außerbörslich zu handeln.</i>	☑
H4	<i>Je höher die Bedeutung von Vertrauen, desto geringer ist die Neigung außerbörslich zu handeln.</i>	☑
H5	<i>Je höher die Involviertheit in den Handelsprozess, desto höher ist die Neigung außerbörslich zu handeln.</i>	✗
H6	<i>Je höher die Bedeutung der Individualisierung des Handelsprozesses, desto niedriger ist die Neigung außerbörslich zu handeln.</i>	✗
H7	<i>Je höher die Bekanntheit des Angebots außerbörslich zu handeln, desto höher ist die Neigung außerbörslich zu handeln.</i>	☑

Tab. 4: Übersicht über die Annahme bzw. Verwerfung der Hypothesen

5.3 Interpretation

Bisherige theoretische Überlegungen gehen davon aus, dass explizite Transaktionskosten bei der Handelsplatzwahl besondere Beachtung finden. [MeWi94] Das Ergebnis der Regressionsanalyse stützt die Hypothese, dass Online-Investoren, die den Transaktionskosten eine besondere Bedeutung beimessen, eher zum außerbörslichen Handel tendieren. Der Einfluss der „Bedeutung der expliziten Transaktionskosten“ auf die „Neigung außerbörslich zu handeln“ scheint auch aus Praxisgesichtspunkten plausibel, da geringere Gebühren eines der zentralen Marketingargumente der Online-Broker für den außerbörslichen Emittentenhandel sind. Daneben werden auch ganz gezielt No-Fee-Aktionen angeboten, die zum außerbörslichen Handel animieren sollen. Die Börsen haben darauf unter anderem mit einem Courtage-Cap reagiert.

Der moderierende Effekt, der von einem positiven Zusammenhang zwischen der „Tradeanzahl“ und der „Bedeutung der expliziten Transaktionskosten“ ausgeht, musste abgelehnt werden. Es lassen sich durchaus Argumente gegen die Hypothese finden: Zum Beispiel ist denkbar, dass Online-Investoren im Allgemeinen sehr preissensitiv und die expliziten Transaktionskosten daher für alle gleichermaßen wichtig sind.

Schierreck leitet die Bedeutung der Handelszeit für die Handelsplatzwahl theoretisch ab. [Schi95] Die Hypothese bzgl. der Bedeutung der Handelszeit musste nicht verworfen werden. Damit wird die in der Theorie zur Handelsplatzwahl angenommene Bedeutung der Handelszeit

auch für den außerbörslichen Emittentenhandel bestätigt. Die Möglichkeit länger handeln zu können, wird von Online-Investoren als wichtiger Grund für den außerbörslichen Emittentenhandel gesehen.

Der Zusammenhang, dass eine hohe „Bedeutung von Vertrauen“ mit einer geringen „Neigung außerbörslich zu handeln“ einhergeht ist zum 0,01%-Niveau signifikant. Das bedeutet, dass Online-Investoren, denen der Sicherheits- bzw. Vertrauensaspekt besonders wichtig ist, eher Abstand vom außerbörslichen Emittentenhandel nehmen. [BaKM03] beschreiben das Vertrauensverhältnis zwischen Online-Investor und Online-Broker bzw. zwischen Online-Investor und Handelsplatz. Dieses Vertrauensverhältnis scheint auch beim außerbörslichen Emittentenhandel eine Rolle zu spielen. Börsen könnten dies im Wettbewerb für sich nutzen und noch stärker auf ihre Vorteile bezüglich Sicherheit und Regulierung hinweisen.

Die „Involviertheit in den Handelsprozess“ hat keinen signifikanten positiven Einfluss auf die „Neigung außerbörslich zu handeln“. Ein möglicher Grund für die Ablehnung der Hypothese könnte sein, dass bei der Operationalisierung des Konstruktes verschiedene, (zu) unterschiedliche Aspekte einbezogen wurden. So wurde zum einen die schnelle Depotverbuchung thematisiert. Zum anderen wurde der Aspekt der direkten Interaktion mit einem Handelspartner berücksichtigt. Die Zusammenfassung verschiedener Aspekte unter dieses Konstrukt könnte somit für die Ablehnung der Hypothese verantwortlich sein.

Die „Bedeutung der Individualisierung des Handelsprozesses“ hat keinen signifikanten negativen Einfluss auf die „Neigung außerbörslich zu handeln“. Dies könnte dadurch begründet sein, dass Online-Investoren, denen Orderzusätze wichtig sind, auch weiterhin den börslichen Handel nutzen und den außerbörslichen Emittentenhandel nur dann in Anspruch nehmen, wenn sie ohne Orderzusätze handeln wollen. Ob eine situationsabhängige Nutzung der Handelsplätze vorliegt, wurde im Rahmen der Befragung nicht analysiert.

Wie in vielen Beiträgen der Marketing-Literatur besprochen, hat die Bekanntheit eines Angebots einen positiven Einfluss auf die Nutzung dieses Angebots. [HoBr90] Dies gilt zunächst einmal unabhängig von der Angebotsqualität. Mit der durchgeführten Regression wurde dieser Zusammenhang auch für den außerbörslichen Emittentenhandel bestätigt. Je bekannter der außerbörsliche Emittentenhandel bei Online-Investoren ist, desto eher wird er genutzt.

Die Modellgüte kann als mittelmäßig bezeichnet werden. Es erscheint plausibel anzunehmen, dass weitere Faktoren einen Einfluss auf die Neigung außerbörslich zu handeln haben. Darauf deutet auch der relativ hohe Wert der Konstanten der Regressionsgleichung hin. Mögliche Ein-

flussfaktoren sind die in Abschnitt 3.2 vorgestellten, nicht berücksichtigten Faktoren, welche bei institutionellen Investoren teilweise eine große Bedeutung – bspw. die Liquidität [Schi95] – haben. Eine zusätzliche Erklärung für die nicht zufrieden stellende Modellgüte liefert eine weitergehende Analyse der gewonnenen Daten: Hierbei zeigt sich, dass Online-Investoren dazu neigen, speziell beim Handel von Hebelprodukten *außerbörslich zu kaufen*, um dann – möglicherweise unter Nutzung von entsprechenden Orderzusätzen (bspw. Stop-Loss) – über die *Börse zu verkaufen*. Besonders deutlich wird dies bei außerbörslich Erfahrenen. Diese wählen für Knock-out-Produkte bei Käufen zu 80% hauptsächlich den außerbörslichen Handelsweg, während es beim Verkauf derselben Produkte nur 73% sind. Eine solch situative Nutzung der Handelsplätze ist im vorgestellten Modell nicht abgebildet. Zudem könnte auch die „Preissicherheit“ eine bedeutende Rolle spielen. In einer direkten Frage wurde diese als dritt wichtigster Hauptgrund für die Nutzung des außerbörslichen Emittentenhandels genannt. Dies mag auch ein Indiz für die Bedeutung des Marktmodells bei der Handelsplatzwahl sein.

6 Zusammenfassung und Ausblick

Der vorliegende Artikel liefert einen Beitrag zur Erklärung der Handelsplatzwahl von Online-Investoren im Hinblick auf den außerbörslichen Emittentenhandel. Zur Beantwortung der Forschungsfrage wurden unter Verwendung bisheriger Theorie und bereits vorliegender empirischer Ergebnisse zur Handelsplatzwahl sechs direkte Einflussfaktoren sowie ein moderierender Effekt vermutet. Die Daten für den Test des unterstellten Modells wurden mittels einer Online-Umfrage erhoben, bei der 1.190 Personen die Fragen zu den Konstrukten beantworteten. Vier der sieben aufgestellten Hypothesen konnten mittels einer multiplen Regressionsanalyse mit Interaktionseffekt gestützt werden. Demnach haben die „Bedeutung der expliziten Transaktionskosten“, die „Bedeutung der Handelszeit“, die „Bedeutung von Vertrauen“ und die „Bekanntheit“ einen Einfluss auf die Neigung von Online-Investoren außerbörslich zu handeln. Das getestete Modell kann 35 Prozent der Varianz erklären und ist damit als mittelmäßig einzustufen. Es ist anzunehmen, dass neben den modellierten Einflussfaktoren auch noch andere, nicht berücksichtigte Einflussfaktoren eine Rolle spielen (vgl. Abschnitte 3.2 und 5.3).

Neben der Herausarbeitung weiterer Einflussfaktoren auf die Neigung von Online-Investoren außerbörslich zu handeln, könnten zukünftige Studien das gesamte Spektrum der angebotenen Handelsplätze - inklusive ausländischer Handelsplätze - für Online-Investoren untersuchen. Vor

allem erscheint interessant, wie sich Online-Investoren situativ für einen Handelsplatz entscheiden (bspw. in Abhängigkeit von Kauf/Verkauf oder des gehandelten Produkts). Kann dies auch Auswirkungen auf das Marktergebnis haben? Welche Rolle spielt das Marktmodell bei der Handelsplatzwahl? Um einen Gestaltungsbeitrag im Sinne eines Financial Market Engineering [Holt04, WeHN03] leisten zu können, sind daher weitere Untersuchungen vorzunehmen.

Literaturverzeichnis

- [ADM01] A.D.M.: Standards zur Qualitätssicherung für Online-Befragungen. <http://www.adm-ev.de/>, 2001, Abruf am 17.02.2005.
- [Aver98] *Averdiek-Bolwin, C.*: Die Effizienz von Aktienbörsen. Oldenbourg, München, 1998.
- [BaKM03] *Balasubramanian, S., Konana, P., Menon, N.M.*: Customer Satisfaction in Virtual Environments: A study of Online Investing. In: Management Science 49 (2003) 7, S. 871-889.
- [BaOd02] *Barber, B.M., Odean, T.*: Online Investors: Do the Slow Die First?. In: Review of Financial Studies 15 (2002) 2, S. 455-487.
- [Euwa05] Euwax: Geschäftsbericht 2005.
- [Gerk93] *Gerke, W.*: Computerbörse für den Finanzplatz Deutschland. In: Die Betriebswirtschaft 53 (1993) 6, S. 725-748.
- [Glas03] *Glaser, M.*: Online Broker Investors: Demographic Information, Investment Strategy, Portfolio Positions, and Trading Activity. Working Paper, University of Mannheim, 2003.
- [Gomb00] *Gomber, P.*: Elektronische Handelssysteme – Innovative Konzepte und Technologien im Wertpapierhandel. Physica, Heidelberg, 2000.
- [Heil02] *Heilmann, K.*: Erfolgsfaktoren von Wertpapierbörsen im internationalen Wettbewerb Analyse auf der Basis einer weltweiten empirischen Erhebung. Deutscher Universitäts-Verlag, Wiesbaden, 2002.
- [HiSt96] *Hielscher, U., Stenzel, S.K.*: Umfang und Ursachen des außerbörslichen Aktienhandels. In: Die Betriebswirtschaft 56 (1996) 4, S. 523-540.

- [Holt04] *Holtmann, C.:* Organisation von Märkten - Market Engineering für den elektronischen Wertpapierhandel. Dissertation, Fakultät Wirtschaftswissenschaften, Universität Karlsruhe (TH), (2004).
- [HoBr90] *Hoyer, W.D., Brown, S.P.:* Effects of Brand Awareness on Choice for a Common, Repeat Purchase Product. In: Journal of Consumer Research 17 (1990), S. 141-148.
- [KuKr06] *Kundisch, D., Krammer, A.:* Transaktionshäufigkeit als Indikator für die Angebotsgestaltung bei deutschen Online-Brokern. In: Der Markt 46 (2006) 176, S. 20-38.
- [MaSh00] *Macdonald, E.K., Sharp, B.M.:* Brand Awareness Effects on Consumer Decision Making for a Common, Repeat Purchase Product: A Replication. In: Journal of Business Research 48 (2000) 1, S. 5-15.
- [MeWi94] *Meyer, F., Wittrock, C.:* Marketing-Strategien für die deutschen Börsen. In: Zeitschrift für das gesamte Kreditwesen 47 (1994) 11, S. 536-542.
- [NFO03] NFO Infratest Finanzforschung: NFO Online Brokerage Monitor Germany 2003.
- [Nunn67] *Nunnally, J.C.:* Psychometric theory. McGraw-Hill, New York, 1967.
- [Onvi04] *OnVista:* OnVista Derivate-Umfrage 2004, Ergebnisfoliensatz.
- [PiBR96] *Picot, A., Bortenlänger C., Röhr, H.:* Börsen im Wandel, Fritz Knapp, Frankfurt am Main, 1996.
- [Schi95] *Schiereck, D.:* Internationale Börsenplatzentscheidungen institutioneller Investoren. Gabler, Wiesbaden, 1995.
- [StRa04] *Strader, T.J., Ramaswami, S.N.:* Investor Perceptions of Traditional and Online Channels. In: Communications of the ACM 47 (2004) 7, S. 73-76.
- [Thei98] *Theissen, E.:* Organisationsformen des Wertpapierhandels: Gesamtkursermittlung, kontinuierliche Auktionen und Market-Maker-System. Gabler, Wiesbaden, 1998.
- [Theo00] *Theobald, A.:* Determinanten des Online Research. In: Planung & Analyse, 27 (2000) 5, S. 72-76.
- [WeHN03] *Weinhardt, C., Holtmann, C., Neumann, D.:* Market-Engineering. In: Wirtschaftsinformatik 45 (2003) 6, S. 635-640.

Travo-Projekt – Der Handel mit Aktionärsstimmrechten

Christoph Lattemann, Hartmut Klemens

Juniorprofessur für Corporate Governance und E-Commerce
Universität Potsdam
14482 Potsdam
lattema@uni-potsdam.de, hartmut.klemens@klemens.com

Michael Durica

Eurex Frankfurt AG
60485 Frankfurt am Main
mdurica@web.de

Abstract

In den letzten Jahren waren weit unter 50% der stimmberechtigten Anteilscheine auf den Hauptversammlungen der DAX-30-Unternehmen vertreten. Dies wirkt sich negativ auf Corporate Governance und Unternehmenskontrolleffizienz aus.

Das TRAVO-Projekt (*Trading with Voting Rights*) stellt einen innovativen Ansatz dar, dem fallenden Trend zur Stimmrechtsausübung entgegenzuwirken. Er fokussiert nicht wie die herkömmlichen Ansätze auf die Reduktion der Transaktionskosten, sondern auf die effiziente Allokation von Stimmrechten über Marktmechanismen.

Die vorliegende Arbeit hat zum Ziel, die Problematik, die mit einem Handel mit Stimmrechten einhergeht, darzustellen und ein an die Marktbedürfnisse angepasstes Marktmodell aufzuzeigen. Sie ist Teil eines Forschungsprojektes, das in weiteren Stufen die prototypische Implementierung des Marktmodells vorsieht, um hiermit experimentelle Untersuchungen zum Beitrag des Stimmrechtehandels zu guter Corporate Governance durchzuführen.

1 Einleitung

Nach Erhebungen der Deutsche Schutzvereinigung für Wertpapierbesitz [DSW05] waren in den letzten Jahren bei fallender Tendenz weit unter 50% der stimmberechtigten Anteilscheine auf den Hauptversammlungen der DAX-30-Unternehmen vertreten. 26% aller Aktionäre, zumeist

Kleinaktionäre, zeigen kein Interesse an der Ausübung ihrer Stimmrechte auf Hauptversammlungen [ErGP05]. Diese schwache Präsenz auf den Hauptversammlungen liegt darin begründet, dass Anleger die Rolle des Trittbrettfahrers [EaFi83, GrHa88a] einnehmen oder vor dem Hintergrund der Rationalen Apathie keinen ökonomischen Anreiz darin sehen, ihr Stimmrecht auf der Hauptversammlung auszuüben [Körb89].

Durch die Nichtteilnahme an Hauptversammlungen ergeben sich Probleme: Erstens führt die schwache Präsenz des stimmberechtigten Kapitals zur Gefahr der Konzentration der Entscheidungsmacht auf der Hauptversammlung, da zum Beispiel die Gesellschaft mit vergleichsweise geringem Aufwand in Abhängigkeiten gebracht kann [Noac05]. Zweitens verfällt bei Nichtausübung des Stimmrechts der Wert des Stimmrechts, und damit erfolgt die Wertvernichtung eines ökonomischen Gutes.

Zur Darstellung des Stimmrechtehandels werden im folgenden Literaturüberblick zunächst die zwei Ansätze zur Lösung dieser Probleme beleuchtet und es wird der Stand der aktuellen Diskussion um die Stimmrechtsausübung vorgestellt. Überlegungen zur Identifikation der Anbieter- und Nachfragerseite und zum Potential eines Marktes für Aktienstimmrechte leiten nachfolgend über zur Skizzierung eines möglichen Marktmodells für den elektronischen Handel mit Aktienstimmrechten. Es wird ein innovatives funktionales Konzept für den Stimmrechtehandel dargestellt, das die Basis für eine marktliche Koordination von Stimmrechtsüberhängen und -nachfragen vorsieht. Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick ab.

2 Literaturüberblick

Eine Vielzahl wissenschaftlicher wie praxisorientierter Arbeiten hat sich in den letzten Jahren mit der Thematik der Erhöhung der Stimmrechtspräsenz in den Hauptversammlungen beschäftigt. In der Literatur zeichnen sich als Lösungsansätze zur Erhöhung der Präsenzen zwei sich nicht zwingend gegenüberstehende Ansätze ab. Der erste Forschungszweig untersucht die Möglichkeiten zur Reduktion von Transaktionskosten bei der Ausübung von Stimmrechten in Hauptversammlungen (z.B. durch die Entbindung des Aktionärs von der Anwesenheit in der Hauptversammlung mit Hilfe von Proxy Votings und der Elektronisierung der Hauptversammlung [Bhat97; LaNi03; Noac05; Zetz02]) oder etwa durch die Etablierung von Bonussystemen [Noac05; Perl06]. Hierdurch soll die Präsenz des Abstimmungskapitals erhöht werden. Diese Maßnahmen werden hier unter dem *Kostenansatz* subsumiert. Obwohl vereinzelt durch solche

Maßnahmen eine Zunahme der Präsenz zu verzeichnen ist [FAZ06], zeigen breiter angelegte empirische Studien, dass solche Maßnahmen nur von geringer nachhaltiger Wirkung sind [Latt05; Noac05].

Der zweite - und der für diese Arbeit zentrale Forschungszweig - besteht darin, Aktie und Stimmrecht zu trennen und die Stimmrechte über einen Markt in die Hände der an der Unternehmenskontrolle interessierten Marktteilnehmer zu übertragen (Reallokation). Er wird in dieser Arbeit als *Handelsansatz* bezeichnet. Entsprechend der Ergebnisse verwandter Arbeiten (z.B. von [BIGG89; Mann75; Schü79; NeOr03]) können neben der Erhöhung der Stimmrechtspräsenz darüber hinaus auch Ineffizienzen im bisher existierenden Allokationsmechanismus überwunden und die Effizienz der Unternehmenskontrolle verbessert werden.

Grundlage aller Diskurse über die Reduktion der Transaktionskosten wie auch zur Etablierung eines Stimmrechtshandels stellen theoretische wie empirische Analysen zum Wert von Stimmrechten dar. Mit der Werthaltigkeit von Stimmrechten besteht ein ökonomisches Gut, das durch Nichtausübung qua definitio zu Marktineffizienzen führt, da der Wert bei Nichtausübung verfällt. So weisen aktuellere Analysen (u.a. von [DaEh02; ErGP05; Nico98; Rydq96]) einen ökonomischen Wert des Stimmrechts nach. In ihren Erklärungsansätzen zum Wert eines Stimmrechts gehen die Autoren von der Annahme aus, dass das Management zum Beispiel durch feindliche Übernahmen und damit über die Ausübung von Stimmrechten diszipliniert werden kann (u.a. [KiKM04; Nico98; Rydq96; Zing95]). Nach [Doer92] kann sogar bereits die Gefahr bzw. die Drohung einer Übernahme bei schlechter Performance des Unternehmens das Management disziplinieren. Unterschiedliche Stimmrechtsausstattungen bei den Aktien eines Unternehmens (z.B. bei Stammaktien, Vorzugsaktien) führen demnach zu unterschiedlichen Kurssteigerungen, wobei die Aktien mit Stimmrechten aufgrund des intensiven Bieterwettbewerbs hohe Aufschläge verzeichnen und Aktien ohne Stimmrechte im Gegensatz dazu von den Bietern nicht nachgefragt werden (u.a. [Rydq96]).

Diverse Autoren (u.a. [DaEh02; Doer92; Hoff99; RoMK94]) haben empirische Analysen zum Wert des Stimmrechts und dessen Einflussfaktoren durchgeführt. [DeDe85] sowie [Zing95] weisen zum Beispiel nach, dass die Werte von Stimmrechten bei feindlichen Übernahmen steigen. Ausgangspunkt der Überlegungen sind die in der Realität zu beobachtenden Kursdifferenzen zwischen Stamm- und Vorzugsaktien. Der hieraus abzuleitende Wert des Stimmrechts wird durch mehrere Faktoren, wie z.B. feindliche Übernahme, das Rechtssystem, der damit verbundene Rechtsschutz der Kleinaktionäre [LLSV99] sowie die Höhe der Stimmrechtsanteile ver-

schiedener Großaktionäre [DaEh02] beeinflusst. Trotz der zum Teil unterschiedlichen Ergebnisse hinsichtlich der Prämien für das Stimmrecht weisen die meisten Untersuchungen einen positiven Wert des Stimmrechts auf (u.a. [Horn88]). Nach [Rydq96] liegt die Differenz zwischen Stamm- und Vorzugsaktien in Deutschland zwischen 1956 und 1998 im Mittel bei etwa 17% und damit im Rahmen der auch für andere Länder beobachteten Werte, die zwischen 5% und 23% schwanken.

Die Trennung von Aktie und Stimmrecht wird aus ökonomischer wie aus rechtlicher Sicht kontrovers diskutiert. [Mann75] spricht sich für einen Handel mit Stimmrechten aus und misst der Möglichkeit zur Kontrolle eines Unternehmens direkt einen eigenen monetären Wert zu, da hiermit gleichzeitig die Verfügungs- und Dispositionsmöglichkeiten über die Unternehmensressourcen verbunden sind. Nach [Els88] haben die Trennung der Aktie vom Stimmrecht und der separate Stimmrechtshandel aber nicht nur positive Wirkung. Der Managementkontrolle könnte mit einer solchen Maßnahme entgegen gewirkt werden und wohlfahrtserhöhende Transaktionen könnten beeinträchtigt werden. Andererseits verhindert die erzwungene Koppelung von Aktie und Stimmrecht eine wesentliche ökonomische Signalwirkung [Mann75].

[EaFi83] weisen ebenfalls als Gegner des separaten Stimmrechtshandels darauf hin, dass nur beim gekoppelten Erwerb von Aktien und Stimmrecht gewährleistet sein kann, dass die Inhaber bei der Ausübung des Stimmrechts auch die daraus resultierenden Konsequenzen für ihr Vermögen, beispielsweise in Form der Aktienkursänderungen oder Dividendenzahlungen, mittragen müssen und dass so die entschädigungslose Beeinträchtigung der Interessen anderer Aktionäre ausgeschlossen werden könnte. Unter anderen vertreten [GrHa88b] die Meinung, dass durch einen separaten Stimmrechtshandel aus der Aktie nur negative Konsequenzen für Unternehmen und Aktionäre resultieren.

Befürworter eines Stimmrechtshandels wie [BIGG89] oder [Clar79] empfehlen hingegen die Einführung gesetzlicher Regelungen zum Stimmrechtshandel, denn die Gefahr von Plünderungen ist zwar nach ihrer Meinung nicht von der Hand zu weisen, doch die Vorteile einer effizienten Allokation der Stimmrechte und einer verbesserten Unternehmenskontrolle könnten die Nachteile kompensieren. Zu den vorgeschlagenen Lösungsansätzen zum Handel mit Aktienstimmrechten gehören beispielsweise die Gleichbehandlung aller verkaufswilligen Stimmrechtsinhaber zu im Vorfeld fest vereinbarten Konditionen, inhaltliche oder zeitliche Beschränkungen bei der Ausübung des Stimmrechts, Rückkaufsmöglichkeiten, zeitlich befristete Über-

lassung des Stimmrechts im Rahmen einer Wertpapierleihe [CGMR04; Davo02] und die Einführung zusätzlicher Kontroll- und Sanktionsrechte der Aktionäre [Els88].

Im aktuellen rechtlichen Diskurs geht es vor allem um die gesetzlichen Gestaltungsoptionen zum Stimmrechtehandel, insbesondere vor dem Hintergrund, dass die gegenwärtige Rechtslage die freie und getrennte Verfügung aller Aktionärsrechte blockiert (u.a. [Els88; Noac05]).

Empirische Studien über Wertpapierleihegeschäfte zeigen auf, dass ein *indirekter* Handel der Stimmrechte über den Markt für Wertpapierleihe bereits realisiert wird [CGMR02], diesbezügliche einschränkende Regelungen gelockert wurden [CGMR04] und ein Handel der Stimmrechte in Verbindung mit einer Wertpapierleihe in Australien bereits legal praktiziert wird [HaRa88]. Umso interessanter erscheint hier der Gedanke, einen solchen *Quasi-Handel* mit einem dedizierten Marktmodell zum Handel von Aktienstimmrechten zu standardisieren.

Trotz dieser Vielzahl von verwandten Forschungsrichtungen in dem Kontext des Stimmrechtehandels existieren bisher keine Arbeiten, die die erforderlichen Rahmenbedingungen eines *standardisierten* Stimmrechtehandels und ein mögliches Marktmodell explizit aufzeigen. Aus diesem Grund wird im Folgenden nach der Beschreibung der Besonderheiten des Stimmrechtehandels ein Marktmodell abgeleitet und beschrieben, das den Anforderungen an einen Stimmrechtehandel entspricht.

3 Der Markt für Aktienstimmrechte

Ein möglicher Stimmrechtehandel wäre per se nicht nur auf professionelle Finanzakteure beschränkt. Vielmehr spielen Kleinanleger eine bedeutende Rolle im Handel mit Stimmrechten. Dies lässt sich vor allem aus der mangelnden Teilnahme dieser Investorengruppe auf Hauptversammlungen ableiten, die aus einem ungünstigen Kosten-/Nutzenverhältnis in der Ausübung der Stimmrechte resultiert. Sie stellen zusammen mit den Stimmen der rational handelnden Aktionäre, die bislang ihr Stimmrecht an Depotstimmrechtsvertreter kostenfrei abgegeben haben, und den Stimmen der „HV-Touristen“ die potentielle Angebotsseite dar (s. Abb. 1).

Um eine erste Einschätzung des realen Marktpotentials für einen Stimmrechtehandel zu erhalten, wird zunächst zur Vereinfachung auf die Unternehmen des DAX-30 und auf das Verhalten der Investoren bei Hauptversammlungen deutscher Aktiengesellschaften zurückgegriffen

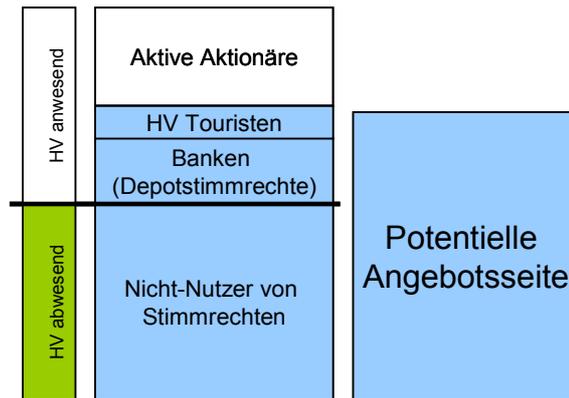


Abbildung 1: Gruppen von Stimmrechtsinhabern

Nach aktuellen Analysen des Deutschen Aktieninstituts [DAI05] treten in Deutschland rund 50% des stimmberechtigten Grundkapitals auf Hauptversammlungen auf. Werden durch einen Stimmrechtehandel lediglich 10 % (5% absolut) der bisher nicht ausgeübten Stimmrechte aktiviert, so ist für die 12 Mrd. Aktien [DeutoJ] des Deutschen DAX-30-Index mit einem Marktpotenzial von 600 Mio. gehandelten Stimmrechten zu rechnen. Über die Verrechnung des empirisch ermittelten Barwertes der Stimmrechte in Höhe von rund 20% der Unternehmenswerte [Rydq96] (bei 5% Diskontierung sind dies 1% vom Unternehmenswert für das einzelne Stimmrecht) mit der Marktkapitalisierung deutscher börsennotierter Unternehmen in Höhe von etwa 855 Mrd. Euro [DAI04] kann das jährliche Handelsvolumen allein auf dem deutschen Aktienmarkt auf etwa 400 Millionen Euro geschätzt werden. Diese Zahl würde sich deutlich erhöhen, wenn ein aktiver periodischer Stimmrechtehandel mit Weiterverkauf stattfinden würde.

Nachgefragt werden Stimmrechte bereits heute insbesondere im Rahmen feindlicher Übernahmen und Proxy Fights - was nichts anderes als eine Form des Marktes für Stimmrechte darstellt. Für diese Nachfragergruppe ist ein liquider Stimmrechtemarkt von besonderem Interesse.

4 Das Marktmodell¹ im TRAVO-Projekt

4.1 Rechtliche Aspekte

Gegenstand des Stimmrechtehandels ist die isolierte Übertragung von Stimmrechten vom Verkäufer zum Käufer. Nach aktueller Gesetzgebung ist der Handel von isolierten Stimmrechten in

¹ Nach Gomber stellt ein Marktmodell „die Gesamtheit der konkreten Ausprägungen der Strukturmerkmale einer Marktstruktur...“ dar [Gomb00; S.11].

vielen Ländern aufgrund des Abspaltungsverbots jedoch nicht möglich, und die Stimmrechtsausübung bleibt zumindest in Deutschland an den Besitz der Aktie gebunden (§12 Abs.1 AktG). Obwohl der deutsche Gesetzgeber bereits die restriktive Handhabung der Stimmrechtsabgabe bzw. -vertretung gelockert hat, ist die Ausübung des Stimmrechts nach wie vor dem Besitzer der Aktie zum Zeitpunkt des Record Date (stichtagsbezogene Hinterlegungspflicht, in der Regel ca. 7 Tage vor der HV) vorbehalten. Das Stimmrecht selbst stellt kein selbstständiges Recht dar und zählt im Gegensatz etwa zum Bezugsrechtehandel nicht zu den Wertpapieren im Sinne des § 11 BörsZulV.

Die Prämissen des Handelsansatzes können aufgrund dieser gesetzlichen Restriktionen nicht erfüllt werden. Bei der Realisierung eines Handels mit Stimmrechten nach heutigem Stand könnte in der Praxis allerdings auf Kassamarktgeschäfte mit Aktien zurückgegriffen werden. Denn es ist trotz der genannten rechtlichen Restriktionen möglich, durch einen gleichzeitig gezielten Kauf einer *Cum-Vote-Aktie* (mit Stimmrecht) und den Verkauf einer *Ex-Vote-Aktie* (*ohne Stimmrecht*) das Stimmrecht indirekt zu erwerben. Diese beiden Kassamarktgeschäfte könnten in einem Wertpapierleihegeschäft [Diwa99] gebündelt werden. Dieser Variante liegt die Idee zugrunde, den Aktienkauf vor der Ausübung des Stimmrechts mit einem festgelegten Rückkauf nach dessen Ausübung zu kombinieren, so dass die Aktie zusammen mit dem Stimmrecht am Record Date Eigentum des Käufers ist. Rechtlich ist die indirekte Realisierung eines Stimmrechtehandels auf diesen Wegen möglich. In Australien wird nach [CGMR04] ein solches Verfahren bereits indirekt (nicht über ein dediziertes Handelssystem) praktiziert. Allerdings verwehrt eine solche rein institutionelle und im OTC-Markt durchgeführte Wertpapierleihe beispielsweise den Kleinanlegern weitestgehend den Marktzugang.

4.2 Fungibilität der Stimmrechte

Ein Handel mit Stimmrechten setzt neben der Etablierung einer Marktorganisation auch die Marktfähigkeit der Stimmrechte voraus, die durch Fungibilität und Liquiditätsgrad determiniert wird [Geld00]. Da die in den Aktien einer Gesellschaft verbriefenen Stimmrechte typisiert sind und gesetzlich normierte Einwirkungsmöglichkeiten implizieren, sind diese grundsätzlich als standardisierte und handelbare Objekte anzusehen, und eine Fungibilität ist gegeben, sofern die Stimmrechte in symmetrischer Weise auf die Aktien verteilt sind [Geld00].

Dem Aktienstimmrecht kann also durchaus Fungibilität zugeschrieben werden. Gleiches gilt für die dem Aktienstimmrecht zugrunde liegende börsengehandelte Aktie.

Hinsichtlich der Liquidität eines Stimmrechtehandels führen [GeFo99] als beschränkende Faktoren den nur teilsaisonalen Charakter des Stimmrechts, eine begrenzte Nachfragermenge sowie die fehlende Kontrahentenauswahl auf. Bezüglich des teilsaisonalen Charakters und der begrenzten Nachfragermenge muss diesen Aspekten im Rahmen der phasenspezifischen Modellierung des Handels Rechnung getragen werden (s. Abschnitt 4.6). Mit der Einbindung einer Kontrahentenauswahl in das Marktmodell [Gomb00] würde neben Preis und Menge eine höhere Komplexität einhergehen, die zu erheblichen Einbußen in der Liquidität führen könnte und auch die Systemkomplexität immens steigern würde.

4.3 Die Ausprägung des Handels mit Stimmrechten

Bei Vernachlässigung rechtlicher Restriktionen stellt der Handel des isolierten Stimmrechts die effizienteste Alternative zur Realisierung eines Stimmrechtehandels dar. Dem Kassamarkt-basierten Aktienhandel kann nur eine stark eingeschränkte Eignung für einen indirekten Stimmrechtehandel zugesprochen werden, da der Handel der ganzen Aktie nicht nur mit Transaktionskosten verbunden ist, sondern der Käufer auch das Aktienkursrisiko für den Besitzzeitraum trägt. Dieses Risiko könnte allerdings durch die Minimierung der Haltedauer reduziert sowie über den Terminmarkt abgesichert werden [CGMR04].

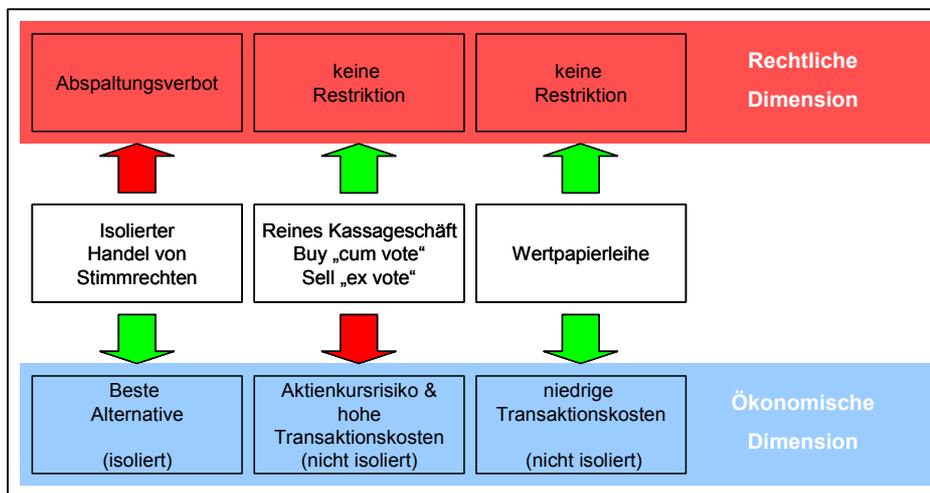


Abb. 2: Rechtliche und ökonomische Dimension eines Stimmrechtehandels

Abbildung 2 stellt die drei grundsätzlich möglichen Formen des direkten und indirekten Handels mit Stimmrechten dar, die sich aus dem Spannungsfeld zwischen rechtlichen und ökonomischen Rahmenbedingungen ergeben. Bei Berücksichtigung der rechtlichen Restriktionen kann die Wertpapierleihe zur *First-Best-Lösung* für den Stimmrechtehandel werden: Die bei einer Umsetzung über den Kassamarkt notwendigen Hedgegeschäfte entfallen, und das Marktrisiko

wird aufgrund des festgelegten Verkaufspreises eliminiert [CGMR04]. Eine solche Lösung muss dann im Rahmen der Standardisierung eines neuen Wertpapierleihegeschäfts voraussetzen, dass die Aktie mitsamt ihrer Vermögens- und Verwaltungsrechte an den Stimmrechtskäufer übertragen wird und die anfallenden Erträge aus den Vermögensrechten (z.B. Dividenden und Bezugsrechte) an den Verkäufer der Stimmrechte fließen. Nach Ablauf der Leihe wird die Aktie zurück übertragen. So kann das Ziel eines bestmöglichen isolierten Stimmrechtehandels unter Ausschluss etwaiger Preisverzerrungen realisiert werden (s. Abb. 3).

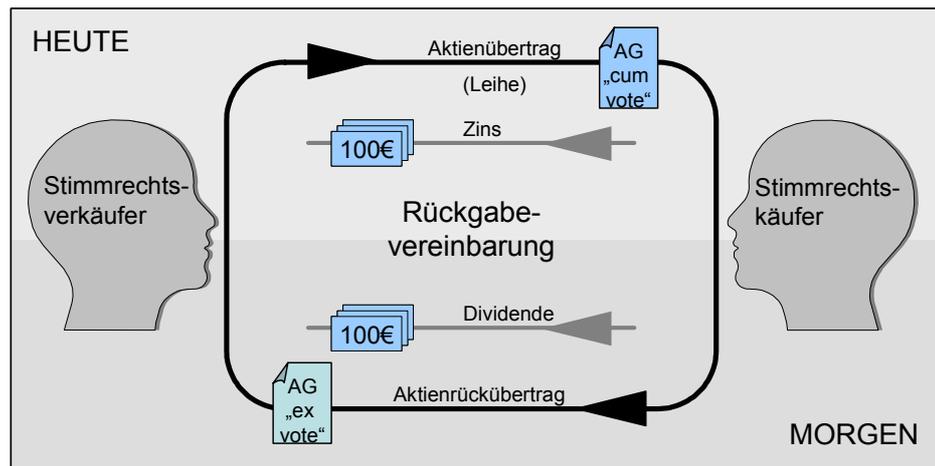


Abb. 3: Stimmrechtehandel basierend auf der Wertpapierleihe

Im vorgestellten Modell wird mit Geschäftsabschluss der Stimmrechtspreis in Form der Leihgebühr für den vereinbarten Zeitraum fällig. Zum Laufzeitende der Wertpapierleihe übereignet der Stimmrechtskäufer als derzeitiger Aktienbesitzer nach Ausübung seines Stimmrechts die Aktien in gleicher Qualität an den Stimmrechteverkäufer zurück. Der Zeithorizont der Wertpapierleihe muss sich hierbei nach der minimal notwendigen Haltefrist richten, die es erlaubt, dass der Stimmrechtskäufer zur Hauptversammlung stimmberechtigt ist.

Die Laufzeiten entsprechender Kontrakte werden dann im Rahmen der regulatorischen Vorgaben standardisiert. In der Regel ist bei Publikumsaktiengesellschaften eine satzungsmäßige Hinterlegung von Aktien üblich [Büsc01]; eine Hinterlegungsfrist wird zumeist auf eine Woche vor dem Tag der Hauptversammlung festgesetzt. Die Hauptversammlung wird in der Regel 5 bis 6 Wochen vor dem geplanten Versammlungstermin einberufen. Nach den derzeitigen Usancen müsste ein Handel des Stimmrechts aus organisatorischen Gründen spätestens am letzten Hinterlegungstag enden, d.h. in der Regel fünf Werkstage vor der Hauptversammlung. Unter Berücksichtigung eines Abwicklungszeitraums zur Gewährleistung des Eigentumsübergangs an

der Aktie muss der Handel mit Stimmrechten um weitere zwei Tage verkürzt werden und sieben Werktagen vor der Hauptversammlung enden (s. Abb. 4).

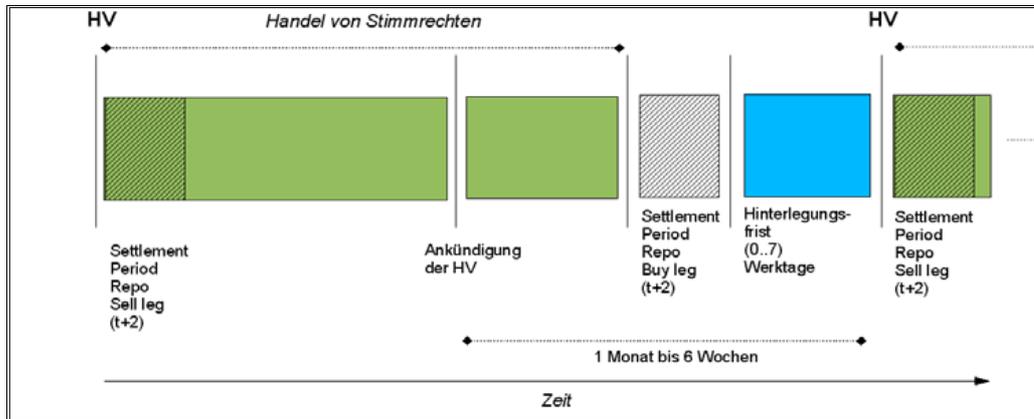


Abb. 4: Zeitliche Handelbarkeit des auf der Wertpapierleihe basierenden Stimmrechts

4.4 Orderbuch

Abbildung 5 zeigt den möglichen Aufbau eines Orderbuches für den Handel mit Stimmrechten. Die Darstellung orientiert sich an herkömmlichen Handelssystemen wie dem Kassamarktsystem Xetra oder dem Derivatehandelssystem Eurex der Deutsche Börse AG.

Contract	LALV	3206	UNIT	100	days	27									
LstPrc	0.57	LstQty	10	NetChg	-0.01	High	0.58	Low	0.56	Volume	20,000	Phase	TRAD	Xchg	ESE
BidAvg	0.5543	BidAcc	46	BidQty	46	Bid	0.5543	Ask	0.5700	AskQty	38	AskAcc	38	AskAvg	0.5700
	0.5468	174	128	0.5441	0.5802	141	179	0.5780							
	0.5424	265	91	0.5339	0.5904	78	257	0.5818							
	0.5343	467	202	0.5237	0.6006	145	402	0.5886							
	0.5318	530	63	0.5135	0.6108	143	545	0.5944							
	0.5242	722	192	0.5033	0.6210	112	657	0.5989							
	0.5200	835	113	0.4931	0.6312	96	753	0.6031							
	0.5161	933	98	0.4829	0.6414	162	915	0.6098							
	0.5083	1,137	204	0.4727	0.6516	132	1,047	0.6151							
	0.5022	1,314	177	0.4625	0.6618	147	1,194	0.6209							

Abbildung 5: Beispiel Handelsbildschirm

In dem dargestellten Fall wird für eine Aktie der Allianz AG (WP-Kürzel ALV) das Stimmrecht LALV (Lending ALV) angeboten. Dieses Stimmrecht hat eine Laufzeit „3206“, d.h. in der 32. Kalenderwoche des Jahres 2006 findet eine Hauptversammlung statt, und der Rückübertrag der Aktie an den Stimmrechtsverkäufer erfolgt unmittelbar nach Abschluss der Hauptversammlung. 27 Tage ist das Handelsprodukt noch handelbar. Der aktuelle Marktpreis (LstPrc) liegt bei 0,57 €, der aktuelle Bid-Ask-Spread liegt bei 0,5543 zu 0,5700 €.

4.5 Die technologische Plattform

Bisher existiert kein Referenzhandelssystem für Stimmrechte. Eine Realisierung auf Basis der Wertpapierleihe ist durch Implementierung in gängige Kassamarkthandelssysteme jedoch durchaus denkbar.

Die aktuelle informationstechnische Unterstützung im Umfeld des Aktienstimmrechts orientiert sich am gesetzlich möglichen Rahmen und deckt lediglich die elektronische Aktienregisterführung, Investor Relations, und die Unterstützung der Hauptversammlung ab [LaNi03], so dass bspw. Live-Anbindungen an den öffentlichen Teil der Hauptversammlung, aber auch insbesondere die elektronische Erteilung von Vollmachten und Weisungen an den Stimmrechtsvertreter ermöglicht werden.

Aufgrund der mit den Geschäften einhergehenden physischen Lieferung stellen sowohl die Generierung von Lieferinstruktionen wie auch deren Weiterleitung an die entsprechenden Settlementinstitutionen über Schnittstellen integrale Bestandteile der Kassamarkthandelssysteme dar. Um auf solchen Kassamarktsystemen einen Stimmrechtehandel aufzusetzen, sollte zusätzlich zur Lieferinstruktionsgenerierung auch eine entsprechende Verwaltung von Lieferinstruktionen aus dem später folgenden Gegengeschäft der Wertpapierleihe zur Verfügung gestellt werden. Weiterhin ist diese unter ökonomischen Aspekten auch um eine Netting-Funktionalität zu erweitern, um die Anzahl der tatsächlich notwendigen physischen Lieferungen zu reduzieren. Darüber hinaus sollte das System in der Lage sein, die mit den neuen Produkten einhergehenden erweiterten Kontraktsspezifikationen abzubilden.

4.6 Handelsphasenspezifische Betrachtung

Nach [ScLi98] besteht der Handel auf elektronischen Marktplätzen grundsätzlich aus den drei Phasen Informationssuche, Geschäftsabschluss und Abwicklung der Transaktion. Die mit dem Handel mit Stimmrechten einhergehenden Besonderheiten im Stimmrechtehandel im Vergleich zum traditionellen Aktienhandel werden anhand dieser drei Phasen im Folgenden dargestellt.

4.6.1 Informationsphase

Die Gelegenheit einer effizienten Informationsbeschaffung und -auswertung über mögliche Handelsplätze, Produkte, Kontrahenten und Teilnehmer ist eine essentielle Voraussetzung für funktionierende Wertpapiermärkte und stellt somit den ersten Schritt im Marktprozess dar. Je

nach Risikoneigung des Marktteilnehmers variiert hierbei der Informationsbedarf hinsichtlich Informationstiefe und Zeitnähe [GeRa94].

Im Rahmen des elektronischen Stimmrechtehandels wird sich die Informationsbeschaffung und -auswertung der marktexogenen Seite primär auf die Beurteilung der Managementleistung des betreffenden Unternehmens konzentrieren. Sie kann dadurch gekennzeichnet sein, dass unter Inanspruchnahme externer Informationslieferanten eine bestimmte Managementleistung in Abhängigkeit zu einem mit dem Stimmrechtspreis negativ korrelierenden Index gemessen wird, der als Branchenindex gewählt wird, um eine möglichst objektive Bewertung zu gewährleisten [GeFo99]. Darüber hinaus können derartige Informationsdienstleister Informationen zu anstehenden regulären oder außerordentlichen Hauptversammlungen, zu Übernahmeangeboten sowie zu Tagesordnungspunkten kommunizieren.

4.6.2 Abschlussphase

Im Rahmen eines Stimmrechtehandels ist der Nutzen des Stimmrechts an die Ausübung und diese an die Hauptversammlung gebunden. In der Periode zwischen zwei Hauptversammlungen sollte daher eine Sofortigkeit des Stimmrechteerwerbs umso eher gewährleistet sein, je mehr man sich dem Zeitpunkt der Stimmrechtsausübung nähert. Dies liegt darin begründet, dass die Liquidität des Handels in einem spezifischen Stimmrecht dann umso unzulänglicher ist, je weiter der Zeitpunkt der nächsten Stimmrechtsausübung entfernt ist und bis zu diesem Zeitpunkt wenig Informationen über die Leistung des Managements vorliegen [Geld00].

Die Preisermittlung nach dem Auktionsprinzip ermöglicht durch die Ansammlung von Kauf- und Verkauforders und durch die Gesamtpreisermittlung nach dem Meistausführungsprinzip eine verbesserte Markttiefe, diese allerdings zulasten der sofortigen Ausführbarkeit.

4.6.3 Abwicklungsphase

Die Erfüllung der abgeschlossenen Geschäfte vollzieht sich in der Abwicklungsphase, indem die an der jeweiligen Transaktion beteiligten Parteien über die konkreten Handelsdaten informiert werden, die wechselseitigen Verpflichtungen festgestellt („Clearing“) und die Übertragung der Eigentumsrechte bzw. die Lieferung des Wertpapiers sowie die Zahlung des Kaufbetrages durchgeführt werden („Settlement“). Ein Stimmrechtskontrakt gilt im Rahmen des Settlements dann als erfüllt, wenn sowohl beide Eigentumsübertragungen als auch die während der Leihe vereinbarten monetären Zahlungen vollständig erfolgt sind [Büsc01]. Als weiteres optionales Strukturmerkmal des Marktes könnte die Anbindung an die elektronische Aktienregister-

verwaltung des Unternehmens angedacht werden. Das elektronische Aktienregister könnte dann eine Verwaltung der Stimmrechte und ein Autorisierungskonzept zur Online-Stimmabgabe vorsehen. Da die Settlementfrist einen Zeitraum darstellt, in dem beide Geschäftspartner dem Kontrahentenrisiko ausgesetzt sind [Gomb00], kann ein zentraler Kontrahent neben seiner Funktion als Clearingstelle auch für sämtliche Geschäfte die Rolle der Gegenpartei einnehmen und Handelsdaten zur Verfügung stellen. Die Absicherung des Kontrahentenrisikos erfolgt durch Sicherheitsleistungen der Marktteilnehmer, die nicht nur auf der Basis aktueller Kurse errechnet werden, sondern auch zukünftige mögliche Kursrisiken abdecken [Eure00]. Dem zentralen Kontrahenten kommt im Rahmen des Stimmrechtshandels eine bedeutende Rolle zu, da ein Ausfall des Kontrahenten den vollständigen Wertverlust des Stimmrechts aufgrund des Termincharakters zur Folge hätte.

5 Zusammenfassung und Ausblick

Die Präsenz des stimmberechtigten Kapitals auf den Hauptversammlungen deutscher DAX-30-Unternehmen nimmt kontinuierlich ab und liegt mittlerweile bei unter 50%. Dies führt einerseits zu einer Konzentration der Entscheidungsmacht auf das vertretene Kapital, und andererseits verfallen jährlich etwa 6 Mrd. Stimmrechte mit dem entsprechenden monetären Gegenwert. Es wurden zwei Ansätze vorgestellt, die dieser Entwicklung entgegen wirken können. Im Rahmen des *Kostenansatzes* wird versucht, die Transaktionskosten der Stimmrechtsabgabe zu reduzieren, was jedoch nicht erfolgversprechend ist. Mit dem Kostenansatz werden darüber hinaus keine zwangsläufig besseren Entscheidungen im Sinne einer optimalen Ressourcenallokation herbeigeführt. Der *Handelsansatz* hingegen verfolgt das Ziel, Stimmrechte frei handelbar zu machen und den Stimmrechtspreis auf einer Handelsplattform durch Angebot und Nachfrage zu bilden. Es zeichnet sich ab, dass ein beträchtliches Angebots- und Nachfragepotential besteht, das zur Erhöhung der Stimmrechtspräsenz führen kann. Der isolierte Stimmrechteerwerb kann weiterhin den Effekt einer effizienteren Ressourcenallokation zur Folge haben. Aufgrund derzeitiger rechtlicher Restriktionen kann ein solcher Stimmrechtshandel nach dem Handelsansatz jedoch nur über den Umweg der Wertpapierleihe realisiert werden. Die derzeit gängige Form der Wertpapierleihe muss hierzu modifiziert und standardisiert werden. Dies beinhaltet die Beseitigung der Zugangsbarrieren für Kleinanleger, die Produktstandardisierung auch von kleineren Kontraktgrößen sowie den Rückfluss der Erträge aus den Vermögensrechten an den Ver-

käufer der Aktie. Nur so kann unter den gegebenen rechtlichen Restriktionen über den Weg der Wertpapierleihe in Deutschland ein weitgehend isolierter Handel des Stimmrechts realisiert werden. Hierfür wurde in diesem Artikel ein im Rahmen des TRAVO-Projektes entwickeltes Marktmodell für den Stimmrechtehandel vorgestellt, das in weiteren Schritten im Rahmen einer prototypischen Umsetzung validiert werden soll. Aus Experimenten, die auf dieser Plattform durchgeführt werden, könnten Erkenntnisse über die Auswirkungen auf die Gestaltung des Stimmrechtspreises und auf den Markt der Unternehmenskontrolle im Allgemeinen gewonnen werden, die unter Umständen aufgrund von Marktfriktionen den ursprünglichen Annahmen entgegenwirken. Für Ende 2006 ist die prototypische Umsetzung der Plattform geplant, auf der im zweiten Quartal 2007 Simulationen durchgeführt werden können.

Literatur

- [AktG65] Aktiengesetz in der Fassung vom 6. September 1965, (BGBl I 1965).
- [Bhat97] Bhattacharya, U.: Communication Costs, Information Acquisition, and Voting Decisions in Proxy Contests, in: Review of Financial Studies, Vol. 10, Heft 4, 1997, S. 1065-1098.
- [BIGG89] Blair, D., Golbe, D., Gerard, J.: Unbundling the Voting Rights and Profit Claims of Common Shares, in: Journal of Political Economy, Vol. 97, Nr. 2, 1989, S. 420-443.
- [Büsc01] Büschgen, H.: Das kleine Börsenlexikon, 2. Aufl., Düsseldorf, 2001.
- [CGMR02] Christoffersen, S., Geczy, C., Musto, D., Reed, A.: The Market for Record-Date Ownership, in: Social Science Research Network, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=302522, 2002, Abruf am 03.03.2006.
- [CGMR04] Christoffersen, S., Geczy, C., Musto, D., Reed, A.: How and Why do Investors Trade Votes, and What does it Mean?, in: Scientific Series, <http://www.cirano.qc.ca/pdf/publication/2004s-23.pdf>, 2004, Abruf am 13.06.2005.

- [Clar79] Clark, R.: Vote Buying and the Corporate Law, in: Case Western Reserve Law Review, Vol. 29, 1979, S. 776-807.
- [DaEh02] Daske, S. / Ehrhardt, O.: Kursunterschiede und Renditen deutscher Stamm- und Vorzugsaktien, in: Financial Markets and Portfoliomangement, Vol. 16, 2002, S. 179-207.
- [DAI04] Deutsches Aktieninstitut: Factbook 2004, Eigenverlag, Frankfurt am Main, 2004.
- [DAI05] Deutsches Aktieninstitut: Kurzstudie 1/2005, Zahl der direkten Aktienanleger gestiegen - Zahl der Fondsanleger rückläufig, [http://www.dai.de/internet/dai/dai-2-0.nsf/LookupDL/CD16F46C0258BFC6C1256F880032785D/\\$File/DAI-KS-1-2005.pdf](http://www.dai.de/internet/dai/dai-2-0.nsf/LookupDL/CD16F46C0258BFC6C1256F880032785D/$File/DAI-KS-1-2005.pdf), 2005, Abruf am 2006-11-13.
- [Davo02] D'Avolio, G.: The Market for Borrowing Stock, in: Journal of Financial Economics, Vol. 66, Nr. 2-3, 2002, S. 271-306.
- [DeDe85] De Angelo, H. / De Angelo, L.: Managerial Ownership of Voting Rights, in: Journal of Financial Economics, Vol. 14, 1985, S. 33-69.
- [DeDe89] De Angelo, H. / De Angelo, L.: Proxy Contests and the Governance of Publicly held Corporations, in: Journal of Financial Economics, Vol. 23, 1989, S. 29-59.
- [DeutoJ] Deutsche Börse AG, <http://www.deutsche-boerse.com>, Abruf am 2006-03-27.
- [Diwa99] Diwald, H.: Zinsfutures und Zinsoptionen. Erfolgreicher Einsatz an internationalen Terminmärkten, Vahlen 1999.
- [Doer92] Doerks, W.: Der Kursunterschied zwischen Stamm- und Vorzugsaktien in der BRD: Eine empirische Untersuchung, Köln 1992.
- [DSW05] Deutsche Schutzvereinigung für Wertpapierbesitz e.V.: Hauptversammlungspräsenzen sinken weiter, http://www.dsw-info.de/Hauptversammlungspraesenzen_sin.538.0.html, 2005-07-07, Abruf am 2006-06-27.

- [EaFi83] Easterbrook, F., Fischel, D.: Voting in Corporate Law, in: Journal of Law & Economics, Vol. 26, 1983, S. 395–427.
- [Els88] Elschen, R.: Die getrennte Handelbarkeit von Stimmrechten, in: Zeitschrift für betriebswirtschaftliche Forschung, Jg. 40, Nr. 11, 1988, S. 1009-1037.
- [ErGP05] Ernst, E., Gassen, J., Pellens, B.: Verhalten und Präferenzen deutscher Aktionäre - Eine Befragung privater und institutioneller Anleger zu Informationsverhalten, Dividendenpräferenz und Wahrnehmung von Stimmrechten, in: Studien des Deutschen Aktieninstituts, Nr. 29, Frankfurt am Main 2005.
- [Eure00] Eurex Frankfurt AG, http://www.eurexchange.com/download/documents/xpand/d_xpand_200027.pdf, 2000-11, Abruf am 2006-11-13.
- [FAZ06] FAZ v. 11.5.2006, S. 25.
- [FeOb86] Fehl, U. / Oberender, P.: Unternehmensverfassung, Kapitalmarktordnung und Wettbewerb: Zum Einfluß gesellschaftsrechtlicher Dimensionen der Kapitalmarktordnung auf den Wettbewerbsprozeß, in: Leipold, H., Schüller, A. (Hrsg.): Zur Interdependenz von Unternehmens- und Wirtschaftsordnung, Stuttgart 1986, S. 137-151.
- [GeFo99] Geldmacher, D., Foit, K.: Der Markt für Stimmrechte als Instrument einer verbesserten Managementkontrolle in Publikumsaktiengesellschaften, <http://www.uni-koeln.de/wisofak/bankseminar/veroeff/mub/81/geldmacher.pdf>, Abruf am 2005-06-03.
- [Geld00] Geldmacher, D.: Marktorientierte Managerkontrolle – Stimmrechte als Kontrollinstrument, Deutscher Universitäts-Verlag Wiesbaden 2000.
- [GeRa94] Gerke, W., Rapp, H.-W.: Strukturveränderungen im internationalen Börsenwesen, in: Die Betriebswirtschaft (DBW), 1, 1994, S. 5-23.
- [Gomb00] Gomber, P.: Elektronische Handelssysteme - Innovative Konzepte und Technologien im Wertpapierhandel, Physica Verlag Heidelberg, 2000.

- [GrHa88a] Grossman, S. J., Hart, O. D.: Takeover Bids, the Free-Rider Problem and the Theory of the Corporation, in: *British Journal of Economics and Political Science* Vol.11, 1980, S. 42-64.
- [GrHa88b] Grossmann, S. J., Hart, O. D.: One Share-One Vote and the Market for Corporate Control, in: *Journal of Financial Economics*, Vol. 20, 1988, S. 175–202.
- [HaRa88] Harris, M., Raviv, A.: Corporate Governance: Voting Rights and Majority Rules, in: *Journal of Financial Economics*, Vol. 20, 1988, S. 203-235.
- [Hoff99] Hoffmann-Burchardi, U.: Corporate Governance Rules and the Value of Control – A Study of German Dual-class Shares, Discussion Paper 315, London School of Economics, Financial Markets Group, 1999.
- [Horn88] Horner, M.: The Value of the Corporate Voting Right, in: *Journal of Banking and Finance*, Vol. 12, 1988, S. 69-83.
- [KiKM04] Kini, O., Kracaw, W., Mian, S.: The Nature of Discipline by Corporate Takeovers, in: *Journal of Finance*, Vol. 59, Nr. 4, 2004, S. 1511-1552.
- [Körb89] Körber, U.: Die Stimmrechtsvertretung durch Kreditinstitute - Überlegungen zur Fortentwicklung des deutschen Rechts unter Berücksichtigung der Vorschläge der Europäischen Kommission, Berlin 1989.
- [LaNi03] Lattemann, C., Niedermeyer, P.: Elektronische Stimmrechtsausübung zur Verbesserung der Corporate Governance in Deutschland?, in: *Zeitschrift für das gesamte Kreditwesen*, Jg. 56, 2003, S. 1415-1423.
- [Latt05] Lattemann, C.: Use of ICT in Annual Shareholder Meeting and Investor Relations on Returns on the German Stock Market, in: *Corporate Reputation Review – An International Journal*, Vol. 8 No. 2, 2005, S. 110-120.
- [LLSV99] La Porta, R., Lopez-De-Silanes, F., Shleifer, A., Vishny, R.: Corporate Ownership around the World, in: *Journal of Finance*, Vol. 54, 1999, S. 471-517.

- [Mann75] Manne, H.: Some Theoretical Aspects of Share Voting, in: Manne, H. (Hrsg.): The Economics of Legal Relationships. Readings in the Theory of Property Rights, 4. Reproduktion der Ausgabe von 1975, St. Paul, 1982, S. 534-554.
- [NeOr03] Neeman, Z., Orosel, G.: On the Efficiency of Vote Buying, <http://homepage.univie.ac.at/Gerhard.Orosel/Vote-Buying-Dec03.pdf>, 2003-12, Abruf vom 2005-06-01.
- [Nico98] Nicodano, G.: Corporate Groups, Dual-class Shares and the Value of Voting Rights, in: Journal of Banking and Finance, Vol. 22, 1998, S. 1117-1137.
- [Noac05] Noack, U.: Handgeld für Teilnahme an Hauptversammlungen, im Betriebsberater Nr. 42, 2005
- [Perl06] Perlitz, J.: Präsenzboni bei Hauptversammlungen, in: Finanzplatz (5) 2006.
- [RoMK94] Rothauge, F. A., Menkhoff, L., Krahen, J. P.: Erklärt Übernahmespekulation die Preisbildung von Vorzugsaktien?, in: Die Bank, Nr. 4., 1994, S. 239-244.
- [Rydq96] Rydqvist, K.: Takeover Bids and the Relative Prices of Shares that Differ in their Voting Rights, in: Journal of Banking and Finance, Vol. 20, S. 1407-1425.
- [Schü79] Schüller, A.: Eigentumsrechte, Unternehmenskontrollen und Wettbewerbsordnung, in: ORDO, Jg. 30, 1979, S. 325-346.
- [ScLi98] Schmid, B., Lindemann, M.A.: Elements of a Reference Model for Electronic Markets, in: Sprague, E. (Hrsg.): Proceedings of the 31st Hawaii International Conference on Systems Sciences (HICSS'98). Nummer 4 (1998) S. 193-201.
- [Zetz02] Zetzsche, D. (Hrsg.): Die virtuelle Hauptversammlung: Aktionärsbeteiligung via Internet aus juristischer und betriebswirtschaftlicher Sicht mit Erfahrungsberichten, Berlin 2002.
- [Zing95] Zingales, L.: What Determines the Value of Corporate Votes?, in: Quarterly Journal of Economics, Vol. 60, 1995, S. 1047-1073.

OR-basierte Persönliche Finanzplanung

Zulässige und optimale Finanzpläne

Oliver Braun

Lehrstuhl für Informations- und Technologiemanagement
Universität des Saarlandes
66123 Saarbrücken
ob@itm.uni-sb.de

Abstract

Auf Basis von Techniken und Methoden des Operations Research (OR) wird ein Modell für das Problem der Persönlichen Finanzplanung formuliert. Dazu wird zunächst unter entscheidungsorientierter Sicht eine Zielbildung für das Problem der Persönlichen Finanzplanung vorgenommen. Neben Algorithmen zur Herstellung zulässiger Finanzpläne wird ein Goal Programming Modell der Persönlichen Finanzplanung beschrieben.

1 Einleitung

Persönliche Finanzplanung ist die gedankliche Vorbereitung zielgerichteter Entscheidungen bezüglich persönlicher finanzieller Angelegenheiten. Planung ist das Bindeglied zwischen Zielsetzung und kalkuliertem Handeln. Zielbildung, Planung und Entscheidung sind eng miteinander verknüpft. Der vorliegende Artikel soll, basierend auf Methoden des *Operations Research*, zur Diskussion einer wissenschaftlichen Auseinandersetzung mit dem Thema der *Persönlichen Finanzplanung* anregen. *Operations Research* (OR) befasst sich mit der „angemessenen Modellierung von Entscheidungsproblemen. Die Formulierung angemessener Entscheidungsmodelle für zu lösende Problemstellungen ist eine Voraussetzung für das Finden optimaler oder befriedigender Problemlösungen.“ [Zimm05, S.12]. Eine sorgfältige und wissenschaftlich fundierte Auseinandersetzung mit dem Thema der *Persönlichen Finanzplanung* ist bislang nur unzureichend erfolgt. Im Bereich der *Persönlichen Finanzplanung* gibt es, mit Ausnahme der auf das Problem der *Persönlichen Finanzplanung* anzuwendenden umfang-

reichen Theorie des *Portfolio-Managements* sowie versicherungsmathematischer Berechnungen, keine Anwendung mathematischer (und insbesondere von OR) Methoden. Insbesondere fehlt ein Entscheidungsmodell der *Persönlichen Finanzplanung*, auf dessen Basis Entscheidungsprobleme im Rahmen der *Persönlichen Finanzplanung* beispielsweise mit Hilfe von Methoden des OR gelöst werden können. *Persönliche Finanzplanung* wird mehr als Kunst verstanden denn als wissenschaftliche Disziplin. Es gibt zwar zahlreiche Lehrbücher zur *Persönlichen Finanzplanung* [u.a. BoSt03; Keow03; NiRR04; Woer02], die als Ratgeber rund um Fragen der *Persönlichen Finanzplanung* genutzt werden können, allerdings nicht wesentlich mehr beinhalten als eine Sammlung von Anekdoten, Daumenregeln und persönlichen Erfahrungswerten. Keine der internationalen Zeitschriften, die sich umfassend mit Themen der *Persönlichen Finanzplanung* befassen (u.a. *Financial Services Review*, *Journal of Personal Financial Planning*, *Journal of Personal Finance*) befindet sich aktuell (Stand: Juli 2006) im *Social Sciences Citation Index* (SCSI). Die Zeitschrift *Financial Services Review* befindet sich im Begutachtungsprozess [Wars05].

Literaturüberblick: Unterscheidet man Betriebe in *Unternehmen* (Betriebe zur Fremdbedarfsdeckung) und in *Haushalte* (Betriebe zur Eigenbedarfsdeckung), so kann man feststellen, dass finanzwirtschaftliche Fragestellungen, die für Unternehmen gelten, in ähnlicher Weise auch für Haushalte Anwendung finden können. Dementsprechend wird in der Literatur als Pendant zum Begriff *Finance* (Finanzwirtschaft) der Begriff *Personal Finance* (Persönliche Finanzwirtschaft) verwendet. *Personal Financial Planning* (Persönliche Finanzplanung) ist Teil des Forschungsgebietes *Personal Finance*. Innerhalb des Forschungsgebietes *Finance* stellt heute die *Kapitalmarkttheorie* ein eigenständiges und anerkanntes wissenschaftliches Forschungsgebiet mit großem Einfluss auf die Praxis (Börsen und Investment-Industrie) dar.

Einen Literaturüberblick zur Anwendung mathematischer (und insbesondere von OR) Methoden in der Finanzwirtschaft geben [AsBD88; McCa82; Mert95]. Bis auf versicherungsmathematische Modelle (einen Überblick gibt [OBri92]) haben bis Anfang der 1950er Jahre keine mathematischen Modelle breitere Anwendung im Forschungsgebiet *Finance* gefunden. Die Arbeiten von Markowitz [Mark52] und Roy [Roy52] markieren den Beginn der *Modernen Portfoliotheorie* und stellen den Anfang einer Entwicklung weg von einer rein deskriptiven hin zu einer theoretischen, kapitalmarktorientierten Finanzierungslehre dar. Basierend auf diesen grundlegenden Arbeiten haben Sharpe [Shar64], Lintner [Lint65] und Mossin [Moss66] die Bewertung von risikobehafteten Kapitalanlagen im Marktgleichgewicht untersucht, und ihr

Capital Asset Pricing Model (CAPM) wurde das grundlegende quantitative Modell zur Messung und Bewertung von Einzelwertrisiken. Anfang der 1970er Jahre veröffentlicht, hat das *Black-Scholes-Modell* zur Preisfindung für Optionen [BlSc73] sowohl in der wissenschaftlichen Diskussion als auch in der praktischen Anwendung herausragende Auswirkungen. [SpSZ05] stellen fest, dass seit Mitte der 1970er Jahre die Forschung zur Anwendung von OR-Methoden im Forschungsgebiet *Finance* in typischen OR-Zeitschriften (namentlich u.a. *European Journal of Operational Research*, *Operations Research*, *Journal of the Operational Research Society*) fortgeführt wird und die Beiträge in klassischen finanzwirtschaftlichen Zeitschriften (namentlich *Journal of Finance*, *Journal of Financial and Quantitative Analysis*, *Journal of Financial Research*) sich in eine Richtung bewegt haben, in der mathematische Modelle wieder durch verbale Modelle ersetzt werden in Richtung einer positiven bzw. deskriptiven Finanztheorie.

Die OR-Methode der *Linearen Programmierung* hat im Bereich der Investitions- und Finanzplanung breite Anwendung gefunden. [ChCM59] beschreiben eine Anwendung der Linearen Programmierung auf die Finanzplanung. [Bern69; Carl69] geben einen Überblick über in den 1960er Jahren erstellte Modelle der mathematischen Programmierung für die Investitionsplanung (darunter [Wein63; BaQu65]). In der deutschsprachigen Literatur sind erstmals Anfang der 1960er Jahre grundlegende Beiträge zur Anwendung von *Linearer Programmierung* in der Investitions- und Finanzplanung erschienen [Alba60; Alba62; Hax64; Moxt63].

Gliederung der Arbeit: Im Kern der *Wirtschaftsinformatik* kann die Beantwortung der Frage gesehen werden, wie mit Hilfe von Techniken und Methoden der Informatik (dazu können Techniken und Methoden des *Operations Research* gezählt werden) betriebswirtschaftliche Probleme beschrieben (modelliert) und besser als ohne Anwendung dieser Techniken und Methoden gelöst werden können. Im Sinne der Modellierung und Entwicklung eines Entscheidungsunterstützungssystems wird in Abschnitt 2 zunächst herausgearbeitet, worin die Zielsetzung der *Persönlichen Finanzplanung* liegt. Im Anschluss daran werden in Abschnitt 3 Modelle beschrieben, mit deren Hilfe die Zulässigkeit von Finanzplänen überprüft und ggf. hergestellt werden kann. Bei der Betrachtung realer Problemstellungen aus dem Bereich der *Persönlichen Finanzplanung* liegt häufig nicht nur ein einziges Optimierungsziel vor, sondern es müssen im Allgemeinen mehrere Optimierungsziele gemeinsam betrachtet werden. Daher befasst sich Abschnitt 4 mit einem *Goal Programming Modell* der *Persönlichen Finanzplanung*. In Abschnitt 5 werden die Ergebnisse des Beitrags zusammengefasst und weitere Forschungsmöglichkeiten im Umfeld einer OR-basierten *Persönlichen Finanzplanung* diskutiert.

2 Zielsetzung der Persönlichen Finanzplanung

In Anlehnung an [Wöhe05, S.96], in der Planung als gedankliche Vorbereitung zielgerichteter Entscheidungen definiert wird, definieren wir *Persönliche Finanzplanung* als gedankliche Vorbereitung zielgerichteter Entscheidungen bezüglich persönlicher finanzieller Angelegenheiten. *Planungsgegenstand* sind somit die persönlichen finanziellen Angelegenheiten, *Planungsdaten* sind Vermögen und Verbindlichkeiten sowie Einzahlungen und Auszahlungen vergangener Perioden, zum aktuellen Zeitpunkt und in zukünftigen Perioden. Der *Planungszeitraum* kann unterteilt werden in strategisch (5 Jahre und mehr), taktisch (2-5 Jahre) und operativ (max. 1 Jahr). Unter Einbeziehung der Zielbildung ergibt sich für die Planung folgendes Phasenschema: Zielbildung, Problemanalyse, Alternativenermittlung, Alternativenbewertung. Weitere grundlegende Definitionen *Persönlicher Finanzplanung*, des *Prozesses der Persönlichen Finanzplanung*, von *Financial Modelling* sowie *finanzwirtschaftlicher Entscheidungsprobleme* sind [CFP06; ISO04; Sch06; SpHa97, S.113-114; SpSZ05, S.801] zu entnehmen.

2.1 Grundlegende Begriffe

Im Allgemeinen versteht man unter der *Finanzplanung einer Unternehmung* alle Berechnungen und Dispositionen im Hinblick auf die Abstimmung der kurzfristig zu erwartenden Einzahlungs- und Auszahlungsströme mit dem *Ziel der Erhaltung der Liquidität* und die mittel- und langfristige Abstimmung von Kapitalbedarf und Kapitaldeckungsmöglichkeiten (Finanzierungsmöglichkeiten) [BiKu00b, S.2]. *Persönliche Finanzplanung* ist eine Finanzdienstleistung. Das Gut, auf das sich Finanzdienstleistungen beziehen, ist Geld. Informationen sind der zentrale Produktionsfaktor der Finanzdienstleister. Informationsprodukte stehen im Mittelpunkt der Leistungserstellung. *Persönliche Finanzplanung* besteht aus Informationsbeschaffung und Informationsverarbeitung sowie Informationsweiterleitung an den Kunden (Mandanten). Entscheidungen sind vom Mandanten zu treffen, Entscheidungsunterstützung kann vom Finanzplaner, der als Berater fungiert, geleistet werden.

2.2 Zielbildung

Ähnlich wie bei Unternehmen muss in der *Persönlichen Finanzplanung* bei jeder Entscheidung über finanzwirtschaftliche Vorgänge die Frage beantwortet werden, welche *Zielsetzungen* zu Grunde gelegt werden. Erst aufgrund gesetzter Ziele als Angabe des gewünschten Zustandes der Realität ist der Entscheidungsträger in der Lage, die zur Verfügung stehenden Handlungs-

alternativen zu bewerten und so zu einer Entscheidung zu gelangen. Entscheidungsträger können also erst dann eine rationale Entscheidung treffen, wenn mittels eines vorgegebenen Zielkriteriums festgestellt werden kann, ob die durch die Entscheidung ausgelöste Maßnahme einen Beitrag zur Erreichung dieses Zieles liefert [BiKu00a, S.10, in Anlehnung an Süch95, S.295]. *Persönliche Finanzplanung* sollte dabei nicht lediglich die Frage behandeln, wie Vermögen maximiert werden kann, sondern, wie Vermögen und künftige Einzahlungen genutzt werden können, um damit ein erfülltes und glückliches Leben führen zu können. Das allgemeine Ziel einer *Persönlichen Finanzplanung* kann also darin gesehen werden, die persönliche Lebensqualität zu maximieren und dabei die Frage zu beantworten, wie eine Person ihr Vermögen und ihre künftigen Einzahlungen auf die für sie beste Art und Weise einsetzen kann. Ziele lassen sich nach unterschiedlichen Klassifikationsmerkmalen einteilen [Wöhe05, S.92 ff.; in Anlehnung an Mach03, S.187 ff.]. Bezüglich des *Zielinhalts* unterschieden zwischen *Formalzielen* und *Sachzielen*. Formalziele bestimmen die Grundlinie des Handelns, Sachziele haben Instrumentalcharakter. Hinsichtlich des *Zielausmaßes* unterscheidet man zwischen begrenzten Zielen (z.B. Verringerung der Lebenshaltungskosten um 10%) und unbegrenzten Zielen (z.B. Maximierung der Eigenkapitalrendite). Daneben wird zwischen lang-, mittel- und kurzfristigen Zielen (*Zeitbezug*), komplementären, konkurrierenden und indifferenten Zielen (*Zielbeziehungen*) und Ober-, Zwischen- und Unterzielen (*Rangordnung*) unterschieden. Offensichtlich ist, dass ein Haushalt immer mehrere Ziele zu gleicher Zeit verfolgt. Daraus entsteht die Notwendigkeit, die Ziele in eine sinnvolle Ordnung (in ein Zielsystem) zu bringen, d.h. den Zielplanungsprozess zu organisieren. Dabei sollen zumindest folgende Grundsätze beachtet werden: Motivationsfunktion, Realitätsbezug, Widerspruchsfreiheit, Verständlichkeit, Kontrollierbarkeit. Wichtig ist, dass jedes Unterziel operational ist. Im Entscheidungsprozess werden Ziele, Handlungsalternativen und Umweltbedingungen analysiert und aufeinander abgestimmt.

2.2.1 *Formalziele der Persönlichen Finanzplanung*

Die wichtigsten Formalziele der Persönlichen Finanzplanung sind Rentabilität, Sicherheitsstreben und Liquidität [BiKu00a, S.10ff]. Die *Rentabilität* (Ziel: Maximierung der Rendite) ist eine Messgröße, sie wird in Kennzahlen ausgedrückt und spiegelt das Verhältnis zwischen Kapitaleinsatz und dem Überschuss wider, der mit diesem Kapitaleinsatz erzielt wird. Der Einsatz der Rentabilität als Zielbildungsgröße ermöglicht es, Entscheidungen unter Zuhilfenahme von Plan- und Daten zu treffen, um gleichzeitig die Zielerreichung durch Steuerung und Überwachung zu gewährleisten. Rentabilitäten können für einzelne investive Maßnahmen aber auch für eine

Person als Gesamtheit ermittelt werden. Als Berechnungszeitraum können eine Periode oder die Gesamtdauer einer Investition herangezogen werden. Unter *Sicherheitsstreben* (Ziel: Vermeidung von Risiko) versteht man das Ziel, das für eine Investition zur Verfügung gestellte Kapital uneingeschränkt zu erhalten. Konsequenterweise würde dies der Verzicht auf jegliches Risiko bedeuten. Umgekehrt gilt: Je höher das Risiko einer Aktion ist, desto höher kann in der Regel der Gewinn aus dieser Aktion und damit die Steigerung der Rentabilität ausfallen. Der Eintritt eines Misserfolgs wirkt sich dagegen negativ auf die Rentabilität aus. *Liquidität* (Ziel: Erhaltung des finanziellen Gleichgewichts) fordert, dass die auf eine Person zukommenden Zahlungsverpflichtungen jederzeit erfüllt werden können [PeSt99, S.15-16]. Die Formalziele Rentabilität, Sicherheitsstreben und Liquidität widersprechen sich in dem Sinne, dass eine Verbesserung hinsichtlich eines Ziels eine Verschlechterung eines anderen Ziels bewirkt. Weitere Formalziele können aus dem Bereich sozialer Ziele (Spenden, Streben nach Ansehen, Prestige, Macht) sowie aus dem Bereich öffentlicher (zum Beispiel ökologischer) Ziele (Ressourcenschonung), abgeleitet werden, sollen hier aber nicht weiter betrachtet werden.

2.2.2 Sachziele der Persönlichen Finanzplanung

Sachziele der *Persönlichen Finanzplanung* lassen sich unterteilen in *Finanzwirtschaftliche Sachziele* (beziehen sich auf die Frage: Wie sicher ist die zukünftige Zahlungsfähigkeit?) und in *Erfolgswirtschaftliche Sachziele* (beziehen sich auf die Frage: Wie groß ist die zukünftige Ertragskraft?). In Abbildung 1 ist das Zielsystem entsprechend abgebildet.

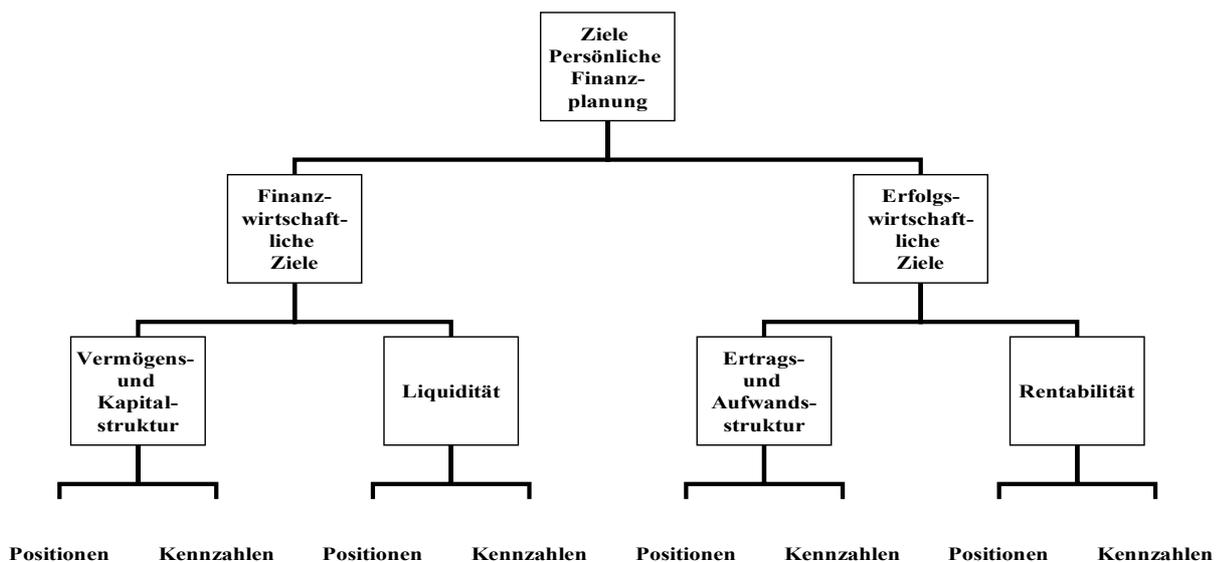


Abb. 1: Zielsystem Persönliche Finanzplanung

Finanzwirtschaftliche Sachziele beziehen sich im Kern auf die Vermögens- und Kapitalstruktur und die Liquidität. Dabei werden aus der Vermögensstruktur Aussagen über die künftige Zahlungsfähigkeit getroffen (z.B. $\text{Finanzanlagenintensität} = \text{Finanzanlagevermögen} / \text{Gesamtvermögen}$), Finanzierungsrisiken betrachtet (z.B. $\text{Verschuldungsgrad} = \text{Fremdkapital} / \text{Eigenkapital}$) und die Frage untersucht, inwieweit das Liquiditätspotential ausreicht, gegebenen Zahlungsverpflichtungen nachzukommen (z.B. $\text{Liquidität I. Grades} = \text{Zahlungsmittel} / \text{kurzfristige Verbindlichkeiten}$). *Erfolgswirtschaftliche Sachziele* beziehen sich auf die Ertrags- und Aufwandsstruktur und die Rentabilität. Dabei werden Ziele bezüglich der Ertragslage gesetzt (z.B. Aufwand-Ertrag-Relationen) und Ergebnis- mit Kapitalgrößen in Verbindung gesetzt (z.B. $\text{Eigenkapitalrentabilität} = \text{Gewinn} / \text{Eigenkapital}$). Hilfsmittel zur Festlegung und Überprüfung der Sachziele stellen die *Vermögensbilanz*, die *Cash-Flow-Rechnung* und die *Erfolgsrechnung* dar. Die *Vermögensbilanz* ist eine *Zeitpunktbilanz*. Eine *Zeitraumbilanz* bezieht sich dagegen auf einen Zeitraum und weist Strömungsgrößen aus. Zeitraumbilanzen verfolgen das Ziel, die Bilanzadressaten über Herkunft und Verwendung finanzieller Mittel während des Betrachtungszeitraums zu informieren. Sie werden auch als Kapitalflussrechnungen bezeichnet, die entweder vergangenheitsorientiert sind und somit auf Istgrößen basieren oder die als Zukunftsrechnungen auf Plangrößen beruhen [Wöhe05, S.831]. Zeitraumbilanzen im Rahmen der Persönlichen Finanzplanung sind die *Cash-Flow-Rechnung* (weist Einzahlungen und Auszahlungen einer Periode aus) und die *Erfolgsrechnung* (Gewinn- und Verlustrechnung, weist Erträge und Aufwendungen einer Periode aus). Zur (bilanzorientierten) Zielbildung können nun konkrete Vorgaben bezüglich einzelner Positionen (z.B. bzgl. der Vermögens- und Kapitalstruktur: Die Position Aktien soll 15% des gesamten Portfolios ausmachen, bzgl. der Liquidität: Die Position Auszahlungen für Lebenshaltungskosten soll weniger als 1600 Euro monatlich betragen) und Kennzahlen (z.B. bzgl. der Rentabilität: Die Kennzahl *Eigenkapitalrentabilität* soll bei mindestens 7% liegen) herangezogen werden

In Anti-Bilanz-Konzeptionen [u.a. Moxt84; Buss66] wird vorgeschlagen, den betrieblichen Jahresabschluss durch einen *prospektiven Finanzplan* zu ersetzen, da so die Informationsbedürfnisse externer Adressaten weitaus besser erfüllt werden könnten. Auf Grund dieser Überlegungen und da *Liquidität* unabdingbare Voraussetzung für die Überlebensfähigkeit einer Person ist (in dem Sinne, dass auf eine Person zukommende Zahlungsverpflichtungen jederzeit erfüllt werden können), wird in Abschnitt 3 das Ziel der Erhaltung der Liquidität in den Mittelpunkt der Überlegungen gestellt.

3 Zulässige Finanzpläne

Zentrales Steuerungsinstrument der *Persönlichen Finanzplanung* ist der *Finanzplan* als tabellarische, zeitliche und sachliche Aufgliederung des in Zukunft erwarteten Stroms der Ein- und Auszahlungen. In unserem Grundmodell ist ein *Finanzplan* mit $n+1$ Einzahlungen $E_i > 0$ ($i = 0, \dots, n$) und $n+1$ Auszahlungen $A_i > 0$ ($i = 0, \dots, n$) sowie ein Zinssatz z , mit dem die Einzahlungsüberschüsse verzinst werden können, gegeben (Tabelle 1).

n	wir betrachten $n+1$ Zeitpunkte $t = 0, \dots, n$. Die Differenz zwischen zwei Zeitpunkten t und $t+1$ entspricht dabei einer Periode. Wenn nichts Anderes festgelegt wird, entspricht eine Periode einem Jahr.
$E_i \geq 0$ ($i=0, \dots, n$) $A_i \geq 0$ ($i=0, \dots, n$)	$n+1$ Einzahlungen $E_i \geq 0$ und $n+1$ Auszahlungen $A_i \geq 0$ ($i = 0, \dots, n$). Eine Einzahlung E_i oder Auszahlung A_i entspricht dabei der Summe der jährlichen Einzahlungen oder Auszahlungen in i Jahren. Die Differenz $E_0 - A_0$ entspricht dem aktuellen liquidierten Vermögen. Wir nehmen an, dass gilt: $E_0 > 0$ und $A_0 = 0$.
z	Zinssatz z gibt die durchschnittliche jährliche Verzinsung von Einzahlungsüberschüssen an.

Tab. 1: Notation

Damit alle Auszahlungen durch die Einzahlungen gedeckt werden können, muss für alle Zeitpunkte $t=0, \dots, n$ gelten: *Summe aller Einzahlungen bis zum Zeitpunkt t sind größer oder gleich Summe aller Auszahlungen bis zum Zeitpunkt t . Oder äquivalent: Summe aller Einzahlungsüberschüsse bis zum Zeitpunkt t sind größer oder gleich 0.* Zusätzlich können Einzahlungsüberschüsse mit einem jährlichen Zinssatz von z % verzinst werden, also muss gelten: *Summe aller auf den Zeitpunkt t mit dem Zinssatz z aufgezinsten Einzahlungsüberschüsse bis zum Zeitpunkt t sind größer oder gleich 0.* Die Aufnahme eines Kredits (Fremdfinanzierung) kommt dabei einer zusätzlichen Einzahlung gleich, die gleichzeitig zusätzliche Auszahlungen mit sich bringt. Formal ist ein zulässiger Finanzplan wie folgt definiert: *Die Zulässigkeit eines Finanzplans ist genau dann erreicht, wenn zu jedem beliebigen Zeitpunkt $t=0, \dots, n$ gilt: Die Summe aller mit dem Zinssatz z auf den Zeitpunkt t aufgezinsten Einzahlungsüberschüsse bis zum Zeitpunkt t sind größer oder gleich 0.* Um die *Zulässigkeit eines Finanzplans* zu prüfen, müssen also folgende $n+1$ Bedingungen überprüft werden:

$$\sum_{\substack{i=0 \\ t=0 \dots n}}^t (E_i - A_i)(1+z)^{t-i} \geq 0$$

Daraus ergibt sich folgender Algorithmus zur Prüfung eines Finanzplans auf Zulässigkeit.

```

Input:       $E_i (i=0, \dots, n), A_i (i=0, \dots, n), z$ 
Schritte:
    for (  $t = 0..n$  )
    {
         $S_t := 0$  ;
        for (  $i = 0..t$  )
             $S_t := S_t + (E_i - A_i)(1+z)^{t-i}$  ;
        if (  $S_t < 0$  ) then
            print ("Finanzplan ist unzulässig") ; stop ;
    }
    print ("Finanzplan ist zulässig") ;
Output:    Ausgabe "Finanzplan ist zulässig" oder Ausgabe "Finanzplan ist unzulässig".

```

Ergibt die Zulässigkeitsprüfung als Ergebnis Unzulässigkeit vom Grad 0, so können folgende Maßnahmen ergriffen werden: Berechnung eines (risikoreicheren) Zinssatzes z , für den der Finanzplan zulässig ist: Abschnitt 3.1. Erhöhung der regelmäßigen Einzahlungen (beispielsweise Partner geht auch arbeiten) oder Verringerung der regelmäßigen Auszahlungen (beispielsweise Senkung der Lebenshaltungskosten): Abschnitt 3.2. Erhöhung einzelner Einzahlungen oder Verringerung einzelner Auszahlungen (beispielsweise kleineres Auto anschaffen): Abschnitt 3.3. Zeitliche Verschiebung von Einzahlungen nach vorne (beispielsweise Auszahlung einer Lebensversicherung) oder zeitliche Verschiebung von Auszahlungen nach hinten (beispielsweise Anschaffung eines neuen Autos verschieben): Abschnitt 3.4.

3.1 Variabler Zinssatz

Besteht die Bereitschaft, ein höheres Risiko einzugehen, einhergehend mit der Hoffnung auf eine höhere Rendite, so kann Zulässigkeit eines Finanzplans dadurch hergestellt werden, dass der Zinssatz z , mit der die Einzahlungsüberschüsse jährlich verzinst werden, erhöht wird. Zu beachten ist, dass stets gilt: $E_0 > 0$ und $A_0 = 0$. Da der Finanzplan unzulässig ist, muss es wenigstens einen Zeitpunkt t geben mit $\sum_{i=0}^t (E_i - A_i)(1+z)^{t-i} < 0$. Führe nun für jedes dieser t folgendes durch: Berechne Zinssatz z_t , so dass gilt $\sum_{i=0}^t (E_i - A_i)(1+z_t)^{t-i} = 0$. z_t kann beispielsweise mit Hilfe des Newton-Verfahrens zur Berechnung der Nullstellen von Polynomen berechnet werden. Der Zinssatz z , mit der die Einzahlungsüberschüsse jährlich verzinst werden müssten, um einen zulässigen Finanzplan zu gewährleisten, ist dann das Maximum aller z_t .

Input: $E_i (i=0, \dots, n), A_i (i=0, \dots, n), z$
 Schritte:

```

for (  $t = 0..n$  )
  {
     $S_t := 0 ; z_t := 0 ;$ 
    for (  $i = 0..t$  )
       $S_t := S_t + (E_i - A_i)(1+z)^{t-i};$ 
    if(  $S_t < 0$  ) then
      Berechne Zinssatz  $z_t$  (z.B. mit dem Newton-Verfahren), so dass gilt:
       $(E_0 - A_0)(1+z_t)^t + (E_1 - A_1)(1+z_t)^{t-1} + \dots + (E_t - A_t)(1+z_t) = 0 ;$ 
    }
  }
   $z := \max \{ z_0..z_n \} ;$ 
  Output: Zinssatz  $z$ , mit dem die Einzahlungsüberschüsse verzinst werden müssten, so dass der
  Finanzplan zulässig ist.
  
```

3.2 Erhöhung regelmäßiger jährlicher Einzahlungen oder Verringerung regelmäßiger jährlicher Auszahlungen

Bei gegebener jährlicher Verzinsung der Einzahlungsüberschüsse kann Zulässigkeit eines Finanzplans hergestellt werden, indem die regelmäßigen jährlichen Einzahlungen erhöht und/oder die regelmäßigen jährlichen Auszahlungen verringert werden. Eine Fragestellung, die so zu beantworten wäre, ist: „Um wieviel müssten die jährlichen Lebenshaltungskosten reduziert werden, um bei einem durchschnittlichen jährlichen Zinssatz von z % Zulässigkeit vom Grad 0 zu erreichen?“. Berechne nun für jeden Zeitpunkt t mit $\sum_{i=0}^t (E_i - A_i)(1+z)^{t-i} < 0$ einen Betrag d_t , so dass gilt: $\sum_{i=0}^t (E_i - A_i + d_t)(1+z_t)^{t-i} = 0$. Gesucht ist also ein Betrag $d = \max\{d_0..d_n\} \geq 0$, so dass der Finanzplan zulässig ist. d gibt die konstante Veränderung der jährlichen Einzahlungen und Auszahlungen an.

Input: $E_i (i=0, \dots, n), A_i (i=0, \dots, n), z$
 Schritte:

```

for (  $t = 0..n$  )
  {
     $S_t := 0 ;$ 
    for (  $i = 0..t$  )
       $S_t := S_t + (E_i - A_i)(1+z)^{t-i};$ 
    if(  $S_t < 0$  ) then
       $d_t := - S_t / t;$ 
    }
  }
   $d := \max \{ d_0, \dots, d_n \};$ 
  Output: Konstante Veränderung  $d$  der jährlichen Einzahlungen und Auszahlungen, so dass der
  Finanzplan zulässig ist.
  
```

3.3 Erhöhung einzelner jährlicher Einzahlungen oder Verringerung einzelner jährlicher Auszahlungen

Unzulässigkeit eines Finanzplans kann durch eine einzelne hohe Auszahlung eintreten. In diesem Fall bieten sich mehrere Lösungen an, um durch eine wertmäßige Veränderung von Einzahlungen oder Auszahlungen einen zulässigen Finanzplan herzustellen. Beispielsweise könnten die Einzahlungen erhöht oder die Auszahlungen gekürzt werden. Im Folgenden betrachten wir lediglich die Möglichkeit, Auszahlungen zu kürzen. Dabei soll der Betrag, um den die Auszahlungen gekürzt werden sollen, möglichst gering sein. Führe nun für jeden Zeitpunkt t mit $\sum_{i=0}^t (E_i - A_i)(1+z)^{t-i} < 0$ folgendes durch: Ersetze die Auszahlung A_t durch A_t' mit $A_t' = A_t + \sum_{i=0}^t (E_i - A_i)(1+z)^{t-i}$, so dass gilt: $\sum_{i=0}^t (E_i - A_i')(1+z)^{t-i} = 0$. Die Auszahlung A_t wird also entsprechend gekürzt und ist auch bei den weiteren Zulässigkeitsprüfungen entsprechend gekürzt zu berücksichtigen.

```

Input:       $E_i (i=0, \dots, n), A_i (i=0, \dots, n), z$ 
Schritte:
    for (  $t = 0..n$  )
    {
         $S_t := 0$  ;
        for (  $i = 0..t$  )
             $S_t := S_t + (E_i - A_i)(1+z)^{t-i}$  ;
        if(  $S_t < 0$  ) then
             $A_t := A_t + S_t$  ;
    }
Output:    Liste der verringerten Auszahlungen  $A_t$  .

```

3.4 Zeitliche Veränderungen einzelner Auszahlungen oder Einzahlungen

Unzulässigkeit eines Finanzplans kann durch eine ungünstige zeitliche Reihenfolge von Einzahlungen und Auszahlungen eintreten. In diesem Fall bieten sich zwei Lösungen an, um durch eine zeitliche Veränderung von Einzahlungen oder Auszahlungen einen zulässigen Finanzplan herzustellen. Entweder werden Auszahlungen nach hinten verschoben oder Einzahlungen vorgezogen. Im Folgenden betrachten wir lediglich die Möglichkeit, eine Auszahlung nach hinten zu verschieben. Dabei soll der Zeitpunkt, zu dem die Auszahlung getätigt werden soll, so nahe wie möglich am ursprünglich geplanten Zeitpunkt liegen. Führe nun für jeden Zeitpunkt t mit $\sum_{i=0}^t (E_i - A_i)(1+z)^{t-i} < 0$ folgendes durch: Setze $A_t := 0$ und verschiebe die Auszahlung A_t so lange nach hinten bis zu einem Zeitpunkt t' gilt: $\sum_{i=0}^{t'} (E_i - A_i)(1+z)^{t'-i} \geq A_t$. Die Auszahlung A_t wird also zu A_t' hinzuaddiert und es gilt $A_t = 0$ und $A_t' = A_t' + A_t$.

```

Input:       $E_i (i=0, \dots, n), A_i (i=0, \dots, n), z$ 
Schritte:
     $H := 0 ;$ 
    for (  $t = 0..n$  )
    {
         $S_t := H ;$ 
        for (  $i = 0..t$  )
             $S_t := S_t + (E_i - A_i)(1+z)^{t-i};$ 
        if(  $S_t < 0$  ) then
             $H := - A_t ; A_t := 0 ;$ 
        else
             $A_t := A_t + H ;$ 
    }
Output:    Liste der zeitlich verschobenen Auszahlungen  $A_t$ .

```

4 Goal Programming

Im Allgemeinen lassen sich Einzahlungen und Auszahlungen weder betragsmäßig noch zeitpunktbezogen exakt voraussagen, so dass die zur Zulässigkeitsprüfung notwendige Betrags- und Zeitpunktsgenauigkeit nicht gegeben ist. Daher stellt sich die Frage, wie eine Zulässigkeitsprüfung mit in Betrag und Zeitpunkten variablen Einzahlungen und Auszahlungen durchgeführt werden kann. Darüber hinaus stehen sich beispielsweise gewünschte Ausgaben konkurrierend gegenüber (z.B. Porsche für 80.000 Euro oder Immobilie für 200.000 Euro). Um solche reale Problemstellungen mit mehreren gemeinsam zu betrachtenden Optimierungszielen zu bearbeiten, hat sich im *Operations Research* die *Multikriterielle Analyse* (*Multicriteria analysis*, synonym zu verwenden *Multiple criteria decision making* (MCDM) oder *Multicriteria decision aid* (MCDA) [Zopo99, S.405]) als geeignetes Hilfsmittel herausgestellt. [HaSp02] beschreiben die Relevanz multikriterieller Analyse für finanzielle Entscheidungsprobleme. [StNa03] geben einen umfangreichen aktuellen Überblick über *Multiple criteria decision making combined with finance*. Finanzielle Entscheidungsprobleme mit mehreren Optimierungszielen betrachten beispielsweise [LeLe74; BhMc83; SpHa97]. Das Konzept des *Goal Programming* geht auf Charnes, Cooper und Ferguson [ChCF55] zurück. Einen Überblick liefern [Rome86; TaJE95; TaJR98]. Weitere Artikel zur multikriteriellen Analyse im Forschungsgebiet Finance sind den Übersichtsartikeln [SpSZ05; StNa03; ZoDo02; Zopo99] zu entnehmen. *Goal Programming* ermöglicht sowohl die Berücksichtigung einseitig als auch zweiseitig angestrebter Ziele [HiLi05, S.15 ff.]. Bei einem einseitigen Ziel soll ein Zielwert erreicht oder angenähert, keinesfalls aber über-, bzw. unterschritten werden (Satisfizierungsziele). Bei

einem zweiseitigen Ziel soll ein Zielwert möglichst exakt erreicht werden, Abweichungen nach unten oder oben sind möglich. In einem *Goal Programming Modell* werden dazu untere und obere (nicht negative) Abweichungsvariablen d_i^- und d_i^+ ($i=0, \dots, n$) eingeführt.

Bei dem Modell in diesem Abschnitt beschränken wir uns auf den Fall, lediglich Ziele bezüglich Einzahlungen und Auszahlungen zu betrachten. Zunächst soll der Fall betrachtet werden, dass nicht mehr betragsgenaue Einzahlungen und Auszahlungen gegeben sind, sondern dass obere (untere) Schranken, in denen sich die Auszahlungen (Einzahlungen) bewegen dürfen, mit Hilfe von Parametern $d_i^- \geq 0$ und $d_i^+ \geq 0$ ($i=0, \dots, n$) wie folgt vorgegeben sind. d_i^- entspricht dem Betrag, um den die Auszahlung A_i gekürzt werden kann, d_i^+ entspricht dem Betrag, um den die Einzahlung E_i gekürzt werden kann. Ziel ist, einen zulässigen Finanzplan zu finden, in dem $\sum d_i^- + d_i^+$ ($i=0 \dots n$) möglichst klein ist. Da einzelnen Einzahlungen und Auszahlungen unterschiedlich hohe relative Bedeutungen beigemessen werden können (z.B. „Immobilie ist mir wichtiger als Porsche“), werden die zu minimierenden Abweichungen $d_i^- + d_i^+$ jeweils mit $w_i \in \mathbb{R}, 0 \leq w_i \leq 1$, gewichtet. Eine Beurteilung der unterschiedlich hohen relativen Bewertungen kann beispielsweise mit dem *Analytical Hierarchy Process* (AHP) [Saat86; Saat01; VaKu06] erfolgen. Im Rahmen des AHP lassen sich die vergleichenden Bedeutungen einzelner Einnahmen oder Ausgaben durch paarweise Vergleiche (Ordinalskala mit Werten von 1 bis 9) beurteilen. Beispielsweise bedeutet für das Paarvergleichsurteil a_{ij} der Eintrag einer 3 in die Evaluationsmatrix A , dass Ausgabe i etwas bedeutender als Ausgabe j eingeschätzt wird. Die Bedeutungsgewichte w_i können durch die Berechnung des normalisierten Eigenvektors der Matrix A bestimmt werden. Zusammenfassend kann folgendes mathematische Programm zur Minimierung der betragsmäßigen Zielabweichungen zukünftiger Einzahlungen und Auszahlungen formuliert werden.

$$Z = \sum_{i=0}^n w_i (d_i^- + d_i^+) \longrightarrow \min$$

u.d.N.

$$(1..n+1) \quad \sum_{i=0}^t (E_i - A_i + d_i^- - d_i^+) (1+z)^{t-i} = 0, \forall t = 0, \dots, n$$

$$(n+2..2n+2) \quad d_i^-, d_i^+ \geq 0, \forall i = 0, \dots, n$$

Der Fall, dass nicht mehr zeitpunktgenaue Einzahlungen und Auszahlungen gegeben sind, sondern Zeiträume, in denen sich die Einzahlungen und Auszahlungen bewegen dürfen, kann äh-

lich gehandhabt werden. Das mathematische Programm kann auf einfache Art und Weise derart erweitert werden, dass *einzelne Investitionen* entweder als oberes einseitiges Ziel (z.B. „maximal 80000 Euro“) oder als zweiseitiges Ziel („ca. 80000 Euro“) berücksichtigt werden können.

5 Zusammenfassung und Ausblick

Bislang wird unter *Persönlicher Finanzplanung* mehr eine Kunst als eine Wissenschaft verstanden. Bis auf das Problem des *Portfolio-Managements*, das lediglich Teil einer umfassenden *Persönlichen Finanzplanung* darstellt, hat eine wissenschaftliche Methodik kaum Einklang gefunden bei der Untersuchung des Problems der *Persönlichen Finanzplanung*. Der vorliegende Artikel soll, basierend auf Techniken und Methoden des *Operations Research*, zur Diskussion einer wissenschaftlichen Auseinandersetzung mit dem Thema der *Persönlichen Finanzplanung* anregen. Dazu werden Probleme der *Persönlichen Finanzplanung* modelliert, mit Hilfe mathematischer Methoden gelöst und im Rahmen der Entwicklung eines webbasierten Systems zur *Persönlichen Finanzplanung* implementiert. Dabei kann beispielsweise eine Zulässigkeitsprüfung im Rahmen der Erstellung eines Sollkonzepts durchgeführt werden. Dazu werden alle erwarteten zukünftigen Einzahlungen und Auszahlungen periodenbezogen saldiert dargestellt (Abbildung 2).

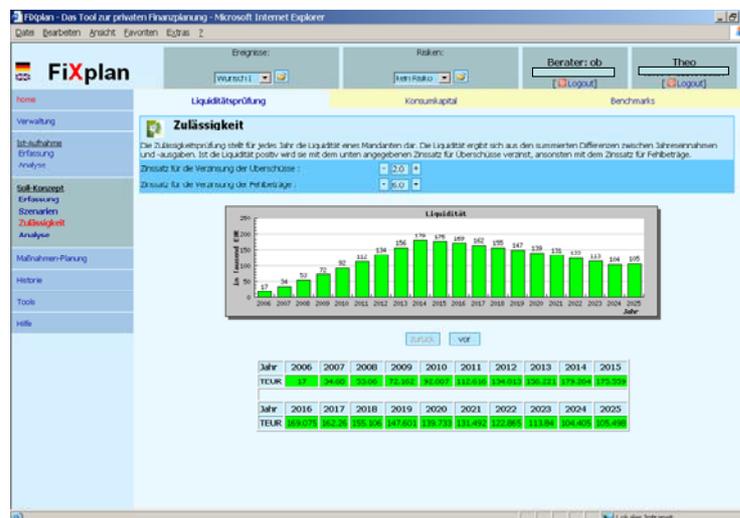


Abb. 2. Zulässigkeitsprüfung im Rahmen des Sollkonzepts

Grundlegend für eine *Persönliche Finanzplanung* ist eine Planung der zukünftigen Einzahlungen und Auszahlungen einer Person. Welche Einzahlungen und Auszahlungen eine Person in der Zukunft plant, ist klarerweise sehr individuell. In jedem Fall muss dabei die Liquidität

sichergestellt werden, damit die Person jederzeit ihren Zahlungsverpflichtungen nachkommen kann. Daher erhält die Sicherstellung der Liquidität die größte Bedeutung bei der Erstellung eines Finanzplans. Ein zulässiger Finanzplan kann ggf. hergestellt werden durch zeitpunktbezogene oder betragsmäßige Veränderungen zukünftiger Einzahlungen und Auszahlungen. Welche Auszahlungen eine Person in Zukunft reduzieren möchte oder auf welche Auszahlungen eine Person in Zukunft gar verzichten kann, ist wiederum sehr individuell. Alternativ können zur Herstellung eines zulässigen Finanzplans auch Einzahlungen erhöht oder vorgezogen werden. Ist Zulässigkeit eines Finanzplans hergestellt, kann, abhängig von den persönlichen Zielen, mit der Erstellung eines bezüglich einer (zu definierenden) *Nutzenfunktion* der Person optimalen Finanzplans begonnen werden. Zu berücksichtigen hierbei ist, dass viele Menschen offensichtlich Entscheidungen treffen, die nicht automatisch ihren Wohlstand oder ihr Einkommen maximieren. Diese Entscheidungen sind aus ihrer Sicht rational und wichtig für ihr Leben. *Persönliche Finanzplanung* sollte also nicht lediglich die Frage behandeln, wie Vermögen maximiert werden kann, sondern, wie Vermögen und Einkommen genutzt werden können, um damit ein erfülltes und glückliches Leben führen zu können. Weitergehende Forschungsmöglichkeiten bestehen darin, den Ansatz *multikriterieller Analyse* (Abschnitt 4) weiter auszubauen. Unsicherheit bzgl. künftiger Einzahlungen und Auszahlungen könnten mit Hilfe *stochastischer Prozesse* und Methoden der Behandlung von *Online-Problemen* modelliert werden [Ely98]. Die Formulierung von *Fuzzy Modellen* und *Fuzzy Goal Programming Modellen* [TaGu00; Rom04; Chan06] stellen weitere interessante Möglichkeiten dar, das in diesem Beitrag beschriebene Modell der *Persönlichen Finanzplanung* zu erweitern.

Literaturverzeichnis

- [Alba60] Albach, Horst: Lineares Programmieren als Hilfsmittel betrieblicher Investitionsplanung. Zeitschrift für handelswissenschaftliche Forschung. 12.Jg. (1960), S. 526-549.
- [AsBD88] Ashford, Robert W.; Berry, Robert H.; Dyson, Robert G.: Operational research and financial management. European Journal of Operational Research 36:2 (1988), S. 143–152.
- [BaQu65] Baumol, William J.; Quandt, Richard E.: Investment and discount rates under capital rationing – a programming approach. The Economic Journal 75:198 (1965), S. 317-329.
- [Bern69] Bernhard, Richard H.: Mathematical programming models for capital budgeting: A survey, generalization, and critique. Journal Financial Quantitative Analysis 4:2 (1969), S. 111-158.

- [BhMc83] Bhaskar, Krish; McNamee, Patrick: Multiple objectives in accounting and finance. *Journal of Business Finance & Accounting* 10:4 (1983), S. 595-621.
- [BiKu00a] Bieg, Hartmut; Kußmaul, Heinz: *Investitions- und Finanzierungsmanagement*. Band I: Investition. Vahlen, München 2000.
- [BiKu00b] Bieg, Hartmut; Kußmaul, Heinz: *Investitions- und Finanzierungsmanagement*. Band III: Finanzwirtschaftliche Entscheidungen. Vahlen, München 2000.
- [BlSc73] Black, Fischer; Scholes, Myron: The pricing of options and corporate liabilities. *Journal of Political Economy* 81:3 (1973), S. 637-654.
- [BoSt03] Böckhoff, Michael; Stracke, Guido: *Der Finanzplaner*. 2. Auflage. Sauer, Heidelberg 2003.
- [Buss66] Busse von Colbe, Walther: Aufbau und Informationsgehalt von Kapitalflußrechnungen. *Zeitschrift für Betriebswirtschaft* (1966), S.88-97.
- [Carl69] Carleton, Willard T.: Linear programming and capital budgeting models: A new interpretation. *Journal of Finance* 24:5 (1969), S. 825 – 833.
- [CFP06] Certified Financial Planner's (CFP) Board of Standards Inc (2006). Retrieved August 14, 2006, from <http://www.cfp.net>
- [Chan06] Chang, Ching-Ter: Binary fuzzy goal programming. *European Journal of Operational Research*, in Press, corrected proof, available online 15 May 2006.
- [ChCM59] Charnes, A.; Cooper, W.W., Miller, M.H.: Application of linear programming to financial budgeting and the costing of funds. *Journal of Business* 32:1 (1959). S. 20-46.
- [ElY98] El-Yaniv, Ran: Competitive solutions for online financial problems. *ACM Computing Surveys* 30:1 (1998), S. 28-69.
- [HaSp02] Hallerbach, Winfried; Spronk, Jaap: The relevance of MCDM for financial decisions. *Journal of Multi-Criteria Decision Analysis* 11 (2002), S. 187-195.
- [Hax64] Hax, Herbert: *Investitions- und Finanzplanung mit Hilfe der linearen Programmierung*. *Zeitschrift für betriebswirtschaftliche Forschung*. 16.Jg. (1964), S. 430-446.
- [HiLi05] Hillier, Frederick S.; Lieberman, Gerald J.: *Introduction to Operations Research*. 8. Auflage. McGraw-Hill, Boston 2005.
- [ISO04] International Organization for Standardization (ISO/TC 222) (2004) ISO/DIS 22222-1 of the Technical Committee ISO/TC 222, Personal financial planning.
- [Keow03] Keown, Alex J.: *Personal Finance: Turning Money into Wealth*. Prentice Hall 2003.
- [LeLe74] Lee, Sang M.; Lerro, A.J.: Capital budgeting for multiple objectives. *Financial Management* 3:1 (1974), S. 58-66.

- [Lint65] Lintner, John: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47:1 (1965), S. 13-37.
- [Mach03] Macharzina, Klaus: Unternehmensführung. 4. Auflage. Gabler, Wiesbaden 2003.
- [Mark52] Markowitz, Harry M.: Portfolio Selection. *Journal of Finance* 7:1 (1952), S. 77-91.
- [McCa82] McInnes, J. Morris; Carleton, Willard J.: Theory, Models and Implementation in Financial Management. *Management Science* 28:9 (1982), S. 957-978.
- [Mert95] Merton, Robert C.: Influence of Mathematical Models in Finance on Practice: Past, Present, and Future. *Financial Practice and Education* 5:1 (1995), S. 7-15.
- [Moss68] Mossin, Jan: Optimal multiperiod portfolio policies. *Journal of Business* 41 (1968), S.215-229.
- [Moxt63] Moxter, Adolf: Lineares Programmieren und betriebswirtschaftliche Kapitaltheorie. *Zeitschrift für handelswissenschaftliche Forschung*. 15.Jg. (1963), S. 285-309.
- [Moxt93] Moxter, Adolf: Bilanzlehre. Band I: Einführung in die Bilanztheorie. Gabler, Wiesbaden 1993.
- [NiRR04] Nissenbaum, Martin; Raasch, Barbara J.; Ratner, Charles L.: Ernst & Young's Personal Financial Planning Guide, 5. Auflage, John Wiley & Sons, Hoboken NJ 2004.
- [OBri92] O'Brien, Chris: Actuarial methods in finance. IN: Peter Newman, Murray Milgate, John Eatwell (Eds.), *The new palgrave dictionary of money and Finance*, London, Macmillan Press Limited, S. 17-20.
- [PeSt99] Perridon, Louis; Steiner, Manfred: Finanzwirtschaft der Unternehmung. 10. Auflage. Vahlen, München 1999.
- [Rome86] Romero, Carlos: A survey of generalized goal programming (1970-1982). *European Journal of Operational Research* 25 (1986), S. 183-191.
- [Romm04] Rommelfanger, Heinrich J.: The advantages of fuzzy optimization models in practical use. *Fuzzy Optimization and Decision Making* 3:4 (2004), S. 295-309.
- [Roy52] Roy, A.D.: Safety first and the holding of assets. *Econometrica* 20 (1952), S. 431-449.
- [Rubi02] Rubinstein, Michael: Markowitz's portfolio selection: a fifty year retrospective, *Journal of Finance* 17 (2002), S. 1041-1045.
- [Saat01] Saaty, Thomas L.: Decision making for leaders – The analytic hierarchy process for decisions in a complex world. 2. Auflage. RWS Publications, Pittsburgh 2001.
- [Sch06] Schmidt, Günter: Persönliche Finanzplanung. Springer, Berlin 2006.
- [Shar64] Sharpe, William F.: Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19:3 (1964), S. 425-442.

- [SpHa97] Spronk, Jaap; Hallerbach, Winfried G.: Financial modelling: Where to go? With an illustration for portfolio management. *European Journal of Operational Research* 99:1 (1997), S. 113-125.
- [SpSZ05] Spronk, Jaap; Steuer, Ralph E., Zopounidis, Constantin: Multicriteria Decision Aid / Analysis in Finance. In: Figuera, J., S. Greco and M. Ehrgott (Eds.): *Multiple criteria decision analysis: State-of-the-art surveys*, Springer Science, 2005, S. 799-857.
- [StNa03] Steuer, Ralph E.; Na, Paul: Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of Operational Research* 150 (2003), S. 496-515.
- [Süch95] Süchting, Joachim: *Finanzmanagement – Theorie und Politik der Unternehmensfinanzierung*. 6. Auflage. Gabler, Wiesbaden 1995.
- [TaGu00] Tarrazo, Manuel; Gutierrez, Luis: Economic expectations, fuzzy sets and financial planning. *European Journal of Operational Research* 126 (2000), S. 89-105.
- [TaJE98] Tamiz, Mehrdad; Jones, Dylan; El-Darzi, Elia: A review of goal programming and its applications. *Annals of Operations Research* 58 (1995), S. 39-53.
- [TaJR98] Tamiz, Mehrdad; Jones, Dylan; Romero, Carlos: Goal programming for decision making: An overview of the current state-of-the-art. *European Journal of Operational Research* 111 (1998), S. 569-581.
- [VaKu06] Vaidya, Omkarprasad; Kumar, Sushil: Analytic hierarchy process: An overview of applications. *European Journal of Operational Research* 169:1 (2006), S. 1-29.
- [Wars05] Warschauer, Tom: email vom 2005-12-15.
- [Wein63] Weingartner, H. Martin: *Mathematical Programming and the analysis of capital budgeting problems*. Englewood Cliffs, NJ 1963.
- [Wöhe05] Wöhe, Günter: *Einführung in die Allgemeine Betriebswirtschaftslehre*. 22. Auflage. Vahlen, München 2005.
- [Woer02] Woerheide, Walt: *Core concepts of Personal Finance*. John Wiley & Sons, New York, NY 2002.
- [Zimm05] Zimmermann, Hans-Jürgen: *Operations Research*. Vieweg, Wiesbaden 2005.
- [ZoDo02] Zopounidis, Constantin; Doumpos, Michael: Multi-criteria decision aid in financial decision making: Methodologies and literature review. *Journal of Multi-Criteria Decision Analysis* 11 (2002), S. 167-186.
- [Zopo99] Zopounidis, Constantin: Multicriteria decision aid in financial management. *European Journal of Operational Research* 119 (1999), S. 404-415.

An Order-Channel Management Framework for Institutional Investors

Bartholomäus Ende, Peter Gomber, Adrian Wranik

Chair of Business Administration, especially e-Finance

E-Finance Lab

Johann Wolfgang Goethe-University

60054 Frankfurt, Germany

{ende,gomber,wranik}@wiwi.uni-frankfurt.de

Abstract

Efficient *Order-Channel Management*, i.e. the process of information gathering, evaluation, decision and control regarding the setup of the overall trading infrastructure and the actual order routing implementation plays a crucial role for trading success as well as the competitiveness of Institutional Investors. This article introduces a framework intended to support Institutional Investors in establishing an individual *Order-Channel Management* (OCM). For this overall goal, OCM is decomposed into its *strategic* and *operational* constituents and the involved *key entities*, *parameters*, *processes* and their interdependencies are outlined. Based on the identified properties, a *framework* is derived that aims at identifying a suitable mapping from order characteristics to execution venues.

1 Introduction

Investment decisions of Institutional Investors i.e. of buy-side¹ companies typically initiate the process of security trading. Within their quest for liquidity it is essential for execution success and the competitiveness of Institutional Investors to enforce an allocation process that identifies suitable venues as well as execution strategies before orders can be communicated down the value chain. The channeling decision itself is addressed mainly on two levels: First, on a strategic level, a setup for accessing execution venues and building up required infrastructure in terms of people and technology has to be established. Second, on an order by order basis, a suitable venue from this pool has to be selected.

With new upcoming execution venues in the security trading industry, the demand for order-channelling solutions has intensified [Rami06]. Primarily, the changing intermediation relationships, driven by technical innovations within electronic trading, create new pools of execution

¹Buy-side refers to investment management companies that are "buying" trading services from the sell-side, i.e. investment banks and brokers [Harr03].

opportunities. Thus, Institutional Investors can choose to execute their orders bilaterally with their brokers at Regulated Markets, Alternative Trading Systems (e.g. Crossing Networks) or via new electronic execution concepts like Smart Order Routing, Direct Market Access (DMA) and Algorithmic Trading. Within this range of execution opportunities the two main entities, execution venues and orders, involved in the process of order-channeling can be described by a bundle of characteristics and interdependencies. Execution venues for instance can be determined by fixed (e.g. market model) and temporary (e.g. market situation and volatility) parameters whereas actual orders typically face a trade-off between urgency and costs. Altogether these bundles of characteristics cause the order-channeling process to become a complex, multidimensional task.

At the same time, new technology-driven solutions enable Institutional Investors to add value to their order processing and thus offer them the opportunity to outperform competitors. This potential for differentiation receives increased attention with the changes of the European regulation in securities trading. Within these changes that will take effect with the implementation of MiFID in November 2007, the topic Best Execution plays a major role as it requires investment firms to set up an individual 'Best Execution Policy' and to realize the best possible result for customer orders according to this policy [GoSe05]. The new regulation enables the buy-side to request evidence of best execution.

In order to support Institutional Investors with the decisions involved in order-channeling, this paper aims at introducing the concept of Order-Channel Management (OCM) and at outlining key parameters for its strategic and operational decisions. Altogether a framework for OCM is set up by identifying and analyzing the key considerations and decision parameters of traders based both on a literature review and an industry screening via interviewing industry representatives. Institutional Investors can utilize the presented results as a structural approach for implementing their own, individual OCM strategy.

To achieve this goal, the remainder of this article is organized as follows: Section 2 presents a brief overview of related work. Based on this, section 3 introduces the concept of OCM by outlining its strategic and operational aspects, their interdependencies, as well as by identifying their key decision parameters. Then, section 4 illustrates the day-to-day handling of operational OCM that maps particular orders to suitable execution venues. Finally section 5 concludes.

2 Related work

On a conceptual level, the overall set of available strategic and operational decision parameters for Institutional Investors and their interdependencies have not been investigated yet. Academic literature focuses on rather specific aspects of the securities trading value chain like *trading styles*, empirical analysis of *markets*, *execution quality* and *order routing decisions*, *market models* as well as *execution costs*:

Behavior of institutional traders, their *trading styles* and related transaction costs are analyzed e.g. in [KeMa97], using proprietary transaction data. There, various hypotheses regarding trading characteristics like the choice of order type, trade duration and immediacy demand are validated. Further, differences between trades initiated by value, index and technical investors are outlined. Focus on informed investors' order types and trading patterns is drawn by [LeFL01, LLRS04, AnWe04, BIOS05, AnCM05], where evidence for the application of hidden limit orders as well as their performance are presented. The impact of order aggressiveness on execution performance is investigated in [GSTW00].

Markets and the dimensions of execution quality and costs are addressed by [BaHo01]. Comparisons of European markets include an analysis of trading costs at the Paris Bourse and London's SEAQ-I [dJNR95]. Similar comparisons for US markets are provided by [BaHJ00, HaBJ01], containing an analysis of market order execution quality [Boeh05] after the introduction of decimals. An overview of the upstairs market for trading of block orders at Paris Bourse is given by [BeVe04]. Altogether an apparent trade-off between costs and execution speed is revealed, emphasizing the demand for models with multiple dimensions of execution quality.

Other investigations address *order flow* and *order routing decisions*. Indications of order flow stickiness to venues despite changes in transaction costs can be found in [AhCC98]. Opportunities to strategically route limit orders to improve execution quality are shown by [BGHJ02]. The negative impact of order flow fragmentation on market quality is depicted in [BeWe06]. Further, a competition-for-order-flow model based on liquidity provision is presented by [PaSe03]. These research results outline the importance of non-price dimensions for execution quality.

Beside these findings, related research also focuses on different *market models*, e.g. the central limit order book [BiHS95, GrHR03], the convergence to order-driven markets in Europe [DeFo98] and the relative advantages of floor versus electronic trading systems [KeKo98]. Market experiments for a comparison of call markets, continuous auctions and dealer markets are conducted by [Thei00]. Performance improvement of floor-based trading systems through information sharing among floor brokers can be found in [FoLe01].

Institutional *execution costs* across major US exchanges are compared by [JoLi99], suggesting that institutions consider characteristics of the used markets. A dynamic model of an order-driven market populated by discretionary liquidity traders that have to trade but can choose their strategy is developed by [FoKK01]. This generates a set of predictions on the relation between market parameters, time to execution and spreads. A model for strategic trading is developed by [HoRa02], where traders have to learn about liquidity from past prices and trading volume. The model implies that strategic trades and market statistics are path-dependent on past market outcomes. The decision of traders to supply or to demand liquidity in a limit order market is modeled by [HMSS02]. Simulations of alternative trading strategies based on a detailed data set from a large US investor indicate that the strategy of initially trying to cross all stocks is cost effective [NaSk03]. A look at best execution obligations can be found in [McC104].

Beside these rather singular investigations, the contribution of this article can be compared best to the work of Wagner describing a hierarchy of trading decisions [Wagn06]. The framework derived in this article goes beyond Wagner’s operational decision tree model as it creates a generic setup including a strategic level. Another related article that is focused on the operative level is [YaJi06] where a quantitative approach for the selection of the most suitable Algorithmic Trading Solution is derived.

3 Introducing the concept of Order-Channel Management

With the evolution of new execution opportunities, the security trading industry has undergone massive changes in recent years. Order execution transforms itself from a broker intermediated market access to one which is controlled mainly by electronic means at the buy-side trading desk. Furthermore, new execution venues (e.g. Alternative Trading Systems), trying to meet the requirements of institutional order flow, have been launched. Altogether these changes offer Institutional Investors potential for cost-savings and improvements in order execution quality.

Definition: Order-Channel Management

Order-Channel Management (*OCM*) is the process of information gathering, evaluation, decision and control of Institutional Investors concerning the setup of the overall trading infrastructure (strategic OCM) and the actual order routing implementation (operational OCM).

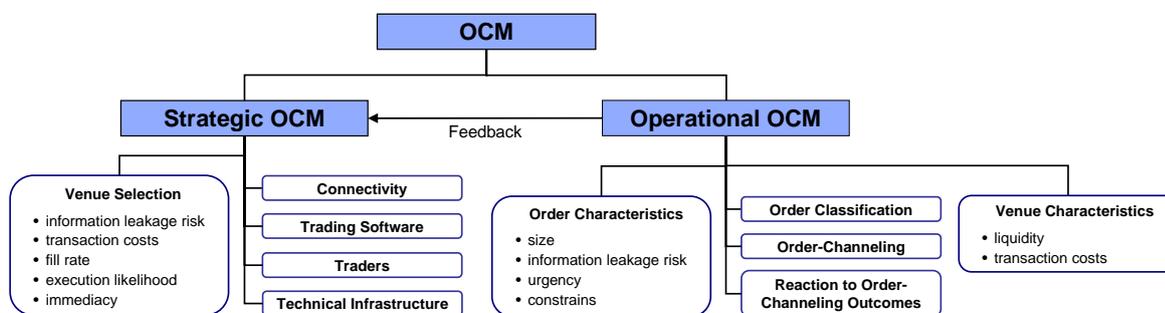


Figure 1: Decomposition of Order-Channel Management responsibilities.

OCM focuses on two interdependent levels that are depicted in figure 1: First, on the strategic level the focus is laid on a pre-selection from a pool of accessible venues. For this purpose, an introspection of the investment strategy is required in order to identify the expected order flow which provides the basis for the pre-selection. Further, within the strategic level the required personnel skills of traders, the technical and trading software infrastructure for the usage of new execution concepts like Algorithmic Trading and the connectivity to sell-side² companies and to markets have to be determined and set up. Second, within the environment defined by

²Sell-side refers to firms that trade for customers and earn money with fees, commissions and research [Harr03].

strategic OCM, on the operational level the actual routing of orders to the pre-selected venues has to be managed on an order by order basis. This is established by a comparison of venue characteristics (e.g. liquidity and transaction costs) with actual order parameters (e.g. size, information leakage risk and urgency) as well as execution constraints to be fulfilled. In order to achieve sound routing decisions, an analysis of order characteristic by combining Pre- and Post-Trade Analysis shall be incorporated. These two types of analyses provide important feedback information for future adjustments of the decisions within the strategic level.

3.1 Strategic Order-Channel Management

Traditionally, the infrastructure setup of Institutional Investors for the implementation of their investment decisions refers to their business relationships to brokers. The buy-side traders are responsible for order specifications, order releases to brokers and for phone-based over-the-counter trading, while brokers execute these orders at exchanges or OTC. New execution venues and access channels as well as Information Technology (IT) solutions expand the decision set and thus require a structured approach. Strategic decisions include the overall setup of a trading desk as well as its equipment, technological choices (e.g. usage of DMA or Algorithmic Trading), relationship management with execution brokers and the selection of execution venues as targets for actual investment decisions. This involves a make-or-buy decision for infrastructure provision and the services necessary to setup execution channels.

Definition: Strategic Order-Channel Management

Strategic Order-Channel Management (*strategic OCM*) is the process of information gathering, identification, selection and decision for implementation regarding execution venues, their connectivity, trading software, traders as well as the stipulation of technical infrastructure.

Figure 2 on the next page gives an overview of the aspects of strategic OCM that are described in detail in the following sections. Within the considerations of strategic OCM the investment strategy of a buy-side company is an important factor influencing the outcome of trades as well as execution costs [KeMa97] and thus has to be reflected in the strategic managerial decisions. Therefore, the first step is to determine the expected order flow via analysis of historical order data, interviewing fund managers or an introspection of quantitative investment models. Further, insights to venue performance collected by Post-Trade Analysis give important feedback for the selection of the trading setup and for its evaluation as well as future adjustments.

3.1.1 Execution Venues

The value propositions of execution venues regarding *information leakage risk*, *transaction costs*, *fill rate*³, *execution likelihood* and *immediacy* are of high importance for the strategic se-

³Fill rate refers to the amount of purchased (sold) shares in relation to order size [Harr03].

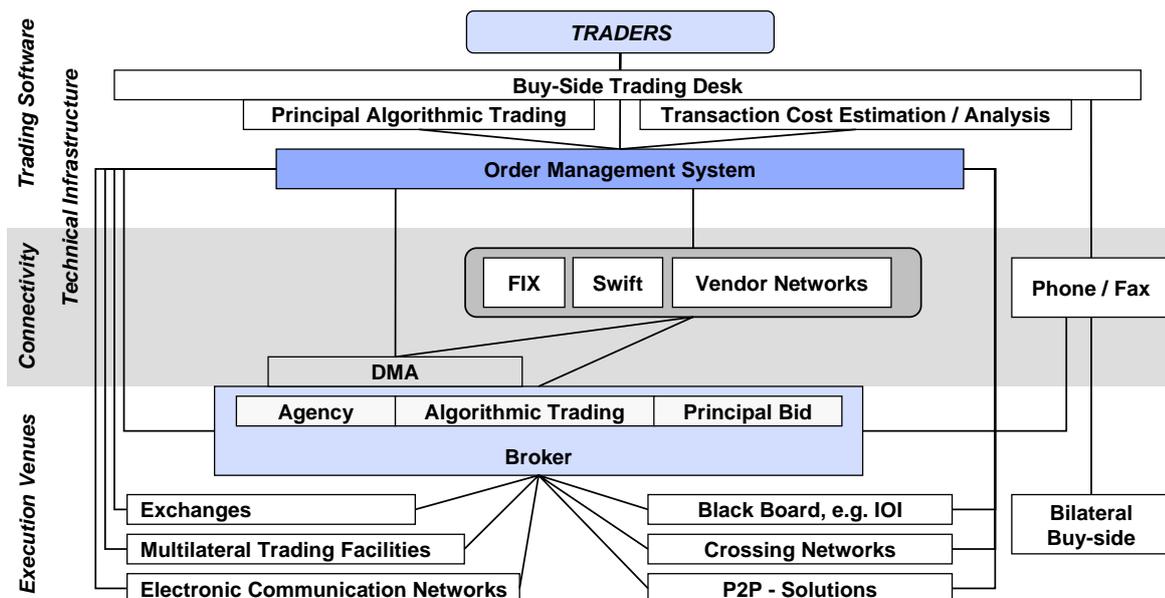


Figure 2: Decision Parameters in Strategic Order Channel Management

lection of venues by Institutional Investors. *Information leakage risk* is addressed by anonymity in one or several phases within a transaction and other functionalities that are offered to camouflage the information provided by order size (e.g. via iceberg orders) [Chak01]. Parameters like low fees and commissions as well as a high level of liquidity⁴ [O’Ha98, Kind05] decrease *transaction costs* and are therefore an important proposition to reduce the implementation shortfall. Liquidity also influences *fill rates* and the *likelihood of execution* that is tightly related to order routing and counterparty search.

Further, *immediacy* for executable orders is determined by the degree of automatization as well as the access channel, e.g. electronic access to an electronic market enables faster execution.

The first traditional channel is direct trading, i.e. to trade *bilaterally with another buy-side investor*, which typically involves direct communication between the two trading desks. A drawback of this solution are its search costs within the order routing phase, leading to slow execution and high internal costs of manual processing through negotiation and reconciliation.

Another traditional execution path is the delegation of orders to an agency *broker* who acts on behalf of the investor. The broker chooses a venue that is available to him or identifies trading opportunities with other brokers. An advantage of this channel is the specialization and know-how of the brokerage company improving fill rate and execution likelihood. Its disadvantage is increased information leakage risk which is caused by interest conflicts that arise from broker relationships to multiple investors. On top of that, this kind of execution provides lower immediacy because orders are worked successively throughout a trading period.

There exists also the possibility of a *principal bid*, where a sell-side broker guarantees full execution of an order at a given price for a negotiated commission. However, as the commissions

⁴Schwartz defines liquidity as the ability to trade whenever one wants to trade [ScFr04]. A comprehensive overview of liquidity measures can be found in [Kind05].

or the net price provided by brokers compensates them for taking the position as well as the risk on their books, the transaction costs tend to be higher than those for other channels [KiGI03]. To overcome especially the transaction cost issue of these bilateral solutions, buy-side investors can use electronic venues. A first alternative is provided by black board tools, i.e. Indication of Interest (IOI) messages, which allow to locate counterparties' willingness to trade in a particular stock. However, as IOIs only represent indications rather than executable quotes, likelihood of execution is low. More advanced solutions are Crossing Networks like Posit, which are able to match large order sizes by applying closed order books and without dismantling the investment interest to other parties. As Crossing Networks execute orders at the midpoint imported from a reference market and therefore without any market impact, transaction costs are comparably low. Nevertheless, these closed order books have very limited likelihood of execution as well as fill rates. An extension of Crossing Networks is offered by Liquidnet which searches for liquidity using a Peer-to-Peer approach. Once the size on the opposite side has been found, both investors are informed and can bilaterally and anonymously negotiate trade price and volume. Further, orders can be executed on exchanges, Multilateral Trading Facilities (MTFs) or Electronic Communication Networks (ECNs). Today's exchanges and MTFs enable electronic access and fast, automated execution. Additionally, these markets offer the possibility to forecast the execution price of orders. Unfortunately, exchanges and MTFs incur higher transaction costs, which attributes especially to market impact and immediacy costs. Therefore, market participants slice their orders over time in order to exploit market resiliency. *Algorithmic Trading* [GoGW05] can be distinguished as another execution venue, which can be developed in-house, bought from a third party or used as a service from a sell-side provider. Finally, the use of sell-side connectivity for *Direct Market Access* to exchanges while retaining the trading strategies at the buy-side should also be taken into consideration.

3.1.2 Connectivity

Various *connectivity* options exist that enable Institutional Investors to place their orders and receive execution confirmations. Some of them are based on industry standards and are independent from actual execution venues whereas others are proprietary to the respective venues. This requires an own infrastructure, causing operational costs, membership fees and data subscriptions. Standardized connectivity solutions are e.g. the Financial eXchange Protocol (FIX) and third party connectivity infrastructure like S.W.I.F.T's secure IP network to manage the various channels and to transport orders to various venues. Further, as mentioned before, sell-side connectivity like DMA can be used. These connectivity options can be combined depending on the selected channels.

3.1.3 Trading software

Trading desks can utilize various software solutions with different features. The basic software are 'plain vanilla' trading screens that are often offered by venues at no additional costs to the access fee. These solutions provide core functions like order entry, receipt of execution status as well as single venue market data. More advanced solutions are Order Management Systems (e.g. Sungard Decalog or Simcorp Dimension) which allow for integrating multiple venues within a single front-end and additionally offer features for inventory management on quantity and value basis as well as reporting functions. Sophisticated software suites include e.g. Algorithmic Trading engines, Pre-Trade Analysis tools for the prediction of transaction costs, volatility and liquidity development based on historical data as well as tools for position and risk management.

3.1.4 Traders

Even with sophisticated IT support, the need for experienced human traders will prevail. Especially for large orders or orders in illiquid securities human traders provide additional value. This highly skilled type of staff executes strategies for more difficult orders. Additionally, their experience is used to parameterize existing software and to develop new automated strategies. The number and skill level of traders is directly linked to the choice of execution venues, i.e. when DMA or Algorithmic Trading are used, significantly higher trader skills are required than in case of brokers as primary channels.

3.1.5 Technical infrastructure

The infrastructure for trading consists of generic information system components. Because the data processed by these systems represents monetary value, these components have to meet high quality standards. Further constraints for infrastructure result from venues which define authorized components (e.g. network components). Additionally, infrastructure has to meet the criteria of scalability, performance, security and reliability. Especially automated venues like Algorithmic Trading require high computing power to handle real-time market data, which leads to bandwidth requirements in order to ensure real-time data receipt and processing. For example, 1.2 billion trades have been executed over the course of 252 trading days in 2005⁵ on one exemplary venue, the New York Stock Exchange, leading to a corresponding number of updates to be processed by the investor's infrastructure, if this venue is used regularly. Thus an infrastructure for multiple venues represents a significant IT investment as well as the corresponding total costs of ownership.

⁵See statistics at World Federation of Exchanges <http://www.world-exchanges.org>.

3.2 Operational Order-Channel Management

Based on the setup defined by the strategic OCM process, *operational OCM* provides a framework for the actual order-channeling decision on an order by order basis. Therefore, it outlines order characteristics that can be utilized for the identification of suitable venues and access strategies according to the venue characterization outlined in section 3.1.1. Further, it provides important feedback about the performance of each venue that is to be considered in future adjustments of the setup established by strategic OCM. As the individual decisions within operational OCM are supposed to reflect *order constraints*, operational OCM is a constrained optimization process.

Definition: Operational Order-Channel Management

Operational Order-Channel Management (*operational OCM*) is the decision process concerning the execution of individual orders reflecting order characteristics, constraints as well as access strategies based on the setup established by strategic OCM.

For the identification of relevant *order characteristics* the focus is drawn on the investment cycle depicted in figure 3, where an order represents the outcome of an investment decision. Because order execution is supposed to take place at venues which meet specified requirements and at the same time are cost-effective, additional information is required, specifying the actual execution characteristics for each order. Within the investment cycle this information can be provided by a combination of Pre-Trade Analysis concerning the venue accessible from the strategic OCM setup and a Post-Trade Analysis after the completed trade [KiG103].

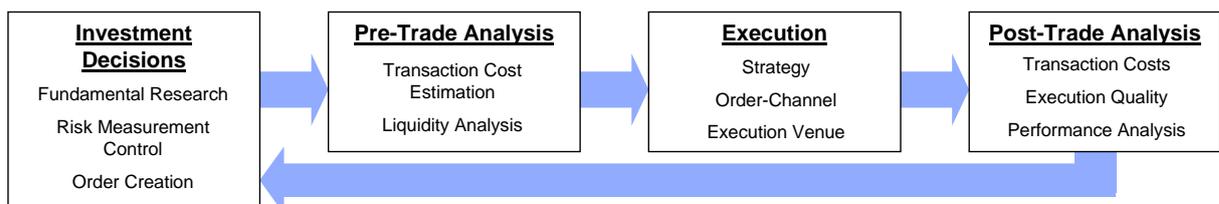


Figure 3: Investment Cycle adopted from [Madh02]

3.2.1 Key characteristics for operational Order-Channel Management

As actual transaction costs play a crucial role for execution success, they constitute the starting point for further investigation. These costs are defined as those associated with the implementation of the investment decision [KiG103]. They can be further split up into visible costs (e.g. commissions, fees, taxes, spreads) and hidden costs like price appreciation⁶, market impact, timing risk and opportunity costs. Hidden costs make the lion share of overall costs, especially

⁶Accordingly to Kissel price appreciation represents the costs of the natural price movement without market impact, i.e. the cost to buy (sell) in a rising (falling) market [KiG103].

Order Size	Difficulty Level
0 – 5%	Easy: one day
5 – 15%	Relatively Easy: one day with some work
15 – 25%	Difficult: may require multiple days
25% and more	Very Difficult: recommend multiple days

Table 1: Order Size and Difficulty Level from [KiGl03]

when it comes to large orders or block trades [ScFr04]: First of all, large trades create market pressure and thus lead to market impact e.g. because they sweep the book in an order-driven market. One common technique to avoid this market impact is to slice a large order and to trade more passively over multiple periods [KiGl03]. This slicing solution leads to other problems. The enlarged trading period leads to timing risk and an increased risk for price appreciation as prices can develop in an unfavorable direction. Furthermore, the motive can become obsolete or, because of information leakage, other market participants might anticipate the order and perform front-running. Thus, *order size*, *information leakage risk* and the *level of execution urgency* are the relevant characteristics for *operational OCM*.

The crucial role of *order size* arises from the fact that market impact costs are a convex function in order size [BSHvdS06]. This convexity can be outlined by liquidity-measures like the eXchange Liquidity Measure (XLM)⁷. Typically these measures quantify the round trip⁸ costs for a specific order size in a security [GoST04]. For instance the round trip costs on XETRA for the DAX listed security Deutsche Bank in June 2006 are 1.4 *bps* for a 100k € order and raise up to 33.3 *bps* for a 2mn € order⁹. As the assessment of *order size* depends on the liquidity of the traded instrument, an estimation of the market's ability to execute the desired *order size* with little or no price movement should incorporate market liquidity statistics like XLM mentioned above. Liquidity is also strongly related to market capitalization [KiGl03]. The most common measure for *order size* that allows comparisons among different securities is the Average Daily Volume (ADV) [KiGl03]. Table 1 depicts different size categories as well as their implications for order execution.

Depending on the information other market participants can collect about the order as well as its motivation the risk of front-running arises. Hence, the *information leakage risk* is tightly related to the motive of the order. If it is initiated by liquidity motives like cash in and out flows or the requirement to track an index it will face lower *information leakage risk*. For orders based on private information the situation is contrary. Orders issued by Institutional Investors rebalancing portfolios accordingly to their research results encounter higher *information leakage risk* because their private information might be figured out. Thus, informed traders, especially prominent Institutional Investors, have to pay appropriate attention to *information leakage risk* in order to avoid other market participants gaining profits from their trading.

⁷XLM is a trademark of Deutsche Börse Group.

⁸A round trip is a purchase and immediate sale of a particular security or vice versa [GoST04].

⁹Data provided by Deutsche Börse Group.

Factor	Traders	CIOs	Factor	Weight
Lowest execution costs	3.53	3.39	Little or no market impact	3.95
Rewarding good research	3.39	3.42	Speed	3.42
Fastest possible execution	3.37	3.24	Not revealing the full size of order to market	3.40
Soft commission obligations	2.45	2.44	Not revealing the identity of company or fund	3.21
Portfolio manager direction	2.39	N/A	Within the current market inside spread	3.06
			Price better than the VWAP	2.93
			Low or no commission	1.29

Scale: 1 (never) to 5 (very frequently, or 75 to 100 percent of the time)

Table 2: On the left key factors determining how institutions choose brokers are highlighted. The table on the right presents factors important to chief investment officers in judging the quality of execution for large orders. Both tables are adopted from [ScFr04].

Several studies reveal that Institutional Investors commonly possess only trading-related reasons for urgent orders, but in fact do not receive immediacy [ScSt02, ChLa95]. For instance Chan and Lakonishok find out that only about 20% of the value of institutional buy orders are completed within one day, and less than half within four days [ChLa95]. Thus, the *execution urgency level* is also tightly related to information leakage. In this context, most attention is paid to the estimated time that is necessary for the motivation of a trade to become public knowledge [KiGl03]. Hence trades initiated by transient, private information are executed with higher urgency because this allows exploiting knowledge before it is reflected by market prices. This holds especially for human intermediated markets because of the risk that intermediaries like agency brokers might inform other clients about trading intentions [ScFr04]. Furthermore, the expected price appreciation has to be considered because it might convey information about the momentum of the security to be traded. Finally this information can be enriched by statistics for volatility as well as stock classifications like the affiliation to momentum or value growth stocks or the membership to indices that can lead to rapid price movements preceding index reconstructions [KiGl03]. The importance of the three order dimensions *order size* for market impact costs, *information leakage risk* as well as *urgency* in form of execution speed are further stressed by the results of the Schwartz and Steil survey [ScSt02] concerning the assessment of execution quality by chief investment officers which is depicted on the right of table 2.

3.2.2 Order classification within operational Order-Channel Management

With the three classification parameters at hand, we group orders to a total of six classes depicted on the next page in table 3. In a first step small orders are separated because they require less care to prevent market impact. Among these *low touch orders* there is no need for differentiation by information leakage risk because of their low information content. Thus only two further subcategories remain: The first contains *passive low touch orders* with a low level of execution urgency, that can be implemented via liquidity providing means like limit orders. The second class constitutes *active low touch orders* with a higher level of execution urgency which

		leakage risk			
		low	high	low	high
size	small	passive low touch order		active low touch order	
	large	strategic trading	high touch order	not applicable	urgent high touch order
		low urgency		high urgency	

Table 3: Characterization of orders. On the left the order dimension *low urgency* is highlighted, whereas on the right the order dimension *high urgency* is presented.

implies active trading. Among large orders with a low level of execution urgency two additional classes can be specified: Orders with a low leakage risk belong to the class of orders resulting from *strategic trading* like those for share buy-back programs. Orders with higher leakage risk constitute the class of *high touch orders* because they require much attention during their implementation. Finally large orders with a high level of execution urgency can be subdivided into two additional categories: While the parameter setting of high urgency and low leakage risk is reasonably not existent (*not applicable*), orders with a high leakage risk constitute to the class of *urgent high touch orders* being the toughest order type.

3.2.3 Order constraints

Orders might also incorporate constraints which can be constituted already by the order data itself. These constraints narrow the number of possible execution paths or trading venues as well as the available trading models for a stock. In this context benchmarks like the Volume Weighted Average Price (VWAP) or Arrival Price (AP) are common parameters which are used especially by buy-side companies for an internal execution evaluation or when routing their orders to brokers. Reference prices are also used to measure implementation shortfall [Lehm03], but are not without critique [ScWo03], especially when the overall market moves in an unfavorable direction. Finally, orders can contain restrictions in form of predefined execution styles (e.g. provided by the portfolio managers of an investment fund company), dependencies on venues and proprietary handling instructions. For example in the left part of table 2 the fact is outlined that traders direct 26% of their orders to brokers as a means of rewarding them for non-trading related services like good research [ScSt02].

4 Illustration of actual order handling in operational Order-Channel Management

For the day-to-day handling of operational OCM we propose a subdivision into three phases that are passed by each order. First, *orders have to be classified* according to the three key order characteristics as already depicted in section 3.2.2. Second, the actual *order-channeling*

decision has to take place. In the final step order execution should be controlled which enables *reactions to observed outcomes*.

4.1 Order classification

The first step is the concrete classification of orders based on the three order characteristics defined in section 3.2.1 and the scheme depicted in section 3.2.2. For this purpose, rules as well as processes should be established that try to achieve a non-overlapping segregation. Here, IT infrastructure plays a crucial role for the processing of market data like daily volumes (e.g. for the comparison of order size to ADV), volatility as well as liquidity which are all required within the order classification.

A tight coordination between fund management and the trading desk enables an integrated optimization of this part of the security value chain and thus helps to achieve higher trading success. First, it increases the traders' insight to the motivation of investment decisions which helps them to quantify and mitigate the information leakage risk more appropriately. Second, as the emphasis concerning short and long-term alphas becomes available to the trading desk also the determination of the level of execution urgency should improve.

Together with further instructions (e.g. target trading strategies), a Pre-Trade Analysis should be performed and orders are supposed to be mapped to the corresponding classes. During this step, IT-based tools like artificial neural networks might also be incorporated in order to enforce automation in the classification process. Depending on the class, the number of suitable as well as usable venues is narrowed. The final mapping on an order by order basis and the 'channeling' is done in the following step.

4.2 Actual order-channeling

The actual order-channeling depends on the assignment of individual orders to the order classes. Processing of large *order size* requires usage of multiple liquidity pools. Therefore, state-of-the-art technology allows liquidity consolidation concerning location and time. The former is enabled by advanced Smart Order Routing software seeking hidden liquidity pools [HaId03] whereas the latter can be accomplished by manual as well as automated slicing strategies. In this context, Domowitz and Yegerman have shown that current Algorithmic Trading solutions are not suitable for all kinds of orders yet, as their investigations have identified a performance breakdown for order sizes above 10% of the ADV [DoYe05].

A common strategy to reduce the *information leakage risk* is to hide the complete trading interest or to show only smaller parts. This can be established by using stealth trading techniques that are supported by agency brokers or Algorithmic Trading. Another applicable technique is to select venues which offer an appropriate value proposition, e.g. pre-trade, trade as well as post-trade anonymity.

Finally, for orders with a high *level of execution urgency* a Pre-Trade Analysis shall be utilized to calculate or at least estimate a trade-off between immediacy and opportunity costs and thus to determine an optimized execution strategy.

Based on these general remarks, we focus on the order/strategy types identified in section 3.2.2. *Passive low touch orders* allow the usage of all venues. To optimize the achieved price and therefore trading revenues, passive strategies via limit orders or venues providing price improvement opportunities might be incorporated. In contrast, *active low touch orders* require a more aggressive execution via market or marketable limit orders¹⁰ on venues offering immediacy. Further, *strategic trading* can also benefit from a passive realization throughout a longer period across several venues. As a single release of a *high touch order* would cause a significant market impact, one might initially try to cross them [NaSk03] and if this fails use slicing techniques, splitting the actual order into e.g. hourly or daily packages that can be handled similar to low touch orders. However, an extended execution period bears the risk of opportunity costs and thus has to be continuously tracked. As buy orders tend to convey more information than sell orders [ChLa95], their implementation should incorporate additional techniques to reduce information leakage. Finally for *urgent high touch orders*, constituting the hardest type, automated strategies are not suitable yet. Instead, the full order size or at least large parts of these orders shall be delegated to a broker to whom a trusted relationship has been build up and who either provides a principal bid or who is sophisticated enough to work the order or to 'smoke out' desired liquidity via IOIs within the given time frame.

4.3 Reaction to order-channeling outcomes

As, due to their size, orders from Institutional Investors are far from fire-and-forget tasks *continuous tracking* till their final completion is required. A *readjustment* of a stealth execution strategy becomes necessary when information disseminates or the order cannot be filled. Further exceptional market changes require also strategy reviews. Under such circumstances brokers typically inform their clients while some automated solutions might fail to achieve this and thus require manual tracking and intervention capabilities.

Beside strategy revisions, order-channeling outcomes should be incorporated in a comprehensive *Post-Trade Analysis* that evaluates execution quality relative to the predefined price benchmarks and adjusts the parameters for the actual strategy selection. Simulations based on historical market data allow to evaluate alternative channels. An example in this context is the Penn-Lehman Automated Trading Project that uses real-time data from US ECNs for the investigation of automated trading strategies [KeOr03].

¹⁰buy (sell) order with a limit equal or above (below) the best offer(bid)

5 Conclusion

For Institutional Investors, new technology-driven execution opportunities allow for self-directed trading and a greater independence from their brokers, their traditional channels for order execution. Thus, the complexity of their trading desks' tasks and infrastructure increases as they face upcoming execution venues, technology developments as well as new trading strategies. The management of this complexity requires a structured approach.

Our paper extends the existing literature on Institutional Equity Trading by introducing the concept of *Order-Channel Management* (OCM) providing a framework for Institutional Investors both on a *strategic* and on an *operational level* (section 3). First, strategic OCM addresses management issues regarding execution venues, connectivity, trading software, traders as well as technical infrastructure (section 3.1) and thus provides the framework for operational OCM on an order by order basis in daily operations. For the latter we have introduced a classification scheme that maps orders into five classes along the three dimensions of order size, information leakage risk and level of execution urgency (section 3.2). Finally, we have outlined how operational OCM can be implemented within three phases (section 4).

As a future research topic, we will empirically validate our framework - that was derived based on bilateral interviews and industry screening - via a series of structured case studies. Further, we intend to analyze strategic and operational topics like Pre- and Post-Trade Analysis for the evaluation of execution quality, especially on multiple venues.

Acknowledgment:

We thankfully acknowledge the support of the *E-Finance Lab*, Frankfurt for this work.

References

- [AhCC98] *Ahn, H.J.; Cao, C.Q.; Choe, H.*: Decimalization and competition among stock markets: Evidence from the Toronto Stock Exchange cross-listed securities. In: *Journal of Financial Markets* 1, (1998), pp. 51–87.
- [AnCM05] *Anand, A.; Chakravarty, S.; Martell, T.*: Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders. In: *Journal of Financial Markets* 8, (2005), pp. 289–309.
- [AnWe04] *Anand, A.; Weaver, D.G.*: Can order exposure be mandated? In: *Journal of Financial Markets* 7, (2004), pp. 405–426.
- [BaHJ00] *Battalio, R.; Hatch, B.; Jennings, R.*: Dimensions of Best Execution for Market Orders: Assessing Differences between the NYSE and the Nasdaq Third Market. Working Paper, 2000.

- [BaHo01] *Battalio, R.; Holden, C.W.*: A simple model of payment for order flow, internalization, and total trading cost. In: *Journal of Financial Markets* 4, (2001), pp. 33–71.
- [BeVe04] *Bessembinder, H.; Venkataraman, K.*: Does an electronic stock exchange need an upstairs market? In: *Journal of Financial Economics* 73, (2004), pp. 3–36.
- [BeWe06] *Bennett, P.; Wei, L.*: Market structure, fragmentation, and market quality. In: *Journal of Financial Markets* 9, (2006), pp. 49–78.
- [BGHJ02] *Battalio, R.; Greene, J.; Hatch, B.; Jennings, R.*: Does the Limit Order Routing Decision Matter? In: *The Review of Financial Studies* 15, (2002), pp. 159–194.
- [BiHS95] *Biais, B.; Hillion, P.; Spatt, C.*: An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse. In: *The Journal of Finance* 50, (1995), pp. 1655–1689.
- [BIOS05] *Bloomfield, R.; O’Hara, M.; Saar, G.*: The ”make or take” decision in an electronic market: Evidence on the evolution of liquidity. In: *Journal of Financial Economics* 75, (2005), pp. 165–199.
- [Boeh05] *Boehmer, E.*: Dimensions of execution quality: Recent evidence for US equity markets. In: *Journal of Financial Economics* 78, (2005), pp. 553–582.
- [BSHvdS06] *Bikker, J.; Spierdijk, L.; Hoevenaars, R.; van der Sluis, P.J.*: Forecasting market impact costs and identifying expensive trades. Working paper, 2006.
- [Chak01] *Chakravarty, S.*: Stealth-trading: Which traders’ trades move stock prices? In: *Journal of Financial Economics* 61, (2001), pp. 289–307.
- [ChLa95] *Chan, L.K.C.; Lakonishok, J.*: The Behavior of Stock Prices Around Institutional Trades. In: *The Journal of Finance* 50, (1995), pp. 1147–1174.
- [DeFo98] *Demarchi, M.; Foucault, T.*: Equity Trading Systems in Europe. A survey of recent changes. Working Paper, 1998.
- [dJNR95] *de Jong, F.; Nijman, T.; Röell, A.*: A comparison of the cost of trading French shares on the Paris Bourse and on SEAQ International. In: *European Economic Review* 39, (1995), pp. 1277–1301.
- [DoYe05] *Domowitz, I.; Yegerman, H.*: The Cost of Algorithmic Trading: A First Look at Comparative Performance. In: *Algorithmic Trading - Precision, Control, Execution* , (2005), pp. 26–34.
- [FoKK01] *Foucault, T.; Kadan, O.; Kandel, E.*: Limit Order Book as a Market for Liquidity. Working Paper, 2001.
- [FoLe01] *Foucault, T.; Lescourret, L.*: Information Sharing, Liquidity and Transaction Costs in Floor-Based Trading Systems. Working Paper, 2001.
- [GoGW05] *Gomber, P.; Gsell, M.; Wrانik, A.*: Algorithmic Trading - Maschinen auf Finanzmärkten. In: *Die Bank Sonderausgabe zur E.B.I.F.*, (2005), pp. 40–45.

- [GoSe05] *Gomber, P.; Seitz, J.*: Neue Transparenzregeln für den Wertpapierhandel in Europa. In: DIW - Quarterly Journal of Economic Research , (2005), pp. 153–166.
- [GoST04] *Gomber, P.; Schweickert, U.; Theissen, E.*: Zooming in on Liquidity. In: 31st Annual Meeting of the European Finance Association, 2004.
- [GrHR03] *Grammig, J.; Heinen, A.; Rengifo, E.*: An analysis of order submissions on the Xetra trading system using multivariate time series of counts. Working Paper, 2003.
- [GSTW00] *Griffiths, M.D.; Smith, B.F.; Turnbull, D.A.S.; White, R.W.*: The costs and determinants of order aggressiveness. In: Journal of Financial Economics 56, (2000), pp. 65–88.
- [HaBJ01] *Hatch, B.; Battalio, R.; Jennings, R.*: Post-Reform Market-Order Execution Quality: Multidimensional Comparisons Across Market Centers. In: The Financial Review 38, (2001), pp. 123–152.
- [HaId03] *Hallam, N.; Idelson, N.*: Breaking the Barriers A Technological Study of the Obstacles to Pan-European Best Execution in Equities. Tech. rep., Traderserve Limited, 2003.
- [Harr03] *Harris, L.*: Trading and Exchanges: Market Microstructure for Practitioners. Oxford University Press, New York, 2003.
- [HMSS02] *Hollifield, B.; Miller, R.A.; Sandas, P.; Slive, J.*: Liquidity Supply and Demand in Limit Order Markets. CEPR Discussion Paper No. 3676, 2002.
- [HoRa02] *Hong, H.; Rady, S.*: Strategic trading and learning about liquidity. In: Journal of Financial Markets 5, (2002), pp. 419–450.
- [JoLi99] *Jones, C.M.; Lipson, M.L.*: Execution Costs of Institutional Equity Orders. In: Journal of Financial Intermediation 8, (1999), pp. 123–140.
- [KeKo98] *Kempf, A.; Korn, O.*: Trading System and Market Integration. In: Journal of Financial Intermediation 7, (1998), pp. 220–239.
- [KeMa97] *Keim, D.B.; Madhavan, A.*: Transaction costs and investment style: an inter-exchange analysis of institutional equity trades. In: Journal of Financial Economics 46, (1997), pp. 265–292.
- [KeOr03] *Kearns, M.; Ortiz, L.*: The Penn-Lehman Automated Trading Project. In: IEEE Intelligent Systems , (2003), pp. 22–31.
- [KiGl03] *Kissell, R.; Glantz, M.*: Optimal Trading Strategies: Quantitative Approches for Managing Market Impact and Trading Risk. AMACOM, New York, 2003.
- [Kind05] *Kindermann, S.*: Liquiditäts- und Effizienzmessung im Aktienhandel. Deutscher Universitätsverlag , 2005.
- [LeFL01] *Lee, Y.T.; Fok, R.C.; Liu, Y.J.*: Explaining Intraday Pattern of Trading Volume from the Order Flow Data. In: Journal of Business Finance and Accounting, 28, (2001), pp. 199–230.

- [Lehm03] *Lehmann, B.N.*: What we measure in execution cost measurement. In: *Journal of Financial Markets* 6, (2003), pp. 227–231.
- [LLRS04] *Lee, Y.T.; Liu, Y.J.; Roll, R.; Subrahmanyam, A.*: Order imbalances and market efficiency: evidence from the Taiwan stock exchange. In: *Journal of Financial and Quantitative Analysis* 39, (2004), pp. 327–342.
- [Madh02] *Madhavan, A.*: Implementation of Hedge Fund Strategies. In: *Hedge Fund Strategies* Fall Issue, (2002), pp. 74–80.
- [McCl04] *McCleskey, S.*: *Achieving Market Integration. Best Execution, Fragmentation and the Free Flow of Capital.* Butterworth-Heinemann, Oxford, 2004.
- [NaSk03] *Naes, R.; Skjeltorp, J.A.*: Equity trading by institutional investors: Evidence on order submission strategies. In: *Journal of Banking and Finance* 27, (2003), pp. 1779–1817.
- [O’Ha98] *O’Hara, M.*: *Market Microstructure Theory.* Blackwell Publishers, 1998.
- [PaSe03] *Parlour, C.A.; Seppi, D.J.*: Liquidity-Based Competition for Order Flow. In: *The Review of Financial Studies* 16, (2003), pp. 301–343.
- [Rami06] *Ramistella, A.*: *Crossing Networks: Bringing Back Large Block Trades to Institutional Trading.* Tech. rep., Tower Group, Inc, 2006.
- [ScFr04] *Schwartz, R.A.; Francioni, R.*: *Equity Markets in Action: The Fundamentals of Liquidity, Market Structure & Trading.* Wiley Trading, New Jersey, 2004.
- [ScSt02] *Schwartz, R.A.; Steil, B.*: Controlling Institutional Trading Costs. In: *The Journal of Portfolio Management* Spring, (2002), pp. 39–49.
- [ScWo03] *Schwartz, R.A.; Wood, R.A.*: Best Execution - A candid analysis. In: *The Journal of Portfolio Management* Summer, (2003), pp. 37–48.
- [Thei00] *Theissen, E.*: Market structure, informational efficiency and liquidity: An experimental comparison of auction and dealer markets. In: *Journal of Financial Markets* 3, (2000), pp. 333–363.
- [Wagn06] *Wagner, W.H.*: Creating A Hierarchy of Trading Desicions. In: *The Journal of Trading* 1, (2006), pp. 6–11.
- [YaJi06] *Yang, J.; Jiu, B.*: Algorithm Selection: A Quantitative Approach. In: *Algorithmic Trading II - Precision, Control, Execution* , (2006), pp. 26–34.

IT-Architekturmanagement in Banken

Ergebnisse einer leitfadengestützten Expertenbefragung

Christian Schmidt, Peter Buxmann
Fachgebiet Information Systems
Technische Universität Darmstadt
Hochschulstr. 1, 64289 Darmstadt
{schmidt, buxmann}@is.tu-darmstadt.de

Zbynek Sokolovsky
iED Consulting GmbH
Karl-Hermann-Flach-Str. 36, 61440 Oberursel
zbynek.sokolovsky@ied-consulting.de

Abstract

In der Praxis setzt sich zunehmend die Auffassung durch, dass die Zukunftsfähigkeit komplexer Informationssystemlandschaften auf Dauer nur durch ein kontinuierliches Management auf Architekturebene zu gewährleisten ist. Viele Banken investieren daher verstärkt in den Aufbau eines entsprechenden Architekturmanagements (AM). In diesem Beitrag werden die Ergebnisse einer Expertenbefragung zum Entwicklungsstand von AM in der deutschsprachigen Kreditwirtschaft vorgestellt. Danach ist AM fast in allen befragten Häusern als eigene Funktion etabliert. Im Hinblick auf Ziele und Aufgaben existiert bisher jedoch noch kein einheitliches Verständnis. Dennoch sind in vielen Fällen bereits erste Erfolge sichtbar. Dies gilt vor allem im Hinblick auf ein nachhaltiges Komplexitäts- und Kostenmanagement.

1 Einführung

Viele Unternehmen verfügen heute über komplexe und heterogene Informationssystemlandschaften. Dies gilt in besonderer Weise für Banken, die aufgrund der Informationsintensität ihrer Produkte und Prozesse traditionell eine hohe Durchdringung mit Informationstechnologie (IT) aufweisen.

In der Praxis schlägt sich die Komplexität der bankbetrieblichen IT-Landschaften oft in erheblichen Wartungs- und Betriebskosten, einem hohen Änderungsaufwand (häufig mit mangelnder "Flexibilität" umschrieben) sowie letztlich in einem schlechten Business-IT-Alignment¹ nieder. Die Entwicklung der Bank-IT kann daher in vielen Fällen nicht mehr mit den bankpolitischen Anforderungen schritthalten.² Diese Problematik konnte bisher auch durch die Auslagerung von Teilen der Informationsverarbeitung sowie durch einen stärkeren Einsatz von Standardsoftware nicht grundsätzlich beseitigt werden.

In Banken setzt sich daher zunehmend die Auffassung durch, dass die Zukunftsfähigkeit der IT auf Dauer nur durch eine systematische und langfristig orientierte Planung und Steuerung auf Gesamtarchitekturebene gewährleistet werden kann. Viele Institute haben hierzu innerhalb der letzten Jahre in den Auf- bzw. Ausbau eines (IT-) Architekturmanagements (AM) investiert.

AM ist eine vergleichsweise junge Disziplin, für die sich bis heute keine allgemein anerkannten Standards herausgebildet haben. In der Praxis mangelt es daher oft an einem gemeinsamen Verständnis und einer klaren Orientierung. Eine theoretische Aufarbeitung des Themengebietes hat bislang nur in Ansätzen stattgefunden; zudem gibt es nur wenige empirische Untersuchungen zur Verbreitung und Ausgestaltung von AM in der Praxis sowie zu den dabei erzielten Wirkungen. Letzteres wiegt umso schwerer vor dem Hintergrund der oft erheblichen Anlauf- und Overheadkosten von AM.

Das vorliegende Forschungsvorhaben soll dazu beitragen, diese Lücke zu schließen. In einem ersten Schritt wurde daher eine Expertenbefragung in Unternehmen der deutschsprachigen Kreditwirtschaft durchgeführt. Ziel der Untersuchung war eine explorative Analyse des Entwicklungsstandes von AM in der Praxis. Dabei sollten insbesondere auch die bisherigen Erfahrungen im Hinblick auf Wirkungen, Probleme und Erfolgsfaktoren von AM ermittelt werden. Neben einer deskriptiven Bestandsaufnahme sollten die Untersuchungsbefunde darüber hinaus als Ausgangspunkt für die Formulierung von Hypothesen dienen.

Im Folgenden werden zentrale Ergebnisse der Untersuchung vorgestellt. Dabei wird in Abschnitt 2 zunächst ein Überblick zu den bereits bestehenden Forschungsarbeiten gegeben. Anschließend wird in Abschnitt 3 das Untersuchungsdesign erläutert. Es folgt die Darstellung der empirischen Befunde (Abschnitt 4). Der Beitrag schließt mit einer Diskussion der Ergebnisse sowie einigen Schlussfolgerungen.

¹ Der Begriff Alignment wird hier im Sinne von "Kohärenz zwischen Unternehmensstrategie und IT" gebraucht.

² So gelten die bestehenden Informationssystemlandschaften von Banken auch als ein wesentliches Hindernis für eine weitere Industrialisierung der Finanzwirtschaft.

2 Literaturübersicht

Das Themengebiet (Unternehmens-) Architektur ist traditionell durch eine gewisse begriffliche Unschärfe und Heterogenität gekennzeichnet. Eine Ursache hierfür kann in der vergleichsweise großen inhaltlichen Vielfalt gesehen werden, die sich auch in den unterschiedlichen Forschungsströmungen widerspiegelt. So stehen etwa bei den zahlreichen Ansätzen zur Unternehmensmodellierung [z. B. Sche91; FeSi95] integrierte Modelle des Unternehmens bzw. des Unternehmens-Informationssystems im Mittelpunkt der Betrachtung. "Unternehmensarchitekturen" werden hier primär als strategisches Planungsinstrument angesehen. Andere Strömungen fokussieren stärker auf Aspekte der Interoperabilität und Standardisierung [z. B. BYDH03]. Gemäß dieser Sichtweise werden Unternehmensarchitekturen als Rahmenwerke innerbetrieblicher Richtlinien und Standards für den IT-Einsatz aufgefasst. Breite Anwendung erfährt der Architekturbegriff schließlich auch auf dem Gebiet des Software Engineering [z. B. IEEE00]. Neben einer statischen Betrachtung rückt in Literatur und Praxis zunehmend das kontinuierliche Management von IT- bzw. Unternehmensarchitekturen in den Blickpunkt [Wint05]. Nach [Hafn05, 79ff.] kann ein solches *Architekturmanagement (AM)* als Teilfunktion des Informationsmanagements angesehen werden, welche die langfristige Steuerbarkeit der Informationssystemlandschaft sicherstellen soll.

Im Mittelpunkt der wissenschaftlichen Forschung standen in der Vergangenheit vor allem die Entwicklung von Architektur-Frameworks sowie entsprechender Modellierungsmethoden [z. B. Zach87; Sche91]. Empirische Untersuchungen sind dagegen bisher kaum zu finden. Diese sind zudem meist auf Teilaspekte wie die Analyse bestehender Ist-Architekturen [z. B. Mehl03], die Verbreitung bestimmter Architekturformen [z. B. BChP05] oder die Anwendung von Integrationstechnologien [z. B. AiSc05] beschränkt. Nur wenige Arbeiten haben sich bisher mit AM als Disziplin bzw. als betrieblicher Funktion beschäftigt. Dabei überwiegen Erfahrungsberichte aus der Unternehmenspraxis [z. B. Gaer04, Penz04]. Darüber hinaus wurden von einigen Forschern Fallstudien durchgeführt. So analysiert etwa [Hafn05] die Ansätze zum AM in drei Unternehmen der Finanzindustrie im Hinblick auf die dort eingesetzten Vorgehensmodelle. Ferner betrachtet [Pulk06] die Entscheidungsprozesse bei der Unternehmensarchitekturplanung von drei finnischen Unternehmen. Weiterhin untersuchen [MaGH04] den Einsatz von Enterprise Architecture Frameworks im Bereich der öffentlichen Verwaltung. [BYDH03] schließlich analysieren anhand von zwei Fallstudien aus der Finanzwirtschaft den Einfluss bestimmter organisatorischer Rahmenbedingungen auf die Erreichung der verfolgten Architekturziele.

Unter einer *Architektur* wird im Folgenden in Anlehnung an [IEEE00] die grundlegende Organisation eines existierenden oder geplanten Systems im Hinblick auf einen oder mehrere relevante Systemaspekte verstanden. Mit *IT-Architektur* wird die Architektur der gesamten Informationssystemlandschaft bezeichnet ("Architecture in the Large"). Die Begriffe *Enterprise Architecture* bzw. *Unternehmensarchitektur* werden dagegen im Sinne von "Architektur des Unternehmens" verwendet. Unter *AM* wird schließlich eine Teilfunktion des Informationsmanagements verstanden, welche die ganzheitliche Steuerung der IT-Architektur zum Gegenstand hat.

3 Forschungsdesign

Um der explorativen Zielsetzung der Untersuchung und der Heterogenität des Untersuchungsgegenstandes gerecht zu werden, wurde ein qualitativer Forschungsansatz auf Basis leitfadengestützter Experteninterviews [GILa04] gewählt. Die Identifikation der Experten erfolgte dabei mit Hilfe einer Suchmaschinen-Recherche. Auf diese Weise konnten insgesamt 28 Kandidaten ermittelt werden. Diese wurden anschließend per E-Mail über das Forschungsvorhaben informiert und um einen entsprechenden Interviewtermin gebeten. Von den angeschriebenen Personen erklärten sich 14 zu einer Studienteilnahme bereit.

Obwohl die Untersuchungsgruppe keine repräsentative Stichprobe darstellt, konnte mit ihr dennoch ein relativ breites Spektrum der deutschsprachigen Kreditwirtschaft abgedeckt werden. So haben insbesondere Vertreter aller drei großen deutschen Bankengruppen (Sparkassensektor, Genossenschaftssector, privater Sektor) an der Studie teilgenommen. Die Untersuchungsgruppe umfasste darüber hinaus drei Institute aus der Schweiz. Zu den Teilnehmern gehörten sowohl Universal- als auch verschiedene Spezialbanken; sechs Institute verfügten zum Zeitpunkt der Befragung über eine Bilanzsumme von mehr als 300 Mrd. €. Abgerundet wird das Spektrum durch drei auf Banken spezialisierte IT- bzw. Outsourcing-Dienstleister.³

Bei den Interviewpartnern handelte es sich meist um die jeweiligen AM-Leiter⁴ der betreffenden Unternehmen. In einigen Fällen waren die Teilnehmer auch Leiter von Teilfunktionen innerhalb des AM. Bei großen Konzernen existiert oft kein konzernübergreifendes AM. Die Interviewpartner repräsentieren hier zum Teil nur spezifische Sparten bzw. Regionen.⁵ Die übrigen

³ Streng genommen handelt es sich hierbei um branchenfremde Unternehmen. Diese wurden aufgrund der vergleichbaren Problemstruktur (Management bankbetrieblicher IT-Architekturen) mit in die Studie einbezogen.

⁴ Z. B. "Leiter Architekturmanagement", "Leiter IT Architektur" oder "Leiter Enterprise Architecture".

⁵ Die Untersuchungsergebnisse beziehen sich hier entsprechend auch nur auf diese Konzernteile.

Teilnehmer kamen überwiegend aus dem höheren IT-Management bis hin zum Gesamtleiter IT. Die Studiendurchführung erfolgte zwischen Dezember 2005 und März 2006. Im Verlauf wurden 14 Einzelinterviews mit einer Dauer von ca. 1,5 bis 3 Stunden geführt. Die Gesprächspartner wurden dabei zunächst über Ziel und Ablauf des Interviews aufgeklärt. Der anschließende formale Teil wurde mit Hilfe eines Leitfadens strukturiert. Dieser umfasste unter anderem die Themenfelder Organisation, Ziele, Aufgaben, Instrumente, Wirkungen, Probleme und Erfolgsfaktoren des AM. Dabei wurden vor allem offene Fragen gestellt. Auf diese Weise sollten möglichst viele unterschiedliche (und ggf. unerwartete) Informationen erfasst werden.

Die Dokumentation der Interviews erfolgte über Notizen in einem Dokumentationsbogen und anschließendes Gedächtnisprotokoll. Auf eine Tonbandaufzeichnung wurde aufgrund der zu erwartenden schlechten Akzeptanz verzichtet. Für die Datenauswertung wurde eine Form der qualitativen Inhaltsanalyse eingesetzt [GILa04, 191ff.].

4 Empirische Ergebnisse

AM ist in allen befragten Häusern ein wichtiges und sehr aktuelles Thema. Entsprechende Funktionen wurden oft erst innerhalb der letzten ein bis drei Jahre aufgebaut. Die Entwicklung ist dabei in vielen Fällen noch nicht abgeschlossen. Einige Institute verfügen auch bereits über längerjährige Erfahrungen. Im Laufe der Zeit kam es dort aber aufgrund einer zu großen Praxisferne häufig zu einem sukzessiven Akzeptanz- und Bedeutungsverlust des AM ("Elfenbeinturmbildung"). Erst innerhalb der letzten Jahre erfolgte dann eine Wiederbelebung.

Die von den einzelnen Unternehmen verfolgten Ansätze unterscheiden sich zum Teil erheblich. Unterschiede gibt es nicht nur bezüglich der Form der Implementierung (Abschnitte 4.1 und 4.3) und der eingesetzten Mittel (Abschnitt 4.4), sondern bereits auf der Ebene der verfolgten Ziele (Abschnitt 4.2). So orientieren sich bisher auch nur wenige Häuser an bestehenden Standards wie z. B. *TOGAF* [TheO03]. Interessanterweise stuft die Mehrheit der Teilnehmer die Reife ihrer AM-Implementierung trotzdem als mittel bis hoch ein. Dies erstaunt umso mehr als spezielle Reifegradmodelle des AM so gut wie unbekannt sind und bisher in keinem der befragten Unternehmen zum Einsatz kommen.

Trotz aller Unterschiede wird AM von allen Teilnehmern als ein Kernprozess angesehen, welcher nicht ausgelagert werden kann.

Im Folgenden werden die Ergebnisse ausgewählter Themenfelder im Detail vorgestellt.

4.1 Organisation

Fast alle befragten Unternehmen verfügen über ein institutionalisiertes AM (12 Fälle). In Großkonzernen bestehen dabei oft weitgehend autonome Teil-Organisationen für verschiedene Konzernbereiche. Dies überrascht angesichts des ganzheitlichen Anspruches von AM.

Die verschiedenen Organisationsformen lassen sich zunächst danach klassifizieren, inwieweit das AM über eine eigene Primär- und / oder Sekundärorganisation verfügt (vgl. Tab. 1).

Anzahl Fälle			Sekundärorganisation		Σ
			mit Entscheidungsgremien	ohne Entscheidungsgremien	
Primär- organisation	mit regulären Organisations- einheiten	zentrales Team	1	2	3
		föderative Struktur	4	2	6
		dezentrale Teams	0	1	1
	ohne reguläre Organisationseinheiten	2	2	4	
Σ			7	7	14

Tab. 1: Organisationsformen des Architekturmanagements

Von den teilnehmenden Instituten besitzen derzeit zehn eine eigene Primärorganisation. Diese ist in der Regel als IT- bzw. Operations-Querschnittsfunktion ausgelegt. Die Zahl der regulär beschäftigten Mitarbeiter liegt dabei typischerweise zwischen fünf und zehn. In Einzelfällen existieren auch AM-Organisationen mit bis zu 60 Mitarbeitern. Kleinere Häuser verfügen dagegen oft nur über einen einzelnen "IT-Architekten", welche alle Aufgaben in sich vereint.

Die Ansätze mit Primärorganisation können weiterhin nach dem Grad und den Kriterien der Zentralisierung bzw. Dezentralisierung differenziert werden. Bei den befragten Unternehmen lassen sich hier drei spezifische Ausprägungen unterscheiden: Im einfachsten Fall existiert lediglich eine integrierte Organisationseinheit. Vor allem in größeren Häusern werden die zentralen Teams darüber hinaus oft durch dezentrale Architekten (bzw. Architekturgruppen) ergänzt. Diese werden meist bestimmten Teilsegmenten der IT-Landschaft zugeordnet und der zentralen Einheit funktional unterstellt ("föderative Organisation"). In einer dritten Variante schließlich wird das AM in unabhängige Teilgruppen für die verschiedenen Architekturebenen unterteilt.

Die Hälfte der befragten Unternehmen verfügt weiterhin über eine Sekundärorganisation bestehend aus ein oder mehreren Entscheidungsgremien (z. B. "Architekturboard"). Viele andere Institute planen die Einführung solcher Gremien. Diese werden vorwiegend zur Verabschiedung von Architekturänderungen (z. B. Aufnahme eines neuen Standards, Abnahme eines Bauungsplanes) sowie für die Entscheidung über die Architekturkonformität von Änderungsvor-

haben eingesetzt. Sie sollen die Einbindung aller für eine erfolgreiche Implementierung relevanten Stellen ermöglichen und überdies eine querschnittliche Sichtweise fördern.

Beim Versuch, die Organisationskonzepte der befragten Unternehmen anhand der vorgenannten Kriterien in Gruppen einzuteilen, zeigt sich ein vergleichsweise heterogenes Bild (vgl. Tab. 1). Am weitesten verbreitet ist unter den teilnehmenden Instituten ein Ansatz mit föderativer Primärorganisation und zusätzlichen Entscheidungsgremien (aktuell 4 Fälle, mehrere in Planung). Als ein wesentlicher Vorteil dieses "schwergewichtigen" Ansatzes können die gute "Verankerung" des AM im Unternehmen sowie eine hohe Skalierbarkeit gelten. Dem steht jedoch auch ein vergleichsweise großer Personal- und Koordinationsaufwand entgegen.

Ein alternatives Modell stellt der zur Zeit von zwei Teilnehmern verfolgte "leichtgewichtige" Ansatz dar. Hierbei wird bewusst auf eine Primärorganisation verzichtet und stattdessen lediglich eine Koordination der ansonsten dezentral zuständigen IT-Einheiten auf Gremienebene vorgenommen. Dieser Ansatz kann als eine Reaktion auf negative Erfahrungen mit zu restriktiven und praxisfernen Architekturgruppen verstanden werden. Dem Vorteil geringer Kosten und einer hohen Problemorientierung stehen hier allerdings auch gewichtige Nachteile entgegen. So räumt eines der betroffenen Unternehmen ein, dass aufgrund der fehlenden Primärorganisation oft notwendige operative Aufgaben nicht (bzw. nicht fristgerecht) erledigt werden können.

4.2 Ziele

Die befragten Unternehmen verfolgen mit dem AM zum Teil sehr unterschiedliche Ziele. Oft herrscht diesbezüglich auch eine gewisse Unsicherheit. Stringente Zielsysteme können nur von den wenigsten Häusern angegeben werden. Dies überrascht angesichts knapper IT-Budgets und der ohnehin bestehenden Legitimationsproblematik von AM (vgl. Abschnitt 4.6).

Als Hauptziele des AM werden am häufigsten eine Senkung der IT-Kosten, eine Erhöhung der IT-Effizienz und / oder eine Steigerung der IT-Produktivität angegeben (zusammen 8 Fälle). Sechs Unternehmen nennen weiterhin eine Erhöhung der IT-Flexibilität als wichtiges AM-Ziel. Damit soll vor allem der Zeitaufwand zur Implementierung neuer fachlicher Anforderungen (und insbesondere von Produktinnovationen) nachhaltig gesenkt werden ("Time-to-Market").

Viele weitere der genannten Ziele lassen sich ebenfalls direkt oder indirekt auf das Bestreben zur Verbesserung von IT-Effizienz und / oder IT-Flexibilität zurückführen. Dies gilt etwa für eine Reduktion der IT-Komplexität (4 Fälle), eine stärkere Standardisierung (2 Fälle), die Erhöhung der Wiederverwendung (2 Fälle) oder eine Verkürzung der Entwicklungszeiten (1 Fall).

Während viele Teilnehmer angeben, mit dem AM der eher kurzfristig und auf partikulare Interessen ausgerichteten Business-Sichtweise eine systemübergreifende und zukunftsorientierte Betrachtung entgegensetzen zu wollen, betonen andere explizit die Business-Orientierung ihres Ansatzes. Für diese Unternehmen stellt daher vor allem auch die Erhöhung des Business-IT-Alignments ein Kernziel von AM dar (2 Fälle).

Einige Teilnehmer nennen schließlich generische Ziele wie eine ganzheitliche Planung und Optimierung des IT-Einsatzes oder die Umsetzung der IT-Strategie (jeweils 2 Fälle). Weiterhin wird die Schaffung von Transparenz als ein wichtiges Teilziel von AM angesehen (4 Fälle).

4.3 Aufgaben

Zur Erreichung der verfolgten Ziele wird von den für das AM zuständigen Stellen und Gremien eine Vielzahl unterschiedlicher Aufgaben wahrgenommen. Die Angaben der Interviewpartner wurden hier zu den in Abb. 1 dargestellten Aufgabenkategorien verdichtet. Dabei erfolgte in Anlehnung an [Niem05, 24] eine Unterteilung in operative und strategische Aufgaben.

4.3.1 Strategische Aufgaben

Eine wichtige Aufgabe des AM sehen viele Teilnehmer zunächst in der Dokumentation ihrer IT-Architektur (9 Fälle). Hierbei werden üblicherweise unterschiedliche Architektursichten unterschieden (vgl. Abschnitt 4.4.1). In acht Unternehmen wird darauf aufbauend eine systematische Architekturplanung durchgeführt. Dies beinhaltet meist die Entwicklung entsprechender Soll-Architektur-Modelle. Davon ausgehend werden in einigen Unternehmen spezielle Architekturprogramme zur gezielten Transformation der IT-Architektur geplant und gesteuert.

Als weitere wichtige Aufgaben von AM werden die Mitwirkung an der IT-Planung (und hier insbesondere dem Portfoliomanagement) sowie die Informationsversorgung gegenüber IT- und Business-Management genannt. In manchen Häusern umfasst das AM darüber hinaus Aufgaben aus dem Bereich des Technologiemanagements (z. B. Technology-Life-Cycle-Management).

4.3.2 Operative Aufgaben

Fast alle Institute nehmen weiterhin einer Regulation ihrer Prozesse zur Entwicklung bzw. Bereitstellung neuer Anwendungslösungen vor. So werden in 13 der 14 Unternehmen Standards definiert, welche von den durchgeführten Entwicklungs- bzw. Veränderungsvorhaben einzuhalten sind (vgl. Abschnitt 4.4.3). Die Verbindlichkeit der Vorgaben und die vorhandenen Kontrollmechanismen unterschieden sich dabei von Haus zu Haus. Zum Teil existieren hierzu for-

male Prüf- und Abnahmeprozesse, welche zu bestimmten Meilensteinen entsprechende "Architektur-Checks" vorschreiben.⁶ Eine vollständige Abdeckung wird dabei jedoch nur von wenigen Häusern erreicht; zum Teil wird bewusst eine Quote von nicht mehr als 70 - 80 % angestrebt. Im Hinblick auf die Einhaltung der Standards müssen weiterhin häufig Zugeständnisse gemacht werden.⁷ Verstöße werden nur selten formell sanktioniert.

Strategische Aufgaben		
<ul style="list-style-type: none"> ■ Dokumentation und Analyse der Ist-Architektur (+++) ■ Planung der Soll-Architektur (++) 	<ul style="list-style-type: none"> ■ Planung und Steuerung der Architektur-Transformation (++) ■ Informationsversorgung / Mitwirkung an der IT-Planung (++) 	<ul style="list-style-type: none"> ■ Technologie-management und Umfeldanalyse (+) ■ Definition des Architekturframeworks / Metaaufgaben (++)
Operative Aufgaben		
<ul style="list-style-type: none"> ■ Definition von Architekturstandards (+++) ■ Entwicklung und Pflege wiederverwendbarer Komponenten (+) 	<ul style="list-style-type: none"> ■ Architekturkommunikation (Publikation, Schulung, Beratung) (++) ■ Projektarbeit / Entwicklung von Projektarchitekturen (++) 	<ul style="list-style-type: none"> ■ Prüfung der Architekturkonformität von Änderungsvorhaben (+++)

+ bis zu 4 Nennungen ++ zwischen 5 und 8 Nennungen +++ mehr als 8 Nennungen

Abb. 1: Aufgaben des Architekturmanagements

Um die Anwendung und Akzeptanz der Standards zu fördern, werden oft auch verschiedene Maßnahmen der Kommunikation ergriffen (z. B. Publikationen, Schulungen).⁸ Zum Teil wird darüber hinaus operative Projektarbeit geleistet. Das AM kann dabei insbesondere aktiv an der Architekturentwicklung auf Projektebene ("Architecture in the Small") mitwirken.⁹ In einigen Häusern werden durch das AM schließlich auch wiederverwendbare Komponenten bereitgestellt (z. B. bankfachliche Komponentenbibliotheken, Services, Frameworks etc.). Insgesamt gesehen setzen die Unternehmen bei der Ausgestaltung des AM unterschiedliche Schwerpunkte. Das volle Aufgabenspektrum wird bisher nur von wenigen Instituten abgedeckt.

⁶ Z. B. wurden in einer Großbank im vergangenen Jahr über 300 Projekte auf Architekturkonformität geprüft.

⁷ So lag die Ausnahmekquote in einer teilnehmenden Großbank im vergangenen Jahr bei ca. 3 %. Ein vergleichbares Institut nennt sogar eine Quote von ca. 20 %.

⁸ In einem Unternehmen werden sogar monatlich stattfindende Informations-Events veranstaltet. Ferner ist dort die Einführung von "Pflichtschulungen" geplant.

⁹ In manchen Fällen werden für Projektarbeit sogar über 50 % der AM-Kapazitäten allokiert.

4.4 Instrumente

Bei der Aufgabenerfüllung bedienen sich die befragten Unternehmen unterschiedlicher Konzepte, Methoden und Werkzeuge. Diese wurden hinsichtlich ihres primären Verwendungszweckes in die Gruppen allgemeine Ordnungsinstrumente, Instrumente der Dokumentation und Planung sowie Instrumente der Standardisierung unterteilt.

4.4.1 Allgemeine Ordnungsinstrumente

Fast alle befragten Institute verwenden einen konzeptionellen *Ordnungsrahmen* zur Strukturierung ihrer Aktivitäten im AM. Am weitesten verbreitet ist dabei eine Unterteilung nach Techniknähe in die Ebenen Businessarchitektur, Anwendungsarchitektur und Infrastrukturarchitektur. Zum Teil werden auch weitere – hierzu orthogonale – Sichten verwendet.¹⁰

Der gewählte Ordnungsrahmen bildet in der Regel die Grundlage für die Architekturplanung und -dokumentation. Beim Einsatz eines Metadaten-Repositories (s. Abschnitt 4.4.2) determiniert er daher auch die Grobstruktur eines entsprechenden Metamodells. Er dient darüber hinaus oft als Vorlage für die Binnenorganisation des AM. So werden z. B. häufig Untergruppen oder Rollen nach den Elementen des Ordnungsrahmens definiert (z. B. "Anwendungsarchitekt").

Unabhängig von dem verwendeten Ordnungsrahmen nehmen neun der 14 befragten Unternehmen eine (vertikale) Segmentierung ihrer IT-Landschaft vor. Die Segmente werden dabei in der Regel als *Domänen* bezeichnet. Die Domänenbildung erfolgt meist nach fachlichen Kriterien. Sie dient in erster Linie der Komplexitätsbeherrschung. So werden die Domänen meist als Planungseinheiten verwendet, für welche separate Architekturpläne erstellt werden können. Weiterhin werden den einzelnen Domänen oft spezifische Architekturstandards zugeordnet. Vor allem in großen Unternehmen wird die Verantwortung für die Architekturkonformität dabei häufig auf spezielle Mitarbeiter (bzw. Rollen) übertragen ("Domänenarchitekten").

Die Segmentierung bildet schließlich oft auch die Grundlage für eine technische Entkopplung der einzelnen Landschaftssegmente. Hierzu werden entsprechende Dienste-Schnittstellen auf den Domänengrenzen definiert. Auf diese Weise soll insbesondere die Wartbarkeit der Informationssystemlandschaft erhöht werden. Ferner soll so die Umsetzung strategischer Management-Entscheidungen (z. B. Verkauf von Unternehmensteilen) erleichtert werden.¹¹

¹⁰ Gebräuchlich ist insbesondere die Betrachtung von Informations-, Integrations-, Sicherheits-, Entwicklungs- sowie Betriebsarchitekturen.

¹¹ Dabei sollen die Clustergrenzen im Idealfall als "Sollbruchstellen" dienen. Die Praktikabilität dieses Konzeptes ist unter den Teilnehmern allerdings umstritten.

Das Prinzip der *Serviceorientierung* findet auch über die Domänenentkopplung hinaus breite Anwendung. So werden in vielen Instituten zentrale Dienste definiert, die von anderen Anwendungen bzw. Diensten genutzt werden können. Damit soll insbesondere die Wiederverwendung gefördert werden. Außerdem sollen auf diesem Wege ähnliche Daten und Funktionen zentralisiert und entsprechende Redundanzen in der Systemlandschaft abgebaut werden.

4.4.2 *Instrumente der Dokumentation und Planung*

Als Grundlage der Architekturdokumentation und -planung verwalten alle befragten Unternehmen strukturierte (Meta-) Daten über die Elemente und Beziehungen innerhalb ihrer IT- bzw. Unternehmensarchitektur. Art und Umfang dieser *Metadatenverwaltung* unterscheiden sich von Haus zu Haus. Am unteren Ende der Skala rangieren einfache – in der Regel isolierte – Verzeichnisse der eingesetzten Architekturkomponenten. So verfügen etwa alle teilnehmenden Institute über ein Inventar ihrer Anwendungssysteme. Verbreitet sind außerdem Verzeichnisse der im Einsatz befindlichen Infrastrukturelemente sowie der geltenden Standards.

Das obere Ende der Skala bilden weitgehend integrierte Datenbestände unter Einbeziehung der Businesssebene. Dabei werden insbesondere auch die Querbeziehungen zwischen den verschiedenen Architekturebenen erfasst (z. B. von Anwendungen unterstützte Prozesse). Zudem können Entwicklungsprojekte und die durch diese verursachten Änderungen abgebildet werden. Dieser Entwicklungsstufe lassen sich bislang allerdings nur wenige Institute zurechnen.

Als Werkzeuge zur Metadatenverwaltung verwenden die Unternehmen in erster Linie Eigenentwicklungen. Dabei handelt es sich oft um relativ einfache Datenbankanwendungen. Kommerzielle *Enterprise Architecture Tools* sind dagegen bis jetzt kaum im Einsatz.¹² Sie werden darüber hinaus nur selten voll ausgenutzt. Die Datenpflege erfolgt in der Regel manuell; nur wenige Häuser verfügen über eine (teil-) automatisierte Datenversorgung.

Die meisten Unternehmen verwenden weiterhin grafische *Architekturmodelle* zur Visualisierung verschiedener Sichten der Ist- und / oder Soll-Architektur. Diese stellen nicht zuletzt auch ein wichtiges Mittel der Kommunikation zwischen den verschiedenen Stakeholdern dar. Je nach Gegenstand und Zielgruppe kommen zum Teil sehr unterschiedliche Darstellungsformen zum Einsatz. So stehen etwa auf der Ebene der Anwendungsarchitektur Cluster- und Matrix-Modelle im Vordergrund.¹³ Cluster-Modelle werden vor allem zur grafischen Repräsentation der bestehenden Domänenstruktur (s. Abschnitt 4.4.1) verwendet. Die einzelnen Informationssysteme

¹² Als ein Grund hierfür werden unter anderem zu hohe (nicht vertretbare) Lizenzkosten genannt.

werden dabei durch Rechtecke symbolisiert, welche räumlich zu Clustern gruppiert werden. Matrix-Modelle erlauben dagegen eine strukturierte Darstellung einer Anwendungslandschaft entlang von zwei nominal-skalierten Attribute-Dimensionen. Die einzelnen Informationssysteme werden hierzu ebenfalls in Form von Rechtecken symbolisiert, welche anhand ihrer spezifischen Merkmalsausprägung auf einer zweidimensionalen Matrix positioniert werden. Am weitesten verbreitet sind dabei unter den befragten Unternehmen Gliederungen nach unterstützten Prozess- bzw. Wertschöpfungsstufen einerseits sowie Bankprodukten andererseits.



Abb. 2: Instrumente der Standardisierung

Auf den übrigen Architekturebenen überwiegen einfache, graphenorientierte Modelltypen. Die Architekturelemente werden dabei in der Regel durch Rechtecke und die zwischen diesen bestehenden Beziehungen durch entsprechende Verbindungslinien repräsentiert. UML oder spezielle Architektursprachen kommen so gut wie nicht zum Einsatz.

Als Modellierungswerkzeuge nutzen die meisten Unternehmen einfache Office-Tools. Modellierung und Modellpflege erfolgen dabei in der Regel manuell. Die Ablage der Modelle geschieht meist unstrukturiert auf einem File-Server.¹⁴ Nur in wenigen Fällen werden die Architekturmodelle mit Hilfe der eingesetzten Metadaten-Repositories erzeugt und verwaltet.

4.4.3 Instrumente der Standardisierung

Die befragten Institute nutzen weiterhin verschiedene Instrumente zur Standardisierung ihrer Entwicklungsprozesse. Hinsichtlich des Standardisierungsgegenstandes lassen sich dabei Input-, Output- und Prozess-Standards unterscheiden (s. Abb. 2) [s. auch BaAX04, 3756].

Fast alle befragten Unternehmen verwenden Architekturprinzipien (13 Fälle). Hierbei handelt es sich um allgemeine Grundsätze und Verfahrensregeln, welche bei der Durchführung von Ände-

¹³ [LaMW05] verwenden für solche Modelle den Oberbegriff "Softwarekarten".

¹⁴ In einer Bank erstrecken sich allein die Modelle der Infrastruktur-Architektur auf über 150 Präsentationsfolien.

nungsvorhaben einzuhalten sind. Architekturprinzipien werden oft über mehrere Hierarchieebenen operationalisiert und können alle drei Standardisierungsdimensionen adressieren. Die meisten Unternehmen verwenden weiterhin Standardkataloge (12 Fälle). Darin werden die im Rahmen der (Neu-) Entwicklung (bzw. Beschaffung) zulässigen Technologien und Produkte festgelegt. Knapp die Hälfte der befragten Unternehmen nutzt darüber hinaus Referenzarchitekturen (6 Fälle). Hierunter werden vordefinierte Architektur-Templates für bestimmte Systemklassen verstanden (z. B. Referenzarchitektur für Vertriebssysteme). In einigen Häusern wird schließlich auch die Entwicklungsmethodik durch das AM festgelegt. Zum Teil werden Prinzipien, Standardkomponenten, Referenzarchitekturen und Entwicklungsmethoden zu ganzen "Standardplattformen" verknüpft (z. B. Java-Plattform, Mainframe-Plattform). Erklärtes Ziel ist dabei eine weitgehende Standardisierung von Entwicklungs- und Produktionsprozessen.

4.5 Wirkungen

Als beobachtete Wirkung von AM nennen sechs der befragten 14 Unternehmen eine Reduktion der Komplexität ihrer IT-Landschaft. So konnten in einigen Fällen die Anzahl von Anwendungen, Schnittstellen und / oder Plattformen deutlich verringert werden. Hierzu wurden teilweise aber auch entsprechende Konsolidierungsprojekte durchgeführt. In anderen Häusern konnte immerhin eine Begrenzung des Komplexitätswachstums erreicht werden.

Sechs Unternehmen berichten weiterhin über eine Senkung der IT-Kosten. So konnten in einem Fall über einen Zeitraum von ca. fünf Jahren Netto-Einsparungen in Höhe von 100 - 200 % des währenddessen eingesetzten AM-Aufwandes realisiert werden. In zwei anderen Häusern sind dagegen bisher vorwiegend zusätzliche Kosten entstanden. Dies wird einerseits auf Anlaufkosten, andererseits aber auch auf Wirkungsverzögerungen zurückgeführt.

Als Erfolg von AM wird außerdem häufig eine Erhöhung der Transparenz angegeben (6 Fälle). Eine Ursache hierfür ist in den durch das AM aufgebauten Dokumentationsbeständen zu sehen. In einigen Unternehmen konnte mit Hilfe des AM ferner eine Steigerung der Entwicklungseffizienz bzw. eine Reduktion der Durchlaufzeiten bewirkt werden (3 Fälle). Von einer Erhöhung der IT-Flexibilität wird hingegen nur in einem Fall explizit berichtet.

Insgesamt fällt es den befragten Unternehmen schwer, konkrete Wirkungen des AM anzugeben. Zum einen wird hierfür der zurückliegende Wirkungszeitraum oft noch als zu kurz angesehen. So erwarten einige Teilnehmer nachhaltige Ergebnisse erst mit einer Wirkungsverzögerung von fünf bis zehn Jahren. Zum anderen haben die Unternehmen meist Schwierigkeiten bei der Messung des Architektur Erfolges. Probleme bereiten hier sowohl die Operationalisierung der oft

qualitativen Architekturziele als auch die effiziente Erhebung der für eine regelmäßige Messung erforderlichen Daten. Bei der großen Mehrheit der befragten Institute wird daher bisher keine systematische Erfolgsmessung durchgeführt. In den anderen Fällen überwiegen einfache Kennzahlen (z. B. Anzahl technischer Plattformen, Anzahl wiederverwendbarer Services). Dass eine weitergehende Operationalisierung und Messung von Architekturzielen prinzipiell möglich ist, zeigt das Beispiel einer Großbank. Hier wurde eine integrierte Architektur-Scorecard entwickelt, welche unter anderem Kennzahlen aus den Bereichen IT-Effizienz (z. B. Entwicklungskosten pro Use Case Point), IT-Flexibilität (z. B. Entwicklungsdauer pro Use Case Point) und Wiederverwendung (z. B. Anzahl Nutzer pro Service) beinhaltet. Darüber hinaus werden interne Steuerungsgrößen wie etwa die operative Architekturkonformität (Anteil der architekturkonformen Projekte pro Kalenderjahr) oder die Prozessleistung des AM (z. B. durchschnittliche Review-Dauer, Anzahl betreuter Projekte) betrachtet.

4.6 Probleme und Erfolgsfaktoren

Die Schwierigkeiten bei der Erfolgsmessung und die auftretenden Wirkungsverzögerungen schlagen sich oftmals in einem gewissen Legitimationsdruck des AM nieder. Verstärkt wird dies meist noch durch ein geringes Architekturverständnis auf Seiten der Fachbereiche sowie eine allgemeine Kultur der Kurzfristorientierung. Einen Ansatzpunkt zur Lösung dieser Problematik sehen viele Teilnehmer in verstärkten Bemühungen zur Messung und Kommunikation des AM-Erfolges. Andere Interviewpartner halten dies für unpraktikabel und empfehlen stattdessen eine konsequente Business-Orientierung des AM.

In vielen Unternehmen hat das AM weiterhin mit Akzeptanzproblemen bei den von seinen Maßnahmen betroffenen Mitarbeitern zu kämpfen. Als wichtigen Erfolgsfaktor nennen daher die meisten Häuser eine konsequente Praxisorientierung. Nach Ansicht vieler Teilnehmer spielt dabei neben intensiver Kommunikation vor allem auch die konkrete Unterstützung von Entwicklungsprojekten eine wesentliche Rolle. Das AM kann so insbesondere Feedback-Informationen hinsichtlich der praktischen Anwendbarkeit der definierten Standards aufnehmen. Als erfolgskritisch werden außerdem ein pragmatischer und kompromissbereiter Umgang mit den Stakeholdern sowie eine moderate Restriktivität angesehen.

Zur effektiven Durchsetzung der Architekturziele halten darüber hinaus viele Teilnehmer einen nachhaltigen Rückhalt durch IT und Business-Management für unverzichtbar. Als wichtige Erfolgsfaktoren gelten weiterhin eine angemessene Ressourcenausstattung und Organisation des AM, eine hohe Mitarbeiterqualifikation sowie ein effizienter Werkzeugeinsatz.

5 Diskussion und Schlussfolgerungen

Gegenstand der vorliegenden Studie war eine explorative Untersuchung zum Status Quo des AM in der deutschsprachigen Kreditwirtschaft. Hierzu wurden insgesamt 14 Experteninterviews mit Vertretern aus der Praxis durchgeführt. Der gewählte Forschungsansatz ermöglichte es dabei, einen vergleichsweise großen Ausschnitt der Wirklichkeit zu erfassen, ohne jedoch auf die Vorteile qualitativer Methoden verzichten zu müssen. Aufgrund des Auswahlverfahrens und der geringen Fallzahl sind die Ergebnisse allerdings nicht ohne Weiteres verallgemeinerbar.

AM ist unter den befragten Instituten bereits weit verbreitet. Es ist dabei fast in allen Häusern als eigenständige Funktion etabliert und institutionalisiert. Gleichwohl stellt es eine relativ junge Disziplin dar, deren Entwicklung noch nicht abgeschlossen ist. So unterscheiden sich die einzelnen Implementierungen zum Teil erheblich. Mögliche Ursachen hierfür können in einem differierenden Architekturverständnis sowie in unterschiedlichen Rahmenbedingungen liegen.

Der gemeinsame Kern aller Ansätze kann in der Wahrnehmung einer zentralen Koordinationsfunktion im Hinblick auf eine ganzheitliche und zukunftsorientierte Gestaltung und Steuerung der IT-Architektur gesehen werden. Die Unternehmen scheinen dabei in Abhängigkeit ihrer Situation jeweils unterschiedliche Koordinationsziele in den Vordergrund zu stellen. Am häufigsten werden in diesem Kontext die Erhöhung bzw. Aufrechterhaltung von IT-Effizienz und / oder IT-Flexibilität genannt. Primär auf die Lenkbarkeit der IT ausgerichtete Zielsysteme, wie sie z. B. von [Hafn05, 79ff.] vorgeschlagen werden, lassen sich somit nur zum Teil bestätigen.

Zur Erreichung der verfolgten Ziele werden von den befragten Unternehmen im Wesentlichen zwei (zueinander komplementäre) Koordinationsmechanismen eingesetzt. So ist auf der strategischen Ebene eine Koordination durch Planung vorherrschend. Gegenstand dieser Architekturplanung ist die Entwicklung integrierter Modelle der Soll-Architektur. Die Koordinationswirkung wird dabei vor allem über eine entsprechende Beeinflussung der Vorhabenplanung erzielt. Auf der operativen Ebene steht demgegenüber eine Koordination durch Standardisierung im Vordergrund. Hierzu werden allgemeine Regeln definiert, welche von den durchgeführten Änderungsvorhaben einzuhalten sind. Die Koordinationswirkung wird dabei entsprechend über eine Kontrolle der Standardkonformität erreicht.¹⁵ Zum Zwecke der Komplexitätsbewältigung kommen darüber hinaus in beiden Fällen systemtechnische Konzepte "im Großen" zur Anwendung (z. B. Modularisierung, Kapselung, Entkopplung).

¹⁵ Dabei spielt es keine Rolle, ob die entsprechenden Leistungen intern oder extern erbracht werden.

Erste Erfolge in den befragten Unternehmen deuten darauf hin, dass AM einen Beitrag zur Lösung der bestehenden Probleme leisten kann. Dies gilt vor allem im Hinblick auf ein nachhaltiges Komplexitäts- und Kostenmanagement. Auswirkungen auf die Flexibilität bzw. Steuerbarkeit der IT können auf Basis der vorliegenden Daten hingegen nicht belegt werden. Möglicherweise ist hierfür der Wirkungszeitraum noch zu kurz. Insgesamt scheinen bei der Implementierung von AM erhebliche Verzögerungen aufzutreten. Entsprechende Investitionen sollten daher als mittel- bis langfristig eingestuft werden.

Die Erfahrungen einiger Studienteilnehmer legen den Schluss nahe, dass der Erfolg von AM wesentlich von einer intensiven Kommunikation mit den Stakeholdern und einer konsequenten Praxisorientierung abhängt. Im Falle einer Überregulation droht dagegen die Gefahr eines Akzeptanz- und Bedeutungsverlustes. Ein nicht zu vernachlässigender Faktor kann darüber hinaus in der richtigen Organisation des AM gesehen werden. In größeren Unternehmen hat sich hierfür ein föderativer Ansatz mit zentralem Team und übergreifendem Architekturboard bewährt.

Bisher werden die Potentiale von AM in der Praxis kaum ausgeschöpft. Hierzu erscheint vor allem eine klarere Positionierung des AM innerhalb des Informationsmanagements notwendig. Dies könnte unter anderem durch die Definition stringenter Zielsysteme und den Einsatz entsprechender Kennzahlen zur Messung der Zielerreichung bewirkt werden. Auf diese Weise ließe sich die Rolle des AM stärken und die inhärente Legitimationsproblematik entschärfen. Optimierungspotential gibt es darüber hinaus vor allem im Bereich der Werkzeugunterstützung. So könnte die Effizienz des AM durch eine Integration der bisher oft isolierten Daten- und Dokumentationsbestände in einem zentralen Repository in vielen Häusern deutlich erhöht werden.

Tab. 2 zeigt beispielhaft einige vorläufige Arbeitshypothesen, welche aus den Studienergebnissen abgeleitet werden können. Diese sollen in einem nächsten Schritt zusammen mit weiteren Hypothesen einem empirischen Test unterzogen werden. Hierzu ist die Durchführung einer quantitativen Breitenstudie geplant.

H1	Unternehmen verfolgen mit AM unterschiedliche Koordinationsziele. Am weitesten verbreitet sind dabei eine Erhöhung bzw. Aufrechterhaltung von IT-Effizienz und / oder IT-Flexibilität.
H2	Unternehmen, welche die Entwicklung ihrer IT-Architektur in stärkerem Umfang durch Architekturplanung und Standardisierung koordinieren, weisen höhere Erreichungsgrade im Hinblick auf die verfolgten Architekturziele auf.
H3	Der Einfluss von Architekturplanung und Standardisierung auf die Erreichungsgrade der verfolgten Architekturziele ist von der jeweiligen Anwendungsdauer abhängig.

Tab. 2: Vorläufige Arbeitshypothesen (Auszug)

Literaturverzeichnis

- [AiSc05] Aier, S.; Schönherr, M.: EAI als integrierendes Element einer nachhaltigen Unternehmensarchitektur. In: Aier, S.; Schönherr, M. (Hrsg.): Unternehmensarchitekturen und Systemintegration. Gito, Berlin 2005, S. 3-56.
- [BaAX04] Ball, N. L.; Adams, C. R.; Xia, W.: IS/IT Architecture: An Integrated View and Typology. In: Proc. 10th Americas Conference on Information Systems, New York, 2004. S. 3753-3761.
- [BCHP05] Baskerville, R.; Cavallari, M.; Hjort-Madsen, K.; Pries-Heje, J.; Sorrentino, M.; Virili, F.: Extensible Architectures: The Strategic Value of Service-Oriented Architecture in Banking. In: Proc. 13th European Conference on Information Systems, Regensburg, 2005. S. 1-12.
- [BYDH03] Boh, W. F.; Yellin, D.; Dill, B.; Herbsleb, J. D.: Effectively Managing Information Systems Architecture Standards: An Intra-Organization Perspective. In: Proc. MISQ Special Issue Workshop: Standard Making: A Critical Research Frontier for Information Systems, Seattle, 2003. S. 171-187.
- [FeSi95] Ferstl, O. K.; Sinz, E. J.: Der Ansatz des Semantischen Objektmodells (SOM) zur Modellierung von Geschäftsprozessen. In: Wirtschaftsinformatik 37 (1995) 3, S. 209-220.
- [Gaer04] Gaertner, W.: Ansatz für eine erfolgreiche Enterprise Architecture im Bereich Global Banking Division / Global Transaction Banking IT and Operations der Deutschen Bank. In: Wirtschaftsinformatik 46 (2004) 4, S. 311-313.
- [GILa04] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. VS Verlag für Sozialwissenschaften, Wiesbaden 2004.
- [Hafn05] Hafner, M.: Entwicklung einer Methodik für das Management der IS-Architektur im Unternehmen. Dissertation, Hochschule St. Gallen, 2005.
- [IEEE00] IEEE Computer Society (Hrsg.): IEEE Recommended Practice for Architectural Description of Software-Intensive Systems – IEEE Std 1471-2000.

- [LaMW05] Lankes, J.; Matthes, F.; Wittenburg, A.: Softwarekartographie: Systematische Darstellung von Anwendungslandschaften. In: Ferstl, O. K.; Sinz, E. J.; Eckert, S.; Isselhorst, T. (Hrsg.): Wirtschaftsinformatik 2005 – eEconomy, eGovernment, eSociety. Physica, Heidelberg 2005, S. 1443-1462.
- [MaGH04] Martin, N.; Gregor, S.; Hart, D.: Using a Common Architecture in Australian E-Government: The Case of Smart Service Queensland. In: Proc. 6th International Conference on Electronic Commerce, Delft, 2004. S. 516-525.
- [Mehl03] Mehla, J. I.: IT-Architekturen von Finanzdienstleistern. In: Bartmann, D. (Hrsg.): Bankinformatik 2004 – Strategien, Konzepte, Technologien für das Retail Banking. Gabler, Wiesbaden 2003, S. 203-219.
- [Niem05] Niemann, K. D.: Von der Unternehmensarchitektur zur IT-Governance – Bausteine für ein wirksames IT-Management. Vieweg, Wiesbaden 2005.
- [Penz04] Penzel, H. G.: Architekturmanagement aus der Sicht einer Großbank. In: Moor- mann, J.; Fischer, T. (Hrsg.): Handbuch Informationstechnologie in Banken. 2. Aufl., Gabler, Wiesbaden 2004, S. 111-130.
- [Pulk06] Pulkkinen, M.: Systemic Management of Architectural Decisions in Enterprise Architecture Planning. Four Dimensions and Three Abstraction Levels. In: Proc. 39th Hawaii International Conference on Systems Science, Kauai, 2006. S. 179.
- [Sche91] Scheer, A.-W.: Architektur integrierter Informationssysteme: Grundlagen der Unternehmensmodellierung. Springer, Berlin et al. 1991.
- [The003] The Open Group (Hrsg.): TOGAF (The Open Group Architecture Framework) V. 8.1 "Enterprise Edition". www.opengroup.org/togaf, Abruf am 2006-05-10.
- [Wint05] Winter, R.: Architektur braucht Management. In: Wirtschaftsinformatik 46 (2004) 4, S. 317-319.
- [Zach87] Zachman, J. A.: A Framework for Information Systems Architecture. In: IBM Systems Journal 26 (1987) 3, S. 276-292.

MiFID-Readiness

- Die Umsetzung der MiFID “Markets in Financial Instruments Directive“ in der deutschen Finanzindustrie -

Peter Gomber, Markus Gsell

Lehrstuhl für BWL, insb. e-Finance
Johann Wolfgang Goethe-Universität Frankfurt am Main
60054 Frankfurt/Main
{gomber,gsell}@wiwi.uni-frankfurt.de

Claudia Reininger

Goethestr. 34
63674 Altenstadt
claudia.reininger@t-online.de

Abstract

Aktuell stellen die neue EU Richtlinie „Markets in Financial Instruments Directive“ (MiFID) und deren Umsetzungserfordernisse ein wichtiges Thema für die Finanzindustrie in Europa dar, da diese Richtlinie die Erbringung von Wertpapierdienstleistungen und den Wertpapierhandel europaweit grundlegend neu regelt und harmonisiert. Die MiFID ist ab dem 01. November 2007 anzuwenden und erfordert die Anpassung von Prozessen und IT-Systemen. Der vorliegende Beitrag präsentiert die Ergebnisse einer Umfrage unter deutschen Wertpapierfirmen bezüglich des Umsetzungsstatus der Projekte zur MiFID und die Haltung der Häuser zu den strategischen Potenzialen sowie den Nutzen bzw. Kosteneinschätzungen bezüglich der neuen Richtlinie.

1 Einleitung

“Any ‘Lamfalussy’ Directive is complicated. With the MiFID, the stakes are high. The sums involved big. This has given the debate an added edge. Indeed, my services tell me that this may be the toughest securities law dossier they have ever had to deal with!” [McCr05]. EU Binnenmarktkommissar McCreevy beschreibt hiermit nicht nur die Sicht der Kommission,

sondern er trifft auch die Einschätzung vieler Marktteilnehmer zur neuen „Richtlinie über Märkte für Finanzinstrumente“ („Markets in Financial Instruments Directive“, MiFID) [Euro04]. Diese ist am 30. April 2004 in Kraft getreten und wird nach Umsetzung in das jeweilige nationale Recht der Mitgliedsstaaten ab 01. November 2007 von Marktteilnehmern, d. h. den Wertpapierfirmen und Börsen, anzuwenden sein. Die MiFID ist die Kernmaßnahme des Financial Services Action Plan der EU. Ziele der MiFID sind Integration und Harmonisierung der Finanzmärkte in Europa, Stärkung des Investorenschutzes, die Sicherung von Marktintegrität und -transparenz sowie eine Ausdehnung der einbezogenen Finanzinstrumente und Wertpapierdienstleistungen im Vergleich zur Vorgängerrichtlinie der MiFID, der „Investment Services Directive“ (ISD) aus dem Jahr 1993 [Euro93]. Damit soll die neue Richtlinie die Veränderungen der letzten fast 15 Jahre in Bezug auf neue Finanzinstrumente, auf neue z. B. internetbasierte Geschäftsmodelle im Banking und Brokerage sowie in Bezug auf technologiebasierte Handelsmodelle/-plattformen auf regulatorischer Seite nachvollziehen. Mit dieser Maßnahme wird der Wertpapierhandel in Europa grundsätzlich neu geregelt. Hieraus ergeben sich neue Chancen, aber auch Anpassungsbedarf bezüglich der Prozesse und IT-Systeme von Wertpapierfirmen.

Die 73 Artikel der MiFID adressieren die Themenblöcke Wertpapierhandel und Markttransparenz, Organisations- und Compliance-Anforderungen sowie Wohlverhaltensregeln und Operations von Wertpapierfirmen und Marktplatzbetreibern. Im Themenblock Wertpapierhandel – der im Zentrum der MiFID und der nachfolgenden Ausführungen steht – wird erstmals eine umfassende und europaweit gültige Verpflichtung von Wertpapierfirmen zur „kundengünstigsten Ausführung von Aufträgen (*Best Execution*)“ eingeführt. Weiterhin werden für viele Mitgliedsstaaten (insbesondere Deutschland) neue Transparenzanforderungen speziell für die außerbörsliche Orderausführung definiert. Diese verlangen neben der Veröffentlichung und Meldung der Transaktionsdaten abgeschlossener Transaktionen (Post-Trade Transparenz) auch die Offenlegung von Daten zu Orders und Quotierungen (Pre-Trade Transparenz) und beinhalten neue Anforderungen an Prozesse und Marktdatensysteme.

Die Umsetzung der Richtlinie erfolgt nach dem so genannten Lamfalussy-Gesetzgebungsverfahren, das die Implementierung regulatorischer Maßnahmen im Bereich Finanzmärkte in vier Stufen (*levels*) gliedert. Mit der Veröffentlichung der MiFID als Rahmenrichtlinie am 30. April 2004 wurde Level 1 und damit der Basisrechtsakt auf EU-Ebene

abgeschlossen. Am 10. August 2006 wurden die technischen Durchführungsbestimmungen¹ in der zweiten Stufe (Level 2), die in einem intensiven Abstimmungsprozess zwischen den Gesetzgebungsorganen der EU, den nationalen Wertpapieraufsichtsbehörden und den Marktteilnehmern definiert wurden, erlassen. Parallel zu den Prozessen auf europäischer Ebene hat die Umsetzung der Richtlinienbestimmungen durch die Mitgliedstaaten über nationale Gesetzgebungsverfahren² zu erfolgen und soll bis zum 31. Januar 2007 abgeschlossen sein³. Die Marktteilnehmer müssen die Anforderungen ab dem 01. November 2007 erfüllen. EU Binnenmarktkommissar McCreevy unterstreicht, dass sich die Marktteilnehmer rechtzeitig vorbereiten müssen [Exch06]: *„MiFID is a ground-breaking piece of legislation. It will transform the landscape for the trading of securities and introduce much needed competition and efficiency. [...] All firms in the business should now prepare.“*

Ziel dieses Beitrags ist es, vor dem Hintergrund des näher rückenden Umsetzungstermins, den Umsetzungsstatus der MiFID in der deutschen Finanzindustrie aufzuzeigen und die Einschätzung deutscher Wertpapierfirmen bezüglich der Nutzeffekte der Kernanforderungen der MiFID einerseits und bezüglich der Umsetzungserfordernisse im Hinblick auf Prozesse und IT-Systeme andererseits darzustellen. Weiterhin werden der Informationsstand der Häuser, der Projekt-Aufbau sowie die erwarteten Auswirkungen auf interne Prozesse und IT-Strukturen erhoben. Methodisch basiert diese Arbeit auf einer empirischen Studie, die zwischen Dezember 2005 und Februar 2006 mittels eines strukturierten Fragebogens unter 193 deutschen Wertpapierfirmen durchgeführt wurde. Die Ergebnisse wurden den Teilnehmern am 24. April 2006 übermittelt und dienen dem Benchmarking des eigenen Umsetzungsstands.

Kapitel 2 stellt kurz die existierende Literatur sowie existierende europäische Studien zur Umsetzung der MiFID-Anforderungen dar und erläutert den Aufbau der hier durchgeführten Studie zur MiFID-Readiness der deutschen Finanzindustrie. Kapitel 3 präsentiert die Ergebnisse der Umfrage zu den Projektrahmendaten und Kostenschätzungen sowie Einschätzungen bezüglich der Kernthemen Best Execution und Markttransparenz. Abschließend fasst Kapitel 4 die Ergebnisse zusammen und gibt einen Ausblick auf weitere Forschungsschritte.

¹ [Euro06a] und [Euro06b]

² Hierzu veröffentlichte im September 2006 das Bundesfinanzministerium einen Referentenentwurf [BuMF06].

³ Level 3 dient der Gewährleistung einheitlicher Umsetzung und Anwendung durch Empfehlungen zu Auslegungsfragen durch das „Committee of European Securities Regulators“ (CESR), während in Level 4 die EU Kommission die Einhaltung der Rechtsvorschriften durch die Mitgliedstaaten überprüft und gegebenenfalls Vertragsverletzungsverfahren einleitet.

2 Existierende Literatur und Aufbau der MiFID-Studie für die deutsche Finanzindustrie

Naturgemäß existieren aufgrund der Umsetzung der MiFID ab 01. November 2007 und der noch fehlenden empirischen Datengrundlage bisher nur wenige wissenschaftliche Arbeiten, die sich mit Inhalten und Auswirkungen der MiFID befassen. Die vorhandene Literatur in diesem Kontext fokussiert daher auf die Darstellung der inhaltlichen Anforderungen und Auslegungen der MiFID-Vorschriften (z. B. [GoSe05], [CaLa06]) oder versucht durch Analogieschluss die Konsequenzen anderer marktstruktureller Veränderungen auf die MiFID zu übertragen (z. B. [AyDe05]). Erste Erkenntnisse bzgl. des Umsetzungsstands in den Unternehmen wurden bisher durch Umfragen von Industrieinitiativen und Beratungsunternehmen gewonnen. Insbesondere in Großbritannien wurden verschiedene Umfragen zur MiFID durchgeführt: Im September 2005 zeigte die von der „MiFID Joint Working Group“ durchgeführte Studie unter Londoner Investmentbanken [Full05] eine große Unsicherheit auf. Lediglich 24% verstanden die inhaltlichen Anforderungen der Regulierung hinreichend, um z.B. notwendige Maßnahmen für die Entwicklung und Umsetzung der Best Execution Anforderungen zu ergreifen. 80% der antwortenden Firmen hatten noch keine technologische Strategie für die Anforderungen der Markttransparenz-Vorschriften entwickelt. Die im November 2005 von Liquidnet Europe [Liqu05] unter seinen Mitgliedern durchgeführte Studie setzte den Fokus auf das Kosten-/Nutzenverhältnis der MiFID. Mehr als 55% der Befragten sahen in der MiFID hohe Umsetzungskosten und geringe Nutzeffekte für ihre Firma. Insgesamt betrachteten lediglich 39% die MiFID als eine mögliche Verbesserung im Vergleich zur aktuellen Wettbewerbssituation. Eine Studie von KPMG [KPMG06] im Februar 2006 unter 199 europäischen Wertpapierfirmen unterstrich den noch geringen Vorbereitungsstatus: Erst 48% der antwortenden Firmen hatten sich mit der MiFID beschäftigt, 44% erwarteten dennoch, die MiFID bis November 2007 umgesetzt zu haben. Der größte Nutzen der MiFID wurde von 37% der Firmen im Angebot grenzüberschreitender Wertpapierdienstleistungen gesehen, 24% sahen für ihre Firma keine Vorteile durch die neue EU Richtlinie.

Die Stichprobe der vorliegenden MiFID-Studie für die deutsche Finanzindustrie umfasst 193 Wertpapierfirmen, die sich aus den 100 größten Kreditinstituten (Jahr 2004) sowie den Handelsteilnehmern der drei umsatzstärksten deutschen Börsen (Frankfurt, Stuttgart,

München)⁴ zusammensetzen. Der Fragebogen umfasst 33 Fragen mit bis zu neun Unterfragen. Verständlichkeit und Handhabbarkeit der Fragen wurden durch fünf Pre-Tests⁵ überprüft. Die Befragung startete am 12. Dezember 2005 mit einem festgesetzten Auswertungstichtag am 10. Februar 2006 (Feldzeit: neun Wochen). Von den versendeten Fragebögen wurden 55 zurückgeschickt⁶, 28 davon in elektronischer Form, 27 Antworten schriftlich, was einer Rücklaufquote von 28,5% entspricht. Tabelle 1 zeigt den Rücklauf der Studie im Überblick:

Anzahl der kontaktierten Firmen	Anzahl der versandten Fragebögen	Fragebogenrücklauf (Absolut-/Prozentwert)
193	131	55 ≈ 28,5 %

Tab. 1: Rücklauf der MiFID-Studie

Ein Test der Repräsentativität bezüglich der Größenverteilung der Wertpapierfirmen in der Stichprobe ist positiv und diese kann damit bezüglich der Verteilung der Bilanzsumme als repräsentativ für die Grundgesamtheit eingestuft werden⁷. Abbildung 1 stellt die Struktur des Rücklaufs bezüglich der Geschäftstypen der befragten Häuser dar:

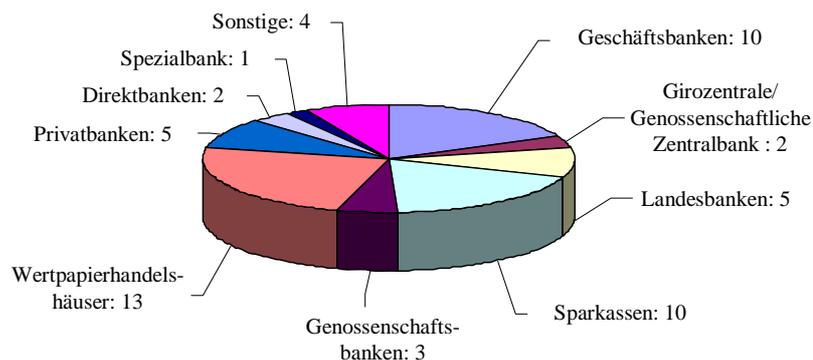


Abb. 1: Zusammensetzung der Wertpapierfirmen innerhalb des Rücklaufs

3 Zentrale Ergebnisse der Studie

Für die Auswertungen der Studie, wurden die teilnehmenden Wertpapierfirmen anhand zweier Dimensionen klassifiziert: i) die Größe anhand der Bilanzsumme der Unternehmen und ii) die

⁴ Vgl. [Kars05]; vgl. [DeBöoJ]; vgl. [BöMüoJ]; vgl. [BöStoJ]

⁵ Dadurch sollte erreicht werden, dass die Fragen für Wertpapierfirmen, die sich bereits mit der MiFID befasst haben, relevant sind und alle wichtigen Sachverhalte erfassen. Für Wertpapierfirmen, die sich noch nicht bzw. wenig mit der MiFID auseinandergesetzt haben, standen Allgemeinverständlichkeit und Klarheit der Fragestellung im Vordergrund.

⁶ In einer telefonischen Kontaktaufnahme wurde vorab für die Teilnahme an der MiFID-Studie geworben, was in 131 Zustimmungen resultierte, den Fragebogen zu versenden.

⁷ Der Wert der Irrtumswahrscheinlichkeit des Chi-Quadrat-Tests übersteigt mit 0,156 das üblicherweise geforderte Signifikanzniveau von 0,050, damit kann die Nullhypothese einer gleichen Verteilung in Stichprobe und Population nicht verworfen werden.

Selbsteinschätzung der Wertpapierfirmen bzgl. ihres Kenntnisstandes zur MiFID. In beiden Dimensionen wurden 3 Gruppen gebildet. Nach der Bilanzsumme lassen sich kleine (Gruppe K), mittlere (Gruppe M) und große Wertpapierfirmen (Gruppe G) unterscheiden, anhand des MiFID-Kennntnisstandes Wertpapierfirmen, die erst in der Informationssammlung sind (Gruppe A), Wertpapierfirmen, denen die Anforderungen der MiFID im Wesentlichen bekannt sind (Gruppe B), sowie Wertpapierfirmen, die mit der MiFID bereits sehr vertraut sind (Gruppe C). Tabelle 2 verdeutlicht, dass in allen Klassen von Unternehmensgrößen Häuser mit verschiedenen Informationsständen vorhanden sind. Statistisch konnte kein Zusammenhang zwischen den Dimensionen Unternehmensgröße und Informationsstand festgestellt werden⁸.

		Informationsstand			Summe
		A	B	C	
Unternehmensgröße nach Mrd. €Bilanzsumme	< 1 (Gruppe K)	6	7	2	15
	1 - 100 (Gruppe M)	15	10	2	27
	> 100 (Gruppe G)	2	7	4	13
Summe		23	24	8	

Tab. 2: Verteilung nach Informationsstand und Unternehmensgröße

Bei der weiteren Darstellung der Ergebnisse werden sowohl die Ergebnisse über alle antwortenden Häuser als auch relevante Abweichungen in den Ergebnissen bezüglich des MiFID-Kennntnisstandes (Gruppe A einerseits und Gruppen B&C der besser informierten Häuser andererseits) sowie bezüglich der Größenklassen K, M und G dargestellt. Auf Basis der Klassifizierung anhand der Dimensionen Unternehmensgröße und Informationsstand wurden die Antworten der Häuser auf Zusammenhänge mit dieser Klassifizierung hin untersucht. Hierzu wurde Goodman und Kruskals Gamma⁹ und dessen Signifikanz berechnet. Signifikante Ergebnisse werden bei den jeweiligen Antworten erläutert. Für die folgenden Ausführungen werden die Antworten der Häuser zu 19 aus den insgesamt 33 Fragen einbezogen und dabei in die Themen Projektrahmendaten und Kostenschätzungen einerseits und Einschätzungen bezüglich der Kernthemen Best Execution und Markttransparenz andererseits gegliedert.

3.1 Projektrahmendaten und Kostenschätzungen

Bezüglich des geforderten Starttermins am 01. November 2007 stellt die folgende Tabelle 3 dar, welche Vorlaufzeiten die Wertpapierfirmen – aufgegliedert nach Unternehmensgröße – zur

⁸ Für die beiden Klassifizierungsdimensionen Unternehmensgröße und Informationsstand selbst konnte nur ein schwacher positiver, jedoch nicht signifikanter Zusammenhang gefunden werden.

⁹ Gamma ist ein Assoziationsmaß, das auf der Anzahl der Fehlorderungen einer Variablen basiert, die sich ergeben, wenn man nach einer zweiten Variable aufsteigend ordnet. Gamma kann Werte von -1 (totaler negativer Zusammenhang) bis +1 (totaler positiver Zusammenhang) annehmen. Für Gamma = 0 liegt keinerlei Zusammenhang vor. [BüZö02]

Umsetzung der MiFID insgesamt erwarten (Stand Februar 2006). Erwartungsgemäß steigt die notwendige Vorbereitungszeit mit der Unternehmensgröße an und beträgt für 8 Wertpapierfirmen über 18 Monate. 85,2% sehen eine Umsetzung innerhalb von 18 Monaten als realistisch an, d.h. die deutsche Finanzindustrie steht einer pünktlichen Implementierung der Anforderungen optimistisch gegenüber.

Frage: Wie viel Vorlaufzeit (in Monaten) halten Sie für erforderlich, um die Kern-Anforderungen der MiFID zu erfüllen? (n=54)

Bilanzsumme (in Mrd. €)	Vorlaufzeit in Monaten			
	1 – 6	6 – 12	12 – 18	> 18
<1 (Gruppe K)	3	7	5	0
1- 100 (Gruppe M)	3	11	9	3
> 100 (Gruppe G)	0	2	6	5
Randhäufigkeit	6	20	20	8

Tab. 3: Vorlaufzeit zur Umsetzung der MiFID insgesamt¹⁰

Die Dominanz von Softwareanbietern und Beratungsunternehmen als Dienstleister bei der Umsetzung des MiFID-Projekts (siehe Tabelle 4) kann durch die erforderlichen Änderungen in den Handelsprozessen und –systemen begründet werden. 46% der Wertpapierfirmen wollen keine Unterstützung durch externe Dienstleister in Anspruch nehmen.

Frage: Durch welche externen Dienstleister wird das Projekt unterstützt bzw. soll das Projekt unterstützt werden (mehrere Antworten möglich)? (n=50)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
Beratungsunternehmen	11	22,0%	4	20,0%	7	23,3%
Anwaltskanzlei	10	20,0%	2	10,0%	8	26,7%
Softwareanbieter	12	24,0%	6	30,0%	6	20,0%
Informationsvendoren	2	4,0%	2	10,0%	0	0,0%
Makler/Broker	0	0,0%	0	0,0%	0	0,0%
Keine Unterstützung durch externe Dienstleister	23	46,0%	10	50,0%	13	43,3%
Sonstige	13	26,0%	3	15,0%	10	33,3%

Tab. 4: Externe Dienstleister

45 der befragten Wertpapierfirmen konnten bereits eine Schätzung der Gesamtkosten der Umsetzung der MiFID-Anforderungen in ihrem Unternehmen abgeben (siehe Tabelle 5). 57,8% aller Wertpapierfirmen (sogar 77,8% der Wertpapierfirmen aus der Gruppe A) schätzen ihre Umsetzungskosten kleiner als 0,5 Mio. € Insgesamt 80% der Befragten schätzen ihre Umsetzungskosten auf unter 1 Mio. € (94,5% in der Gruppe A). Auch die besser informierten Wertpapierfirmen (Gruppen B und C) schätzen zu 70,3% ihre Umsetzungskosten auf diesem

¹⁰ Hier und im Weiteren wird bei den einzelnen Fragen auch die jeweilige Anzahl n der Wertpapierfirmen, die diese Frage beantwortet haben, angegeben.

niedrigen Niveau ein. Hier konnte ein positiver signifikanter Zusammenhang zwischen dem Informationsstand der antwortenden Unternehmen und den erwarteten einmaligen Umsetzungskosten gezeigt werden (Gamma = 0,616, Signifikanzniveau = 0,002). Lediglich 4,4% der insgesamt antwortenden Wertpapierfirmen schätzten ihre einmaligen Umsetzungskosten auf über 20 Mio. € Die Klassen 5 bis 10 Mio. € und 10 bis 20 Mio. € sind nicht besetzt. Ein etwas anderes Bild ergibt sich, wenn man die Antworten der Wertpapierfirmen nach der Unternehmensgröße getrennt betrachtet. Hierbei schätzen zwar auch 94,3% der kleinen und mittleren Unternehmen ihre einmaligen Umsetzungskosten kleiner als 1 Mio. € von den großen Unternehmen jedoch teilen nur 30% diese Einschätzung. Die Hälfte der großen Unternehmen schätzt die Kosten auf 1 bis 5 Mio. € Lediglich 20% der großen Unternehmen schätzen ihre Umsetzungskosten auf über 20 Mio. € Erwartungsgemäß konnte dabei ein signifikanter positiver Zusammenhang zwischen der Unternehmensgröße und der erwarteten einmaligen Umsetzungskosten (Gamma = 0,8, Signifikanzniveau = 0) festgestellt werden.

Frage: Wie hoch schätzen Sie die Gesamtkosten in EUR (einmalige Umsetzungskosten) inkl. internen und externen Personalkosten für die Umsetzung der MiFID in Ihrem Unternehmen? (n=45)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
< 500.000	26	57,8%	14	77,8%	12	44,4%
500.000 - 1 Mio.	10	22,2%	3	16,7%	7	25,9%
1 - 5 Mio.	7	15,6%	1	5,6%	6	22,2%
5 - 10 Mio.	0	0,0%	0	0,0%	0	0,0%
10 - 20 Mio.	0	0,0%	0	0,0%	0	0,0%
> 20 Mio.	2	4,4%	0	0,0%	2	7,4%

Tab. 5: Einmalige Umsetzungskosten

Knapp 70% der Wertpapierfirmen haben noch keine Budget-Planung für ein MiFID-Projekt vorgenommen (siehe Tabelle 6). Hier ist eine deutliche Diskrepanz zwischen den Wertpapierfirmen der Gruppe A und den Wertpapierfirmen aus den Gruppen B und C festzustellen.

Frage: Für welche Jahre wurden / werden Budgets zur Umsetzung der MiFID bereitgestellt (mehrere Antworten möglich)? (n=53)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
Budget für 2005	1	1,9%	0	0,0%	1	3,2%
Budget für 2006	8	15,1%	2	9,1%	6	19,4%
Budget für 2007	16	30,2%	3	13,6%	13	41,9%
Noch keine Planung vorhanden	37	69,8%	19	86,4%	17	54,8%

Tab. 6: Budgetplanung

Erwartungsgemäß dominiert bei der Frage nach den inhaltlichen Themen, die die Unternehmen mit ihrem MiFID Projekt aktuell adressieren (siehe Tabelle 7), die Informationssammlung zum

Thema MiFID – was insbesondere dadurch begründet ist, dass die Level 2 Finalisierung und der nationale Umsetzungsprozess aktuell noch ausstehen. Auffällig hingegen ist, dass die Analyse der Auswirkungen auf Prozesse und IT über alle Gruppen hinweg von einer größeren Anzahl von Wertpapierfirmen adressiert wurde als die Analyse der Auswirkungen auf die Strategie. Die Wertpapierfirmen sehen die Anforderungen der MiFID bisher eher als eine regulatorische Pflicht, als eine strategische Chance im Wettbewerb.

Frage: Welche inhaltlichen Themen hat Ihr Haus adressiert (mehrere Antworten möglich)? (n=54)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
Informationssammlung zur MiFID	50	92,6%	20	90,9%	30	93,8%
Planungsstatus des Projekts	20	37,0%	4	18,2%	16	50,0%
Analyse der Kundeninformation und rechtlichen Vertragsdokumentation	12	22,2%	2	9,1%	10	31,3%
Analyse der Auswirkungen auf Prozesse und IT	27	50,0%	5	22,7%	22	68,8%
Analyse der Auswirkungen auf die Strategie	17	31,5%	4	18,2%	13	40,6%
Durchführungsstatus des Projekts	6	11,1%	2	9,1%	4	12,5%

Tab. 7: Inhaltliche Themen

3.2 Einschätzungen bezüglich der Kernthemen Best Execution und Markttransparenz

Neben den Einschätzungen der Häuser zu den Projektrahmendaten wurden die Themen im Kontext Wertpapierhandel adressiert und im Folgenden werden die Antworten bezüglich der Kernthemen Best Execution und Markttransparenz aufgezeigt. Wertpapierfirmen werden mit der MiFID zur „kundengünstigsten Ausführung“ von Aufträgen verpflichtet („Best Execution“ gemäß Art. 21 MiFID) [Euro04]. Best Execution bezieht sich dabei nicht nur auf den Preis bzw. Kurs des Wertpapiers, sondern Wertpapierfirmen haben unter Berücksichtigung einer Vielzahl von Parametern (u. a. Kurs, Kosten, Schnelligkeit, Umfang) alle angemessenen Maßnahmen zu ergreifen, um das bestmögliche Ergebnis für ihre Kunden zu erreichen. Dies ist in „Grundsätzen der Auftragsausführung (Best Execution Policy)“ einer Wertpapierfirma zu formulieren, die für jede Gattung von Finanzinstrumenten Angaben zu den Handelsplätzen, an denen die Wertpapierfirma Aufträge ihrer Kunden ausführt, und den Faktoren, die für die Wahl des Ausführungsplatzes ausschlaggebend sind, enthält. Für diese Best Execution Policy ist die vorherige Zustimmung der Kunden einzuholen (Art. 21(3) MiFID) und die Policy ist mindestens jährlich zu überprüfen. Insbesondere muss das Haus seinen Kunden ex post nachweisen können, dass es deren Aufträge im Einklang mit der Policy ausgeführt hat.

In diesem Kontext sehen 30,8% der Wertpapierfirmen in der Best Execution Policy selbst eine Differenzierungsmöglichkeit im Wettbewerb. Der Einsatz neuer Technologien und die Nutzung

neuer Geschäftsmodelle bei der Auftragsausführung werden mit jeweils 26,9% an zweiter Stelle genannt. 57,7% sehen diese Frage als für sie noch nicht relevant an (siehe Tabelle 8).

Frage: Wo sehen Sie Wettbewerbsmöglichkeiten, um sich mit Ihren Grundsätzen der Auftragsausführung gegenüber anderen Wertpapierfirmen differenziert zu positionieren (mehrere Antworten möglich)? (n=52)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
In den Grundsätzen der Auftragsausführung selbst	16	30,8%	4	17,4%	12	41,4%
Anreize setzen, um Kundenweisungen abweichend von den Ausf.-Grundsätzen zu erhalten	6	11,5%	2	8,7%	4	13,8%
Neugestaltung von Preisen und Provisionssätzen	14	26,9%	4	17,4%	10	34,5%
Neue Geschäftsmodelle (z. B. Angebot neuer Wertpapierdienstleistungen)	10	19,2%	3	13,0%	7	24,1%
Einsatz neuer Technologien (z. B. für neue Order Routing Konzepte)	14	26,9%	4	17,4%	10	34,5%
Das ist für uns noch nicht relevant	30	57,7%	16	69,6%	14	48,3%
Eigene Ansätze	2	3,8%	0	0,0%	2	6,9%

Tab. 8: Wettbewerbsmöglichkeiten

Die Schätzungen für den Zusatzaufwand der Best Execution-Anforderungen verdeutlichen (siehe Abbildung 2), dass insbesondere die Umsetzung der Informationspflichten nach Art. 21(3) MiFID einen hohen Aufwand bedeuten - im Speziellen für große Unternehmen. Bezüglich des Aufwands für die Historisierung und Rekonstruktion der Daten zu Orderausführungen zum ex post-Nachweis der Ausführung gemäß Best Execution Policy kommt mehr als die Hälfte aller befragten Wertpapierfirmen zu eher niedrigen oder sehr niedrigen Aufwandsschätzungen.

Frage: Wie schätzen Sie den Zusatzaufwand für Ihr Unternehmen ein bezüglich der Anforderungen für die Grundsätze der Auftragsausführung?

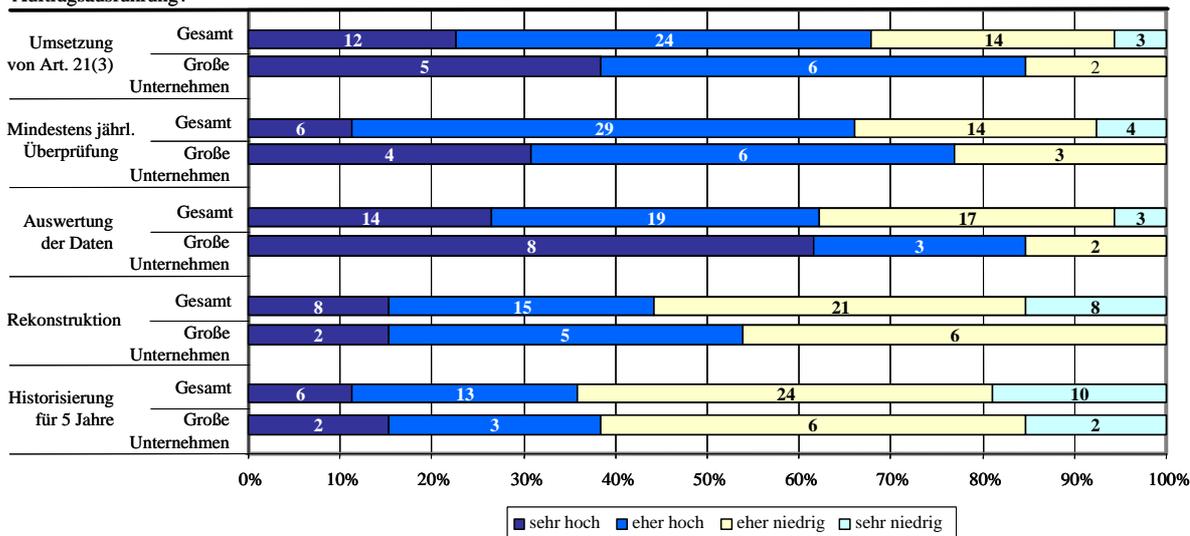


Abb. 2: Zusatzaufwand der Best Execution Anforderungen

Im Kontext des Themas Markttransparenz waren die Vorschriften des Art. 27 MiFID das zentrale Element der politischen Diskussion zur MiFID, da erstmals umfangreiche Transparenzanforderungen für systematische Internalisierer¹¹ vorgesehen werden, was für alle Finanzmärkte aller EU Mitgliedsstaaten eine grundlegende Neuerung darstellt. Gemäß Art. 27 sind systematische Internalisierer verpflichtet, verbindliche und zugängliche Kursofferten (sog. „Quotes“) für Transaktionen in liquiden Aktien zu veröffentlichen.

Die Einschätzungen zur Relevanz/Nützlichkeit der Vorschriften des Art. 27 weisen deutliche Unterschiede auf (siehe Abbildung 3). So stimmen 71,4% der Aussage voll bzw. eher zu, dass die Nützlichkeit von Quote-Daten für den Eigenhandel hoch ist. Die Nutzbarkeit für Privatkunden wird deutlich geringer eingeschätzt. Des Weiteren sehen 76,1% der Wertpapierfirmen die Unterteilung in liquide und weniger liquide Aktien als sinnvoll an.

Frage: Welche Meinung haben Sie zur Relevanz/Nützlichkeit des Art. 27(1) MiFID?

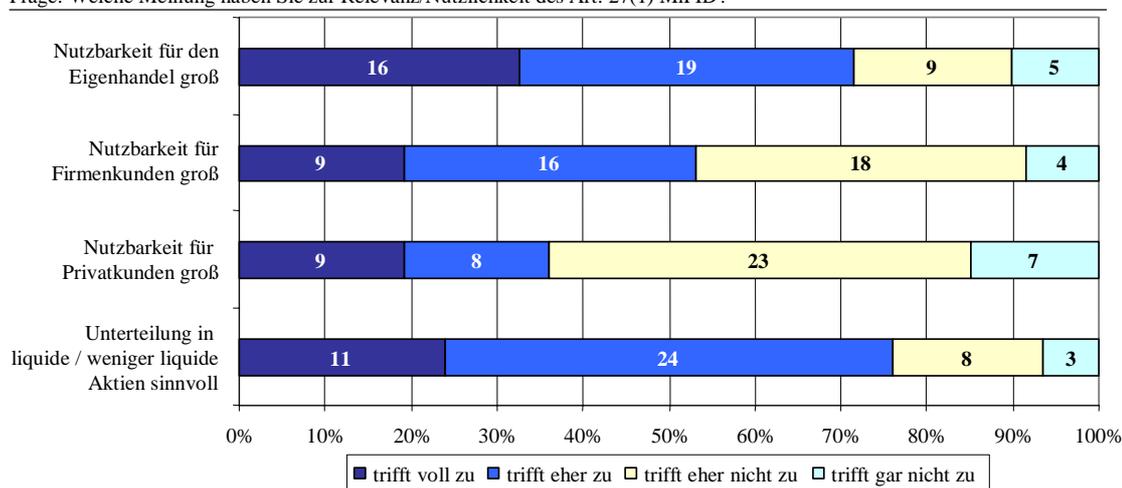


Abb. 3: Relevanz/Nützlichkeit des Art. 27(1) MiFID

Die Erfüllung von vier kumulativ vorliegenden Kriterien¹² ist Basis für die Einstufung einer Wertpapierfirma als systematischer Internalisierer. Aufgrund der regulatorischen Verpflichtungen zur Pre-Trade Transparenz, die mit der Einstufung als systematischer Internalisierer einhergehen, wurde die Selbsteinschätzung der Wertpapierfirmen bezüglich der Erfüllung dieser vier Kriterien erhoben (siehe Abbildung 4): Mehr als die Hälfte (55,1%) gibt an, dass sie über (i) Regeln und Prozesse zur Ausführung von Kundenaufträgen gegen den eigenen Handelsbestand verfügen. Bei 40,8% existiert (ii) Personal oder ein automatisches technisches System zur Ausführung von Kundenorders gegen den eigenen Handelsbestand.

¹¹ Nach Art. 4 Abs. 1 Nr. 7 MiFID ist ein „systematischen Internalisierer“ eine Wertpapierfirma, „die in organisierter und systematischer Weise häufig regelmäßig Handel für eigene Rechnung durch Ausführung von Kundenaufträgen außerhalb eines geregelten Marktes oder MTFs treibt“.

¹² Vgl. [Euro06a], Art. 21(1)

Über ein Viertel der Befragten (26,5%) bestätigte, dass die (iii) Bereitstellung für Kunden auf regelmäßiger und kontinuierlicher Basis stattfindet und 14,3% gaben an, dass die (iv) Internalisierung in ihrem Geschäftsmodell eine materielle kommerzielle Rolle spielt. Aus den Antworten lassen sich Hinweise auf die Zahl der zukünftig in Deutschland agierenden systematischen Internalisierer ableiten.

Frage: In wie weit könnten die folgenden Kriterien auf Ihr Unternehmen zutreffen?

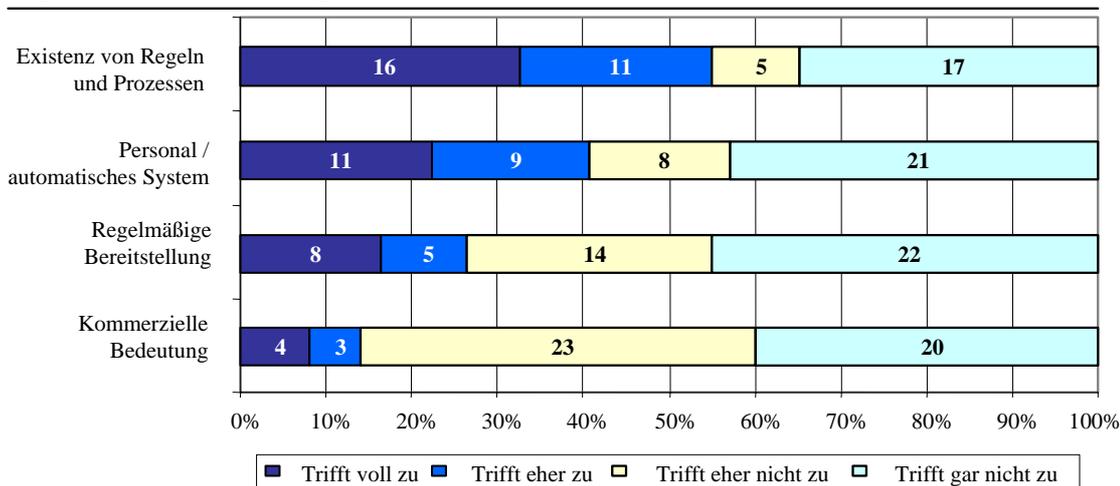


Abb. 4: Kriterien für systematische Internalisierung

Mit der MiFID werden umfassende Post-Trade Transparenz Anforderungen für den außerbörslichen Handel aufgestellt, die für den deutschen Markt neu, in anderen europäischen Märkten aber bereits etablierte Marktpraxis sind (z. B. UK). Für die außerbörsliche Orderausführung sieht Art. 28 MiFID vor, dass alle Geschäfte in Aktien auf Echtzeitbasis mit Umfang, Kurs und Zeitpunkt veröffentlicht werden müssen. Es ist den Wertpapierfirmen überlassen, auf welche Weise sie die Transaktionsdaten verbreiten; laut MiFID können auf der Basis „angemessener kaufmännischer Bedingungen“ hierfür auch Erlöse erzielt werden.

Post-Trade Daten sollen soweit möglich in Echtzeit veröffentlicht werden, in jedem Fall jedoch innerhalb von drei Minuten nach der Transaktion. Für Blocktrades ermöglicht die MiFID eine verzögerte Veröffentlichung. Die folgende Tabelle 9 zeigt die Einschätzungen der Wertpapierfirmen bzgl. der zusätzlichen Kosten, um eine Zeitvorgabe von 3 Minuten für die Veröffentlichung von Post-Trade Daten einzuhalten. Die Ergebnisse werden unterteilt nach der Unternehmensgröße dargestellt. Wie man sieht, schätzen auch 8 von 10 großen Unternehmen diese Kosten auf unter 1 Mio. € Auch hier konnte ein signifikanter positiver Zusammenhang zwischen der Unternehmensgröße und den erwarteten Umsetzungskosten ($\text{Gamma} = 0,767$, $\text{Signifikanzniveau} = 0$) festgestellt werden. Ebenso konnte auch hier ein positiver signifikanter

Zusammenhang mit dem Informationsstand der antwortenden Unternehmen aufgezeigt werden (Gamma = 0,704, Signifikanzniveau = 0,003).

Frage: Wie hoch schätzen Sie die zusätzlichen Kosten (einmalige Umsetzungskosten) in EUR für die Anpassung der Interaktion von Handel, Middle Office und Back Office ein, um z.B. eine Zeitvorgabe von 3 Minuten für die Veröffentlichung von Post-Trade Daten einzuhalten?

Bilanzsumme (in Mrd. €)	Einmalige Umsetzungskosten (in €)					
	< 500.000	500.000 – 1 Mio.	1 – 5 Mio.	5 – 10 Mio.	10 – 20 Mio.	> 20 Mio.
<1 (Gruppe K)	9	1	0	0	0	0
1- 100 (Gruppe M)	16	4	1	0	0	0
> 100 (Gruppe G)	2	6	0	1	1	0
Randhäufigkeit	27	11	1	1	1	0

Tab. 9: Einmalige Umsetzungskosten für Post-Trade Anforderungen

Als Veröffentlichungskanäle für die Post-Trade Daten kommen gemäß Art. 28(3) MiFID das System eines Regulierten Marktes/Börse oder MTFs¹³, die zuständigen Stellen eines Dritten oder eigene Vorkehrungen in Frage. Zwei Drittel der Befragten Wertpapierfirmen haben sich noch für keinen Kommunikationskanal entschieden (siehe Tabelle 10). Aus der Gruppe der besser informierten Wertpapierfirmen haben sich 56,3% noch nicht entschieden. Von den Wertpapierfirmen, die bereits eine Entscheidung getroffen haben, werden am häufigsten die Systeme von Börsen oder Informationsvondoren (z.B. Reuters oder Bloomberg) als mögliche Kommunikationskanäle benannt. Nur sehr wenige Wertpapierfirmen planen bisher die Systeme eines MTFs zur Veröffentlichung zu verwenden oder eigene Systeme zu nutzen.

Frage: Welche Reporting-Kanäle planen Sie in Ihrem Unternehmen für die Bekanntmachung abgeschlossener Transaktionen einzurichten (mehrere Antworten möglich)? (n=54)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
Systeme von Börsen	17	31,5%	4	18,2%	13	40,6%
Systeme von MTFs	1	1,9%	0	0,0%	1	3,1%
Systeme von Informationsvondoren	8	14,8%	2	9,1%	6	18,8%
Eigene Web-Anwendungen	2	3,7%	0	0,0%	2	6,3%
Eigene Marktdatensysteme	3	5,6%	1	4,5%	2	6,3%
Anderes System	0	0,0%	0	0,0%	0	0,0%
Das ist noch offen	36	66,7%	18	81,8%	18	56,3%

Tab. 10: Post-Trade Reporting Kanäle

Bei der Gestaltung des Geschäftsprozesses zur Veröffentlichung von Post-Trade Daten spielt das Kriterium der Anonymität für 57,9% der antwortenden Wertpapierfirmen eine wichtige Rolle, weshalb sie planen, keine eigenen Systeme zur Publikation zu verwenden (siehe Tabelle 11). 39,5% der Wertpapierfirmen möchten von der Möglichkeit zur verzögerten

¹³ „Multilateral Trading Facility (MTF)“ ist ein börsenähnliches, von einer Wertpapierfirma oder einem Marktbetreiber betriebenes, multilaterales System, das bezüglich der regulatorischen Anforderungen weitgehend mit Regulierten Märkten gleichgestellt ist.

Veröffentlichung von Blocktrade-Daten Gebrauch machen. Insbesondere in der Gruppe der besser informierten Wertpapierfirmen (Gruppen B&C) will die Hälfte davon Gebrauch machen.

Frage: Wie planen Sie den Prozess zur Veröffentlichung von Post-Trade Daten zu gestalten (mehrere Antworten möglich)? (n=38)

	Absolute Häufigkeit (Gesamt)	Relative Häufigkeit (Gesamt)	Absolute Häufigkeit (Gruppe A)	Relative Häufigkeit (Gruppe A)	Absolute Häufigkeit (Gruppe B&C)	Relative Häufigkeit (Gruppe B&C)
Nutzung der Möglichkeit zur verzögerten Veröffentlichung bei großen Transaktionen	15	39,5%	3	21,4%	12	50,0%
Wegen des Aufwandes keine Separierung nach Transaktionsgrößen	12	31,6%	5	35,7%	7	29,2%
Wegen fehlender Anonymität keine eigenen Systeme zur Publikation	22	57,9%	10	71,4%	12	50,0%

Tab. 11: Prozess zur Veröffentlichung von Post-Trade Daten

18,6% der Befragten untersuchen, ob sich aus der Veröffentlichung von Post-Trade Daten eine neue Ertragsquelle für Wertpapierfirmen erschließen lässt und arbeiten an einer wirtschaftlichen Bewertung ihrer Transaktionsdaten. Allerdings hat kein befragtes Haus bisher eine Vorstellung bzgl. des Wertes der eigenen Transaktionsdaten. Unabhängig vom Informationsstand und Unternehmensgröße glaubt der überwiegende Teil der Wertpapierfirmen, dass mit den eigenen Transaktionsdaten keine Erlöse zu erzielen sind.

Die Transparenzvorschriften der MiFID in ihrer jetzigen Form gelten nur für Aktien. Die MiFID eröffnet jedoch den EU Mitgliedsstaaten die Möglichkeit, diese Vorschriften auf andere Finanzinstrumente auszudehnen. 51,1% der befragten Wertpapierfirmen gaben an, dass sie solch eine Ausweitung des Anwendungsbereiches als nicht sinnvoll erachten.

Tabelle 12 gliedert die geschätzten Vorlaufzeiten für die Umsetzung der Kernanforderungen der MiFID auf¹⁴:

Frage: Wie viel Vorlaufzeit (in Monaten) halten Sie für erforderlich, um die Kern-Anforderungen der MiFID zu erfüllen?

Anforderung	Vorlaufzeit in Monaten			
	1 – 6	6 – 12	12 – 18	> 18
Entwicklung der Grundsätze der Auftragsausführung	26	22	4	1
Umsetzung der Grundsätze der Auftragsausführung	19	26	7	1
Anforderungen zur Pre-Trade Transparenz	18	16	7	4
Anforderungen zur Post-Trade Transparenz	16	23	9	2

Tab. 12: Vorlaufzeit zur Umsetzung der Kernanforderungen

¹⁴ Bezüglich dieser Frage liegt ein signifikanter positiver Zusammenhang zwischen Unternehmensgröße und erwarteten Vorlaufzeiten (Gamma > 0,55, Signifikanzniveau ≤ 0,01) vor.

Dabei wird deutlich, dass die Anforderungen zur Pre- und Post-Trade Transparenz zeitlich aufwendiger eingeschätzt werden als die Anforderungen zur Entwicklung und Umsetzung der Grundsätze der Auftragsausführung. Dies lässt sich dadurch begründen, dass diese neuen Transparenzanforderungen für deutsche Wertpapierfirmen auch neue Prozesse implizieren: Die Einhaltung der Erfordernisse, Pre-Trade Daten kontinuierlich und Post-Trade Daten „soweit wie möglich in Echtzeit“ (Art.29 [Euro06a]) zu publizieren, erfordern grundlegende Änderungen in der Interaktion zwischen Handel und Back-Office bei außerbörslichen Transaktionen.

Eine zentrale Frage ist die nach möglichen neuen Wettbewerbspotenzialen, die sich mit der MiFID für Wertpapierfirmen ergeben können (siehe Abbildung 5). In der Gesamtbetrachtung fällt auf, dass die Wertpapierfirmen für alle Dienstleistungen ein eher niedriges bzw. sehr niedriges Wettbewerbspotential sehen, wobei in der Bereitstellung von Services für Firmenkunden und in den Grundsätzen der Auftragsausführung die höchsten Wettbewerbspotentiale gesehen werden.

Frage: Wo sehen Sie Wettbewerbspotenzial für Ihre Firma, das sich mit der Umsetzung der MiFID bieten könnte?

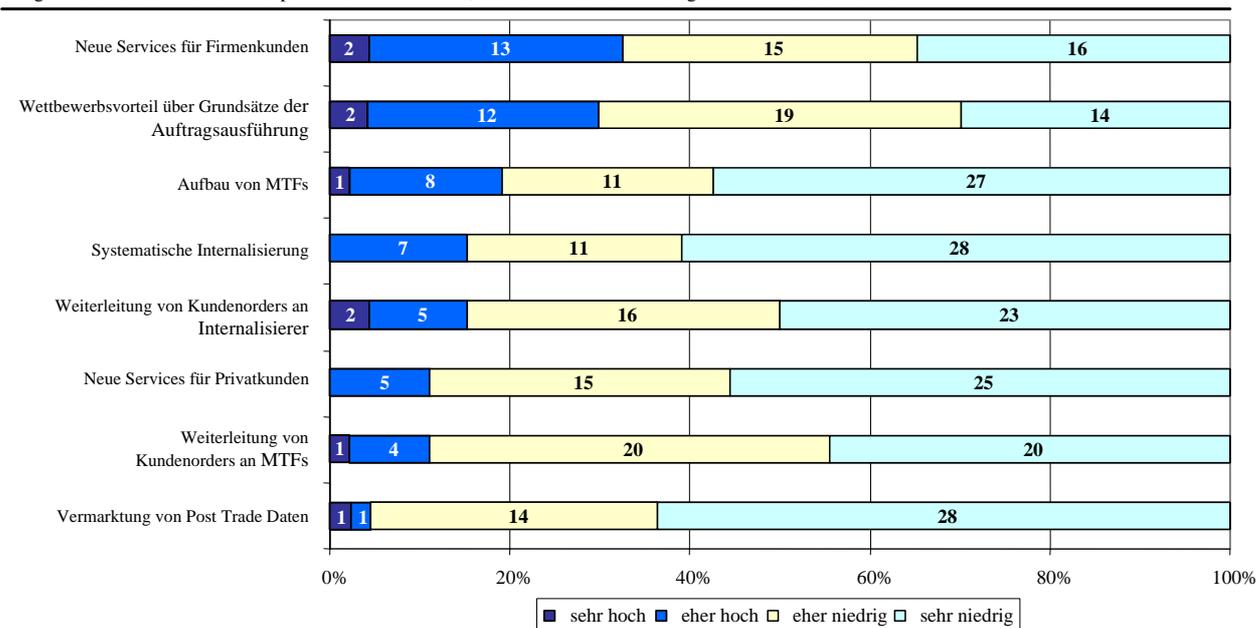


Abb. 5: Wettbewerbspotenzial der MiFID-Umsetzung

4 Zusammenfassung und Ausblick

Für die Umsetzung der Anforderungen der MiFID zum 01. November 2007 werden Wertpapierfirmen weit reichende Änderungen in ihren Prozessen und Systemen vorzunehmen

haben. Dies bezieht sich insbesondere auf die Themen Best Execution und Markttransparenz. Das Thema Best Execution wird dabei sowohl als Herausforderung als auch als Chance im Wettbewerb gesehen. Im Kontext der Best Execution stellt insbesondere die durch die MiFID geforderte Best Execution Policy einen möglichen Wettbewerbs- oder Differenzierungsfaktor dar. Im Kontext der Post-Trade Transparenz ist insbesondere die Umsetzung der Pre- und Post-Trade Publikationsverpflichtungen als neue Herausforderung zu nennen, die relevante Umsetzungserfordernisse in Prozessen und Systemen mit sich bringen werden, da diese Anforderungen für Deutschland, im Gegensatz zu anderen europäischen Staaten, völlig neu sind. Die Tatsache, dass 68,8% der besser informierten Unternehmen antworteten, sie hätten in ihren MiFID-Projekten bereits die Auswirkungen der neuen Anforderungen auf Prozesse und IT adressiert, verdeutlicht, dass die Unternehmen diese Herausforderung ernst nehmen. Deutlich wird das Ausmaß der zur bewältigenden Veränderungen auch durch die Tatsache, dass die Umsetzung der Transparenzanforderungen als zeitlich aufwendiger betrachtet werden, als die Anforderungen zur Entwicklung und Umsetzung der Grundsätze der Auftragsausführung. 42% der befragten Wertpapierfirmen sind noch nicht mit der MiFID vertraut und befinden sich erst in der Informationssammlung und scheinen viele Herausforderungen, die die MiFID mit sich bringen wird, noch zu unterschätzen. Die dargestellten Ergebnisse der Studie machen deutlich, dass die deutsche Finanzindustrie die MiFID bisher eher als regulatorische Pflicht begreift und weniger als strategische Chance im Wettbewerb. 80% der Häuser schätzen die Umsetzungskosten mit unter 1 Mio. € als recht moderat ein. Ebenso ist man optimistisch, die Anforderungen rechtzeitig zum 01. November 2007 umsetzen zu können. Um die Entwicklungen der deutschen Finanzindustrie im Thema MiFID in einer Längsschnittanalyse verfolgen zu können, ist für Anfang 2007, also ein Jahr nach Durchführung der ersten Studie und unmittelbar nach der Umsetzung der Level 2 Richtlinie in nationales Recht, geplant, die Studie zu wiederholen und auch europäische Marktplatzbetreiber/Börsen einzubeziehen.

Literaturverzeichnis

- [AyDe05] Avgouleas, E; Degiannakis, S: The Impact of the EC Financial Instruments Markets Directive on the Trading Volume of EU Equity Markets. Athens University of Economics and Business, Statistics Technical Report No. 219.

- [BöMüoJ] Börse München: Marktteilnehmer. <http://www.boerse-muenchen.de/ueberuns/organisation/marktteilnehmer.html>, Abruf am 2005-11-07.
- [BöStoJ] Börse Stuttgart: Börsenmitglieder Inland. http://www.boerse-stuttgart.de/gruppe_bs/bww/handelsteilnehmer/sub_index.php?headparam=bww, Abruf am 2005-11-07.
- [BöZö02] Bühl, Achim; Zöfel, Peter: SPSS11 – Ein Einführung in die moderne Datenanalyse unter Windows. Addison-Wesley. 2002.
- [BuMF06] Bundesministerium der Finanzen: Referentenentwurf Finanzmarktrichtlinie-Umsetzungsgesetz, 27.September 2006, http://www.bundesfinanzministerium.de/lang_de/DE/Geld__und__Kredit/Aktuelle__Gesetze/005__a,templateId=raw,property=publicationFile.pdf, Abruf am 2006-11-08.
- [CaLa06] Casey, Jean-Pierre; Lannoo, Karel: The MiFID Implementing Measures - Excessive detail or level playing field?; ECMI, Policy Brief, No. 1, Mai 2006, http://shop.ceps.be/BookDetail.php?item_id=1333, Abruf am 2006-07-03.
- [DeBöoJ] Deutsche Börse: Xetra Marktteilnehmer. http://deutsche-boerse.com/dbag/dispatch/de/notescontent/gdb_navigation/trading_members/12_Xetra/80_Xetra_Participants/INTEGRATE/infoxetra?doc=xetra_teilnehmer_l&land=DE&seite=1, Abruf am 2005-11-07.
- [Euro93] Europäische Union: Richtlinie 93/22/EWG des Rates der Europäischen Gemeinschaft vom 10.5.1993 über Wertpapierdienstleistungen, ABl. Nr. L 141 v. 11.6.1993.
- [Euro04] Europäische Union: Richtlinie 2004/39/EG des Europäischen Parlaments und des Rates. http://europa.eu.int/eur-lex/pri/de/oj/dat/2004/l_145/l_14520040430de00010044.pdf, Abruf am 2005-10-07.
- [Euro06a] Europäische Kommission: Verordnung 1287/2006 zur Durchführung der Richtlinie 2004/39/EG. http://eur-lex.europa.eu/LexUriServ/site/de/oj/2006/l_241/l_24120060902de00010025.pdf, Abruf am 2006-11-07.

- [Euro06b] Europäische Kommission: Richtlinie 2006/73/EG zur Durchführung der Richtlinie 2004/39/EG. http://eur-lex.europa.eu/LexUriServ/site/de/oj/2006/l_241/l_24120060902de00260058.pdf, Abruf: am 2006-11-07.
- [Exch06] Ohne Verfasser: Markets In Financial Instruments Directive ("MiFID"): Implementing Measures Close To Adoption. http://www.exchange-handbook.co.uk/news_story.cfm?id=60129, Abruf am 2006-06-27.
- [Full05] Fuller, Robert: MiFID Readiness Survey. http://www.radianz.com/images/PDF/MiFID_Readiness_Survey.pdf, Abruf am 2006-05-23.
- [GoSe05] Gomber, Peter; Seitz, Jochen: How Europe's trading reform is taking shape. In: IFLR, June 2005; pp.28-30.
- [Kars05] Karsch, Wolfgang: Bewegte Bankenwelt – Top 100 der deutschen Kreditwirtschaft 2004. In: Die Bank. Nr. 10/2005. <http://www.die-bank.de/index.asp?issue=102005&art=411>, Abruf am 2005-11-07
- [KPMG06] KPMG, Capturing value from MiFID, http://www.kpmg.at/media/Capturing_value_from_MiFID.pdf, Abruf am 2006-05-23
- [Liqu05] Liquidnet: Liquidnet Europe Survey Gathers Buy-Side Opinion on MiFID and PS 05/9. <http://www.prnewswire.co.uk/cgi/news/release?id=158117>, Abruf am 2006-01-03.
- [McCr05] McCreevy, Charlie: Stock market consolidation and security markets regulation in Europe. Annual Lecture at SUERF, Brussels, 30 November 2005. <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/05/746&format=HTML&aged=0&language=EN&guiLanguage=en>, Abruf am 2006-06-25.

Einführung in den Track

eLearning Geschäftsmodelle

Prof. Dr. Andrea Back

Universität St. Gallen

Prof. Dr. Uwe Hoppe

Universität Osnabrück

Prof. Dr. Michael H. Breitner

Universität Hannover

Dr. Wolfgang Kraemer

imc AG

Integrierte Geschäftsmodelle repräsentieren die kunden-, prozess- und qualitätsorientierte Sicht der modernen Betriebswirtschaftslehre. Markterfolg und Nachhaltigkeit von eLearning Angeboten, d. h. Hard- und Software, Content, Services und Consulting, können so systematisch analysiert und sinnvoll prognostiziert werden. Kritische Erfolgsfaktoren und Risiken können systematisch identifiziert werden. Nach dem eLearning Hype Mitte bis Ende der 90er Jahre mit großzügiger staatlicher und betrieblicher Subventionierung sind heute solche Angebote und Anbieter erfolgreich, die ein ausgewogenes Kosten/Nutzen-Verhältnis nachweisen können (Return on Investment). Zur Senkung der Kosten haben sich z. B. Blended-Learning und Rapid Authoring Angebote etabliert. Für die Nutzenanalyse sind in den letzten Jahren neue Ansätze entstanden, z. B. das 6-Ebenen-Modell von Schenkel oder die eLearning Balanced-Scorecard.

Programmkomitee:

Prof. Dr. Peter Baumgartner, Fernuniversität Hagen

Prof. Dr. Heinz Grob, Lothar Universität Münster

Prof. Dr. Dieter Euler, Universität St. Gallen

Prof. Dr. Lena Suhl, Universität Paderborn

Prof. Dr. Freimut Bodendorf, Universität Erlangen-Nürnberg

Dr. Andreas Knaben, Universität Osnabrück

Prof. Dr. Heimo Adelsberger, Universität Duisburg-Essen

Prof. Dr. Joachim Hasebrook, International School of New Media, Lübeck

Prof. Dr. Michael Kerres, Universität Duisburg-Essen

Prof. Dr. Rolf Schulmeister, Universität Hamburg

E-Learning-Geschäftsmodelle für Hochschulen

Entscheidungsunterstützung bei der strategischen Positionierung

Jan vom Brocke, Christian Buddendick

European Research Center for Information Systems
Westfälische Wilhelms-Universität Münster
48149 Münster
{jan.vom.brocke,christian.buddendick}@ercis.de

Abstract

An Hochschulen sind zunehmend neue Erlösquellen zu erschließen. Der Vermarktung von E-Learning-Angeboten wird hierzu ein hohes Potenzial zugesprochen. Mittlerweile liegt eine Vielzahl unterschiedlicher Geschäftsmodelle vor, so dass die Aufgabe vor allem darin besteht, ein für die situativen Bedingungen einer Hochschule vorteilhaftes Modell auszuwählen. Zur Fundierung von Auswahlentscheidungen sind ausgehend von strategischen Überlegungen auch Kosten- und Erlöswirkungen der Modelle in Betracht zu ziehen, die unter Berücksichtigung von Risiken zu analysieren sind. Mit diesem Beitrag wird ein Methodensystem vorgestellt, das sowohl Unterstützung bei der strategischen Positionierung als auch bei der monetären Bewertung der Konsequenzen bietet. Dieses System wird anschließend durch ein Anwendungsbeispiel veranschaulicht.

1 E-Learning-Geschäftsmodelle – Neue Erlöspotentiale für Hochschulen

Während sich Hochschulen in der Vergangenheit weitestgehend auf eine Finanzierung durch Bundes- bzw. Landesmittel und Drittmittel für Forschungsprojekte verlassen konnten, müssen sie neue Erlösquellen erschließen, um im internationalen Wettbewerb erfolgreich agieren zu können.

Diesbezüglich wird der Vermarktung von E-Learning-Angeboten eine hohe Bedeutung beigemessen [BKS98, GBB05, HB03, KW05, Se01]. Unter dem Begriff E-Learning werden

Potenziale computergestützter Lehr- und Lernsysteme zur Verbesserung von Lernprozessen thematisiert [BKS98, AP02, GB04, SBH01]. Zur Vermarktung dieser Potenziale sind geeignete Geschäftsmodelle zu entwickeln. Allgemein kann unter einem Geschäftsmodell „an architecture for the product, service and information flows, including a description of the various business actors and their roles; and a description of the potential benefits of the various business actors; and a description of the sources of revenues” [Ti98] verstanden werden. Für Hochschulen werden verschiedene Geschäftsmodelle vorgeschlagen, zu denen u. a. Content-Provider, Content-Broker und Full-Service-Provider [KW05, Se01] zählen. Diese Geschäftsmodelle können anhand einer Bildungswertkette strukturiert werden [Se01]. Bei der Auswahl eines spezifischen Geschäftsmodells sind zunächst die in der Hochschule vorhandenen Kernkompetenzen zu identifizieren [PH90]. Diese bestimmen, welche Aktivitäten entlang der Wertkette von der Hochschule in einem Geschäftsmodell integriert werden können. Oftmals halten Hochschulen eine Vielzahl von Kernkompetenzen vor, so dass mehrere Aktivitäten für ein Geschäftsmodell in Frage kommen. Den vorhandenen Kernkompetenzen und korrespondierenden Aktivitäten sind die Ergebnisse einer marktseitigen Analyse entgegen zu stellen. Diese Analyse dient der Identifikation von geeigneten strategischen Positionen im E-Learning-Markt [P0xx]. Nur solche Geschäftsmodelle können langfristig erfolgreich sein, bei denen ein Fit zwischen den in der Hochschule vorhandenen Kernkompetenzen [PH90] und der Positionierung der Leistungen am Markt [Po99] erreicht werden kann. Die eingenommene Position sollte möglichst einzigartig und langfristig zu verteidigen sein, um einen komparativen Konkurrenzvorteil zu erlangen. Aufgrund der vorhandenen Kompetenzen und der möglichen Positionen am Markt, stehen für Hochschulen oftmals verschiedene Geschäftsmodelle zur Auswahl die realisiert werden können.

Im Folgenden wird eine Methode vorgestellt, anhand der diese Alternativen identifiziert und hinsichtlich ihrer monetären Konsequenzen beurteilt werden können. Diese Methode wird anschließend für die Identifikation und Bewertung von zwei alternativen Geschäftsmodellen angewendet.

2 Einführung eines Methodensystems zur monetären Bewertung von E-Learning-Geschäftsmodellen

2.1 Ordnungsrahmen

Ausgehend von den drei Teilmodellen eines Geschäftsmodells, *Marktmodell*, *Aktivitätenmodell* und *Kapitalmodell* [HB03], lässt sich ein Ordnungsrahmen für geeignete Methoden zur Entscheidungsunterstützung entwickeln [GBB05]. Das Marktmodell umfasst die Beschreibung der Struktur des E-Learning-Markts sowie die verschiedenen Akteure und deren Rollen [AP02]. In dem Aktivitätenmodell erfolgt eine Beschreibung der E-Learning-Aktivitäten, die ein Anbieter von E-Learning-Produkten ausführt [GBL01]. Beide Modelle bilden die Basis des Kapitalmodells, das die Beschreibung der durch die Aktivitäten verursachten Kosten und der durch sie zu erwirtschaftenden Erlöse umfasst. Das Aktivitätenmodell liefert das Mengengerüst für die Bestimmung der Auszahlungen, während im Marktmodell über die Grundlage zur Generierung von Einzahlungen disponiert wird. Die Saldierung der Ein- und Auszahlungen in der Zahlungsfolge, sowie deren finanzwirtschaftliche Verrechnung über die Perioden des Planungshorizonts, liefert eine angemessene Fundierung im Kapitalmodell. In dem Kapitalmodell können anschließend die monetären Implikationen der mit dem Markt- und Aktivitätenmodell einhergehenden Chancen und Risiken abgebildet werden. Der von GROB, VOM BROCKE und BENSBERG vorgeschlagene Ordnungsrahmen zur Strukturierung dieses Ansatzes ist in Abb. 1 dargestellt [GBB05].

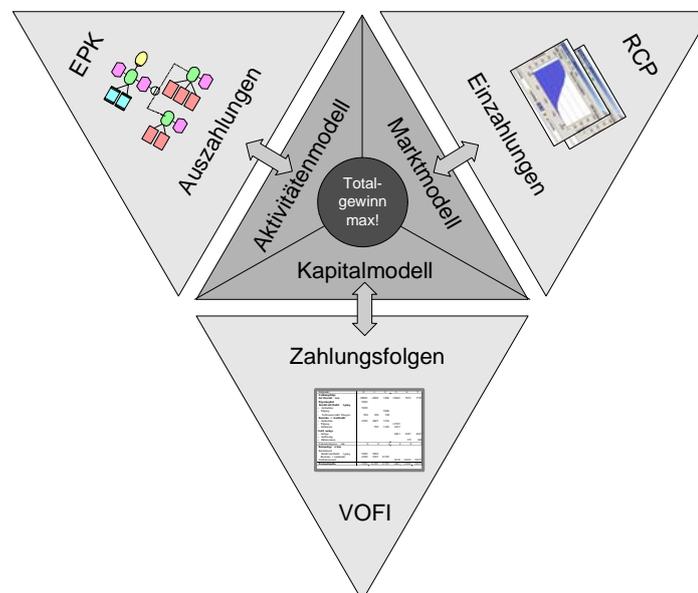


Abb. 1: Ordnungsrahmen zur Bewertung von Geschäftsmodellen im E-Learning [GBB05]

Innerhalb der identifizierten Teilbereiche stellen sich spezifische Anforderungen an das bereitzustellende Methodensystem [Br03]. Im Folgenden werden anforderungsgerechte Methoden vorgestellt.¹

2.2 Bewertung des Marktmodells

Bei der Bewertung des Marktmodells sind die Einzahlungen zu bestimmen, die einem Geschäftsmodell zuzurechnen sind. Hierzu sind diejenigen Märkte zu identifizieren, die aufgrund der vorhandenen Kompetenzen bearbeitet werden können und die gleichzeitig ein hohes Marktpotential (überdurchschnittliche Renditen) ermöglichen. Geeignete Methoden zur Bewertung des Marktmodells sind im Bereich des strategischen Managements, genauer dem Ressource Based View (RBV) und dem Market Based View (MBV), zu finden. Der RBV zielt auf die Identifikation von Kernkompetenzen ab. Kernkompetenzen sind Kompetenzen einer Hochschule, die sich durch drei zentrale Eigenschaften auszeichnen [HP99]: (1) Sie sind zunächst nicht an bestimmte Aktivitäten oder Produkte gebunden, so dass sie in verschiedenen Verwendungsrichtungen eingebracht werden können und den Zugang zu mehreren Märkten ermöglichen; (2) Kernkompetenzen müssen einen signifikanten Beitrag zum wahrgenommenen Kundennutzen leisten und somit in direktem Zusammenhang mit den am Markt angebotenen Leistungen stehen; (3) Sie müssen spezifisch sein, damit es den Konkurrenten nicht möglich ist, diese Kompetenzen in kurzer Zeit anzueignen. Kernkompetenzen können hierbei entweder durch Lernprozesse aufgebaut werden, oder durch die Einstellung von neuen Mitarbeitern.

Hinsichtlich der strategischen Positionierung am Bildungsmarkt sind zunächst diejenigen Teilmärkte des E-Learning-Markts zu bestimmen, in denen überdurchschnittliche Renditen zu erwarten sind. Zur Analyse kann das von Porter vorgeschlagene 5-Forces-Analyseraster [Po99] herangezogen werden. Bei Anwendung dieses Rahmens wird die Attraktivität eines bestimmten Teilmarktes über die herrschenden fünf Wettbewerbskräfte bestimmt: Verhandlungsstärke der Lieferanten, Verhandlungsstärke der Abnehmer, Bedrohung durch neue Konkurrenten, Bedrohung durch Ersatzprodukte und -dienste und die Rivalität unter den bestehenden Unternehmen. Die Stärke jeder dieser Wettbewerbskräfte kann anhand einer Reihe von Einflussfaktoren bestimmt werden. So sind z. B. für die Gefahr durch neue Wettbewerber die bestehenden Markteintritts- und Marktaustrittsbarrieren von zentraler Bedeutung. Ebenso kann

¹ Eine detaillierte Darstellung der einzelnen Methoden findet sich bei Grob et. al. 2005

eine detaillierte Konkurrentenanalyse [Po99], verdichtet zu einem Konkurrentenreaktionsprofil, Aufschlüsse über den Grad der Rivalität innerhalb des Teilmarkts liefern. Die Wettbewerbsanalyse ist für sämtliche Teilmärkte durchzuführen, die aufgrund der vorhandenen Kernkompetenzen bearbeitet werden können. Als Ergebnis der Analysen erhält man eine Liste derjenigen Teilmärkte, die eine hohe Attraktivität aufweisen und die von der Hochschule in einem E-Learning-Geschäftsmodell bearbeitet werden können. Aus dieser Liste sind diejenigen Teilmärkte herauszugreifen, bei denen ein hohes Ausmaß an vorhandenen Kernkompetenzen auf eine hohe Teilmarktattraktivität trifft. Für diese Teilmärkte ist aufbauend eine detaillierte Marktbearbeitungsstrategie zu entwickeln, bei der die distinktive Entscheidung zu treffen ist, ob eine Kosten- oder Qualitätsführerschaft angestrebt wird bzw. ob diese Führerschaft sich auf den gesamten Markt oder nur auf einzelne Nischen erstrecken soll. Bei der Strategiewahl ist eine eindeutige Entscheidung für eine der Optionen unabdingbar, um andere Wettbewerber zu übertreffen [Po99].

Ausgehend von den Teilmärkten auf denen ein komparativer Konkurrenzvorteil erreicht werden kann und der Wettbewerbsstrategie, können die Leistungen definiert werden, die angeboten werden. Für diese Leistungen sind zur Ermittlung der Einnahmen Kundenzahlen und Absatzpreise zu prognostizieren. Aufgrund der vorherrschenden Ungewissheit sind bei der Berechnung der monetären Konsequenzen bestimmter Geschäftsmodelle Methoden bereitzustellen, die Chancen und Risiken der Einnahmenseite abbilden können. Geeignete Methoden werden bei der Beschreibung der Methoden des Kapitalmodells vorgestellt. Zunächst sind jedoch die zur Erstellung der Leistungen notwendigen Aktivitäten zu bestimmen und mit Kosten zu bewerten, um eine Bewertung der E-Learning-Geschäftsmodelle vornehmen zu können.

2.3 Bewertung des Aktivitätenmodells

Zur Bewertung der Auszahlungen werden Methoden benötigt, die Transparenz über die zu erbringenden Aktivitäten schaffen und die zugleich eine Ableitung der mit ihnen verbundenen monetären Konsequenzen ermöglichen. Angesichts der hohen Dynamik der Leistungserstellung im E-Learning [HB03, SBH01] sind dabei effiziente Möglichkeiten zur Adaption des Bewertungssystems zu gewährleisten. Durch einen neuartigen Ansatz kann auf der Grundlage dieser Prozessmodelle die Ableitung der relevanten Auszahlungen vorgenommen werden [GB04]. Diese Ansatz zur Modellierung der Aktivitätenstruktur basiert auf Ereignisgesteuerten Prozessketten (EPK) [KNS92].

Anhaltspunkte über die Erfassung von Auszahlungen auf Basis von EPK liefert das Referenzmodell zum Controlling von Prozessdesigns von GROB/VOM BROCKE [GB04]. Demnach sind die durch eine Funktion als Input genutzten Faktoren zu identifizieren und zu bewerten. Im Hinblick auf die Zurechnung sind Potenzial- und Repetierfaktoren zu unterscheiden. Potenzialfaktoren stellen Ressourcenobjekte dar, die von mehreren Funktionen in Anspruch genommen werden können. Im Aktivitätenmodell von E-Learning-Geschäftsmodellen sind dies vor allem die Arbeitskraft von Mitarbeitern sowie die Nutzung von Hardware- und Softwaresystemen. Repetierfaktoren fließen hingegen als Inputobjekte in die Verarbeitung einer Funktion ein.

Die Auszahlungen einer Funktion setzen sich aus den Auszahlungen für die in Anspruch genommenen Ressourcenobjekte und die in die Funktion eingeflossenen Inputobjekte zusammen. Zur Kalkulation der Inputobjekte ist die in der Funktion verwendete Menge des Objekts zu erfassen, die mit einem Verrechnungspreis des Objekts je Mengeneinheit zu verrechnen ist. Auszahlungen für Ressourcenobjekte werden nach dem Prinzip der Ressourceninanspruchnahme kalkuliert. Die Kalkulation der Ressourceninanspruchnahme kann analog zu dem Vorgehen in der Prozesskostenrechnung vorgenommen werden. Demnach wird die prozentuale Ressourceninanspruchnahme einer Funktion berechnet, indem die von dieser Funktion in Anspruch genommenen Leistungseinheiten einer Ressource in Relation gesetzt werden zu der Summe aller von dieser Ressource an Funktionen abgegebenen Leistungseinheiten. Durch die Bewertung des Aktivitätenmodells werden somit die zur Leistungserbringung auf einem bestimmten Teilmarkt prognostizierten Auszahlungen ermittelt, die bei der Bewertung des Kapitalmodells den Einzahlungen gegenübergestellt werden.

2.4 Bewertung des Kapitalmodells

Ausgehend von den bei der Bewertung des Marktmodells ermittelten Einzahlungen sowie den aus der Bewertung des Aktivitätenmodells resultierenden Auszahlungen werden bei der Bewertung des Kapitalmodells die langfristigen Konsequenzen eines Geschäftsmodells bewertet. Um die periodenindividuellen Zahlungen zu entscheidungsrelevanten Zielwerten zu verdichten, werden Methoden der Investitionsrechnung benötigt. Einzubeziehen sind neben alternativen Formen der Kapitalbeschaffung und -anlage vor allem auch Ertragsteuern. Wegen der Langfristigkeit des Planungshorizonts sind auch Modifikationen der relevanten Einflussgrößen zu erwarten (z. B. Wechsel des Steuersystems). Um die Adaption der

Geschäftsmodelle im E-Learning zu ermöglichen, sollte die Methode transparent und ausbaufähig sein. Da herkömmliche Methoden des Investitionscontrollings diesen Anforderungen nur unzureichend gerecht werden, wird die Verwendung eines vollständigen Finanzplans (VOFI) [Gr01] vorgeschlagen. VOFI ermöglicht die finanzwirtschaftliche Verrechnung der einem Geschäftsmodell über mehrere Perioden hinweg zuzurechnenden Zahlungen. Unter Berücksichtigung differenzierter Konditionen für Finanzierungen, Reinvestitionen und Steuern kann eine breite Palette finanzwirtschaftlicher Zielwerte berechnet werden [Gr01].

Besonders aussagekräftig für den Vergleich alternativer Geschäftsmodelle ist die VOFI-Gesamtkapitalrentabilität, die den Return on Investment (ROI) auf Basis einer dynamischen Zahlenbasis ausdrückt. Die VOFI-Gesamtkapitalrentabilität zeigt die über die Nutzungsdauer realisierte Verzinsung des eingesetzten Kapitals auf, die zur rechnerischen Fundierung der Entscheidung mit den durchschnittlichen Kapitalkosten zu vergleichen ist.

Die Bewertung eines Geschäftsmodells hat die Unsicherheit der zu prognostizierenden Ein- und Auszahlungen zu berücksichtigen. Zwar wird dem E-Learning-Sektor mithin ein großes Marktpotenzial zugesprochen, doch ist weitgehend unklar, wie sich dieses Potenzial aufteilen wird [SG02]. Sowohl die prognostizierten Preise, als auch die prognostizierten Kundenzahlen, die zur Ermittlung der Einzahlungen benötigt werden, sind somit Variablen unter Unsicherheit. Ebenso besteht seitens der Leistungserstellung Unsicherheit. Diese konkretisiert sich z. B. durch Unsicherheit hinsichtlich der Entwicklung der Personalkosten und hinsichtlich der Höhe des zu erwartenden Erfahrungskurveneffekts. Die daher zu gewährleistenden Adaptionmöglichkeiten stellen weitere Anforderungen an die Bewertungsmethoden. Herkömmliche Risikokennzahlen (z. B. Mittelwert, Varianz) erweisen sich hier als unzureichend, um die in mehrfacher Hinsicht variierenden Zahlungen zu verdichten. Die Risiko-Chancen-Analyse ist ein Verfahren, das die Unsicherheit eines Entscheidungsproblems durch Simulation behandelt [Gr01, He64]. Dabei werden mehrere unsichere Einflussgrößen anhand von Verteilungen modelliert und im Zuge des Simulationslaufs auf die Verteilung einer Zielgröße geschlossen. Das bisher beschriebene Instrumentarium dient hierzu als Rechensystem, indem die Auszahlungen und die im Simulationslauf ermittelten Einzahlungen im VOFI zu finanzwirtschaftlichen Zielwerten verdichtet werden. Als Ergebnis des Simulationslaufs liegt z. B. eine Verteilung des Endwerts der Investition oder des sich daraus ergebenden Return on Investments eines Geschäftsmodells vor. Diese Verteilung des Zielwerts wird in ein sog. Risiko-Chancen-Profil transformiert.

Risiko-Chancen-Profile ermöglichen es, die Wahrscheinlichkeit abzulesen, mit der ein Zielwert größer oder gleich einem kritischen Wert ist.

3 Anwendung des Methodensystems

Die Anwendung des Methodensystems wird im Folgenden anhand eines Beispiels voranschaulicht. Ausgangspunkt dieses Beispiels bilden die Planungen eines ein Fachbereichs, seine bereits etablierten E-Learning-Kompetenzen [BBS06], professionell zu vermarkten. Im Rahmen der Planung wurden zunächst die vorhandenen Kernkompetenzen identifiziert. Die Kernkompetenzen des Fachbereichs beziehen sich auf die Konzeption und die Erstellung von Lerninhalten. Ausgehend von diesen beiden Kernkompetenzen wurden verschiedene potentielle Positionen im E-Learning-Markt ermittelt. Grundlage hierfür bildet eine detaillierte Analyse der Wertkette des Bildungsmarkts [Se01]. Diese Wertkette, sowie potentielle Geschäftsmodelle sind in Abb. 2 dargestellt.

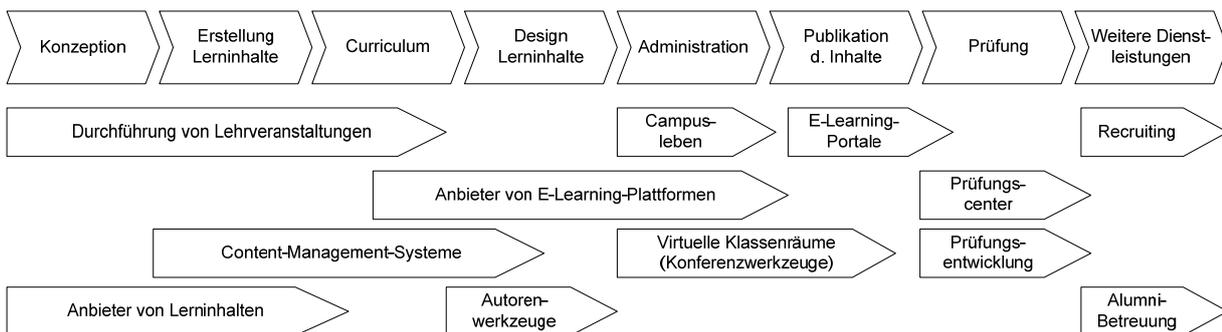


Abb. 2: Wertkette des Bildungsmarkts [Se01]

Von den alternativen Geschäftsmodellen die sich aus der Wertkette ergeben, werden unter Berücksichtigung der vorhandenen Kernkompetenzen zwei Alternativen als vorteilhaft erachtet. Die erste Alternative (GM 1) ist die Positionierung als Anbieter der komplette Lehrveranstaltungen durchführt. Hierbei sollen die selbst erstellten Lerninhalte in Veranstaltungen eingesetzt werden, für die ein eigener Weiterbildungsstudiengang eingerichtet werden soll. Die zweite Alternative (GM 2) ist die Positionierung als Anbieter von Lerninhalten. Diese Lerninhalte sollen über ein Portal vertrieben werden. Ausgehend von einer Wettbewerbsanalyse wurde festgestellt, dass beide Positionen überdurchschnittliche Renditen versprechen. Gründe hierfür sind u. a. die geringe Gefahr durch Substitutionsprodukte, die

hohen Spezifität der angebotenen Leistungen und die daraus resultierenden Markteintrittsbarrieren und die horizontale Integration entlang der Wertkette, da keine Vorlieferanten bei der Erstellung der Leistung involviert sind. Als weitere Planungsgrundlage wurde bestimmt, dass der Fachbereich ausschließlich als Nischenanbieter für das entsprechende Fach und die Fachinhalte auftreten will. Eine umfassende Qualitätsführerschaft bzw. Kostenführerschaft im Gesamt E-Learning-Markt erscheint aufgrund der gegebenen Ressourcenausstattung nicht möglich. In Abhängigkeit der Entwicklung des Geschäfts wird allerdings bereits heute angedacht in Zukunft das Produktangebot durch die Einbindung in Netzwerke auszubauen. Diese Entscheidung soll allerdings erst nach Ablauf der fünfjährigen Pilotphase getroffen werden.

Aktuell stellt sich dem Fachbereich somit die Frage, welches der beiden alternativen Geschäftsmodelle in der fünfjährigen Pilotphase realisiert werden soll. Für die Pilotphase soll zunächst eine GmbH gegründet werden, an der die Mitglieder des Fachbereichs beteiligt sind.

In einem nächsten Schritt wurden zunächst die Aktivitäten bestimmt, die den beiden Geschäftsmodellen zuzuordnen sind und in Prozessmodellen integriert. Beide Geschäftsmodelle verfügen sowohl über gemeinsame Prozesse als auch über distinktive Prozesse. Gemeinsame Prozesse der beiden Geschäftsmodelle betreffen vor allem den Bereich der Konzeption der Lerninhalte und die notwendigen administrativen Prozesse, wie z. B. die Erstellung der jährlichen Steuererklärung. Das GM 1 erfordert ergänzend zu den Aktivitäten des GM 2 auch Prozesse zur Erstellung eines Curriculum sowie die Administration der Teilnehmer der angebotenen Lehrveranstaltungen (Immatrikulation, Erstellen von Leistungsnachweisen, etc.).

In der folgenden Abb. 3 ist der Prozess der Erstellung des Angebots für eine Lehrveranstaltung dargestellt.

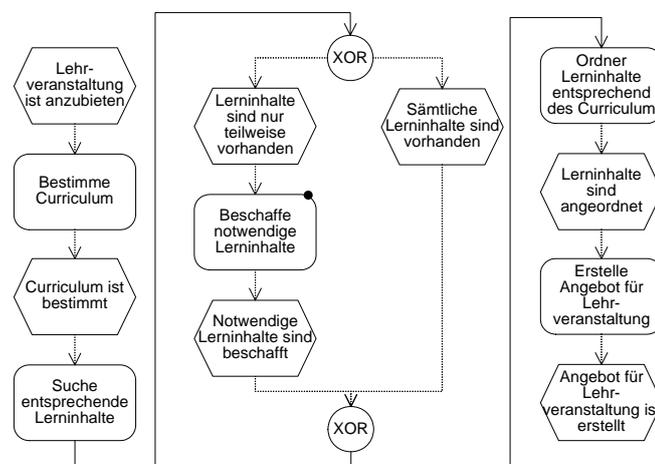


Abb. 3: Prozess GM 1: Erstellen eines Angebots für Lehrveranstaltungen

für den Fall, dass GM 2 realisiert werden soll, sind spezifische Prozesse vorzuzusehen, die die Distribution und Verwaltung der erstellten Lerninhalte in dem Portal betreffen. Typische Prozesse sind hierbei das Einstellen von neuen Lerninhalten in dem Portal sowie die Überwachung des Downloads von Lerninhalten. Diese beiden Prozessmodelle des GM 2 sind in der folgenden Abb. 4 dargestellt worden.

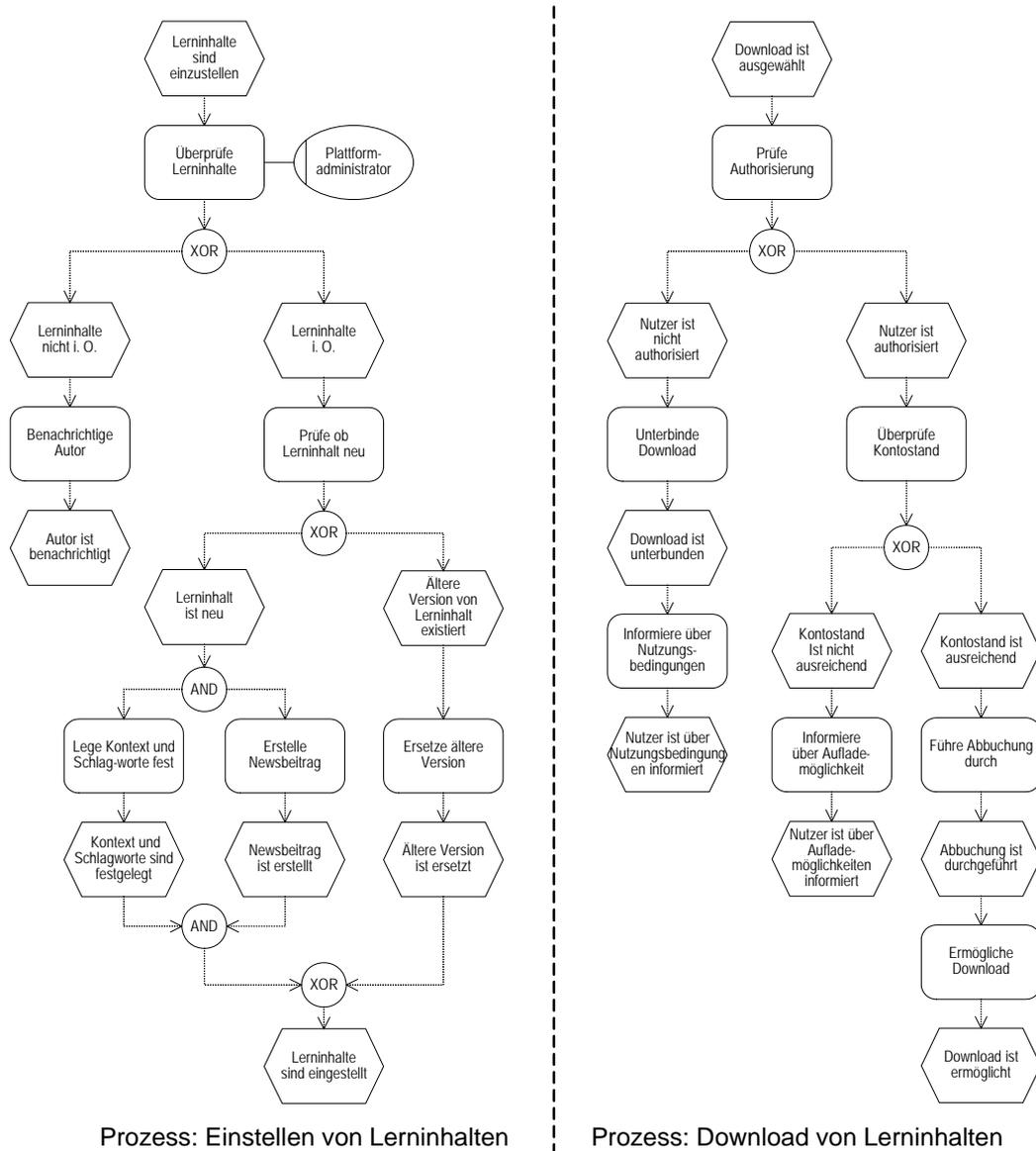


Abb. 4: Geschäftsmodell 2 – Prozesse Einstellen und Download von Lerninhalten

Ausgehend von den Prozessen wurden die Kundenzahlen und damit die Erlöse prognostiziert, die den beiden Geschäftsmodellen zugerechnet werden können. Für GM 1 wurde prognostiziert, dass in der ersten Periode zunächst 50 Personen an dem Weiterbildungsangebot teilnehmen

werden und die Teilnehmerzahlen kontinuierlich anwachsen. Im fünften Jahr wird mit insgesamt 120 Teilnehmern an dem Kurangebot gerechnet. Jeder Teilnehmer soll für die Nutzung des Angebots eine Pauschale in Höhe von 750 Euro entrichten.

Neben den Kundenzahlen und Erlösen wurden ebenfalls die Kosten pro Aktivität in den Prozessen der beiden Geschäftsmodelle ermittelt. Um die Kosten pro Prozessdurchlauf zu ermitteln sind zusätzlich auch die Wahrscheinlichkeiten zu berechnen, mit denen eine Variante des Prozessverlaufs eintritt. So wurde z. B. die Wahrscheinlichkeit, dass die Lerninhalte für ein Lehrangebot vollständig vorhanden sind (vgl. Abb. 3) im GM 1 im ersten Jahr mit 20 % prognostiziert. Durch die kontinuierliche Entwicklung weiterer Lerninhalte wird prognostiziert, dass diese Wahrscheinlichkeit in den folgenden Jahren kontinuierlich steigt und im Jahr fünf bei 80 % liegt. Die anfallenden zahlungswirksamen Kosten pro Prozess werden anschließend mit der Anzahl der Prozessdurchläufe multipliziert, um die prozessabhängigen Kosten pro Periode zu ermitteln. Um die Vorteilhaftigkeit eines Geschäftsmodells beurteilen zu können sind neben diesen Kosten auch die Funktionsunabhängigen Kosten zu ermitteln. Diese Kosten fallen unabhängig von der Anzahl der Prozessdurchläufe an. Zu diesen Kosten zählen z. B. bei dem GM 1 Kosten für die Anmietung eines Raums, in dem die Zertifikate feierlich übergeben werden sowie die Kosten für den telefonischen Support der Kursteilnehmer.

Für GM 2 ist ein zweistufiges Erlösmodell geplant. Neben einer Pauschale für die Nutzung des Portals wird zusätzlich pro Download einer Lehreinheit ein Entgelt von 25 Euro erhoben. Hinsichtlich der Kundenzahlen wird angenommen, dass diese von 250 Kunden im Jahr eins bis auf 600 Kunden im Jahr fünf ansteigen werden. Da zu Beginn des Portalangebots zunächst nur eine beschränkte Anzahl an Lerneinheiten zur Verfügung steht, wird damit gerechnet, dass die Anzahl der Einstellungen von neuen Lerneinheiten im Jahr eins 100 beträgt und kontinuierlich im Zeitablauf sinkt. Als Funktionsunabhängige Kosten sind bei GM 2 neben Raummieten auch die Kosten für die Bereitstellung und Wartung der Portalserver zu berücksichtigen.

In der folgenden Abb. 5 sind die zur Berechnung der Zahlungsfolge der Investition notwendigen Daten dargestellt.

Variablen	P1	P2	P3	P4	P5
Anzahl Kunden	250	325	400	450	600
Anzahl eingestellter Lerneinheiten	100	80	65	40	28
WSK Lerninhalt in Ordnung	0,5	0,55	0,7	0,75	0,9
WSK Lerninhalt neu	0,8	0,7	0,5	0,45	0,4
WSK Nutzer autorisiert	0,6	0,7	0,76	0,78	0,8
WSK Kontostand ausreichend	0,5	0,6	0,7	0,8	0,85
Funktionsunabhängige Kosten					
Betriebskosten					
- Nettwerkkosten	3000	3000	3000	3000	3000
- Energiekosten	2400	2400	2400	2400	2400
- Wartung Server	4800	4800	4800	4800	4800
- Büromiete	10000	10000	10000	10000	10000
Kosten pro Funktion					
Überprüfe Lerninhalte	75	75	75	75	75
Benachrichtige Autor	15	15	15	15	15
Prüfe ob Lerneinheit neu	25	25	25	25	25
Lege Kontext und Schlagworte fest	15	15	15	15	15
Erstelle Newsbeitrag	5	5	5	5	5
Ersetze ältere Version	5	5	5	5	5
Prüfe Autorisierung	0,01	0,01	0,01	0,01	0,01
Verhindere Download	0,01	0,01	0,01	0,01	0,01
Informiere über Nutzungsbedingungen	0,01	0,01	0,01	0,01	0,01
Überprüfe Kontostand	0,01	0,01	0,01	0,01	0,01
Informiere über Auflademöglichkeiten	0,01	0,01	0,01	0,01	0,01
Führe Abbuchung durch	0,01	0,01	0,01	0,01	0,01
Führe Download durch	1,50	1,50	1,50	1,50	1,50
Weitere Kosten					
Konzeption der Lerninhalte	120	120	120	120	120
Erstellung von Lerninhalten	500	500	500	500	500
Erlöse					
Gebühr Download Lerneinheit	25	25	25	25	25
Nutzungspauschale	100	100	100	100	100

Abb. 5: GM 2 – Daten zur Ermittlung der Zahlungsfolge

Durch die Saldierung der periodenspezifischen Ein- und Auszahlungen des jeweiligen Geschäftsmodells können die Elemente der Zahlungsfolge ermittelt werden, die in den VOFI als originäre Zahlungen übernommen werden. Originäre Zahlungen sind diejenigen Zahlungen, die in direktem Zusammenhang mit der Investition stehen. Um die monetären Konsequenzen vollständig bewerten zu können, sind ebenfalls die derivativen Zahlungen zu ermitteln. Diese stehen nicht im direkten Zusammenhang zu der Investition sondern ergeben sich aus den originären Zahlungen. Zu diesen Zahlungen zählen Abschreibungen, Soll- und Habenzinsen sowie Steuerzahlungen.

Die zur Gründung der GmbH notwendige Einlage in Höhe von 25.000 Euro wird durch die Mitglieder des Fachbereichs als Eigenkapital bereitgestellt. Um den Betrieb aufnehmen zu können sind zusätzlich Anschaffungsauszahlungen in Höhe von 30.000 Euro aufzubringen. Diese Anschaffungsauszahlungen beziehen sich z. B. auf die Anschaffung von Büro- und Geschäftsausstattung. Von diesen Auszahlungen sind 5.000 Euro aktivierungspflichtig, die Güter werden über einen Nutzungsdauer von 5 Jahren linear abgeschrieben. Die Anschaffungsauszahlungen sowie notwendige Zwischenfinanzierungen können über einen Kredit mit endfälliger Tilgung (Zinssatz 6,5%, Disagio 10%) oder über einen Kontokorrentkredit in beliebiger Höhe (Zinssatz 13%) finanziert werden. Freie Mittel werden auf einem Geldmarktkonto mit einem Zinssatz von 3% angelegt. In Abb. 4 ist der VOFI des Geschäftsmodells 1 dargestellt.

VOFI						
Zeitpunkt	0	1	2	3	4	5
Zahlungsfolge der Investition	-30000	-6900	7818	23310	33098	52539
Eigenkapital	25000					
Kredit mit						
Endtilgung						
+ Aufnahme (brutto)	5000					
- Disagio	500					
- Tilgung						5000
- Sollzinsen		325	325	325	325	325
Kontokorrentkredit						
+ Aufnahme	500	4233				
- Tilgung			4733			
- Sollzinsen		65	615			
Standardanlage						
- Anlage			40	15011	21518	29285
+ Auflösung						
+ Habenzinsen				1	452	1097
Steuerzahlungen						
- Auszahlungen			2105	7975	11706	19026
+ Erstattung		3057				
Finanzierungssaldo	0	0	0	0	0	0
Bestandsgrößen						
Kredit mit Endtilgung	5000	5000	5000	5000	5000	
Kontokorrentkredit	500	4733				
Guthabenstand			40	15050	36568	65853
Bestandssaldo	-5500	-9733	-4960	10050	31568	65853

Abb. 4: GM 1 – VOFI

Die in der Abb. 4 dargestellten Steuerzahlungen sind durch weitere Nebenrechnungen zu ermitteln. Neben der eigentlichen Körperschaftsteuer (25%) ist auch der Hebesatz der Gemeinde (360%) zu berücksichtigen. Um die Steuerzahlungen zu berechnen sind die periodenindividuellen Elemente der Zahlungsfolge um entstehende Abschreibungen sowie anfallende Soll- und Habenzinsen zu korrigieren.

Ausgehend von den im VOFI ermittelten Endwerten wird anschließend der ROI des Geschäftsmodells ermittelt. Beide Geschäftsmodelle weisen einen positiven ROI auf und sind lukrativer als eine Investition in festverzinsliche Wertpapiere (Zinssatz 8% p.a.). Die Endwerte der beiden Geschäftsmodelle sowie die korrespondierenden ROIs sind in der folgenden Abb. 5 dargestellt.

Endwert (EW ^M)	65853 EUR	Endwert (EW ^M)	61008 EUR
Endwert (EW ^O)	32036 EUR	Endwert (EW ^O)	32036 EUR
ΔEW	33817 EUR	ΔEW	28972 EUR
Eigenkapital (EK)	25000 EUR	Eigenkapital (EK)	25000 EUR
Sollzinsen der Anschaffungsauszahlung (Z ^S)	2282 EUR	Sollzinsen der Anschaffungsauszahlung (Z ^S)	2749 EUR
Anschaffungsauszahlung (a ₀)	30000 EUR	Anschaffungsauszahlung (a ₀)	30000 EUR
Gesamtkapitalrentabilität (g ^M _{GK})	19.24% [in%]	Gesamtkapitalrentabilität (g ^M _{GK})	17.70% [in%]
Referenzzinsfuß (k)	5.11% [in%]	Referenzzinsfuß (k)	5.27% [in%]
Gesamtkapitalrentabilität (g ^M _{GK}) Variante 2	19.24% [in%]	Gesamtkapitalrentabilität (g ^M _{GK}) Variante 2	17.70% [in%]
Eigenkapitalrentabilität der Investition (g ^M _{EK})	21.37% [in%]	Eigenkapitalrentabilität der Investition (g ^M _{EK})	19.53% [in%]
Eigenkapitalrentabilität der Opportunität (g ^O _{EK})	5.08% [in%]	Eigenkapitalrentabilität der Opportunität (g ^O _{EK})	5.08% [in%]
<p>Das Entscheidungskriterium lautet: Investiere, wenn $g^M_{GK} > k$.</p> <p>Somit sollte man in diesem Fall investieren!</p>		<p>Das Entscheidungskriterium lautet: Investiere, wenn $g^M_{GK} > k$.</p> <p>Somit sollte man in diesem Fall investieren!</p>	

Abb. 5: Endwerte und ROIs der beiden Geschäftsmodelle

Durch einen Vergleich der beiden ROIs wurde festgestellt, dass das Geschäftsmodell Durchführung von Lehrangeboten einen höheren ROI (19,24%) aufweist, als GM 2, bei dem Lerninhalte über ein Portal vertrieben werden.

Um der Unsicherheit, die den Planungen anhaftet, gerecht zu werden, wurden aufbauend Chancen und Risiken der beiden Geschäftsmodelle simuliert. Es wurde z. B. ermittelt, welche Auswirkungen eine Variation der Kundenzahlen verursacht. Hierzu wurde eine Normalverteilung der Kundenzahlen angenommen, die in den einzelnen Perioden den folgenden Mittelwert und Standardabweichung aufweist: (1) 250 : 40, (2) 325 : 45, (3) 400 : 50, (4) 450 : 52, (5) 600 : 60. Ausgehend von dieser Verteilung konnte festgestellt werden, dass mit einer Wahrscheinlichkeit von 61,43% GM 1 mindestens den ROI des bislang schlechteren GM 2 (17,70%) aufweist. In Abb. 6 ist die kumulative Häufigkeitsverteilung dargestellt.

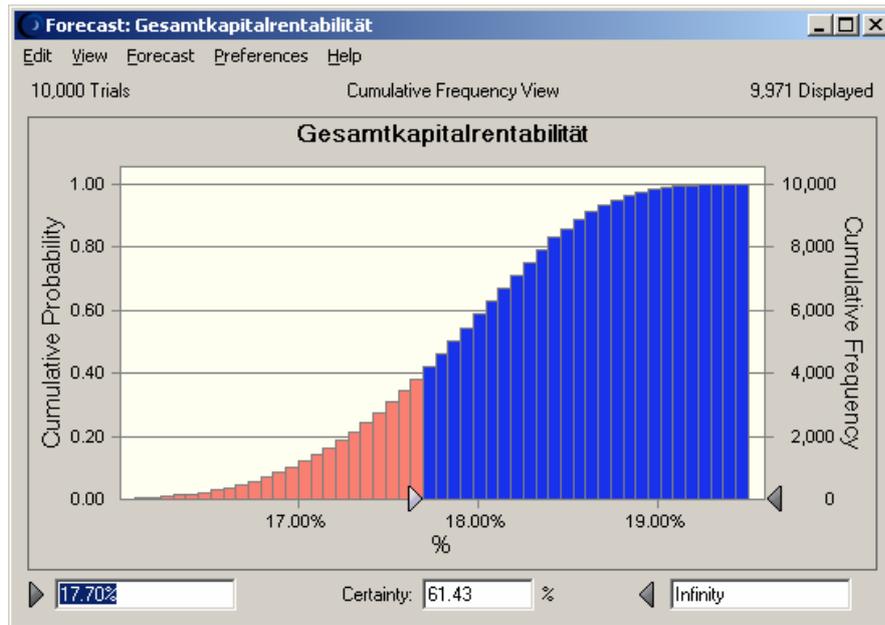


Abb. 6: Kumulative Häufigkeitsverteilung

Ebenfalls wurde anhand einer Dreiecksverteilung simuliert, welche Konsequenzen eine Variation des Erfahrungskurveneffekts hervorrufen würde.

4 Zusammenfassung

In dem vorliegenden Artikel wurde ein Methodensystem für die monetäre Bewertung von E-Learning-Geschäftsmodellen eingeführt. Diese Geschäftsmodelle können zur Erschließung neuer Erlösquellen für Hochschulen beitragen. Hierin besteht eine zentrale Herausforderung, der sich Hochschulen im internationalen Bildungswettbewerb stellen müssen.

Das Methodensystem unterstützt die Entscheidungsfindung, indem die langfristigen ökonomischen Konsequenzen von Geschäftsmodellen berücksichtigt werden. Auch werden die mit der Unsicherheit verbundenen Chancen und Risiken eines bestimmten Geschäftsmodells berücksichtigt. Durch einen ROI-Vergleich der Alternativen kann eine Entscheidungsunterstützung erfolgen.

Zukünftige Arbeiten sollen auf die praktische Anwendbarkeit des Methodensystems fokussieren. Insbesondere ist in diesen Arbeiten zu prüfen, inwiefern eine detaillierte Ermittlung der Auszahlungen pro Funktion in der Praxis möglich ist. Hierzu wurden bereits erste Kooperationen gebildet, in denen das entwickelte Methodensystem zur Beurteilung der monetären Konsequenzen von E-Learning-Geschäftsmodellen eingesetzt wird.

Literaturverzeichnis

- [BKS98] *Back, A.; Kramhöller, S.; Seufert, S.*: Technology enabled Management Education. Die Lernumgebung MBE Genius im Bereich Executive Study an der Universität St. Gallen. In: IO Management (21) Nr. 3, 1998; S. 36-42.
- [GBB05] *Grob, H. L.; vom Brocke, J.; Bensberg, F.*: Finanzwirtschaftliche Bewertung von Geschäftsmodellen im E-Learning, Konzeption, Methoden und Perspektiven. In *Breitner, M. H.; Hoppe, G. (Hrsg.)*: E-Learning, Einsatzkonzepte und Geschäftsmodelle. Physica, Heidelberg, 2005; S. 101-116.
- [HB03] *Hoppe, G.; Breitner, M. H.*: Business Models for E-Learning. Hannover, 2003.
- [KW05] *Kleimann, B.; Wannemacher, K.*: Geschäftsmodelle für E-Learning. Konzepte und Beispiele aus der Hochschulpraxis. In *Tavangarian, D.; Nölting, K., (Hrsg.)*: Auf zu neuen Ufern! E-Learning heute und morgen. Waxmann, Münster, 2005; S. 187-196.
- [Se01] *Seufert, S.*: E-Learning Business Models, Strategies, Success Factors and BestPractice Examples. In *R. De Fillippi, R. W. (Hrsg.)*: Rethinking Management Education. Information Age Press, Greenwich, 2001.
- [AP02] *Adelsberger, H.; Pawlowski, J.*: Electronic Business and Education. In *Adelsberger, H.; Collis, B.; Pawlowski, J. (Hrsg.)*: Handbook on Information Technologies for Education & Training, International Handbook on Information Systems. Springer, Berlin, 2002; S. 653-671.
- [GB04] *Grob, H. L.; vom Brocke, J.*: Referenzmodelle für E-Learning-Systeme. Konzeption und Anwendung für die Produktionswirtschaft. In *Corsten, H.; Brassler, A. (Hrsg.)*: Entwicklungen im Produktionsmanagement. München, 2004; S. 43-62.
- [SBH01] *Seufert, S.; Back, A.; Häusler, M.*: E-Learning, Weiterbildung im Internet. Das „Plato-Cookbook“ für internetbasiertes Lernen. Kilchberg, 2001.

- [Ti98] *Timmers, P.:* Business Models for Electronic Markets. In: *Electronic Markets* (8) Nr. 2, 1998; S. 3-8.
- [PH90] *Prahalad, C. K.; Hamel, G.:* The core competencies of the corporation. In: *Harvard Business Review* (68) Nr. 3, 1990; S. 79-91.
- [Po99] *Porter, M. E.:* Wettbewerbsstrategie. 1999.
- [GBL01] *Grob, H. L.; vom Brocke, J.; Lahme, N.:* Freestyle Learning - Konzept und Entwicklungsprozess. In *Grob, H. L. (Hrsg.): cHL - computergestützte Hochschullehre, Dokumentation zum cHL-Tag 2000. Münster, 2001; S. 41-49.*
- [Br03] *vom Brocke, J.:* Referenzmodellierung. Gestaltung und Verteilung von Konstruktionsprozessen. Dissertation am Institut für Wirtschaftsinformatik, Westfälische Wilhelms-Universität, Münster, 2003.
- [Gr01] *Grob, H. L.:* Einführung in die Investitionsrechnung. Eine Fallstudiengeschichte. Vahlen, München, 2001.
- [GB04] *Grob, H. L.; vom Brocke, J.:* Controlling des Designs von Logistikprozessen. In *Baumgarten, H.; Becker, J.; Wiendahl, H.-P.; Zentes, J. (Hrsg.): Logistik Management, Springer Expertensystem Logistik Management. Berlin, 2004; S. 1-26.*
- [KNS92] *Keller, G.; Nüttgens, M.; Scheer, A. W.:* Semantische Prozeßmodellierung auf der Grundlage „Ereignisgesteuerter Prozeßketten (EPK)“. Saarbrücken, 1992.
- [SG02] *Seufert, S.; Guttman, J.:* Wissens- und Lernportale auf dem E-Learning Markt, dargestellt am Fallbeispiel Siemens AG. In *Pawlowsky, J.; Reinhardt, R. (Hrsg.): Wissensmanagement für die Praxis. Methoden und Instrumente zur erfolgreichen Umsetzung. München, 2002; S. 199-223.*
- [He64] *Hertz, D. B.:* Risk Analysis in Capital Investment. In: *Harvard Business Review* (42) Nr. January-February, 1964; S. 95-106.
- [Wr36] *Wright, T. P.:* Factors affecting the cost of airplanes. In: *Journal of the Aeronautical Science* Nr. 1936; S. 122-128.

Ein Ansatz zur prozessorientierten Evaluation von Learning Management Systemen

Michael Luther

GOPA IT Consultants
65428 Rüsselsheim
Michael.luther@gopa.de

Dr. Matthias Trier

Fachgebiet Systemanalyse und EDV
Institut für Wirtschaftsinformatik und Quantitative Methoden
Technische Universität Berlin
10587 Berlin
Matthias.trier@syesdv.cs.tu-berlin.de

Abstract

Bei der Bewertung und Auswahl von Lernplattformen helfen Kriterienlisten und Usabilitytests bei der Filterung einer großen Anzahl von Lernplattformen. Dabei sind Verfahren, die eine strukturierte Erarbeitung von Bewertungskriterien ermöglichen und die die Effizienz der Arbeit mit den Systemen verlässlich bewerten, nicht Teil des wissenschaftlichen Diskurses.

Vor dem Hintergrund, dass Learning Management Systeme gleiche Aufgaben wie eine Workflowmanagement-Anwendung übernehmen, zeigt dieser Beitrag, wie die Prozessanalyse zu einem strukturierten Auswahlverfahren von Learning Management Systemen entwickelt werden kann. Abschließend wird die Methode am Beispiel einer prozessorientierten Bewertung von Open Source LMS evaluiert.

1 Problemumfeld

Learning Management Systeme (LMS) sind ein zentrales Element einer technologischen Infrastruktur eines Lernszenarios (vgl. [TrWW03, 1]). Die Bewertung von Lernplattformen ist

als vorbereitender Schritt für die Entscheidung für ein bestimmtes Produkt von besonderer Bedeutung bei der Planung eines E-Learning-Angebots. Im aktuellen wissenschaftlichen Diskurs im E-Learning-Bereich werden verschiedene Vorgehen für die Evaluation von LMS diskutiert (vgl. z.B. [Stoc04], [ScTe04] oder [BaHM02]). Eine zentrale Rolle innerhalb des diskutierten Methodenkanons spielen Kriterienlisten und Usabilitytests. Kriterienlisten helfen bei der Filterung einer großen Anzahl von Lernplattformen. Inhaltlich beziehen sie sich auf die Zweckmäßigkeit von Lernplattformen für den spezifischen Einsatzzweck, für den die Evaluation durchgeführt wird. Die Zweckmäßigkeit wird im Sinne Nielsens dadurch bestimmt, in wieweit die Software die Benutzer dazu befähigt, ihre Aufgaben im vollen Umfang zu erfüllen, ohne auf andere Hilfsmittel zurückgreifen zu müssen.

Mit Hilfe von Usabilitytests soll festgestellt werden, ob die Produkte den Ansprüchen der Benutzer gerecht werden.

Innerhalb der einschlägigen Literatur herrscht Einigkeit darüber, dass Evaluationskriterien und –Methoden sich stark an dem jeweiligen Einsatzszenario und der Rolle der Software innerhalb der Lern- und Lehrprozesse orientieren sollen (vgl. [Terg04, 17]). Ein LMS soll die Verwaltung und Koordination des Lernprozesses ermöglichen (den Lernprozess *managen*). Es legt die Ausführungsreihenfolge bei der Bearbeitung bestimmter Aufgaben fest (z.B. erst Kurs initialisieren, Struktur festlegen und dann Inhalte einfügen), ordnet bestimmte Aufgaben bestimmten Personen zu (durch ein rollenbasiertes Rechtesystem, wie es die meisten modernen Lernplattformen aufweisen), hält Daten bereit, die für die Bearbeitung einer Aufgabe benötigt werden (z.B. werden dem Nutzer die E-Mail-Adressen der Kursteilnehmer innerhalb der E-Mail-Funktionalität angezeigt) und bietet die Werkzeuge an, mit denen die Bearbeitung erst möglich wird (z.B. eine Kalenderfunktion zur Koordination kooperativer Aufgabenbearbeitung oder Testwerkzeuge).

Damit erfüllt eine Lernplattform das Aufgabenportfolio, das Stein für *Workflowmanagement-Anwendungen* beschreibt (vgl. [Ste99, 16]). Allgemein gesprochen soll eine Workflowmanagement-Anwendung Prozessabläufe automatisieren und unterstützen (vgl. [MüHa05, 2]). Kriterienlisten genügen dem Anspruch, die Rolle des Learning Management Systems im Lernszenario zu berücksichtigen, nur dann, wenn die Auswahl der Kriterien sich an der Funktion der Lernplattform im Lern- und Lehrprozess orientiert. In den dokumentierten Evaluationsstudien, zum Beispiel in den vielbeachteten Studien von Rolf Schulmeister (vgl. [Schu03]) und von Peter Baumgartner (vgl. [BaHM02]), werden die geplanten Einsatzszena-

rien nicht strukturiert auf ihre Anforderungen an die Lernplattformen hin untersucht. Kriterien werden hier entweder aus Referenzstudien extrahiert oder per Delphi-Methode erarbeitet. Usabilitybefragungen werden im Rahmen von Evaluationsstudien regelmäßig nach einer zeitlich begrenzten Testphase durchgeführt. Während die Vergleichbarkeit der Ergebnisse insgesamt durch dieses Vorgehen erhöht wird (weil die Benutzungsbedingungen innerhalb einer Testphase eher kontrollierbar sind), ist die Effizienz der Arbeit mit einem bestimmten System nach einer Zeit von ca. 4 Monaten (die Tests an Hochschulen erstrecken sich meist auf ein Semester) kaum erfassbar. Nielsen postuliert, dass unter Effizienz der Aufwand eines *Experten* bei der Bearbeitung seiner Aufgaben zu verstehen ist (vgl. [Niel93, 30f.]). Die Benutzer befinden sich aber noch in der Lernphase, wenn die Befragung stattfindet (vgl. [StDZ00, 10]), so dass das effizienzbezogenen Ergebnisse der Befragung vor allem durch die Erlernbarkeit beeinflusst werden. Eine Bewertung der Effizienz kann auf diesem Weg kaum zuverlässige Ergebnisse liefern.

2 Forschungsziel

Vor dem Hintergrund der oben genannten Problemstellung diskutiert dieser Beitrag ein Verfahren zur strukturierten Erarbeitung von Bewertungskriterien für die Evaluation von Learning Management Systemen. Zudem sollen Informationen über die Effizienz der Arbeit bzw. des Lernens mit den jeweiligen Systemen erhoben werden. Dabei wird davon ausgegangen, dass Learning Management Systeme grundsätzlich den Lernprozess managen sollen und daher gleiche Aufgaben wie eine *Workflowmanagement-Anwendung* übernehmen. Daher ist zu untersuchen, inwiefern eine Prozessanalyse als Evaluationsinstrument für die Auswahl von LMS herangezogen werden können, um die Mängel bisheriger Evaluationsansätze zu beheben.

Forschungstheoretisch findet dabei die Methodik der Designwissenschaften (Design Research, vgl. [VaKu04]) Anwendung, da kein empirischer Erkenntnisgewinn über ein Phänomen sondern die Gestaltung und Evaluation (einer Methode) im Mittelpunkt des Interesses steht.

Im folgenden Abschnitt wird zunächst die Prozessanalyse erläutert und deren Anwendung zur Auswahl von Learning Management Systemen hergeleitet, bevor anschließend die

Methode evaluiert wird, in dem sie im Rahmen einer Fallstudie zur Auswahl eines geeigneten Open Source LMS für eine virtuelle Universität eingesetzt wird.

3 Prozessanalyse als Methodik zur Ableitung von Kriterien zur LMS Evaluation

3.1 Prozessorientierung

Das zentrale Element der Prozessorientierung ist die Konzentration auf den Unternehmensprozess. Unter einem Unternehmensprozess wird ein „Bündel von Aktivitäten (verstanden), für das ein oder mehrere unterschiedliche Inputs benötigt werden und das für den Kunden ein Ergebnis von Wert erzeugt“ (vgl. [HaCh95, 52]). Weil es sich bei Universitäten nicht um ein Unternehmen handelt und die Schüler keine Kunden sind, wird in der Folge der allgemeiner gehaltene Begriff *Anwendungsprozess* verwendet. „Bei einem Anwendungsprozess handelt es sich um einen beliebigen Vorgang (...) zur Erfüllung einer Aufgabe.“ (vgl. [Ste99, 10]).

Im Unternehmenskontext werden Prozesse nach ihrer Bedeutung für den Kunden unterteilt. Es werden Leistungs- oder auch Kern- oder Primärprozesse und Unterstützungs- oder auch Sekundärprozesse unterschieden (vgl. [Epp100, 35], [BeKa05, 4]). Eine derartige Unterteilung macht im E-Learning-Bereich ebenfalls Sinn. Prozesse mit direktem Bezug zur Lern- und Lehrtätigkeit werden in der Folge als Primär- oder Kernprozesse bezeichnet, Aktivitäten mit eher administrativem und vorbereitendem Charakter zählen zu den sekundären Prozessen (Supportprozesse). Wie schon erwähnt, handelt es sich nicht um eine relevanzbezogene Einteilung. Vielmehr soll mit dieser Einteilung eine Denkrichtung angedeutet werden: Die Prozesse, die direkt zu einem Lernerfolg führen, werden zuvorderst betrachtet und in Hinblick auf das Lernziel entworfen.

3.2 Prozessanalyse und Evaluation von LMS

Dass einem LMS innerhalb eines komplexen E-Learningszenarios die Funktion einer Workflowmanagement-Anwendung zukommt, wurde bereits erwähnt. In der Folge soll der Zusammenhang zwischen der Prozessanalyse und der Auswahl einer Lernplattform verdeutlicht und das Vorgehen bei der Evaluation erläutert werden.

In dem von den Autoren vorgeschlagenen Vorgehensmodell kommt die Prozessanalyse an zwei Stellen im Evaluationsablauf zum Einsatz: In einem ersten Schritt wird die Prozessanalyse dazu verwendet, funktionale Kriterien für eine Kriterienliste zu erstellen, anhand derer eine kleine Anzahl von Systemen ausgewählt werden kann. Der zweite Schritt besteht in der Entwicklung von systemspezifischen Workflows für die Gruppe ausgewählter Lernplattformen. Die Workflows geben Auskunft über die Effizienz der Arbeit mit der Plattform und lassen Aussagen über die Zweckmäßigkeit des LMS zu.

3.2.1 Ermittlung von Bewertungskriterien durch Prozessanalyse

Da ein Workflow die Teilautomatisierung von Anwendungsprozessen ist (vgl. [MüHa05, 1]), hängt seine Gestalt notwendigerweise sowohl von der Form der Prozesse als auch von der eingesetzten Workflowmanagement-Software ab. Ein optimaler Workflow kommt also nur dann zustande, wenn die als Workflowmanagement-Anwendung fungierende Software optimal zu den durchgeführten Prozessen passt.

Um eine Bewertung verschiedener Softwaresysteme zur Prozessunterstützung vornehmen zu können, müssen also die Prozesse bekannt sein. Zur Mühlen und Hansemann gehen in ihrer Argumentation noch einen Schritt weiter: Nicht nur müssen die Prozesse bekannt sein, sondern sie sollten auch in Hinblick auf die Einführung einer Workflowmanagement-Anwendung neu entworfen werden, sonst „besteht die Gefahr, im Workflowmanagement-System bestehende Ineffizienzen abzubilden und damit künstliche Schwachstellen zu schaffen“ (vgl. [MüHa05, 8]). Tatsächlich bedingt sich der Bedarf eines neuen Prozessentwurfs oft schon durch die Neuartigkeit der möglichen Prozesse mit Hilfe von Lernplattformen. Die Neuartigkeit der Methoden besteht natürlich nicht zwangsläufig. Klingler führt das Scheitern vieler E-Learning-Projekte jedoch darauf zurück, „dass mit eLearning häufig versucht wird, traditionelle Lehr- und Lernmethoden durch Technikeinsatz nachzubilden oder zu optimieren“ (vgl. [Kling04, 1]).¹ Aus diesem Blickwinkel erscheint eine Neuentwicklung der Prozessentwürfe sinnvoll.

Anders als bei Prozessanalysen, die zum Zwecke der Prozessoptimierung, Unternehmensreorganisation oder der Softwareentwicklung durchgeführt werden, sind im hier beschriebenen Anwendungsfall der Prozessanalyse Prozessalternativen zu berücksichtigen, die aufgrund des hypothetischen Einsatzes verschiedener LMS mit unterschiedlichen Funkionali-

¹ Oder auch mit Hammer und Champy – gewohnt einprägsam – „Automating existing processes is analogous to paving cow paths“ [HaCh93, 48].

täten zustande kommen. Schließlich ist ein Ziel der Prozessmodellierung, die alternativen Prozessausprägungen gegeneinander Abwägen zu können, um so Informationen über die Relevanz einzelner Produktmerkmale zu erhalten.

Um aus der Betrachtung der Lern- und Lehrprozesse Anforderungen an die zu bewertenden Lernplattformen ableiten zu können, werden die Vorgänge im Rahmen der vorzustellenden Methode in Form von *erweiterten Ereignisgesteuerten Prozessketten* dargestellt. Das übliche Repertoire an Objekttypen wurde den Bedürfnissen der Evaluation angepasst (vgl. Abbildung 1) und es wurden Modellierungskonventionen eingeführt:

- Komplexe Objekte werden mit eigenen Modellen hinterlegt. Das Modell hat dann denselben Namen wie das Objekt.
- Ressourcenobjekte, die Teile des LMS darstellen (z.B. E-Mail- oder Forumsfunktionalität) werden orange dargestellt und Ressourcen, die nicht Teil des Learning Management Systems sind, in blau. Damit sind die notwendigen Eigenschaften des LMS leicht ablesbar.
- Ressourcenobjekte, die einer Schnittstelle zum LMS bedürfen, werden dunkelblau dargestellt, Ressourcenobjekte die keine Schnittstelle brauchen hellblau.
- Prozessalternativen, die modelliert werden, um unterschiedliche LMS-Eigenschaften zu berücksichtigen, werden mit einem hellblauen Funktionssymbol angekündigt.
- Funktionen werden durch Ereignisse beendet. Daraus folgt, dass auf eine Funktion stets ein Ereignis folgt. Wenn allerdings das Ereignis trivial ist, kann es weggelassen werden, um die Komplexität des Modells zu verringern, so dass zwei Funktionen aufeinander folgen können.
- In modernen LMS werden Stellentypen durch Rollen abgebildet. In diesem Sinne wird in dieser Arbeit statt des Objekttyps Stelle der Objekttyp *Rolle* verwendet.

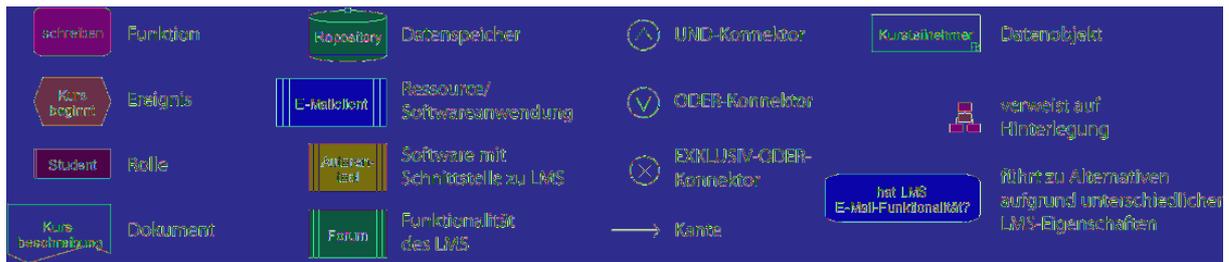


Abbildung 1: Objekttypen der eEPK zum Zwecke der Evaluation von Lernplattformen

Eine weitere Designentscheidung betrifft den Feinheitsgrad, zu dem die Prozesse entworfen werden sollen. Mit dem Detaillierungsgrad steigt der Modellierungsaufwand. Welcher Feinheitsgrad sich wirtschaftlich noch lohnt, hängt von dem Verwendungszweck des Modells ab (vgl. [SpSc05, 2]). Da das Prozessmodell als Basis für die Auswahl von funktionalen Kriterien dienen soll, sollten die Prozesse wenigstens so detailliert dargestellt werden, dass den einzelnen Prozessschritten einzelne Softwarewerkzeuge innerhalb des LMS (wie E-Mail, Dateiapload, Forum oder Autorentools) zugeordnet werden können, die später auch abfragbar sind.

Wurden die Prozessmodelle mit hinreichendem Detaillierungsgrad erstellt, können die Kriterien extrahiert werden. Alle Ressourcen, die in den Prozessmodellen dem LMS zugeordnet wurden (erkennbar an der orangenen Färbung), stellen funktionale Anforderungen dar. Ein System, das gut zu den modellierten Prozessen passt, verfügt über diese Funktionalitäten. Je nach Detaillierungsgrad lassen sich auch noch weitere Eigenschaften der jeweiligen LMS-Funktionalitäten spezifizieren.

Schnittstellenanforderungen mit anderen Softwareanwendungen sind daran erkennbar, dass die gleichen Daten von dem LMS und einem anderen System benutzt werden. Die Anforderung an die Lernplattform besteht dann darin, ein Datenformat zu unterstützen, das auch von der anderen Softwareanwendung unterstützt wird. Anwendungen, die Schnittstellenanforderungen an die Lernplattform stellen, werden dunkelblau dargestellt.

Anforderungen an die Rollenspezifikationen lassen sich an den in der Prozessmodellierung verwendeten Rollen ablesen.

Die Kriterien sind so zu formulieren, dass sie die Existenz oder die Abwesenheit einer bestimmten Eigenschaft abfragen, so dass sie mit ‚ja‘ oder ‚nein‘ beantwortet werden können. Die einzelnen Kriterien sind nicht alle gleich wichtig, sondern einige Softwareeigenschaften tragen einen größeren Teil zur Geeignetheit des System bei als andere. Diesem Umstand soll durch eine Gewichtung der Kriterien Rechnung getragen werden. Aus der Literatur zum

Thema Prozessanalyse und Prozessmanagement konnten keine geeigneten Methoden zur Gewichtung der Kriterien anhand der Prozessanalyse extrahiert werden. In dieser Arbeit wird für die Gewichtung die Methode von Schulmeister (vgl. [Schu03, 67]) angewendet, der damit ebenfalls Lernplattformen evaluierte.

Bei dem von Schulmeister entwickelten Verfahren handelt es sich um eine Abwandlung der Numerischen Gewichtung und Summierung. Die Kriterien werden nach ihrer Relevanz in drei Gruppen eingeteilt: K.O.-, Soll- und Kann-Kriterien. Soll-Kriterien werden mit Werten zwischen eins und fünf gewichtet, Kann-Kriterien erhalten Gewichtungen zwischen 0,1 und 0,5. K.O.-Kriterien werden nicht weiter gewichtet – erfüllt eine Software ein K.O.Kriterium nicht, wird es nicht weiter berücksichtigt. Das Format der Punktwertung entspricht folgendem Muster: XX.YYY.ZZZ. X steht dafür für Anzahl der erfüllten K.O.-Kriterien, Y für die erzielten Punkte bei den Soll- und Z bei den Kann-Kriterien. Diese Methode führt nicht zwangsläufig zu einer eindeutigen Reihenfolge, erlaubt es aber, gute Performer von schlechten zu unterscheiden und so eine Gruppe herauszufiltern, die im nächsten Schritt eingehender betrachtet werden soll.

Harte Kriterien für die Gewichtung selbst können nicht angegeben werden, da die Relevanz nicht unabhängig von den Umständen des Einzelfalls diskutiert werden kann. Es werden stattdessen Richtlinien angegeben, die die Einordnung in K.O.-, Soll- und Kann-Kriterien unterstützen sollen:

- Ein Kriterium ist stärker zu gewichten, wenn es aus mehreren Anwendungsprozessen abgeleitet wurde.
- Ein Kriterium ist stärker zu gewichten, wenn die Erfüllung in hohem Maße zur Koordination eines Prozesses beiträgt.
- Ein Kriterium ist schwächer zu gewichten, wenn es zu dessen Erfüllung eine zweckmäßige Prozessalternative gibt. Dabei sind die Alternativprozesse miteinander auf Effizienz und Effektivität zu vergleichen.
- Kriterien, die sich auf Datenelemente beziehen, sind stärker zu gewichten, wenn sie von Anwendungen benötigt werden, als wenn sie rein informative Funktion besitzen.
- Kriterien sind umso stärker zu gewichten, desto häufiger der Prozess ausgeführt wird, aus dem das Kriterium abgeleitet wurde.

Liegen die gewichteten Kriterien vor, kann eine Bewertung entweder auf Basis eigener Erhebungen oder auch unter Benutzung von online verfügbaren Tools wie dem EduTools Evaluationswerkzeug (vgl. [edut06]) erfolgen. Werden Informationen mit Hilfe externer Dienstleister erhoben, besteht das Problem, dass nur jene Kriterien abgefragt werden können, die auch von den Bereitstellern erhoben wurden.

3.2.2 Workflowanalyse zum detaillierten Vergleich der LMS

Für die ausgewählten Systeme sind die modellierten Anwendungsprozesse in systemspezifische Workflows umzusetzen. Der übliche Weg, von einem Prozess- zu einem Workflowmodell zu kommen besteht, in der Übersetzung in eine workflowspezifische Notation und die Ergänzung workflowrelevanter Daten. Für die Modellierung von Workflows mit dem Ziel der Evaluation von Lernplattformen wird ein anderer Weg vorgeschlagen. Der Grund für die Übersetzung in eine workflowspezifische Notation besteht darin, dass eine Workflowspezifikation von dem WFM-S ausführbar sein muss (vgl. [RuSh95, 4]). Dieser Anforderung muss das Workflowmodell, anhand dessen die einzelnen LMS verglichen werden sollen, nicht genügen. Vielmehr muss das Modell leicht verständlich sein und alle workflowrelevanten Informationen enthalten. Stein gibt den notwendigen Umfang eines Workflowschemas an: Funktionalitäten, Verhalten, Informations- und Datenelemente, Ressourcen und Aufbauorganisation müssen darstellbar sein (vgl. [Ste99]). Diese Elemente sind alle in der im vorigen Kapitel beschriebenen eEPK enthalten.

Für die Modellierung der Workflows werden zudem noch zwei weitere Modellierungskonventionen eingeführt, die die Lesbarkeit und Nachvollziehbarkeit des Modells erhöhen sollen:

- Navigationsschritte werden mit dem *Text des Links* und dem Signalwort *wählen* beschrieben: Soll ein Link mit dem Text *Home* angeklickt werden, wird die Funktion mit „Home wählen“ beschriftet.
- Auswahlmöglichkeiten mit Hilfe von Checkboxen werden mit Hilfe der *Bezeichnung des/der zu wählenden Objekte* und dem Signalwort *auswählen* beschrieben: Soll eine Checkbox, die zu einem Student innerhalb eines Menüs gehört angeklickt werden, wird die Funktion mit “Student auswählen“ beschriftet.
- Werden Objekte erstellt, auf die später per Link zugegriffen werden kann, so werden diese später mit ihrem *Typ* und dem Signalwort *Name* referenziert: Hat ein

Tutor einen Kurs erstellt, auf den der Student später zugreifen soll, so wird die entsprechende Funktion des Zugriffs mit „Kursname wählen“ beschriftet.

Das hier anzuwendende Verfahren unterscheidet sich notwendigerweise von den in der Literatur zum Thema Workflowspezifikation vorgeschlagenen Vorgehensweisen.

Die Workflowmodellierung zum Zwecke der Workflowsteuerung stellt eine Form der Soll-Modellierung dar. Sollen die Workflows als Basis für einen Vergleich mehrerer Softwaresysteme dienen, so handelt es sich dabei um eine Ist-Modellierung der Soll-Prozesse (also die faktische Unterstützung hypothetischer Prozesse). Um die Workflows modellieren zu können, muss untersucht werden, wie die einzelnen Systeme die Umsetzung der Anwendungsprozesse unterstützen. Anhand von Testinstallationen sind die Prozesse durchzuspielen und das resultierende Vorgehen ist zu modellieren.

Die Workflows – und damit die jeweils eingesetzten Lernplattformen – sollen in Bezug auf die Effizienz und Zweckmäßigkeit untersucht werden.

Da die einzelnen Systeme aber unterschiedliche Anwendungsprozesse unterschiedlich gut unterstützen, ist nicht unbedingt mit einem eindeutigen Ergebnis über alle Anwendungsprozesse hinweg zu rechnen. In der Regel werden einige Prozesse am besten von einem System unterstützt, während ein anderes die beste Wahl für einen anderen Prozess darstellt. Es bedarf also einer Handhabe für den Umgang mit solchen Konflikten.

Hier bietet sich die Methode Baumgartners an, QGS, die von Baumgartner ausführlich beschrieben wird (vgl. [BaHM02, 68f.]). Die Methode wird in ihrer Auslegung den Bedürfnissen der Workflowbewertung angepasst und es werden Kriterien für die Gewichtung der Prozesse und Bewertung der Workflows angegeben.

- Die qualitative Gewichtung wird auf die Relevanz eines Teilprozesses bezogen. Die Skala der Gewichtung hat die Einteilung: Essentiell (E), äußerst wichtig (*), sehr wichtig (#), wichtig (+), weniger wichtig (!) und nicht wichtig (0).
- Was bei Baumgartner der Grad der Erfüllung eines Kriteriums ist, wird für diesen Zweck als Ausprägung der Prozesseffizienz bzw. Zweckmäßigkeit definiert.

- Ein LMS kann in Bezug auf einen Prozess höchstens den Wert der Gewichtung des Prozesses für seine Zweckmäßigkeit bzw. Workfloweffizienz zugewiesen bekommen.
- Die Werte für Effizienz und Zweckmäßigkeit werden getrennt betrachtet – schließlich soll über beide Qualitäten eine Aussage getroffen werden. Eine Zusammenfassende Betrachtung der beiden Elemente ist nicht vorgesehen. Sie bleiben beide für sich jeweils ein Teil der Akzeptabilität des LMS.
- Um ein Gesamtergebnis für alle betrachteten Prozesse zu erlangen, werden die jeweiligen Ausprägungen Kriterien addiert. Ein LMS hat also stets ein Ergebnistupel, wobei beide Teile folgendes Format aufweisen:

§ Produkt A: 7 *, 5 #, 2 +, 3 |.

- Wiederum gibt es im Ergebnis nicht unbedingt eine eindeutige Reihenfolge.

Die Gewichtung der einzelnen Prozesse richtet sich vor allem danach, wie häufig der jeweilige Prozess innerhalb eines Semesters stattfindet. Die Zielbeschreibungen der Hochschule selbst sind der zweite wichtige Indikator. Eine Hochschule, die sich das Ziel gesetzt hat, in den nächsten Jahren die Kosten zu reduzieren, muss vor allem auf die administrativen Prozesse achten. Steht für die Hochschule ein umfassender Service im Vordergrund, sind vor allem jene Prozesse stark zu gewichten, die für die Erfüllung dieses Ziels verantwortlich sind. Die Bewertung der Workflows geschieht innerhalb der angegebenen Skala. In der Folge werden die Kriterien für die Bewertung angegeben.

Die Zweckmäßigkeit richtet sich danach, in wie weit die Software den Benutzer in die Lage versetzt, das zu tun was er will oder soll. Dementsprechend sind auch die Bewertungskriterien wie folgt.

- Können alle Aufgaben in vollem Umfang erfüllt werden?
- Müssen andere Werkzeuge benutzt oder der Prozess manuell bearbeitet werden?

Die Effizienz wird durch die Ausprägung der einzelnen Workflows bestimmt. Zur Bewertung der Effizienz wird eine Auswahl der von Krallmann beschriebenen Merkmale dysfunktionaler Prozesse herangezogen (vgl. [KrGF02, 234]):

- ausufernder Informationsaustausch redundanter Daten,

- Mehrfacheingabe von Daten,
- hohes Maß an Kontrolle nötig,
- Nachbearbeitungen und Iteration,
- Komplexität und Ausnahmen.

4 Evaluation

Die vorgestellte Methode wurde im Rahmen einer Diplomarbeit auf das Lernszenario an der Virtual Global University (VGU) angewendet (vgl. [Luth06]). Die drei Open Source Lernplattformen Ilias, ATutor und Moodle wurden auf ihre Zweckmäßigkeit und Effizienz in Bezug auf die Lern- und Lehrprozesse an der VGU hin untersucht.

Die Prozesse wurden mit Hilfe von strukturierten Interviews erfasst und wie oben dargestellt modelliert. Die Bewertungskriterien wurden direkt aus den Prozess- und Datenmodellen abgeleitet. Die Gewichtung wurde unter Berücksichtigung der in Abschnitt 3.2.1 erläuterten Grundsätze vorgenommen. Es wurden 15 K.O.-, 15 Soll- und 16 Kann-Kriterien identifiziert. Die Daten wurden anhand von Testinstallationen ermittelt. Es qualifizierten sich Ilias und Moodle für die zweite Stufe, in der die systemspezifischen Workflows entwickelt wurden.

Die einzelnen Anwendungsprozesse wurden entsprechend der in Kapitel 3.2.2 dargestellten Kriterien in ihrer Relevanz bewertet, und anhand der angegebenen Kriterien auf Effizienz und Zweckmäßigkeit hin bewertet.

Im Ergebnis zeigte sich, dass Moodle zweckmäßiger ist und sein Einsatz zu effizienteren Workflows führt.

Vor der Durchführung der Evaluationsstudie hatte es noch Zweifel gegeben, ob die getrennte Betrachtung von Zweckmäßigkeit und Effizienz für den Fall einer Software mit Workflowmanagement-Funktion überhaupt sinnvoll ist. Da ein Hauptziel der Implementierung von Workflowmanagement-Anwendung in der Erhöhung der Prozesseffizienz liegt (vgl. [MüHa05, 2]), schien es möglich, dass eine höhere Zweckmäßigkeit zwangsläufig auch zu einer höheren Effizienz führen würde. In einzelnen Anwendungsprozessen zeigte sich allerdings, dass es durchaus Fälle gibt, in denen ein System zwar zweckmäßiger ist, aber zu weniger effizienten Prozessen führt. So lagen beide Systeme für den Prozess „Einstellen von

Kursinhalten“ bei der Zweckmäßigkeit gleich auf (Ilias: |, Moodle: |) während die Prozesseffizienz für den Einsatz von Moodle etwas besser war (Ilias: 0, Moodle: |).

In Hinsicht auf das Forschungsziel kann festgestellt werden, dass es gelungen ist, sinnvolle Kriterien aus der Prozessanalyse des geplanten Lernszenarios heraus zu entwickeln. Die Prozessmodelle lieferten auch in Verbindung mit prozessorientierten Richtlinien Ansatzpunkte für die Gewichtung der Kriterien und waren somit als Grundlage für den ersten Evaluationsschritt ein brauchbares Instrument. Ein Vergleich mit anderen Methoden der Kriteriengenerierung wurde nicht angestellt, wäre für die Zukunft aber sinnvoll.

Die Analyse der Effizienz führte ebenfalls zu verwertbaren Ergebnissen und ist, weil es sich um eine analytische Betrachtung der Prozesseffizienz handelt, in seiner Validität nicht von den Fähigkeiten von Testpersonen im Umgang mit der Software abhängig.

Die Prozessanalyse eignet sich allerdings nicht unbedingt als alleiniges Verfahren zur Bewertung von Lernplattformen. Prozessmodelle enthalten letztlich keine Informationen über viele relevante Produkteigenschaften wie beispielsweise den Preis, die Gestaltung der GUI oder dem Lernaufwand. Diese Daten sind dementsprechend auf anderen Wegen zu erheben.

Es stellte sich während der Anwendung der Methode heraus, dass neben dem Prozesswissen ein beträchtliches Produktwissen vonnöten war, um die optimalen Workflows für die jeweiligen Systeme zu erstellen. Es ist allerdings anzunehmen, dass der Einarbeitungsaufwand für andere Evaluationsverfahren kaum geringer ausfällt. Zumal ein einziger Modellierer ausreicht, der die Programme gut kennen muss, während empirische Evaluation die Einarbeitung mehrerer Personen verlangt.

Ob die Prozessanalyse ein geeignetes Instrument der Evaluation darstellt ist zudem abhängig davon, wie stark der Workflowmanagement-Charakter einer Lernplattform innerhalb des Lernszenarios ausgeprägt ist. Je geringer der Anteil der Lehre ist, der über das Internet abgewickelt wird, desto geringer ist der Koordinationsbedarf und desto weniger fällt die Prozessunterstützende Funktionalität ins Gewicht. Für den Fall einer virtuellen Universität greift diese Einschränkung allerdings nicht.

5 Fazit

Dieser Beitrag diskutierte zunächst die Mängel gegenwärtiger Evaluationsmethoden für E-Learning. Es wurde festgestellt, dass insbesondere die Festlegung der Bewertungskriterien

problematisch ist. Zudem erschien die Ermittlung der Effizienz der Arbeit mit den jeweiligen Systemen nicht stichhaltig. Von dieser Situation ausgehend wurde festgestellt, dass die Prozessperspektive auf ein E-Learning System anwendbar ist und ein LMS prinzipiell die gleichen Aufgabentypen wie eine Workflowmanagement-Anwendung erfüllt. Vor diesem Hintergrund liegt es nahe, eine Evaluationsmethode auf Basis der Prozessanalyse daraufhin zu untersuchen, ob sie die erkannten Mängel konventioneller Evaluationen überwinden hilft. Die entwickelte Methodik wurde evaluiert, in dem für spezielle Einsatzbereiche einer privatwirtschaftlichen international agierenden virtuelle Universität (VGU) verschiedene Open Source Learning Management Systeme evaluiert wurden, um zu einem Vergleich und einer Auswahlentscheidung zu kommen.

Die Anwendung der vorgestellten Methode zeigte, dass Kriterien sich bei passender Modifikation der Notationsregeln des Modelltyps eEPK direkt aus den Prozessmodellen ableiten lassen. Eine endgültige Bewertung von Effizienz und Zweckmäßigkeit ließen die software-spezifischen Workflowmodelle zu.

Die Prozessanalyse bietet sich nicht als einziges Evaluationsinstrument an, weil sich in Prozessmodellen viele Produkteigenschaften nicht adäquat abbilden lassen. Es wird ein multimethodisches Verfahren angeraten, um alle relevanten Eigenschaften von Lernplattformen zu erfassen.

Literaturverzeichnis

- [BaHM02] Baumgartner, Peter; Häfele, Hartmut; Maier-Häfele, Kornelia: E-Learning Praxishandbuch: Auswahl von Lernplattformen. Studienverlag, Innsbruck 2002.

- [BeKa05] Becker, Jörg und Kahn, Dieter: Der Prozess im Fokus. In: Becker, J., Kugeler, M., Rosemann, M, (Hrsg): Prozessmanagement. 5. Aufl. Springer, Berlin Heidelberg 2005, S. 373-408.

- [edut06] edutools: CMS – Features List. http://edutools.info/feature_list.jsp?pj=8. Abruf am 2006-06.26.

- [Epl00] Epple, Martin: Prozessqualität – Ansätze zum Qualitätsmanagement bei innerbetrieblichen und zwischenbetrieblichen Prozessen. Difo-Druck, Bamberg 2000.
- [HaCh95] Hammer, Michael und Champy James: Business Reengineering – Die Radikalkur für das Unternehmen. 5. Aufl. Campus, Frankfurt New York 1995.
- [HaCh93] Hammer, Michael und Champy James: Reengineering the Corporation: A Manifesto for Business Revolution. Harper-Business, New York 1993.
- [Kling04] Klingler, Hugo: eLearning ist tot, es lebe eLearning. In: Roters, G., Turecek, O., Klingler, W. (Hrsg): eLearning – Trends und Perspektiven. Vistas, Berlin 2004, S. 41-45.
- [KrGF02] Krallmann, Hermann; Gronau, Norbert; Frank, Helmut: Systemanalyse im Unternehmen. 4. Aufl. Oldenbourg München 2002.
- [Luth06] Luther, Michael: Prozessanalytische Softwareevaluation – Bewertung von Open Source Lernplattformen für den Einsatz im Hochschulbereich. Diplomarbeit, Technische Universität Berlin 2006.
- [MüHa05] zur Mühlen, Michael und Hansmann, Holger: Workflowmanagement. In: Becker, J., Kugeler, M., Rosemann, M, (Hrsg): Prozessmanagement. 5. Aufl. Springer, Berlin Heidelberg 2005, S. 373-408.
- [Niel93] Nielsen, Jacob: Engineering Usability. Academic Press, San Diego 1993
- [ScTe04] Schenkel, Peter und Tergan, Sigmar O.: Qualität von E-Learning – eine Einführung. In: Tergan, S. O., Schenkel, P. (Hrsg): Was macht E-Learning erfolgreich? Springer, Berlin 2004, S. 3-13.
- [Schu03] Schulmeister, Rolf: Lernplattformen für das virtuelle Lernen: Evaluation und Didaktik. Oldenbourg, München 2003.
- [SpSc05] Speck, Mario und Schnetgöke, Norbert: Sollmodellierung und Prozessoptimierung In: Becker, J., Kugeler, M., Rosemann, M, (Hrsg): Prozessmanagement. 5. Aufl. Springer, Berlin Heidelberg 2005, S. 185-220.

- [StDZ00] Steffens, Franz; Dorrhauer, Carsten; Zlender, Andrej: Usability-Tests ausgewählter Geschäftsprozesse. HMD-Praxis der Wirtschaftsinformatik (2000) 212: 57-69
- [Ste99] Stein, Katrin: Integration von Anwendungsprozessmodellierung und Workflow-Management. Dissertation, Universität Erlangen-Nürnberg 1999.
- [Stoc04] Stockmann, Reinhard: Wirkungsorientierte Programmevaluation: Konzepte und Methoden für die Evaluation von E-Learning. In: Meister D.M., Tergan, S. O., Zentel, P. (Hrsg): Evaluation von E-Learning. Waxmann, Münster 2004, S. 23-42.
- [Terg04] Tergan, Sigmar O.: Realistische Qualitätsevaluation von E-Learning. In: Meister, D. M.; Tergan S. O.; Zentel, P. (Hrsg): Evaluation von E-Learning. Waxmann, Münster 2004, S. 131-154.
- [TrWW03] Trahasch, Stephan; Wiedenbruch, Nadine; Wöhrle, Nicole () CampusOnline – E-learning an der Universität Freiburg. In: Bett, K.; Wedekind, J. (Hrsg): Lernplattformen in der Praxis. Waxmann, Münster 2003, S. 15-32.
- [VaKu04] Vaishnavi, Vijay; Kuechler, William: Design Research in Information Systems. <http://www.isworld.org/Researchdesign/drisISworld.htm>, 2004-01-20, Abruf am 2006-01-18.

Modellbasiertes Curriculum-Design für Learning Management Systeme: Ein Integrationsansatz auf Basis von ARIS und IMS Learning Design

Guido Grohmann, Wolfgang Kraemer, Frank Milius, Volker Zimmermann

imc information multimedia communication AG
66115 Saarbrücken

{guido.grohmann,wolfgang.kraemer,frank.milius,volker.zimmermann}@im-c.de

Abstract

Der Artikel diskutiert einen integrativen Ansatz zum modellbasierten Curriculum-Design und einer anschließenden Umsetzung in einem Learning Management System als Laufzeitumgebung. Der Lösungsweg wird anhand der Produkte CLIX[®] der imc AG sowie ARIS Business Architect[®] der IDS Scheer AG aufgezeigt. Er basiert auf Vorüberlegungen zum „Learning Design“ der Standardisierungsinitiative IMS Global Learning Consortium.

1 Werkzeuggestützte Planungs- und Designmethoden für Lernprozesse als Herausforderung

Lehrpläne und Curricula sind das „Gerüst“ jeder Bildungs- und Trainingsmaßnahme. Die Aufgabenstellung Lehrpläne zu entwerfen wird nachfolgend als „Learning Design“ bezeichnet.

Meist zeichnen sich Programmplaner in Kommissionen oder auch Lehrer, Professoren, Dozenten und Trainer selbst für das Learning Design verantwortlich. Mit zunehmender Reife und Verbreitung von Lerntechnologien stehen sie vor der Herausforderung, den Lehrplan und Lernprozess für ihre Zielgruppe unter Nutzung aller technologischen Möglichkeiten optimal zu gestalten. In der Regel führt dies zu einer Kopplung von Präsenztrainingsbausteinen mit eLearning-Modulen und -Methoden entsprechend eines zu erreichenden Lernziels.

Konzepte zur Erstellung und Durchführung von Präsenztrainings sind seit langer Zeit vorhanden und stehen in der wissenschaftlichen Diskussion. Der Diskussion um das Thema „technologiegestütztes Lernen“ fehlte bisher ein systematischer und insbesondere standardisierter An-

satz für das Learning Design. Es ist weniger eine Frage der Ideen, als vielmehr eine Frage der Methodik und der Werkzeuge.

Vergleicht man die Situation mit dem Engineering von Softwaresystemen, das heute meist mit entsprechenden Modellierungsmethoden und Werkzeugen erfolgt, steht man beim Learning Design für technologiegestützte Lernszenarien erst am Anfang: Alle Phasen laufen papiergestützt ab. Erst wenn die Lehrpläne fertig gestellt sind, werden sie im Learning Management System „erfasst“. Das Ergebnis ist, dass nicht alle Möglichkeiten optimal genutzt wurden. Änderungen und Lernerfeedbacks fließen kaum noch in die Lehrpläne ein. Eine Weiterentwicklung unter Berücksichtigung von Änderungen gleicht somit meist einer Neukonzeption.

In Anlehnung an die Methoden des Softwareengineerings müssen deshalb Methoden zum Learning Designs entwickelt und genutzt werden. Zudem sollten Werkzeuge für die Planung in Learning Management Systeme integriert werden. Dieser Beitrag will einen ersten Lösungsansatz für Methoden und Werkzeuge aufzeigen. Dabei ist den Autoren bewusst, dass in diesem Thema erst die Anfänge gemacht sind. Erst langsam kommen unter Stichworten wie „IMS Learning Design“ Methoden auf. Ihre Praxiserprobung erfolgt derzeit in ausgewählten Projekten. Der vorliegende Beitrag soll einen Überblick über die Zielsetzungen geben und aufzeigen, in welche Richtung sich die Diskussion derzeit bewegt.

Hierzu wird in den folgenden Abschnitten wird ein integrativer Ansatz zur methodengestützten Umsetzung der Gestaltung von Lehrangeboten am Beispiel des Learning Management System CLIX[®] der imc AG sowie dem ARIS Business Architect[®] der IDS Scheer AG aufgezeigt. Der Lösungsansatz basiert auf Vorüberlegungen der Standardisierungsinitiative IMS Global Learning Consortium zum Learning Design.

2 Planungsmethoden und Design-Werkzeuge

2.1 IMS Learning Design

Um die Umsetzung pädagogischer Unterrichtskonzepte und -methoden in multimediale Trainingsmaterialien durchgängig zu unterstützen, wurden in den vergangenen Jahren Konzepte, Modellierungsmethoden und -sprachen zur Abbildung von Lerninhalten zur Diskussion gestellt [Groh06, S. 162ff.]. Als führende Entwicklungen auf diesem Gebiet gelten sowohl die Educati-

on Modelling Language (EML) [Kope01] als auch die daraus entstandene IMS Learning Design-Spezifikation, kurz IMS-LD, des IMS Global Learning Consortiums (CENoJ, S. 221ff.).

Die EML beschreibt im Kern den didaktisch kommentierten Aufbau von Lerninhalten, deren Verhalten und Beziehungen im Lehr-/Lernprozess sowie die Interaktion zwischen den Komponenten. Anzumerken ist, dass der Schwerpunkt der EML auf den didaktischen Aspekten liegt und die Beschreibung des Lehrmaterials nur am Rande erfolgt. Ebenso kommt der zugrunde liegenden Architektur der Anwendungssoftware ein untergeordneter Stellenwert zu und ist als nicht verpflichtender Bestandteil anzusehen [LuWS02, S. 212]. Durch die Loslösung des Einsatzes von Software ist die EML für verschiedenste pädagogische Szenarien geeignet und kann für die modellbasierte Konzeption und Beschreibung traditioneller Präsenzlehr- und Blended Learning-Veranstaltungen bis hin zu reinen eLearning-Angeboten genutzt werden. Sie basiert auf einem Metamodell, für welches eine XML-Notation bereitgestellt wird. Diese Notation soll dazu dienen, modellierte Lehr- und Lernszenarien in eine DV-orientierte Sprache zu überführen, um eine informationstechnische Umsetzung der Lernmaterialien in einer elektronischen Lernumgebung zu ermöglichen. Obwohl die EML nicht ausschließlich für die Beschreibung technologiegestützter Trainings konzipiert wurde, ist die Modellierung von eLearning-Content als integrativer Bestandteil der Spezifikation anzusehen [QuSi03, S. 151].

Aufgrund der anhaltenden Standardisierung und Standardisierungsbestrebungen im Bereich des eLearning wurde die Weiterentwicklung der EML als eigenständige Modellierungssprache eingestellt und dem IMS Global Consortium als Spezifikationsvorschlag zur Einbettung in die IMS-Gesamtstrategie übergeben. Als Ergebnis entstand 2003 die erste Version der IMS Learning Design-Spezifikation [KoOA03]. Neben der Erweiterbarkeit des IMS-LD um andere Komponenten und Spezifikation von IMS, wie bspw. IMS Content Packaging oder IMS/LOM Metadata usw., besteht der wesentliche Unterschied zu dem ursprünglich verfolgten Ansatz der EML in dem Verzicht auf die Beschreibung des eLearning-Contents zugunsten einer stärkeren Ausrichtung auf die reinen Lernprozesse sowie deren Ablauflogik bzw. -steuerung.

Grundsätzlich lässt sich feststellen, dass sich mit Hilfe des IMS-LD Lernszenarien für eine Vielzahl lerntheoretischer Ausrichtungen und Medienansätze beschreiben, modellieren und wieder verwenden lassen [JeCu05]. Die Komponenten der Beschreibung entsprechen dabei weitgehend den klassischen didaktischen Kategorien wie Lernzielen, Rollen, Sozialformen, Umgebungsvariablen etc.

Die *Environment* bildet den Rahmen für die Durchführung von Aktivitäten (*Activity*) und beinhaltet neben den reinen Inhaltsobjekten (*Learning Objects*) und Services auch sämtliche notwendige, infrastrukturelle und organisationale Ressourcen.

Analog des Verständnisses der Organisationsmodellierung beschreibt die Klasse *Person* die Einordnung von natürlichen Personen in die (Aufbau-) Organisation. Detaillierte Angaben zur Klasse „Person“ sowie die Ableitung von Rechten und Pflichten erfolgt bei der IMS-LD Spezifikation über das Rollenprofil (*Role*). Die Typklasse *Role* typisiert die Teilnehmer von Lerneinheiten und unterscheidet grundsätzlich zwischen Lernenden (*Learner*) und Mitarbeitern (*Staff*). Diese Subklassen können wiederum einer Tiefengliederung unterliegen. Gängige Ausprägungen von *Staff* sind bspw. Tutor, Mentor, Übungsleiter usw.

Activities sind die Kernelemente des IMS Learning Designs. Sie stellen die Verbindung zwischen den organisationalen Einheiten (*Role*), den anvisierten Zielen (*Learning Objective*), den bereitstehenden Lerneinheiten (*Learning Objects*) sowie den notwendigen Services dar. Innerhalb von IMS-LD erfolgt eine Unterscheidung in Lern- und Unterstützungsaktivitäten.

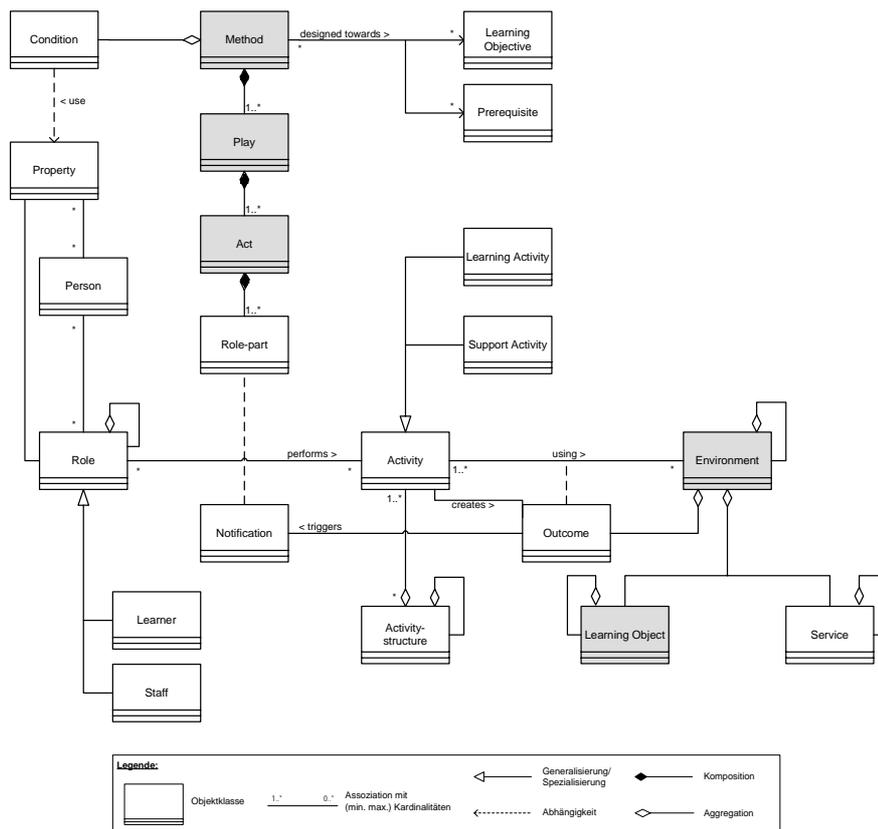


Abb. 1: IMS-LD Metamodell (In Anlehnung an Koper, Olivier, Anderson 2003)

Aktivitäten können innerhalb einer *Activity Structure* aggregiert werden. Die Aggregation dient als Mechanismus zur Strukturierung von Lerneinheiten, der Darstellung der Ablaufreihenfolge und deren Zusammenhänge zu anderen Komponenten des IMS-LD auf.

Learning Objectives definieren die globale Zielsetzung, die durch das erfolgreiche Absolvieren der Lerneinheiten und -veranstaltungen erreicht bzw. durch die methodisch-didaktische Planung intendiert werden.

Lernobjekte werden in der IMS-LD-Spezifikation nicht näher definiert. Stattdessen erfolgt der Verweis auf die Standardspezifikation Learning Object Metadata (LOM) Standards [LTSC05].

Services umfassen sämtliche Dienste und Kommunikationskanäle, die von der Umgebung (Environment) zur Unterstützung der Lernaktivität bereitgestellt werden.

Notifications ermöglichen den rollenbasierten Versand von Benachrichtigungen über neue Aktivitäten (inkl. Subklassen), besondere Ereignisse und Statusänderungen.

Conditions dienen zur Bestimmung personalisierter Lernangebote, sind als Startereignisse zu verstehen und werden zeitlich-logisch vor einer Durch-/Ausführung von Lernaktivitäten geprüft. Sie bestimmen auf Basis der dokumentierten Fähigkeiten/Fertigkeiten des Profils und der Lernhistorie des Lernenden über passende Lernpfade oder alternative Weg. Somit haben Conditions Einfluss auf die Ablaufsteuerung eines Lernarrangements.

Properties dienen der Überwachung und Personifizierung von Lernangeboten. Sie sind eng mit der Rolle (Role), den individuellen und fachlichen Vorgaben der Klassen „Condition“ und „Person“ verbunden. Properties sind durch Lernfortschritte veränderbar und liefern wiederum Daten für zeitlich nachgelagerte bzw. weiterführende Lernaktivitäten. Dieser Einfluss verändert somit die Vorbedingungen (Condition) hinsichtlich ihres Niveaus und verleiht den Properties den Charakter von „Endereignissen“ der Geschäftsprozessmodellierung.

Prerequisites bezeichnen die Anforderungen des Lernenden an eine Lerneinheit im Sinne von Eigenleistungen bzw. des erreichten Niveaus des Lernalters. Sie korrespondieren mit den definierten Lernzielen (Learning Objectives) und gelten als Richtlinien für Ab-/Erarbeitung der Lerninhalte um die Zielsetzungen zu erfüllen.

Methods, *Plays* und *Acts* sind der Kern des IMS-LD. Sie stehen in direktem Zusammenhang zueinander und werden durch die „Conditions“ aufgerufen und gesteuert. Methods können als aggregierte Funktionen beschrieben werden, die jeweils mehrere Plays umfassen. Plays wiederum setzen sich aus mehreren Acts zusammen.

2.2 Architektur integrierter Informationssysteme (ARIS)

Wie oben dargestellt, liefert die Spezifikation von IMS Learning Design die theoretische Grundlage für die Planung von Lernprozessen. In der Praxis ist diese Methodik aufgrund ihrer Abstraktheit sehr schwierig anzuwenden. Deshalb wird vorgeschlagen, eine einfache und praxiserprobte Methode aus der Prozeßmodellierung einzusetzen, diese aber in Analogie zur semantischen Definition von IMS Learning Design anzuwenden.

Nach dem hier vertretenen Ansatz wird zur Modellierung von Lernprozessen die im Rahmen des ARIS-Konzepts entwickelte Methode der Ereignisgesteuerten Prozesskette (EPK) eingesetzt. ARIS ist als Bezugsrahmen für eine systematische und ganzheitliche Modellierung zu verstehen. Die in ARIS verwendeten Modellierungsmethoden, -sprachen und -techniken kennzeichnen die ablauforganisatorischen Problemstellungen durch semiformale Beschreibungsmöglichkeiten [Sche02, S. 1-4]. Zudem helfen sie Prozesse und Prozessschnittstellen sowie organisatorische Zuständigkeiten klar zu definieren.

Zur Modellierung der Geschäftsprozesse wird die im Rahmen des ARIS-Konzepts entwickelte Methode der Ereignisgesteuerten Prozesskette (EPK) eingesetzt. Zentrales Merkmal der EPK bildet die Veranschaulichung der zu einem Prozess gehörenden Funktionen in deren zeitlich-logischer Abfolge. Eingetretene Zustände, die wiederum nachgelagerte Prozessschritte anstoßen können, sowie Bedingungskomponenten werden unter dem zeitpunktbezogenen Konstrukt „Ereignis“ zusammengefasst. Für die Beschreibung der Kontrollflüsse innerhalb einer EPK kommen konjunktive, adjunktive und disjunktive Verknüpfungsoperatoren zur Anwendung.

Funktionen, Ereignisse und Verknüpfungsoperatoren bilden somit das Basismodell der EPK. Weitere semantische Darstellungsobjekte lassen sich an die Funktionen modellieren. Ein solcher Diagrammtyp wird als erweiterte Ereignisgesteuerte Prozesskette (eEPK) bezeichnet und erlaubt bspw. die Abbildung von am Prozess beteiligten Organisationseinheiten. In Abhängigkeit vom gewünschten Abstraktionsniveau lassen sich Ereignisgesteuerte Prozessketten in verschiedenen Granularitätsgraden darstellen. Durch eine stufenweise und strukturierte Hierarchisierung des gesamten Geschäftsprozesses kann dessen Komplexität somit auf ein gewünschtes Maß justiert werden.

Verwendet man diese einfache Notation, so lassen sich sehr leicht Lernprozesse als ereignisgesteuerte Prozesskette abbilden. Zudem sind die meisten Aspekte von IMS Learning Design auch mit dieser einfachen Struktur abbildbar. Im Nachfolgenden wird ein Beispiel modelliert. Hierzu wird die beispielhafte Lehrveranstaltung „Grundlagen des Projektmanagements“ zunächst fach-

lich beschrieben. Hierbei handelt es sich von Struktur und Inhalt um ein überschaubares Szenario, welches die prinzipielle Idee verdeutlichen soll.

Zu Beginn des Kurses wird der Lerner zunächst aufgefordert, einen Einstiegstest „Wissens-Check Projektmanagement“ zu absolvieren, mit dessen Hilfe überprüft wird, ob der Lerner das erforderliche Vorwissen für die Lerninhalte mitbringt. Wird dieser Test mit mindestens 50% korrekten Antworten abgelegt, kann der Lerner am weiteren Verlauf des Kurses teilhaben. Beträgt das Ergebnis weniger als 50%, kann der Lerner den Einstiegstest beliebig oft wiederholen. Um den Test erfolgreich abschließen zu können, werden ihm jedoch nach dem ersten nicht erfolgreichen Absolvieren des Tests Lerninhalte bereitgestellt, die ihn auf die Wiederholung des Eingangstests vorbereiten. Diese Lerninhalte gliedern sich in zwei „Web-based Trainings“ (WBT „Führung und Motivation“ sowie WBT „Harvard Manage Mentor“).

Nach dem Bestehen des Tests wird der Kursteilnehmer zur Teilnahme an einer Präsenzveranstaltung „Veranstaltung zum PM-Kurs“ zugelassen, in welcher der eigentliche Lehrstoff zum Thema Projektmanagement durch einen Tutor vermittelt wird. Die für die Präsenzveranstaltung notwendigen Lernmaterialien stehen dem Lerner ab dem Bestehen des Eingangstests zur Verfügung.

Nach dem Besuch der Präsenzveranstaltung und der Teilnahmebestätigung durch den Tutor, bereitet sich der Lerner mit Hilfe eines weiteren Web-based Trainings auf einen Abschlusstest vor. Sind diese Vorbereitungen abgeschlossen, wird der Abschlusstest (online) durchgeführt.

Wird der Test mit mindestens 60% korrekter Antworten absolviert, gilt er als bestanden und der gesamte Kurs als erfolgreich absolviert. Sind weniger als 60% der Antworten korrekt, hat der Teilnehmer die einmalige Möglichkeit den Test zu wiederholen. Verfehlt der Lerner die 60% ein zweites Mal, gelten Test und Kurs als nicht bestanden.

Bevor der Kurs endgültig beendet wird, ist der Lerner verpflichtet, einen Feedbackbogen zur qualitativen Evaluation der Lehrveranstaltung auszufüllen und einzureichen.

Abbildung 2 enthält ein Modell für das oben beschriebene Beispiel in Notation der eEPK. Die von einer Online-Lernumgebung durchzuführenden Prozessschritte sind ohne Zuordnung von Organisationseinheiten modelliert, die von Lerner oder Tutor durchzuführenden Prozessschritte sind entsprechend gekennzeichnet. Die an einzelne Funktionen modellierten Datenobjekte kennzeichnen Komponenten (Tests, Lerninhalte, Feedbackbögen) auf die zur Durchführung der Funktionen zugegriffen wird.

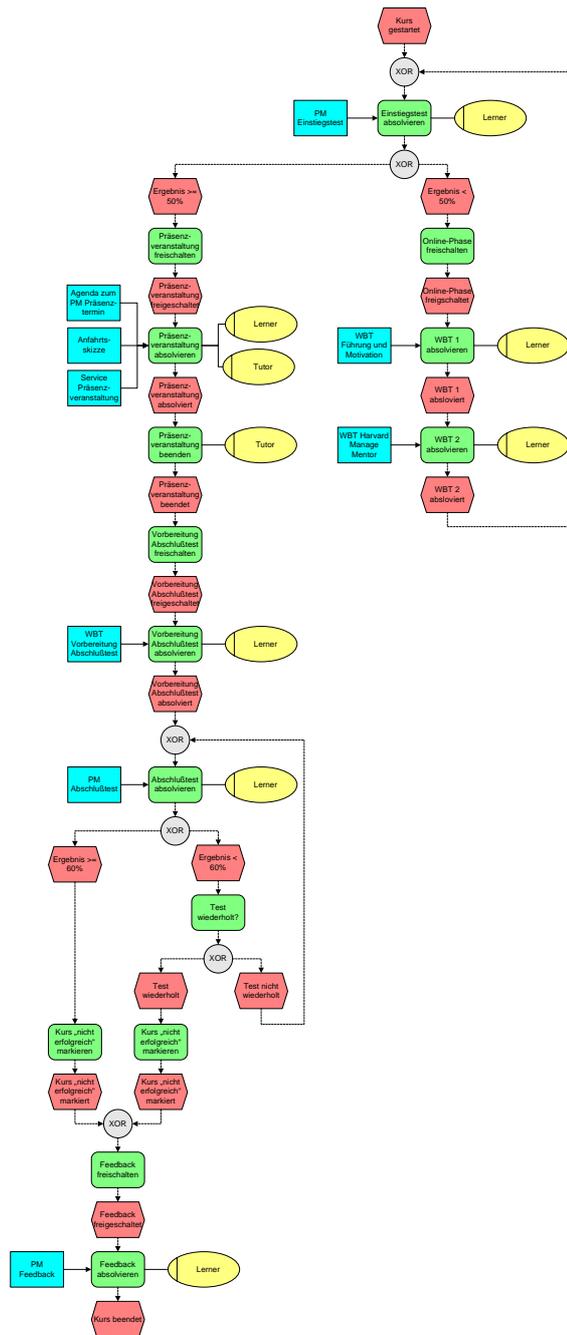


Abb. 2: Gesamtprozess „Grundlagen des Projektmanagements“

2.3 CLIX[®] als Umsetzungswerkzeug für Lernszenarien

Das Anwendungssystem CLIX[®] ist ein so genanntes Learning Management System (LMS), mit dem sämtliche Informations-, Lern- und Wissensprozesse über eine Webapplikation gesteuert werden. Anders formuliert kann CLIX[®] als eine Lernumgebung beschrieben werden, in der alle relevanten Geschäftsprozesse einer medien-basierten Lernwelt unterstützt werden [KrMZ05]. Innerhalb der Lernumgebung treffen Bildungsangebot und -nachfrage aufeinander und somit entstehen Prozesse des Lehrens und Lernens. Diese Prozesse bilden den Rahmen der medien-basierten Lernwelt.

Wird die in Abschnitt 2.1 vorgestellte IMS-LD-Spezifikation ihrem Anspruch gerecht, Lehr- und Lernszenarien aller Art abbilden zu können, so sollten auch unternehmens- und hochschul-spezifische Prozesse im Rahmen von in CLIX[®] durchgeführten Kursen oder Bildungsprogrammen abgebildet werden können. Um diese Annahme zu verifizieren, werden im Folgenden die Bestandteile eines CLIX[®]-spezifischen Lehr- und Lernprozesses zunächst definiert.

2.3.1 Lehr-/Lernszenarien

Innerhalb von CLIX[®] als Lernumgebung werden unterschiedliche Lehr-/Lernszenarien mit Hilfe unterschiedlicher Methodenabläufe umgesetzt. Zu diesen Abläufen zählen:

- Bildungsmaßnahmen bzw. Bildungsprogramme,
- Lehrveranstaltungen bzw. Kurse sowie
- Communities (virtuelle Gemeinschaften).

Während Lehrveranstaltungen ein eher expositorisches (strukturiertes) Lernen unterstützen, wird in virtuellen Gemeinschaften oft mit tutorieller Begleitung vorrangig der informelle Wissensaustausch und das explorative (entdeckende) Lernen unterstützt. Bildungsmaßnahmen sind ebenfalls als strukturierte Lehr-/ Lernszenarien zu verstehen, in denen eine definierte Anzahl von Lehrveranstaltungen zusammengefasst werden (bspw. ein Studiengang).

CLIX[®] unterstützt die Verwaltung von verschiedenen Lehrveranstaltungsvorlagen. Sie dienen als Vorlage im Sinne einer Blaupause, um daraus verschiedene Einzellehrveranstaltungen definieren zu können. Hierbei müssen bestimmte Elemente, welche in zahlreichen Lehrveranstaltungen eingesetzt werden sollen, nicht für jede einzelne definiert werden, sondern werden in der Lehrveranstaltungsvorlage bestimmt. Somit können schnell verschiedene Lehrveranstaltungsva-

rianten erstellt werden, welche alle von einer gemeinschaftlichen Lehrveranstaltungsvorlage abstammen. Die einzelnen Beschreibungsattribute, Inhalte (Komponenten) und der zugehörige Lernpfad (Lernlogik) können lehrveranstaltungsspezifisch angepasst werden, so dass die Vorgaben der Lehrveranstaltungsvorlage nicht obligatorisch sind.

2.3.2 Organisationseinheiten und Rollen

Personen nehmen verschiedene Rollen in den oben genannten Prozessen ein. So wird unter anderem zwischen Lernenden, Trainern, Tutoren, Autoren, Redakteuren und Administratoren unterschieden. Mit CLIX[®] werden all diese Personen in ihren Rollen unterstützt. Neben der Verwaltung individueller Benutzer bietet CLIX[®] hierzu ein umfangreiches Gruppenmanagement. Der Administrator kann die Benutzer in beliebig viele Gruppen einordnen, die mit eindeutigen Rechten versehen werden. Zwischen einzelnen Gruppen können hierarchische Gliederungen mit vererbbaaren Strukturen aufgebaut werden.

2.3.3 Komponenten

Bildungsmaßnahmen und Lehrveranstaltungen lassen sich als strukturierte Lehr-/Lernveranstaltungen mit granular zu beschreibenden Handlungssträngen klassifizieren. Bei Bildungsmaßnahmen werden die Handlungsstränge durch nacheinander zu durchlaufende Komponenten abgebildet. Handlungsstränge von Lehrveranstaltungen und Lehrveranstaltungsvorlagen werden ebenfalls durch einzelne Unterkomponenten repräsentiert. Hierzu zählen in CLIX[®] Strukturierungsordner, Medien, Services, Tests und Feedbacks.

Eine Lehrveranstaltung in CLIX[®] ist somit nicht gleichzusetzen mit einem einzelnen Web-basierten Training oder einem einzelnen Seminar. Erfolgreiches eLearning erfordert vielmehr einen Mix aus verschiedenen Medien wie Diskussionsforen, WBTs, Experten- oder Gruppenchats, Präsenzveranstaltungen und Lernerfolgtests. In CLIX[®] sind Bildungsplaner vollkommen frei in der Wahl der Mittel, aus denen Sie einen Kurs zusammenstellen. Entscheidend ist allein das optimale didaktische Einsatzszenario der zur Verfügung gestellten Komponenten. Sie werden im nachfolgenden erläutert.

Über CLIX[®] können den Lernenden alle gängigen Arten digitaler Medien zur Verfügung gestellt werden. Dies wird ermöglicht durch die Einhaltung von internationalen Standards für Lerninhalte. Folgende Medien stehen in CLIX[®] in Einzelnen zur Verfügung: Animation, Audio, Bild CBT, Dokument, FAQ, Glossar, Linkliste, Präsentation, Tutorial, Video, WBT, AICC-WBT, SCORM-WBT.

Tests sind in CLIX[®] Module, die eigenständig in der Plattform verwaltet werden. Das integrierte Test-Management von CLIX[®] gestattet es, bei der Zusammenstellung von Tests auf einen Pool von vorhandenen Aufgaben zurückzugreifen. Somit müssen Aufgaben, welche bei der Generierung von unterschiedlichen Testszenarien mehrmals verwendet werden, nicht immer wieder neu angelegt werden. Es gibt verschiedene Aufgabentypen, so dass je nach Komplexität der Lernziele adäquate Fragen entworfen werden können, von Multiple Choice und Lückentexten bis hin zu offenen Fragen und Fallstudien. Werden Tests zum Zweck der Zertifizierung durchgeführt, können Zertifikate in der Lernhistorie des Lerners abgelegt werden.

Es gilt jedoch nicht nur den Erfolg der Lerner zu evaluieren, sondern auch die Qualität des gesamten Bildungsangebotes zu überprüfen. Die Qualität und Bedarfsgerechtigkeit der angebotenen Bildungsinhalte sowie die Zufriedenheit der Lernbetreuung kann durch Feedback-Formulare erhoben werden. Im CLIX[®] Feedback-Management können - ähnlich dem Test-Management - Fragen erstellt und bearbeitet und zu Fragebögen zusammengestellt werden.

Über die verschiedenen synchronen und asynchronen Kommunikationsdienste, den so genannten Services in CLIX[®], kann der Lernbegleiter (bspw. Tutor) die Lerner auf vielfältige Weise unterstützen und betreuen. Mögliche Services in CLIX[®] sind Foren, Chats, Nachrichten, Pinnwände, Hinweise auf Präsenzveranstaltungen, Dokumentenarchive. Der Moderator hat dabei die Funktion, die Diskussion zu steuern.

Zur zeitlichen Koordination der Gruppenarbeit werden Termine in einem gemeinsamen Kalender verwaltet. Dies können sowohl Termine für virtuelle Meetings als auch für die Fertigstellung von Gruppenaufgaben sein. So erhalten die Teilnehmer beim Eintritt in ihre virtuelle Lernwelt sofort neueste Nachrichten und sehen die aktuellen Termine ihres Kurses.

2.3.4 Gestaltung eines Curriculums

Curricula werden über Lernziele, Lerninhalte, Lernpfade, Handlungsanweisungen für Lerner bzw. Tutoren und Methoden zur Lernfortschrittskontrolle inklusive Regeln und Zeitparameter definiert. Zur Umsetzung der didaktisch-methodischen Struktur werden unstrukturierte und strukturierte Lernangebote in CLIX[®] systemseitig unterstützt. In unstrukturierten, lernerbestimmten Lernszenarien bestimmen die Lerner die Reihenfolge der Lernschritte selbst, der Tutor hat nur geringe Steuerungschancen. Regelbasierte Lernpfade hingegen sind strukturiert. Hier kann eine Steuerung durch instruktionale Ereignisse, Taktung und/oder Test und Rückmeldung erfolgen.

Adaptivität bezeichnet in einer spezifischen (multimedialen) Lernumgebung die Anpassungsfähigkeit des Lernpfades und der Lernangebote an die Bedürfnisse unterschiedlicher Lerner. Von Adaptivität eines Systems spricht man also, wenn sich das Lernarrangement in Abhängigkeit vom individuellen Lernfortschritt dynamisch anpasst.

Die Definition eines Lernpfades für ein Curriculum in Abhängigkeit mit der gewünschten Lernform umfasst aus technischer Sicht die Struktur bildenden Komponenten Strukturierungsgrad, Zuordnung des Regelwerkes, Adaptivität und Maßnahmen, kurz einer Lernlogik.

Mit Hilfe hierarchischer Kursstrukturen besteht in CLIX[®] die Möglichkeit, Inhalte innerhalb von Kursen nach thematischen, planerischen oder zeitlichen Kriterien zusammenzufassen und anzubieten. Damit steht bereits während der Kursdesignphase eine umfassende Funktionalität zur Gliederung von Inhalten - auch bei einer großen Anzahl - zur Verfügung.

Die Abbildung von Lernzielen, Lernpfaden, Handlungsanweisungen für Lerner und Tutoren, Methoden zur Lernfortschrittskontrolle inklusive Regeln und Zeitparameter - also das Curriculum eines Kurses - erfordert eine softwaretechnische Abbildung. In CLIX[®] steht Ihnen hierfür die Lernlogik zur Verfügung. Die Definition von Lernpfaden wird hierbei durch die Nutzung von Selbst- und Fremdsteuerungsmechanismen in den Ablauf des Curriculums integriert.

Zusammengefasst repräsentiert die Lernlogik den zeitlich-logischen Ablauf eines Lehr-/Lernszenarios und reagiert auf Zustandsänderungen. Ein Startzustand definiert, ob eine nachfolgende Komponente abgearbeitet bzw. durchgeführt werden kann. Ergebnis der Durchführung ist eine Zustandsänderung, die bei positiver Ausprägung in einen neuen Startzustand überleitet. Bei negativer Ausprägung bleibt der vorherige Startzustand entweder erhalten oder führt zu einem negativen Endzustand.

3 ARIS2CLIX mit IMS LD

Mit dem Ansatz „ARIS2CLIX“ sollen die zuvor vorgestellten Ansätze und Werkzeuge integriert werden. In einem Gesamtscenario können so Lehr-/Lernprozesse definiert, mit Hilfe des ARIS-Konzeptes abgebildet, in die Notation von IMS-LD überführt und in CLIX[®] ausgeführt werden.

Um die oben aufgeführten Basiskonzepte mit den Bestandteilen von CLIX[®] in Einklang bringen zu können und die Potentiale bewerten zu können, werden die einzelnen Bestandteile zunächst gegenübergestellt. Dies erfolgt in Tabelle 1.

IMS-LD Komponente	Relevantes CLIX-Objekt	Semantische Vergleichbarkeit zu ARIS
Person	Benutzer	Person
Role Staff Learner	Gruppe / Rolle Tutor / Administrator / ... Lerner	Organisationseinheit / Rolle Rolleninstanz Rolleninstanz
Environment; Learning Objects; Services	Medien, Services, Tests, Feedbacks	Datenobjekte
Activity Learning Activity Support Activity	Arbeitsschritte innerhalb eines Curriculums	Funktion
Activity Structure	Kurshierarchie	Kontrollfluß
Learning Objective	Attribut „Lernziel“	Outputleistung
Notifications	Systemnachrichten	Nachrichtenobjekte
Conditions	Zustände der Lernlogik	Ereignisse in Kombination mit Operatoren
Properties	Zustände der Lernlogik (Werte, auch Werte die die Lernlogik beeinflussen)	„Ergebnisereignisse“
Methods Plays Acts	Kurse, Bildungsmaßnahmen, Communities	Prozesse bzw. Teilprozesse

Tab. 1: IMS-LD 2 CLIX 2 ARIS-Mapping (in Anlehnung an [Mart06])

Zur Reduktion der Komplexität enthält die Tabelle nur die Hauptelemente aus den jeweiligen Bereichen. Sie stellen die Gesamtmenge der für das nachfolgende Beispiel notwendigen Konstrukte dar.

Der Ablauf des oben ausgeführten Beispiels „Grundlagen des Projektmanagements“ im Sinne von einzelnen Arbeitsschritten wird im Folgenden als Umsetzung in CLIX[®] visualisiert.

Im Learning Management System CLIX[®] kann ein solcher Kurs inklusiver der notwendigen Lernkomponenten, Prozesse und Lernlogik-Regeln abgebildet werden. Im herkömmlichen Sinne erstellt der Kursadministrator zunächst die notwendigen Komponenten – im Beispiel die einzelnen Tests, Lernmaterialien sowie den Feedbackbogen, führt diese in einem CLIX[®]-Kurs zusammen und ordnet diesem Kurs die einzelnen Teilnehmer in ihren kursspezifischen Rollen (Lerner, Tutor) zu. Ebenso definiert wer das Curriculum in zeitlich-logischer Reihenfolge und definiert die notwendigen Lernlogik-Regeln.

Um dem Kursadministrator die Erstellungsarbeit zu erleichtern, wird ihm mit dem ARIS Business Architect[®] der IDS Scheer AG (vgl. Abbildung 3) und einem CLIX[®]-spezifischen Modellierungsfiler ein Werkzeug an die Hand gegeben, mit dessen Hilfe er die didaktische Komposi-

tion des Lehr- und Lernszenarios sowie die tatsächliche Umsetzung in einem CLIX®-Kurs kombinieren kann.

Im weiteren Verlauf kann der Bildungsverantwortliche als „Process Owner“ einzelne Modelle nach Erstellung in einem Repository ablegen. Ebenso kann ihm an gleicher Stelle eine Bibliothek an Referenzprozessen zur Verfügung gestellt werden die ihm die Modellierung erleichtern.

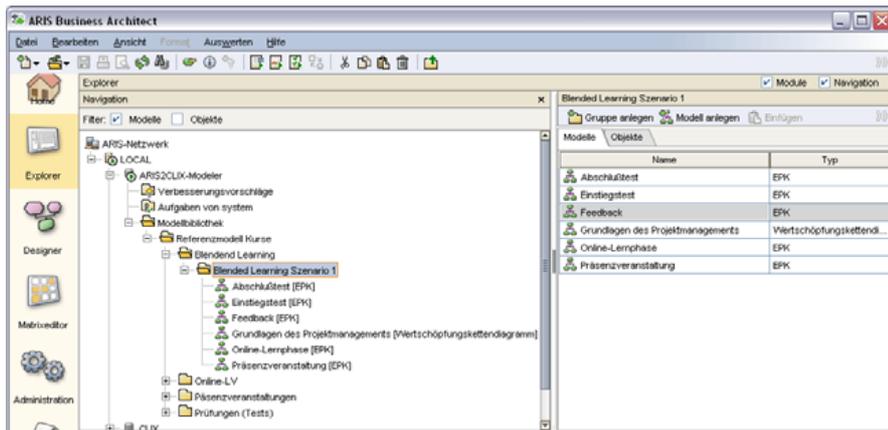


Abb. 3: CLIX® Kursbibliothek im ARIS Business Architect®

Die Erstellung eines solchen Modells soll anhand des oben beschriebenen Beispiels verdeutlicht werden. Der grundsätzliche Aufbau des im Beispiel beschriebenen Kurses kann in Form eines Wertschöpfungskettendiagramms dargestellt werden (vgl. Abbildung 4). Die angedeuteten eEPK-Symbole neben den einzelnen Wertschöpfungskettenkonstrukten symbolisieren dahinter liegende eEPKs, die die einzelnen Komponenten beschreiben.



Abb. 4: WSK-Diagramm als Kursübersicht

Die hinterlegten eEPKs geben den logischen Ablauf der einzelnen Komponenten wieder. Der im Modellierungsbeispiel der Abbildung 2 dargestellte Prozess kann somit in kleinere Teilprozesse zerlegt werden, um insbesondere bei komplexen Lehr- und Lernszenarien die Übersichtlichkeit für den Bildungsverantwortlichen zu erhalten.

Wird der modellierte Kurs in CLIX® umgesetzt, erhält der Lerner nach dem Start des CLIX® Kurses den im folgenden Screenshot der Abbildung 5 umgesetzten Lehrplan.

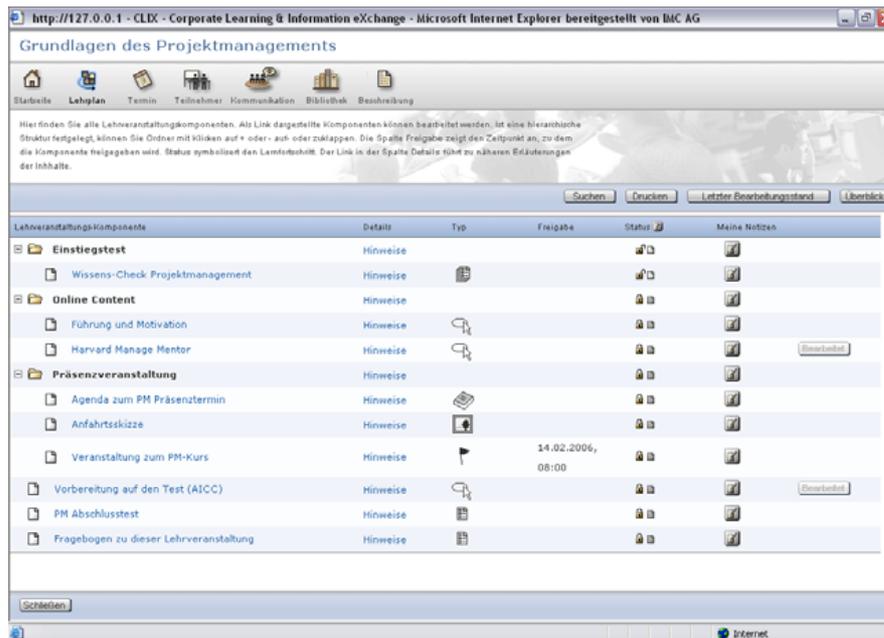


Abb. 4: Lehrplan „Grundlagen des Projektmanagements“

Das Curriculum ist hierbei grundsätzlich mit Ordern und Unterordnern strukturiert. Durch An-klicken der einzelnen Objekte (Komponenten) innerhalb der Ordner besteht die Möglichkeit die Lerninhalte bzw. Tests oder Feedbackbögen aufzurufen. Die Spalte „Typ“ gibt einen Überblick über die Art der Komponenten. Der Stand im zeitlich-logischen Prozess bzw. die momentan geltenden Zustände der Lernlogik-Regeln werden mit Hilfe der Spalte Status visualisiert. Im oben abgebildeten Screenshot wird bspw. angezeigt, dass sich der Kurs noch im Anfangszu-stand befindet. Der Einstiegstest ist frei gegeben, alle anderen Komponenten sind noch nicht zugänglich (ersichtlich durch das nicht geöffnete Schlosssymbol).

Der entscheidende Schritt vom in ARIS modellierten Lehr- und Lernprozess in einen in CLIX[®] abgebildeten Kurs erfolgt mit Hilfe der IMS-LD Spezifikation. Die dazugehörige XML-Notation wurde in einem gemeinschaftlichen Projekt von IDS Scheer AG und imc AG mit den Spezifika des ARIS Business Architect[®] und CLIX[®] in Einklang gebracht. Dies führt zu einem Gesamtszenario, welches in Abbildung 6 visualisiert ist.

Die grundsätzliche Notwendigkeit der Definition eines Curriculums ist Ausgangspunkt der Überlegungen eines Bildungsverantwortlichen. Die Umsetzung des Lehr-/ und Lernszenarios wird dann im ersten Schritt im ARIS Business Architect[®] mit Hilfe des CLIX-spezifischen Mo-dellierungsfilters modelliert. Der ARIS Business Architect[®] ermöglicht dann einen IMS-LD-konformen Export der Modellierungsergebnisse in Form einer XML Datei.

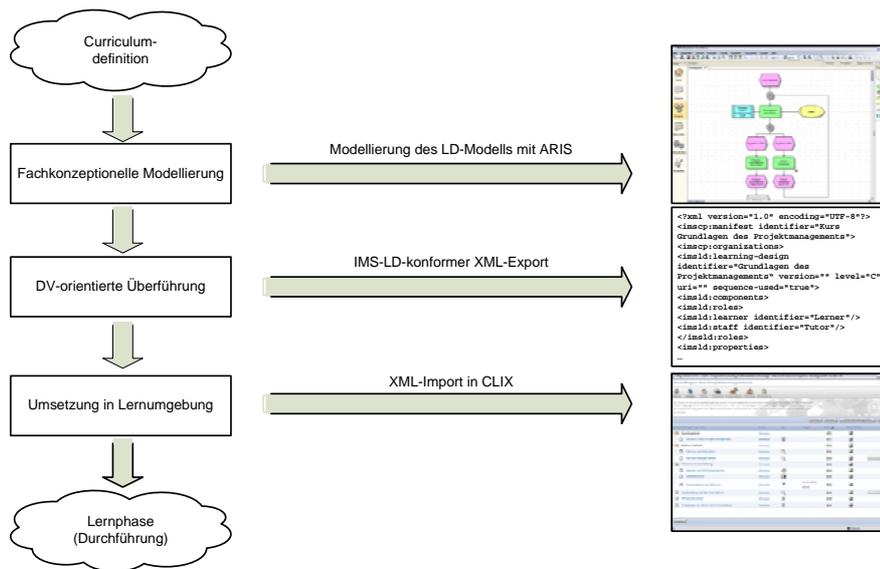


Abb. 6: Umsetzungsszenario

Der Bildungsverantwortliche loggt sich als Administrator in seiner CLIX[®] Lernumgebung ein und importiert die XML-Datei. Mit Hilfe einiger Abfragen zur Feinjustierung in Form eines Wizards kann CLIX[®] automatisch eine Lehrveranstaltung (bspw. einen Kurs) erstellen und dem Bildungsverantwortlichen zur Verfügung stellen. Ist die Lehrveranstaltung lediglich als eine Vorlage für später zu erstellende Instanzen gedacht, sind die Feinjustierungsschritte auf ein Minimum begrenzt. Soll CLIX[®] mit Hilfe des XML-Files eine (Lehrveranstaltungs-)Instanz generieren, hat der Bildungsverantwortliche zusätzliche Angaben zu tatsächlich zuzuweisenden Komponenten, Benutzern usw. zu machen.

4 Zusammenfassung und Ausblick

Mit dem hier vorgestellten Ansatz des modellbasierten Learning Design erhalten Bildungsprogrammplaner in Unternehmen und Hochschulen die Möglichkeit, zukünftig auf ein hilfreiches, visuelles Instrument zurückzugreifen, das den Vorgang vereinfacht und erheblich effizienter macht – mit ähnlichen Effekten wie es die Geschäftsprozessmodellierung für die Analyse von Prozessen ist. Bildungsprogrammplaner können dann grafisch neue ihre Curricula und Kurse entwerfen. Die technische Integration mit LMS Systemen ermöglicht die Übernahme in die Lernumgebung und dortige Ausführung. Sie wird von den Autoren im Rahmen des laufenden EU-Projektes PROLIX durchgeführt (vgl. PROL06). Obgleich noch zahlreiche Forschungsfrage-

gen zu klären sind ist das Ziel eine praxisnahe Umsetzung des Gesamtkonzeptes um einen wertvollen Beitrag für das Learning Design der Zukunft zu leisten.

Literaturverzeichnis

- [CENoJ] Europäisches Institut für Normung (CEN): CEN - Learning Technologies Standards Observatory. <http://www.cen-ltso.net/Users/book/CEN-LTSO-eng.pdf>, Abruf am 2006-07-18.
- [Groh06] Grohmann, Guido: Learning Management. Eul, Lohmar 2006.
- [LTSC05] IEEE Learning Technology Standards Committee: WG 12: Learning Object Metadata. URL <http://ltsc.ieee.org/wg12/>, 2005, Abruf am 2006-07-18.
- [JeCu05] Jeffery, Ann; Currier, Sarah: What is ... IMS Learning Design? In: CETIS Centre For Educational Technology (Hrsg.): CETIS standards briefing series, University of Bolton, Bolton, 2005.
- [Kope01] Koper, Rob: Modeling units of study from a pedalogical perspective: the pedalogical meta-model behind EML. Educational Technology Expertise Center, Open University of the Netherlands, Heerlen 2001.
- [KoOA03] Koper, Rob; Olivier, Bill; Anderson, Thor: IMS learning Design information model. http://www.imsglobal.org/learningdesign/ldv1p0/imslid_infov1p0.html, 2003-01-20, Abruf am 2006-07-18.
- [KrMZ05] Kraemer, Wolfgang; Milius, Frank; Zimmermann, Volker: Von WINFOLine zum Corporate Learning Management. In: Information Management 20 (2005), Nr. Sonderausgabe, S. 50-67.
- [LeHM95] Lehner, Franz; Hildebrand, Knut; Maier, Ronald: Wirtschaftsinformatik: Theoretische Grundlagen. Hanser, München 1995.

- [LuWS02] Lucke, Ulrike; Wiesner, André; Schmeck, Hartmut: XML: Nur ein neues Schlagwort? - Zum Nutzen von XML in Lehr- und Lernsystemen. In: it + ti - Informationstechnik und Technische Informatik 44 (2002), Nr. 4, S. 211-216.
- [Mart06] Geschäftsprozessorientiertes Learning Management: Konzept und Anwendung. Saarbrücken, Universität des Saarlandes 2006.
- [PROL06] PROLIX CONSORTIUM, Requirements Analysis Report, <http://www.prolixproject.org/>, September 2006.
- [QuSi03] Quemada, Juan; Simon, Bernd: A Use-Case Based Model for Learning Resources in Educational Mediators. In: Educational Technology and Society 6 (2003), Nr. 4, S. 149-163.
- [Sche02] Scheer, August-Wilhelm: ARIS - Vom Geschäftsprozess zum Anwendungssystem. 4., durchges. Aufl. Springer, Berlin [u.a.] 2002.

Einflussfaktoren für den Einsatz von Customer Focused E-Learning

Qualitative Erhebung relevanter Faktoren aus Anbietersicht

Kai-Uwe Götzelt, Susanne Robra-Bissantz

Lehrstuhl Wirtschaftsinformatik II
Friedrich-Alexander-Universität Erlangen-Nürnberg
90403 Nürnberg
{goetzelt,robra}@wi2.wiso.uni-erlangen.de

Abstract

Unternehmen stellen zunehmend auch Kunden E-Learning-Angebote zur Wissensvermittlung in verschiedenen Phasen des Customer Buying Cycle zur Verfügung. Der Einsatz von Customer Focused E-Learning (CFEL) ist dabei nicht für jedes Unternehmen Erfolg versprechend. Eine strategische Analyse für CFEL identifiziert relevante Faktoren, die zur Beurteilung der Eignung von CFEL für ein Unternehmen herangezogen werden können. Dazu werden in einer qualitativen Erhebung unter Anbietern von CFEL und Beratungsunternehmen, die an der Planung und Konzeption von CFEL-Angeboten beteiligt waren, Einflussfaktoren für einen Erfolg versprechenden Einsatz von CFEL aus Anbietersicht herausgearbeitet.

1 Motivation und Ziel

Eine Reihe sozio-ökonomischer und technologischer Entwicklungen führen dazu, dass die Aneignung von Wissen ständig notwendig und gefordert ist [Stra05]. Relevantes Wissen wird dabei zunehmend in Echtzeit und auf Abruf abseits von formalen Lehr-Lernangeboten bereitgestellt und durch informelles Lernen im Rahmen von Alltagstätigkeiten erworben [Kirch00; Resn05]. Vor diesem Hintergrund stellen Unternehmen verstärkt auch ihren Kunden E-Learning-Angebote zur Befriedigung von im Kaufprozess entstehenden Wissensbedarfen in verschiedenen Phasen des Customer Buying Cycle zur Verfügung. Customer Focused E-Learning (CFEL) ist gekennzeichnet durch Lernangebote, die sich an Endkunden des

initiierenden Unternehmens richten, intrinsisch motiviertes Lernen im Rahmen des Customer Buying Cycle durch Informations- und Kommunikationstechnologien unterstützen bzw. ermöglichen und der Erreichung von Unternehmenszielen dienen.

CFEL eröffnet Unternehmen die Chance, durch die Bereitstellung eines innovativen E-Services einerseits einem steigenden Wettbewerbsdruck zu begegnen und andererseits neue Zielgruppen für E-Learning zu erschließen. Der Einsatz von CFEL ist allerdings nicht für jedes Unternehmen Erfolg versprechend. Zahlreiche erfolgreiche Beispiele von kundenorientierten E-Learning-Angeboten stehen einer ebenso großen Anzahl von Misserfolgen gegenüber, bei denen bereits wenige Monate nach der Einführung von CFEL das Angebot wieder eingestellt wird. Relevante Einflussfaktoren für einen Erfolg versprechenden Einsatz von CFEL wurden bisher in einer Feldstudie aus Kundensicht untersucht [RoGo05].

Ziel des Beitrags ist es, anhand einer qualitativen Erhebung unter Anbietern von CFEL und Beratungsunternehmen, die an der Planung und Konzeption von CFEL-Angeboten beteiligt waren, Einflussfaktoren für einen Erfolg versprechenden Einsatz von CFEL aus Anbietersicht herauszuarbeiten.

2 Strategische Analyse für CFEL

2.1 Betrachtungsdimensionen und Einflussbereiche

Als Einflussfaktoren sind relevante Informationen zu bestimmen, die für die Beurteilung der Eignung von CFEL für ein Unternehmen eine Aussage über einen Erfolg versprechenden Einsatz zulassen. In Anlehnung an die Ideenbewertung bei Dienstleistungsinnovationen lassen sich die heranzuziehenden Informationen aus einer strategischen Analyse ableiten [ScDa06]. Mithilfe der strategischen Analyse werden gegenwärtige Situation und zukünftige Entwicklungen in der internen und externen Umwelt des Unternehmens betrachtet [StSc93, S. 155], die für den Einsatz von kundenorientiertem E-Learning relevant sind. Die externe Umwelt wird durch Rahmenbedingungen geprägt, die von außen auf ein Unternehmen einwirken und nicht direkt beeinflussbar sind [Hung00, S. 73]. Im Gegensatz dazu beschreiben Einflussfaktoren, die in der internen Umwelt angesiedelt sind, die Kompetenzen des Unternehmens, die für geeignete Maßnahmen oder Handlungen zur Verfügung stehen [Hung00, S. 98]. Da die Betrachtung unabhängig von spezifischen Unternehmen und Branchen erfolgen

soll, ist eine Bestimmung von Faktoren, die in der relevanten Makroumwelt eines Unternehmens angesiedelt sind und sich aus der Wettbewerbsstruktur einer bestimmten Branche durch potenzielle Handlungen der Akteure ergeben, nicht möglich.

Für eine detaillierte Betrachtung der engeren externen und internen Umwelt für CFEL wird daher das Konstrukt des Marktes verwendet, auf dem CFEL angeboten wird. Märkte lassen sich durch Nachfrager mit ihren speziellen Bedürfnissen, angebotene Güter und Leistungen als nutzenstiftende Eigenschaftsbündel sowie Anbieter mit Instrumenten der Nutzungsgestaltung charakterisieren [Baue94, S. 710]. Daraus ergeben sich die in Abbildung 1 dargestellten Betrachtungsdimensionen und Einflussbereiche. Die engere externe Umwelt wird durch die Kunden des Unternehmens, als potenzielle Nachfrager von CFEL-Angeboten, und die auf dem Markt offerierten Kernleistungen betrachtet. Die interne Umwelt für CFEL wird in der Dimension Anbieter untersucht. Bei der Betrachtung der Einflussfaktoren aus Anbietersicht, liegt der Fokus auf einer detaillierten Analyse von Faktoren in der letzten Dimension sowie der Erweiterung und Ergänzung der bereits aus Kundensicht erarbeiteten Einflussfaktoren [RoGo05].

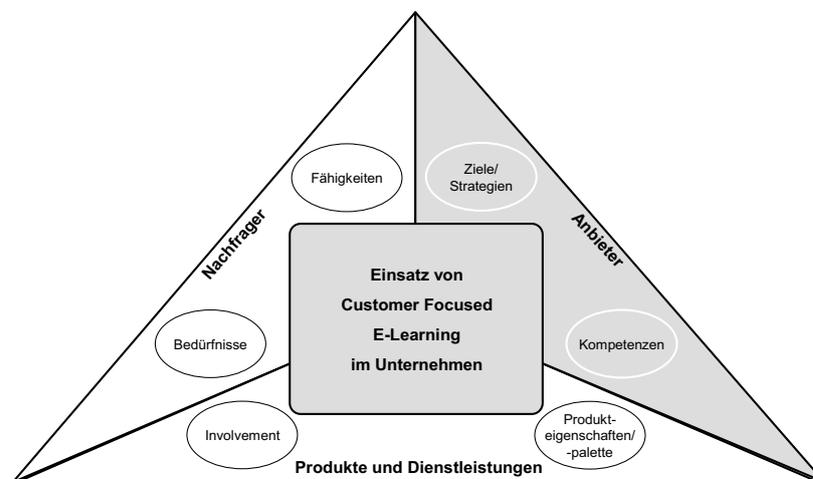


Abbildung 1: Betrachtungsdimensionen und Einflussbereiche für den Einsatz von CFEL

2.2 Untersuchungsansatz

Für die Untersuchung werden in den Betrachtungsdimensionen zunächst relevante Einflussbereiche auf Basis einer theoretischen Analyse näher spezifiziert und abgegrenzt. Im Anschluss werden Einflussfaktoren in den einzelnen Bereichen im Rahmen einer qualitativen

Befragung erhoben und die gewonnenen Daten analysiert, verdichtet und benannt [StCo90, S. 61-74]. Dabei wird das Vorgehen ebenfalls durch eine Literaturanalyse unterstützt.

Als Erhebungsmethode wird das freie und qualitative Interview (Experteninterview) verwendet [BEE96, S. 95]. Wie in der qualitativen Marktforschung üblich, ist die Stichprobengröße klein. Es wurden acht deutsche Unternehmen, die entweder ein gut konzipiertes CFEL-Angebot anbieten oder angeboten haben (comdirect Bank AG, Deutsche Börse AG, Dresdner Bank AG, eBay AG, SAP AG, Sony GmbH, Cortal Consors AG) sowie die zwei Dienstleistungsunternehmen M.I.T und Bassier, Bergmann und Kindler (BB-K), die mehrere kundenorientierte E-Learning-Projekte mitgestaltet haben, befragt. Die Erhebung erfolgte telefonisch im Zeitraum vom 25.01. bis zum 15.02.2005. Die Befragung war halb-standardisiert. Zur Sicherung der Vergleichbarkeit wurde ein sog. Interview-Leitfaden zugrunde gelegt [BEE96, S. 96], der den groben Befragungsrahmen vorgibt, aber entsprechend dem Gesprächsverlauf individuell angepasst wird (elastische Gesprächsführung). Um dem explorativen Charakter der Befragung gerecht zu werden, sind, bis auf zwei, alle Fragen offen formuliert [BEE96, S. 289].

3 Einflussfaktoren des Einsatzes von CFEL

3.1 Anbieterbezogene Faktoren

3.1.1 Verfolgte Ziele und Strategien

Die Einflussbereiche in der Dimension „Anbieter“ beschreiben allgemein, ob ein Unternehmen in der Lage ist, auf Zustände und Entwicklungen in der externen Umwelt zu reagieren sowie Maßnahmen oder Handlungen erfolgreich umzusetzen. Von den Ergebnissen der Analyse und Prognose der Unternehmensumwelt ist auf der Ebene der strategischen Planung für das Gesamtunternehmen abhängig, welche Ziele in Form von gewünschten Soll-Zuständen angestrebt werden und welche Strategie verfolgt wird, um diese zu erreichen. Auf der Ebene der operativen Planung gilt es geeignete Maßnahmen zu wählen, mit denen die instrumentalisierten Ziele realisiert werden können [NDH97, S. 876]. Kundenorientiertes E-Learning als Dienstleistungsinnovation muss dabei konform zu den Zielen und Strategien des Gesamtunternehmens bzw. des verantwortlichen Geschäftsbereich sein. Alle Interviewpartner stimmen der Aussage zu, dass CFEL auf die Unternehmensziele und -strategie auszurichten ist und ordnen diesem Aspekt eine große Bedeutung für einen erfolgreichen Einsatz zu.

Im Unterschied zu klassischen, vom Kunden vorausgesetzten Serviceleistungen unterstützt CFEL als elektronische Dienstleistung nicht nur die Erreichung ökonomischer Zielgrößen sondern bietet als Sekundärleistung in Form eines Value-Added-Service (VAS) auch die Möglichkeit, dem Kunden zusätzlichen Nutzen im Hinblick auf die Primärleistung des Unternehmens zu stiften und so weitere Zielsetzungen zu erreichen [Laak95, S. 2]. In einer offen formulierten Frage nach den mit dem Einsatz von CFEL verfolgten Zielen, wurden die in Tabelle 1 dargestellten Zielsetzungen mit CFEL verbunden.

Anbieterbezogene Einflussfaktoren im Bereich „Ziele und Strategie“	Nennungen Anbieter	Nennungen Dienstleister	Gesamt
Differenzierung gegenüber Wettbewerbern	8	2	10
Steigerung der Kundenzufriedenheit	7	2	9
Steigerung der Kundenbindung	7	2	9
Verbesserung des Image, Branding	3	1	4
Umsatzsteigerung, Aktivierung	3	1	4
Steigerung der Kundengewinnung	3	0	3
Differenzierung der Leistungen	1	2	3
Verbesserung des Vertrauens	2	0	2
Steigerung der Cross-Selling-Rate	2	0	2
Bessere der Nutzung der Hauptleistung	1	1	2
Kostensenkung	0	1	1

Tabelle 1: Unterstützte Ziele und Strategien von CFEL

Dabei wird deutlich, dass der Differenzierung gegenüber den Wettbewerbern sowie der Steigerung der Kundenzufriedenheit und Kundenbindung eine besondere Bedeutung zukommt. Daneben wird CFEL auch eine Verbesserung der Unternehmenserscheinung und Stärkung der Marke zugeschrieben. Durch den Charakter eines VAS kann auch für CFEL allgemein festgestellt werden, dass dessen Einsatz mit den drei Primärzielen Differenzierung gegenüber dem Wettbewerb, Profilierung durch Schaffung von Kundenpräferenzen und Demonstration der Problemlösungskompetenz des anbietenden Unternehmens verbunden ist [Mann98]. Dabei können die Vermutungen von Barron für den Einsatz von CFEL [2000a] „it’s a differentiator that sets us apart from competitors, and [...] adds significant value to our product suite” [Barron 2000a] und dass CFEL andererseits die Möglichkeit „to educate customers and keep them coming back“ [Barron 2000a] bietet, bestätigt werden. Als Sekundärziele können weitere kunden-, wettbewerbs- und leistungsprogrammbezogene sowie ökonomische Zielsetzungen, die konform zu CFEL sind und deren Erreichung CFEL unterstützt, identifiziert werden.

Als Voraussetzung für Kundenbindung und Ergebnis von Kundenzufriedenheit wird die Stärkung des Vertrauens durch CFEL gesehen. CFEL ist geeignet, um Kundenloyalität zu

steigern und langfristige Kundenbeziehungen aufzubauen [Conn01, S. 4; Finn02]. Die Bedeutung von CFEL zur Unterstützung der Neukundengewinnung wird durch den mit dem Angebot oftmals verbundenen Registrierungsprozess geringer eingeschätzt. Vier der betrachteten Angebote können nur nach vorheriger Registrierung eingesehen werden. Allerdings wird auf den positiven Referenzeffekt der Stammkunden verwiesen. Diese weisen eine größere Empfehlungsneigung („Mund-zu-Mund-Propaganda“) auf, was zur Weiterempfehlung durch Privatpersonen führt, die als glaubwürdiger empfunden wird und so wesentlich zur Kundengewinnung beiträgt [Dill95, S. 42]. Die positive Beurteilung der Sekundärleistung strahlt so auf die zugehörige Primärleistung aus [Laak95]. Daneben sind auch das Image des Unternehmens und seine Marke weitere Ansatzpunkte zur Steigerung der Kundenzufriedenheit und -bindung. Das „Mehr“ an Informationen im Rahmen kundenorientierter E-Learning-Angebote erhöht die Reputation und stärkt das Vertrauen in die Leistungsfähigkeit des initiiierenden Unternehmens. Es macht das Unternehmen gegenüber Wettbewerbern glaubwürdiger und reduziert Kaufunsicherheiten beim Kunden.

Im Rahmen der leistungsprogrammbezogenen Zielsetzungen unterstützt CFEL vor allem eine Differenzierung der Leistungen innerhalb des Programms der anbietenden Unternehmen. CFEL wird als sinnvolle Ergänzung zu bestehenden Webangeboten gesehen. Daneben wird CFEL vor allem ein Potenzial zur Sicherstellung sowie Steigerung und Ausweitung der Nutzung der Hauptleistung zugeschrieben. CFEL bietet die Möglichkeit, sowohl das Produkt in Form Zusatznutzenstiftender Informationen aufzuwerten als auch den Faktor Zeit zu beeinflussen, indem es die Einführung neuer Produkte am Markt beschleunigt [Morr03, S. 104].

Im Rahmen ökonomischer Zielsetzungen wird CFEL vor allem ein Beitrag zu Steigerung des Umsatzes und Aktivierung von Kunden zugeschrieben. Daneben kommt durch ein besseres Leistungsverständnis auch der Steigerung der Cross-Selling-Rate eine Bedeutung zu. Die Möglichkeit zur Kostensenkung und damit Gewinnsteigerung wird eher gering eingeschätzt. Zwar lassen sich die bisherigen Servicekosten, z.B. für Hotlines, reduzieren, wenn der Kunde durch das Wissensangebot Fehlbedienungen oder auch Falschbestellungen vermeidet, so die Aussage eines Beratungsunternehmens. Jedoch stehen den Einsparungen in diesem Bereich der Aufwand für Realisierung und Implementierung des CFEL-Angebots gegenüber.

3.1.2 Kompetenzen

In der Kategorie Kompetenzen werden Ressourcen und Fähigkeiten betrachtet, die für einen Erfolg versprechenden Einsatz von CFEL im Unternehmen vorhanden sind oder angepasst bzw.

entwickelt werden müssen [Hung00, S. 98]. Für die Beurteilung des Einsatzes von CFEL sind angrenzende Funktionalbereiche, Prozesse und IT-Systeme relevant, die einen Bezug zur elektronischen Kundenschnittstelle oder E-Learning aufweisen [RoGo05, BBS01]. Der Aussage, dass CFEL insbesondere auf angrenzende Funktional- und Prozessstrategien sowie IT-Systeme ausgerichtet und damit abgestimmt sein muss, stimmen alle Gesprächspartner zu. Dies zielt auf die Bereitstellung und Integration nutzbarer und notwendiger Kompetenzen im Rahmen des weiteren Service Engineering ab [Meir01, BuMe01]. Eine Übersicht zu den innerbetrieblichen Einflussfaktoren im Bereich Kompetenzen für CFEL gibt Tabelle 2.

Anbieterbezogene Einflussfaktoren im Bereich „Kompetenzen“	Nennungen Anbieter	Nennungen Dienstleister	Gesamt
Integration angrenzender Funktionen, Prozesse und Systeme	8	2	10
Integration in bestehendes Geschäftsmodell	7	2	9
Ausgeprägte webbasierte Kundenschnittstelle	7	2	9
E-Learning-Erfahrungen	5	1	6
Schulung von Vertrieb und Service	5	0	5
E-Learning-Angebote für externe Zielgruppen	2	1	3
Nutzung für interne E-Learning-Angebote	3	0	3
Zentrale Erfassung von Kundendaten	2	1	3
Gesamtqualifizierungsstrategie	0	2	2
Vorhandene Lerninhalte/Anwendungen	1	0	1
Etablierung eines neuen Geschäftsmodells	1	0	1

Tabelle 2: Kompetenzen für CFEL

Auch im Rahmen der zuvor offen gestellten Frage nach den innerbetrieblichen Voraussetzungen für einen Erfolg versprechenden Einsatz von CFEL, wird die Integration angrenzender Kompetenzen von allen Befragten genannt. Dabei wird betont, dass CFEL i.d.R. als Customer Service eingesetzt wird und durch den Charakter einer Sekundärleistung in das Gesamtgefüge des Unternehmens entsprechend integriert werden muss und angrenzende Kompetenzen zu nutzen sind, um einen Erfolgsbeitrag leisten zu können. Bei genauerem Hinterfragen wird jedoch deutlich, dass dies in der Praxis häufig keine Anwendung findet und kaum Berührungspunkte zu anderen Funktionen, Prozessen und Systemen im Unternehmen geschaffen werden. Dementsprechend werden auch nur wenige direkt nutzbare Fähigkeiten und Ressourcen des Unternehmens als Einflussfaktoren für einen Erfolg versprechenden Einsatz von CFEL genannt. Dies bedeutet erhebliche Effizienz- und Effektivitätsverluste, die wirtschaftlichen Erfolg sowie dauerhafte Erfolgssicherung und Nachhaltigkeit des Angebots in Frage stellen können. Dementsprechend wird bei der Frage, was die Befragten aus der heutigen

Perspektive bei dem Einsatz von CFEL beachten und verbessern würden, die Verbindung zu anderen Bereichen und unternehmensinterne Zusammenarbeit viermal genannt.

Obwohl die befragten Unternehmen den Einsatz von CFEL durchweg positiv beurteilen und ihre Investitionen jederzeit wiederholen würden, kann darüber hinaus festgestellt werden, dass sich die Verantwortlichen vorab kaum Gedanken über die Integration von CFEL in ein Geschäftsmodell machen, das die gegenwärtige und zukünftige Finanzierung sichert. Als Customer Service wird CFEL lediglich additiv einer Abteilung, i.d.R. Marketing (5 Nennungen), Schulung/Personalentwicklung (2 Nennungen) oder Produktmanagement (1 Nennung), zugeordnet, die das notwendige Budget bereitstellt. Erlöse sowie Wert- und Zielbeitrag werden kaum berechnet und sind vielfach durch die gegebenen Strukturen in den Unternehmen nicht zurechenbar. Es wird lediglich auf Basis positiver Kundenresonanz auf die Erreichung der gesetzten Ziele geschlossen. Der Nachweis eines Ziel- und Wertbeitrags wird im Nachhinein von allen Unternehmen als notwendig eingeschätzt. Da CFEL als Customer Service den besonderen Charakter eines VAS aufweist, ist dies für die Sicherung der Nachhaltigkeit und des Erfolgs der Sekundärleistung besonders wichtig. In diesem Zusammenhang ist auch die Nennung von vorhandenen Instrumenten zur zentralen Erfassung von Kundendaten zu sehen, die eine Voraussetzung zur Messung des Erfolgs ist. Offensichtlich wird die Problematik auch am Beispiel eines befragten Anbieters, dessen von Kunden als sehr gut eingeschätztes CFEL-Angebot auf Grund des fehlenden Nachweises im Zuge von Umstrukturierungs- und Rationalisierungsmaßnahmen im Unternehmen eingestellt wurde. Der Fokus liegt dabei weniger auf der Etablierung eines eigenen Geschäftsmodells für CFEL, das z.B. durch die gleichzeitige Bereitstellung des Angebots für andere Zielgruppen oder den Verkauf an Partner die Finanzierung sicherstellt. Diese Option ist vor allem von den zu vermittelnden Inhalten und der Art des Lernangebots abhängig und konnte nur bei einem der befragten Unternehmen realisiert werden. Vielmehr liegt das Hauptaugenmerk auf der Integration von CFEL in bestehende Geschäftsmodelle, wobei als wesentliche Ansatzpunkte die Einbindung in das Geschäftsmodell der Primärleistung des Unternehmens oder in bestehende Kundenservice- bzw. E-Commerce-Angebote gesehen werden.

Dementsprechend wurde auch der Umfang der Geschäftstätigkeit des Unternehmens im Internet, insbesondere eine ausgeprägte webbasierte Kundenschnittstelle, als entscheidend für den Einsatz von CFEL beurteilt. Neben der Voraussetzung eines frequentierten Webangebots zur Nutzung von CFEL durch die Kunden, wurde dabei auch das Potenzial gesehen, das

vorhandene Webangebot attraktiv aufzuwerten. Wird vom Unternehmen der gesamte Transaktionsprozess mit den Kunden elektronisch unterstützt, bietet CFEL die Möglichkeit, fehlende olfaktorische und taktile Reize beim Onlinekauf durch umfangreiche Informationen auszugleichen sowie Cross-Selling-Potenziale im Anschluss an den Kauf auszuschöpfen [RoGo05]. Insbesondere die befragten Dienstleister sehen bei Anbietern, die das Internet als wesentlichen Vertriebskanal einsetzen, das Potenzial von CFEL besonders hoch, um mögliche Kaufabbrüche zu reduzieren und nach dem Kauf Cross-Selling-Möglichkeiten zu schaffen.

Weiterhin von Bedeutung für den Einsatz von CFEL ist, dass im Unternehmen bereits allgemein Erfahrungen mit dem Einsatz von E-Learning gesammelt werden konnten. Dies wird u.a. als Voraussetzung für die Einschätzung des Kosten-Nutzen-Verhältnisses und für eine erfolgreiche Planung, Gestaltung und Implementierung gesehen. Dabei wird gesondert auf vorhandene Schulungsmaßnahmen für Service und Vertrieb sowie auf Erfahrungen bei der Bereitstellung von E-Learning für externe Zielgruppen verwiesen. Nach Meinung der befragten Dienstleister sollte CFEL generell Bestandteil einer Gesamtqualifizierungsstrategie im Unternehmen sein.

Allgemein werden erworbene Erfahrungen von den Befragten wichtiger eingeschätzt, als vorhandene Lerninhalte, denen eine geringe Bedeutung für einen Erfolg versprechenden Einsatz von CFEL zukommt. Die Lerninhalte werden i.d.R. eigenes für CFEL unternehmensintern entwickelt und in Zusammenarbeit mit einem externen Dienstleister umgesetzt. Mit Ausnahme eines Anbieters hat kein Unternehmen vorhandene Lerninhalte in die Gestaltung des CFEL-Angebots einbezogen. Die befragten Beratungsunternehmen weisen darauf hin, dass prinzipiell kein Unterschied bei der Einführung von E-Learning für Mitarbeiter oder Kunden zu sehen ist. In der Praxis sieht ein Großteil der Befragten die Übernahme von vorhandenen Lerninhalten allerdings als schwierig an, da oftmals ein Know-how-Gefälle zwischen Mitarbeitern und Partnern auf der einen Seite und Kunden auf der anderen Seite besteht. Darüber hinaus wird betont, dass der Fokus von Schulungsmaßnahmen in vielen Unternehmen nicht auf der Vertiefung von fachlichem Wissen, sondern v.a. auf Anwendungsprogrammen und Soft Skills liegt.

Am Beispiel eines befragten Anbieters wird allerdings deutlich, dass es durchaus sinnvoll ist, intern verfügbare Lerninhalte effizient für Kunden aufzubereiten und weiter zu entwickeln. Das Angebot befasst sich allerdings mit der Vermittlung von Basis-Medienkompetenzen. Häufiger zu beobachten ist hingegen, dass die für Kunden qualitativ hochwertig entwickelten Lerninhalte

und E-Learning-Anwendungen anschließend im Rahmen interner Schulungsmaßnahmen eingesetzt werden. Neben dem erwähnten Anbieter wurde dies noch von zwei weiteren Unternehmen realisiert. Dabei wird darauf hingewiesen, dass diese Einsatzoption am besten bereits am Anfang der Planung von CFEL zu berücksichtigen ist.

3.2 Nachfragerbezogene Faktoren

In der Dimension der nachfragerbezogenen Einflussfaktoren ist für einen Erfolg versprechenden Einsatz die Akzeptanz von CFEL bei potenziellen Nachfragern entscheidend [MaWi03, S. 13; AdPa02, S. 669]. Dabei ist neben einer tatsächlichen Steigerung des Kundennutzens durch einen Customer Service für CFEL als Lernangebot insbesondere zu berücksichtigen, dass es dann angenommen wird, wenn die bereitgestellten Dienste und Funktionalitäten auch den Fähigkeiten und Fertigkeiten der Kunden entsprechen [Simo01, S. 104]. Die Betrachtung der Akzeptanz umfasst daher sowohl Einstellungen als auch Verhalten der Kundengruppe und wird vor allem durch deren Fähigkeiten und Bedürfnisse bestimmt [BKM04, S. 5]. In Tabelle 3 sind die nachfragerbezogenen Einflussfaktoren aus Unternehmenssicht dargestellt.

Nachfragerbezogene Einflussfaktoren	Nennungen Anbieter	Nennungen Dienstleister	Gesamt
Qualitäts- und Serviceorientierung	8	2	10
Ausmaß der Kundenintegration	8	2	10
Wertung/Wahrnehmung des Angebots	7	1	8
Motivation der Kunden zum Lernen	5	1	6
Affinität der Kunden zu Neuen Medien	4	1	5

Tabelle 3: Nachfragerbezogene Einflussfaktoren

Im Bereich der Kundenbedürfnisse geben die befragten Anbieter von CFEL allgemein an, dass die Kunden einem derartigen Angebot sehr positiv gegenüberstehen und dieses rege nutzen. Die meisten Unternehmen ziehen ihre Schlüsse aus dem direkten Kontakt mit den Kunden bzw. aus deren Feedback und der Auswertung der Zugriffe auf das Angebot. Nach Aussage der Gesprächspartner hilft CFEL vor allem die Kundenerwartungen an einen qualitäts- und serviceorientierten Anbieter zu erfüllen. Oftmals bestehen allerdings Unstimmigkeiten zwischen den Erwartungen der Kunden an angebotene Dienstleistungen und der Wahrnehmung dieser Erwartungen durch das Unternehmen [PBZ00]. So wird in im Rahmen der Befragung von potenziellen Nachfragern von CFEL ebenfalls bestätigt, dass qualitäts- und serviceorientierte Anbieter durch den Einsatz von CFEL Kundenerwartungen gerecht werden

[RoGo05]. Durch den Charakter eines VAS beeinflusst CFEL so ebenfalls die Einschätzung der Qualität der Kernleistung durch den Kunden positiv.

Der Charakter eines VAS wird auch dadurch deutlich, dass weiterhin die Wahrnehmung und Wertung des CFEL-Angebots durch die Kunden als bedeutend eingeschätzt wird. Es ist entscheidend, dass CFEL als Sekundärleistung des Unternehmens nicht nur die Kundenerwartungen erfüllt, sondern von den Kunden auch nicht als Basisleistung vorausgesetzt wird sowie relevant und wahrgenommen zur Nutzensteigerung beiträgt [Mann98]. Detaillierte Aussagen zu den Eigenschaften potenzieller Nutzer eines CFEL-Angebots können aus Unternehmenssicht nicht getroffen werden, da ein Großteil der Anbieter CFEL ohne vorherige Registrierung jedem Interessierten zur Verfügung stellt und auch in teilweise vorhandenen Registrierungsprozessen der verbleibenden Befragten persönliche Daten der Kunden nicht erfasst werden. Generell wurde aber die Motivation der Kunden zur Aneignung von Wissen als wichtige Voraussetzung für einen Erfolg versprechenden Einsatz von CFEL gesehen. Das Motivationsmodell unterscheidet sich wesentlich von der Bereitstellung von E-Learning für Mitarbeiter. CFEL bedarf einer Grundmotivation auf Seiten der Kunden und die Angebote sind noch ansprechender zu gestalten.

Eine detaillierte Betrachtung der Eignung von CFEL für anvisierte Kunden und Interessenten ist jeweils unternehmensspezifisch vorzunehmen. Aufschluss über Wahrnehmung und Kundennutzen von CFEL gibt eine genaue Typologisierung der Kunden eines Unternehmens, insbesondere anhand von Lebensstil-Typologien [RoGo05]. Da Interesse und Präferenzen der Kunden hinsichtlich E-Learning entscheidend von deren Wertvorstellungen, wie Alltagsbewusstsein, Lebensstil und Lebenszielen beeinflusst werden, reicht eine alleinige Betrachtung sozioökonomischer Kriterien nicht aus. Durch Lebensstil-Typologien werden Personen in Gruppen zusammengeführt, die in ihrer grundsätzlichen Wertorientierung und Lebensauffassung übereinstimmen und anhand sozioökonomischer Kriterien beschrieben [Jaco98, S. 74]. Im Rahmen der Nachfragerstudie wird allgemein festgestellt, dass der Einsatz von CFEL insbesondere Erfolg versprechend ist, wenn die Kunden des Unternehmens über einen eher hohen sozialen Status verfügen und gegenüber neuen Technologien offen eingestellt sind [RoGo05]. Der letzte Aspekt bezieht sich vor allem auf die Fähigkeiten der Kunden, CFEL umfassend zu nutzen und die elektronische Bereitstellung des Lernangebots als geeignet zu betrachten. Hier wurde von den Befragten festgestellt, dass die Kundengruppen grundsätzlich eine Affinität zu neuen Medien und Technologien aufweisen sollten.

Um den Bedürfnissen und Erwartungen der Kunden mit dem Angebot aber auch tatsächlich gerecht zu werden, wurde von allen Gesprächspartnern das Ausmaß der Kundenintegration als bedeutende Einsatzvoraussetzung angesehen. Dabei wird von einem Anbieter betont, dass nur ein aus Nutzersicht entwickeltes Lernangebot die Bedürfnisse bestmöglich abbilden und damit einhergehend Akzeptanz finden kann. Die Kunden wissen demnach am besten, wo ihre Defizite liegen und auf welche Weise sie lernen möchten. Die Kundenintegration umfasst zum einen die Möglichkeit, an den Kundenkontaktpunkten Wissen von und über den Kunden zu erheben bzw. vorhandenes Wissen nutzen. Zum anderen wurde das Bestehen der Option als Voraussetzung genannt, Lead User in den Entwicklungsprozess einzubinden, die trendführende Innovationsbedarfe identifizieren und formulieren [Hipp88]. Bei einem der befragten Anbieter kam den erfolgreichen handelnden sog. „Star Tradern“ eine wichtige beratende Rolle im Rahmen des Entwicklungsprozesses zu, um abzuschätzen, wie hoch der Wissensbedarf auf Kundenseite ist und welche Inhalte überhaupt vermittelt werden sollen.

3.3 Produktbezogene Faktoren

In der Dimension der auf dem Markt angebotenen Produkte und Dienstleistungen wird durch das Involvement das persönliche Interesse der Kunden an den Leistungen des Unternehmens und das mit einer Kaufentscheidung verbundene psychologische (Dissonanzen vor und nach dem Kauf), finanzielle und soziale Risiko (Fremdeinschätzung) determiniert [Pepe01, S. 249]. Es steht im engen Zusammenhang mit den Kunden und ist ein Maßstab für die Intensität der Ich-Beteiligung. Das Involvement ist ausschlaggebend für die Qualität und Tiefe der Informationssuche, -verarbeitung und -speicherung und die Lern- und Wissensbedarfe im Zuge der Kaufentscheidung [KuTo00, S. 64].

Darüber hinaus determinieren die Leistungsmerkmale des Gesamtangebotes in hohem Maße die Beurteilungsmöglichkeiten einzelner Produkt- und Dienstleistungseigenschaften [MeBr00, S. 65]. Je komplexer ein Produkt ist, desto mehr Unsicherheit besteht aufgrund fehlender Informationen während des Kaufs und desto erklärungsbedürftiger wird auch seine Anwendung [HüMa00, S. 365]. Dies hat zur Folge, dass die Konsumenten zum Abbau von Unsicherheiten eine Reihe von Informationen heranziehen und Wissen aufbauen, um die Kaufentscheidungen nicht ohne vollständige Informationen über die Entscheidungsparameter zu treffen. Nach dem Kauf ist der Kunde umso zufriedener mit dem gekauften Produkt, umso besser er dessen Funktionalitäten kennt und beherrscht [Aldr00]. Der Umfang des Lern- und Wissensbedarfs der Kunden ist also weiterhin davon abhängig, welche Eigenschaften den Produkten und

Dienstleistungen zugeordnet werden. In Tabelle 4 sind die Einflussfaktoren für einen Erfolg versprechenden Einsatz von CFEL in der Dimension der angebotenen Produkte und Dienstleistungen dargestellt.

Produktbezogene Einflussfaktoren	Nennungen Anbieter	Nennungen Dienstleister	Gesamt
Komplexe und erklärungsbedürftige Leistungen	8	2	10
Verbindung des Lernens zur Kernleistung	5	0	5
Hohe Variantenvielfalt	3	1	4
Kurze Produktlebenszyklen	2	0	2

Tabelle 4: Produktbezogene Einflussfaktoren

Alle befragten Anbieter geben an, dass sie die von ihnen angebotenen Produkte und Dienstleistungen als komplex und erklärungsbedürftig ansehen. Mit Hilfe des Lernangebotes wird bestehenden und potenziellen Kunden gezielt Wissen zur Reduktion von Unsicherheiten oder zum Abbau von Hemmschwellen vermittelt. Die Kunden sollen durch CFEL mit den Möglichkeiten bzw. Funktionalitäten der jeweiligen Leistung vor und nach dem Kauf vertraut gemacht werden. Die Bedeutung dieses Aspekts wird auch in einer eingangs offen gestellten Frage zu den Hauptgründen der Einführung von CFEL deutlich, bei der sieben Anbieter das variantenreiche Angebot komplexer und erklärungsbedürftiger Leistungen anführten.

Einflussfaktoren im Bereich Involvement konnten nur in der Feldstudie aus Nachfragersicht analysiert werden. Es zeigt sich, dass Leistungen, die sowohl eine hohe Komplexität als auch ein hohes Involvement aufweisen, prädestiniert für den Einsatz von CFEL sind [RoGo05]. Dabei hat allerdings das Ausmaß des Involvements eine wesentlich geringere Bedeutung für den Einsatz von CFEL als die Leistungskomplexität.

Außerdem wird darauf hingewiesen, dass Lern- und Wissensbedarfe der Kunden unmittelbar im Zusammenhang mit der Primärleistung des Unternehmens stehen müssen. Für CFEL-Angebote, die auf Wissensbedarfe im weiteren Leistungsumfeld ausgerichtet sind, wird der Erfolg von CFEL geringer eingeschätzt. Dadurch wird wiederum sichtbar, dass CFEL vor allem als VAS betrachtet wird. Darüber hinaus wurden hohe Variantenvielfalt und kurze Produktlebenszyklen als Einflussfaktoren für einen Erfolg versprechenden Einsatz von CFEL genannt, da diese Faktoren ebenfalls Lern- und Wissensbedarfe bei Kunden entstehen lassen.

4 Fazit und Ausblick

Der Einsatz von CFEL ist insbesondere für Unternehmen, die eine Differenzierungsstrategie verfolgen, sinnvoll und Erfolg versprechend, da vor allem das Erreichen kunden- und wettbewerbsbezogener Zielsetzungen des Unternehmens unterstützt wird. Dabei weist CFEL den Charakter eines Value-Added-Service auf, der den Kunden einen zusätzlichen Nutzen im Zusammenhang mit der Primärleistung des Unternehmens stiftet. Dementsprechend muss CFEL auf angrenzende Funktional- und Prozessstrategien sowie IT-Systeme ausgerichtet und damit abgestimmt sein. Weiterhin ist CFEL in bestehende Geschäftsmodelle der Primärleistungen des Unternehmens einzubinden oder im Zusammenhang mit Aus- und Weiterbildungsmaßnahmen des Corporate E-Learning ein gesondertes Geschäftsmodell, das auf die Distribution des E-Learning-Angebots an verschiedene Zielgruppen ausgerichtet ist, zu etablieren.

Insbesondere qualitäts- und serviceorientierte Anbieter können durch den Einsatz von CFEL Kundenerwartungen gerecht werden sowie einen zusätzlichen und wahrgenommenen Kundennutzen stiften. Voraussetzung ist allerdings eine vorhandene Grundmotivation der anvisierten Kundengruppe zum Lernen und deren Affinität zu neuen Medien und Technologien. Darüber hinaus forcieren insbesondere komplexe und erklärungsbedürftige Leistungen des Unternehmens den Einsatz von CFEL. Die durch CFEL zu deckenden Wissensbedarfe der Kunden müssen in möglichst engem Zusammenhang mit der Primärleistung des Unternehmens stehen.

Geplant, gestaltet und implementiert wird CFEL zurzeit meist als Marketingmaßnahme ohne die Integration bereichsübergreifender Kompetenzen. Die Ausschöpfung der Potenziale, die CFEL bietet, wird nur durch eine umfassende Betrachtung von CFEL mithilfe von Vorgehensmodellen und Standards zur systematischen Entwicklung von Dienstleistungen [ScDa06] möglich. Dadurch können bspw. in der Phase der Ideenfindung während der Entwicklung neue Mobil- und Sensortechnologien, die kontextbasiertes Lernen für Kunden auch on-site in realen, physischen Verkaufsumgebungen möglich machen, berücksichtigt und so neue Einsatzbereiche für CFEL eröffnet werden.

Literaturverzeichnis

- [AdPa02] Adelsberger, H. H.; Pawlowski, J. M.: Electronic Business and Education. In: Adelsberger, H. H.; Collins, B.; Pawlowski, J. M. (Hrsg.): Handbook on Information Technologies for Education and Training. Springer, Berlin 2002, S. 653-671.
- [Aldr00] Aldrich, C.: Customer-Focused E-Learning: The Drivers. <http://www.learningcircuits.org/jul2000/aldrich.html>, Abruf am 2003-08-13.
- [Barr00] Barron, T.: Customer-Focused E-Learning: The Industry. <http://www.learningcircuits.org/2000/jul2000/barron.html>, Abruf am 2004-12-10.
- [Baue94] Bauer, H. H.: Markt. In: Diller, H. (Hrsg.): Vahlens Großes Marketing Lexikon. dtv, München 1994, S. 710-711.
- [BEE96] Berekoven, L.; Eckert, W.; Ellenrieder, P.: Marktforschung. Methodische Grundlagen und praktische Anwendung. Gabler, Wiesbaden 1996.
- [BBS01] Back, A., Bendel, O., Stoller-Schai, D.: E-Learning im Unternehmen. Orell Füssli, Zürich 2001.
- [BKM04] Bürg, O.; Kronburger, K.; Mandl, H.: Implementation von E-Learning in Unternehmen - Akzeptanzsicherung als zentrale Herausforderung. Forschungsbericht Nr. 170, Ludwig-Maximilians-Universität München, Institut für Pädagogische Psychologie, München 2004.
- [BuMe01] Bullinger, H.-J.; Meiren, T.: Service Engineering – Entwicklung und Gestaltung von Dienstleistungen. In: Bruhn, M.; Meffert, H. (Hrsg.): Handbuch Dienstleistungsmanagement, Gabler, Wiesbaden 2001.
- [BuSc96] Bullinger, H.-J.; Schäfer, M.: Die Rolle des Kunden in einem Lernenden Unternehmen. <http://www.htwm.de/bitsea/infoon/lernen/lernen.htm>, 1996, Abruf am 2003-02-20.

- [Conn01] Connell, R.: Educommerce: Online Learning Migrates to the E-Commerce-Arena. <http://www.eduventures.com/pdf/educommerce.pdf>, 2001, Abruf am 2004-12-10.
- [Dill95] Diller, H.: Kundenbindung als Zielvorgabe im Beziehungsmarketing. Universität Erlangen-Nürnberg, Lehrstuhl Marketing, Arbeitspapier Nr. 40, Nürnberg 1995.
- [Finn02] Finn, A.: Trends in E-Learning. <http://www.learningcircuits.org/2002/nov2002/finn.htm>, 2002, Abruf am 2004-12-23.
- [Hipp88] Hippel, E.: The Source of Innovation. Oxford University Press, Oxford 1988.
- [Hube93] Huber, B.: Strategische Marketing- und Imageplanung: Theorie, Methoden und Integration der Wettbewerbsstrukturanalyse aus Imagedaten. Lang, Frankfurt a.M. 1993.
- [HüMa00] Hünerberg, R.; Mann A.: Online-Service. In: Bliemel, F.; Fassott, G.; Theobald, A. (Hrsg.): Electronic Commerce. Gabler, Wiesbaden 2000, S. 357-375.
- [Hung00] Hungenberg, H.: Strategisches Management im Unternehmen. Gabler, Wiesbaden 2000.
- [Jaco98] Jacob, E.: Media-Zukunft für Milieu-Typologien. http://www.mediaundmarketing.de/imperia/md/content/pdfdateien/marktforschung/zielgruppen/typologien/06_072_074_Zielgruppe.pdf, 1998, Abruf am 2004-03-03.
- [Karg01] Karg, M.: Kundenakquisition als Kernaufgabe im Marketing. Universität St. Gallen, St. Gallen 2001.
- [Kirch00] Kirchhoefer, D.: Informelles Lernen in alltäglichen Lebensführungen. QUEM, Berlin 2000.
- [KuTo00] Kuß, A.; Tomczak, T.: Käuferverhalten. Lucius und Lucius, Stuttgart 2000.
- [Laak95] Laakmann, K.: Value-Added Services als Profilierungsinstrument im Wettbewerb: Analyse, Generierung und Bewertung. Lang, Frankfurt a. M. 1995.

- [Mann98] Mann, A.: Erfolgsfaktor Service. Gabler, Wiesbaden 1998.
- [MaWi03] Mandl, H.; Winkler, K.: Auf dem Weg zu einer neuen Weiterbildungskultur. In: Dowling, M.; Eberspächer, J.; Picot, A. (Hrsg.): eLearning in Unternehmen: Neue Wege für Training und Weiterbildung. Springer, Berlin 2003, S. 3-15.
- [MeBr00] Meffert, H.; Bruhn, M.: Dienstleistungsmarketing. Gabler, Wiesbaden 2000.
- [Meir01] Meiren, T.: Entwicklung von Dienstleistungen unter besonderer Berücksichtigung von Human Resources. In: Bullinger, H.-J. (Hrsg.): Entwicklung und Gestaltung innovativer Dienstleistungen. Tagungsband zur Service Engineering 2001, Stuttgart 2001.
- [MIT01] M.I.T newmedia: eduCommerce - Schulung als Marketing-Maßnahme. http://educommerce.mit.de/whitepaper/educommerce_Wpmit.pdf, 2001, Abruf am 2004-09-09.
- [Morr03] Morrison, D.: E-learning strategies: how to get implementation and delivery right first time. Wiley, Chichester 2003.
- [NDH97] Nieschlag, R.; Dichtl, E.; Hörschgen, H.: Marketing. Dunker und Humblot, Berlin 1997.
- [PBZ00] Parasurman, A.; Berry, L. L.; Zeithaml, V. A.: Kommunikations- und Kontrollprozesse bei der Erstellung von Dienstleistungsqualität. In: Bruhn, M.; Stauss, B. (Hrsg.): Dienstleistungsqualität. Gabler, Wiesbaden 2000, S. 115-144.
- [Pepe01] Pepels, W.: Kommunikations-Management. Schäffer-Poeschel, Stuttgart 2001.
- [Resn05] Resnick, M.: Rethinking learning in the digital age. In: Kirkman, G. (Hrsg.): The Global Information Technology Report: Readiness for the Networked World. Oxford University Press, New York 2005.
- [RoGo05] Robra-Bissantz, S.; Götzelt, K.-U.: Erfolgsfaktoren des kundenorientierten E-Learning im E-Commerce. In: Ferstl, O. K.; Sinz, E. J.; Eckert, S.; Isselhorst, T. (Hrsg.): Wirtschaftsinformatik 2005 - eEconomy, eGovernment, eSociety. Physica, Heidelberg 2005, S. 861-880.

- [ScDa06] Schneider, K.; Daun, C.: Vorgehensmodell und Standards zur systematischen Entwicklung von Dienstleistungen. In: Bullinger, H.-J.; Scheer, A.-W. (Hrsg.): Service Engineering. Springer, Berlin, 2006, S. 113- 138.
- [StCo90] Strauss, A.; Corbin, J.: Basics of Qualitative Research. Sage Publications, Newbury Park 1990.
- [Stra05] Straub, R.: Lernen im Kontext - Dynamischen Lernkonzepten gehört die Zukunft. <http://www.competence-site.de/e-learning.nsf/straub>, Abruf am 2005-01-05.
- [StSc93] Steinmann, H.; Schreyögg, G.: Management: Grundlagen der Unternehmensführung. Gabler, Wiesbaden 1993.
- [Simo01] Simon, B.: Wissensmedien im Bildungssektor. http://nm.wu-wien.ac.at/lehre/dpas/Bernd_Simon_Wissensmedien_im_Bildungssektor.pdf, 2001, Abruf am 2005-01-05.
- [Zoll95] Zollner, G.: Kundennähe in Dienstleistungsunternehmen. Gabler, Wiesbaden 1995.

Einführung in den Track

Elektronische Märkte und elektronische Verhandlungen

Prof. Dr. Michael Rebstock

FH Darmstadt

Prof. Dr. Rudolf Vetschera

Universität Wien

Prof. Dr. Mareike Schoop

Universität Hohenheim

Elektronische Märkte haben sich in der industriellen Praxis zu einem wichtigen Instrument entwickelt, um Transaktionen von Unternehmen sowohl mit Endverbrauchern als auch mit anderen Unternehmen zu tätigen. Gegenwärtige Herausforderungen bestehen in methodischen Konzepten zur Entwicklung, Unterstützung und Evaluierung elektronischer Märkte und in einer generellen Serviceorientierung, die sich mit elektronischen Diensten für elektronische Märkte beschäftigt. Elektronische Verhandlungen als Marktmechanismen sind schon seit mehreren Jahren Forschungsgegenstand. Praktisch sind diese Konzepte trotz technischer Möglichkeiten dagegen noch wenig verbreitet, wobei die verschiedenen Verhandlungsformen auf unterschiedliche Akzeptanz stoßen.

Wir laden für diesen Track Papiere der elektronischen Märkte und elektronische Verhandlungen ein, die sich mit theoretischen, methodischen, empirischen und anwendungsorientierten Fragestellungen beschäftigen. Relevante Themen sind unter anderem:

Programmkomitee:

Prof. Dr. Gregory Kersten, Concordia University, Kanada

Prof. Dr. Stefan Klein, University College Dublin, Irland

Dr. Sabine Köszegi, Universität Wien

Prof. Dr. Hans Weigand, Universität Tilburg, Niederlande

Dr. Hans-Dieter Zimmermann, Universität Münster

Strategisches Gebotsverhalten in Name-Your-Own-Price- Auktionen

Oliver Hinz

Fachbereich Wirtschaftswissenschaften
Johann Wolfgang Goethe-Universität Frankfurt am Main
60054 Frankfurt am Main
oliver.hinz@wiwi.uni-frankfurt.de

Abstract

Bei so genannten Name-Your-Own-Price (NYOP)-Auktionen bieten Käufer gegen eine geheime Preisschwelle des Verkäufers. Trifft oder übersteigt das Gebot eines Käufers diese Preisschwelle, so findet der Kauf in Höhe der Käufergebots statt. In dieser Arbeit untersuchen wir, ob Käufer versuchen, ihre Konsumentenrente unter Berücksichtigung ihrer Zahlungsbereitschaft zu maximieren oder ob Käufer unabhängig von ihrer Zahlungsbereitschaft versuchen, ein möglichst gutes Schnäppchen zu machen. In einem ersten Modell, versucht der Käufer ausgehend von seiner Zahlungsbereitschaft und seiner Vermutung über die Preisschwelle, die er als exogen gegeben ansieht, seine erwartete Konsumentenrente zu maximieren. Im zweiten Modell auf Basis von Bargaining-Spielen, lässt der Käufer in seine Überlegungen auch die Erwartungen über das Verhalten des Verkäufers, der die Preisschwelle gesetzt hat, einfließen. Wir entwickeln für beide Fälle ein normatives Modell und testen diese in einem Laborexperiment, um schließlich Implikationen für das Market-Engineering ableiten zu können.

1 Einleitung

Gesunkene Transaktions- und Prozesskosten im Internet haben eine Vielzahl dynamischer Preismechanismen ermöglicht [Bako97]. Als interaktive Preismechanismen sei die Untermenge derjenigen dynamischen Preismechanismen definiert, bei der Käufer durch Abgabe von

Geboten oder das Austauschen von Nachrichten mit dem Verkäufer den Preis aktiv beeinflussen können.

Die Vorteile von interaktiven Preismechanismen sind vielfältig. Der wichtigste Vorteil für Verkäufer besteht in der Möglichkeit der Preisdifferenzierung. Preisdifferenzierung kann durch die Erzielung höherer durchschnittlicher Preise den Gewinn des Verkäufers erhöhen. Des Weiteren kann die Allokationseffizienz des Marktes verbessert werden, da Käufersegmente, die in einem Fixpreisszenario nicht bedient werden, das Produkt zu einem niedrigeren individuellen Preis erhalten können [Bako97].

Interaktive Preismechanismen werden auf zahlreichen Marktplätzen wie eBay oder Amazon eingesetzt. Oft unterscheiden sich diese Marktplätze primär durch den eingesetzten Preismechanismus. Der Marktführer eBay setzt neben klassischen Festpreisen („Sofort Kaufen“) eine Variante der Second Price-Auktion ein. Seit Kurzem ermöglicht eBay Deutschland allerdings Käufern auch, einen Preisvorschlag zu übermitteln, der vom Verkäufer angenommen oder abgelehnt werden kann.

Dieser Preismechanismus ist eng verwandt mit dem in der der Literatur als Name-Your-Own-Price (NYOP) oder Reverse Pricing beschriebenen Mechanismus. NYOP erfreut sich zunehmender Beliebtheit, obgleich er bereits 1998 von dem US-Unternehmen Priceline zum Verkauf von Flügen, Mietwagen und Hotelzimmern eingeführt wurde. Priceline setzte 2005 bereits 962 Millionen USD [Pric05] um und andere Anbieter wie GermanWings und Expedia experimentieren ebenso mit diesem interaktiven Preismechanismus.

Bei NYOP können sowohl der Käufer als auch der Verkäufer eines Produktes Einfluss auf den zu zahlenden Transaktionspreis nehmen. Durch einen vorab festgelegten und geheim gehaltenen Mindestpreis legt der Verkäufer eine untere Preisschwelle für die Transaktion fest, während ein Käufer durch die Abgabe eines Gebotes den Transaktionspreis determiniert, sofern sich dieses Gebot oberhalb des geheimen Mindestpreises des Verkäufers befindet.

Im Unterschied zu gängigen Auktionsmechanismen besteht bei NYOP keine unmittelbare Preiskonkurrenz zwischen den Bietern. Um eine Transaktion zum Abschluss zu bringen, müssen diese lediglich die ihnen unbekannte Preisschwelle eines Anbieters überbieten. Gebote anderer Bieter hingegen haben keinen unmittelbaren Einfluss auf den Preis einer Transaktion. Im Fall einer begrenzt verfügbaren Produktkapazität bei gleichzeitig großer Nachfrage könnten sich Bieter dazu veranlasst sehen, ein Gebot möglichst früh abzugeben, um den Geboten

anderer Bieter zuvor zu kommen. In diesem Fall würde auf Basis eines First-Come/First-Serve-Prinzips an die erfolgreichen Bieter zugeteilt werden.

Eines der kennzeichnenden Merkmale von NYOP ist, dass nach Abschluss der Transaktion Informationen weder über bereits erzielte Preise noch über die geheime Preisschwelle eines Verkäufers bekannt gegeben werden. Die hieraus resultierende Preisintransparenz erlaubt es, Produkte über einen solchen Vertriebskanal auch zu günstigeren Preisen zu verkaufen und somit neue Käufersegmente anzusprechen, ohne hierdurch die Preisstrukturen anderer Vertriebskanäle zu kannibalisieren.

Schließlich eignet sich NYOP insbesondere für den Abverkauf einer großen Anzahl identischer Produkte, da hierfür lediglich ein einziges Angebot erstellt werden muss. Zwar bieten verschiedene Auktionsseiten mittlerweile so genannte „Power“- oder „Dutch“-Auktionen an (z.B. eBay.de, BesteAuktion.de), jedoch wird bei diesen für n Einheiten eines Produktes ein uniformer Preis durch das n -höchste Gebot eines Bieters determiniert. Im Gegensatz hierzu ermöglicht NYOP käuferindividuell differenzierte Preise und somit eine bessere Abschöpfung der Konsumentenrente für einen Anbieter.

Die Forschung hat sich zunächst hauptsächlich mit dem Gebotsverhalten und der optimalen Ausgestaltung eines NYOP-Mechanismus unter stark vereinfachenden Annahmen gewidmet. Die Modelle beruhen auf starken Annahmen und könnten daher das Verhalten in realen Märkten nur unzureichend darstellen. Bisherige Modelle gehen von rationalem Verhalten aus und setzen dabei eine exogen gegebene Preisschwelle voraus. Spieltheoretischen Überlegungen des Käufers und Verkäufers beim Setzen der Preisschwelle, wurde bisher unzureichend in diesen Modellen berücksichtigt.

In dieser Arbeit soll geklärt werden, ob sich Käufer Konsumentenrente-maximierend verhalten, indem sie primär durch ihre Zahlungsbereitschaft bei der Gebotsabgabe geleitet werden oder ob spieltheoretische Überlegungen das Gebotsverhalten weitgehend bestimmen. Diese Erkenntnisse sind sehr wichtig, da bei Konsumentenrente-maximierendem Verhalten über die eingegangenen Gebote auf die wahre Zahlungsbereitschaft des Käufers zurück geschlossen werden kann, während bei spieltheoretischen Verhalten wie wir es aus Verhandlungsspielen (Bargaining Games) kennen, keine Erkenntnisse über die Zahlungsbereitschaft möglich sind. Die Verschleierung der wahren Zahlungsbereitschaft wird in der Auktionstheorie auch „shading“ bzw. „bid shading“ genannt (siehe z. B. [Kris02]). Bisher wurde dieser Aspekt bei NYOP noch nicht berücksichtigt, obwohl er für die richtige Beurteilung dieses

Preismechanismus sehr wichtig ist und damit NYOP-Märkte optimal gestaltet werden können. Für den aktuellen Forschungsstand des Market Engineering siehe [WeHN03].

Die Arbeit ist wie folgt aufgebaut: An diese Einleitung anschließend stellt Kapitel 2 die genaue Funktionsweise und verschiedene Gestaltungsmöglichkeit von NYOP vor, fasst die Literatur zusammen und ordnet diese Arbeit in bisherige Forschungen ein. Kapitel 3 erarbeitet zwei normative Modelle für Gebotsverhalten in NYOP-Auktionen unter unterschiedlichen Annahmen. Kapitel 4 stellt ein Laborexperiment vor, das die beiden Modelle schließlich testet. Kapitel 5 schließt die Arbeit mit einer Zusammenfassungen und Implikationen für das Marktdesign.

2 Name-Your-Own-Price

2.1 Funktionsweise

Als interaktiver Preismechanismus ermöglicht NYOP sowohl dem Käufer als auch dem Verkäufer eines Produktes Einflussnahme auf den endgültigen Preis einer Transaktion. Während der Verkäufer durch die Bestimmung einer geheimen Preisschwelle einen Mindestpreis für diese Transaktion vorgibt, bestimmt ein Käufer die endgültige Höhe des Transaktionspreises durch die Abgabe eines Gebotes. Liegt dieses Gebot oberhalb der geheimen Preisschwelle des Verkäufers, wird die Transaktion in Höhe des vom Käufer genannten Gebotes ausgeführt.

Ein Verkäufer hat dabei die Möglichkeit, NYOP anhand verschiedener Design-Variablen seinen Bedürfnissen gemäß anzupassen. Design-Variablen beschreiben dabei diejenigen Eigenschaften, anhand derer sich eine konkrete Ausprägung des NYOP-Verfahrens bestimmen lässt (z.B. das Preisauswahlverfahren, die Anzahl der möglichen Gebote, die Wartezeit zwischen Geboten eines Käufers oder ein Mindestinkrement für zusätzliche Gebote). [Bern04] geht detailliert auf diese Design-Variablen ein.

Im Gegensatz zu klassischen Auktionsmechanismen besteht bei NYOP keine unmittelbare Preiskonkurrenz zwischen den Käufern. Um eine Transaktion zum Abschluss zu bringen, müssen diese lediglich die ihnen unbekannte Preisschwelle eines Verkäufers überbieten. Gebote anderer Käufer hingegen haben keinen direkten Einfluss auf den Preis einer Transaktion.

Abb. 1 zeigt die Aufteilung der Konsumenten- und Produzentenrente in Falle eines erfolgreichen Gebots: Der Verkäufer hat variable Kosten c für das von ihm angebotene Produkt. Er setzt die Preisschwelle TP im Normalfall auf seine variablen Kosten c oder aber darüber. Das garantiert ihm im Falle eines erfolgreichen Gebots eine Basisrente von BR . Dazu kommt der Betrag, mit der Käufer die Preisschwelle überbietet. Dieser Betrag wird auch als Informationsrente IR bezeichnet. Der Käufer hat eine gewisse Zahlungsbereitschaft und versucht dagegen die Informationsrente IR möglichst klein zu halten. Die Differenz zwischen einem erfolgreichen Gebot b und der Zahlungsbereitschaft WTP ist die Konsumentenrente CS .

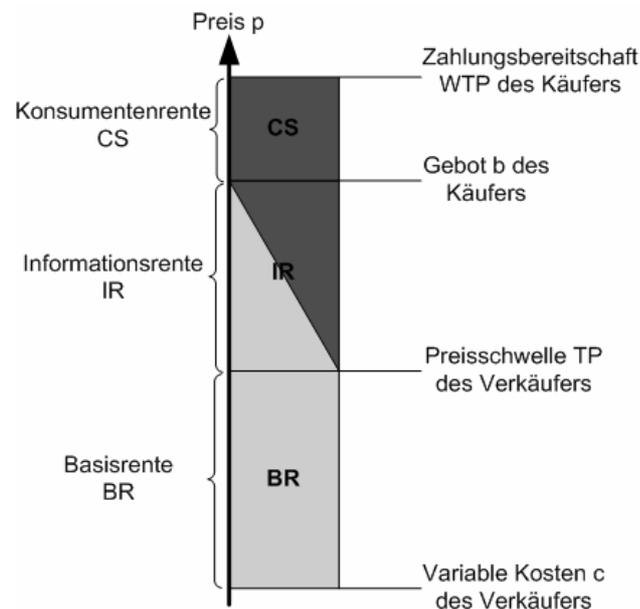


Abb. 1: Aufteilung der Konsumenten- und Produzentenrente

An dieser Abbildung wird deutlich, dass die Informationsrente IR die Verhandlungsmasse darstellt. Auf der einen Seite versucht der Käufer diese möglichst gering zu halten, um sich eine möglichst hohe Konsumentenrente zu sichern, während der Verkäufer natürlich auf ein möglichst deutliches Überbieten durch den Käufer hofft. Schließlich stellt das Überbieten die preisdifferenzierende Komponente dieses Mechanismus dar.

2.2 Literatur

NYOP wurde bisher aus zwei Richtungen analysiert: Eine Forschungsrichtung konzentriert sich auf die Erklärung des Gebotsverhaltens in NYOP-Märkten, während sich die zweite Forschungsrichtung auf die Bestimmung des optimalen Designs fokussiert. Die vorliegende Arbeit ordnet sich demnach in dem erstgenannten Feld ein. An dieser Stelle sollen die

bisherigen Erkenntnisse aus bereits veröffentlichten Arbeiten zusammengefasst werden und die vorliegende Arbeit soll entsprechend abgegrenzt und eingeordnet werden.

Bei der Ausgestaltung des optimalen Designs eines NYOP-Mechanismus schlägt [Cher03] eine interessante Designvariante für den Ein-Gebots-Fall vor, nachdem er die Präferenzen der Käufer in einem Laborexperiment untersucht hat. In diesem Experiment bietet er zwei grundsätzlich unterschiedliche Preisauswahlverfahren an: In einem ersten Fall können Bieter ihr Gebot frei („price generation“) abgeben, während im zweiten Fall dem Bieter lediglich gestattet wird, eine Auswahl aus einer gegebenen Gebotsliste („price selection“) zu treffen. [Cher03] kommt zu dem Ergebnis, dass Käufer lieber aus einer Liste auswählen, als einen Preis frei wählen zu müssen. [Cher03] lässt allerdings offen wie sich die Gebote bilden und ob Käufer eher von ihrer eigenen Zahlungsbereitschaft oder der Vermutung über die Preisschwelle des Verkäufers ausgehen.

[Fay04] leitet aus einem analytischen Modell ab wie sich der Gewinn für einen NYOP-Verkäufer bei einer variierenden Anzahl von zugelassenen Geboten entwickelt. Dabei vergleicht er den Ein-Gebots-Fall mit dem Fall, dass einige erfahrene Käufer sich zusätzliche Identitäten in Form mehrerer Kreditkarten besorgen und somit öfter bieten können. [Fay04] geht allerdings nicht davon aus, dass Käufer Überlegungen des Verkäufers beim Setzen der Preisschwelle in ihr Gebotsverhalten einfließen lassen und vernachlässigt von daher diesen Aspekt.

[HaTe03] untersuchen empirisch das Gebotsverhalten in einem Modell, das wiederholtes Bieten zulässt. Dabei entwickeln sie zuerst ein mikroökonomisches Modell, um das Gebotsverhalten zu erklären und wenden das Modell an, um die Suchkosten der Käufer in einem vorliegenden Datensatz eines NYOP-Verkäufers zu schätzen. [TeSH05] entwickeln dagegen ein Modell, das einem NYOP-Verkäufer bei der Setzung der optimalen Preisschwelle helfen kann. Allerdings gehen [TeSH05] nur von Konsumentenrente-maximierendem Verhalten aus. Für den Ein-Gebots-Fall kann aus dem Modell von [TeSH05] abgeleitet werden, dass es optimal ist, die Preisschwelle auf die variablen Kosten des Verkäufers zu setzen. Denn jedes Gebot oberhalb der variablen Kosten wird angenommen und leistet somit einen Kostendeckungsbeitrag. Legt ein Verkäufer die Preisschwelle höher als die variablen Kosten, so lässt er alle Gebote oberhalb der variablen Kosten aber unterhalb der Preisschwelle unrealisiert. Des Weiteren macht eine Preisschwelle unterhalb der variablen Kosten keinen Sinn, da angenommene Gebote somit zu Verlusten führen könnten. Diese Erkenntnis beruht allerdings auf der Annahme, dass Käufer

versuchen, ihre Konsumentenrente zu maximieren und nicht versuchen, unabhängig von ihrer Zahlungsbereitschaft ein Schnäppchen zu machen, und könnte somit zu suboptimalen Schlussfolgerungen führen.

[SpSS04] entwickeln ebenfalls ein Modell für die mehrfache Gebotsabgabe und schätzen mit Hilfe dieses Modells die Zahlungsbereitschaft der teilnehmenden Bieter. Ferner leiten die Autoren aus ihrem Modell das optimale Gebotsverhalten ab und vergleichen die Gewinne, die bei einfacher Gebotsabgabe resultieren mit den Gewinnen bei mehrfacher Gebotsabgabe.

[Bern04] fasst die möglichen Designvariablen zusammen, während [BeHi05] einen XML-Dialekt entwickeln, mit dem sich die verschiedene Designvarianten von NYOP in XML beschreiben lassen.

Im Gegensatz zu den zuletzt beschriebenen Studien, fokussieren sich einige Arbeiten nicht auf das optimale Design des NYOP-Mechanismus, sondern beschäftigen sich vorwiegend mit dem Konsumentenverhalten in einem NYOP-Kanal:

[DEHS05] untersuchen das Konsumentenverhalten im Ein-Gebots-Fall in der Situation, in der das Gebot abgelehnt wurde und der Bieter das Produkt daher über einen „traditionellen“ Verkaufskanal kaufen muss. Dabei berücksichtigen sie mehrere, aufeinander folgende Perioden und gehen davon aus, dass Bieter nicht nur ihren Nutzen in Form von Geldeinheiten maximieren, sondern dass Käufer auch einen zusätzlichen Nutzen durch das Gewinnen („excitement“) erhalten. Falls das Gebot allerdings abgelehnt wurde, so erleiden die Bieter einen Nutzenverlust („frustration“). [DEHS05] zeigen in diesem Paper analytisch und in einem Laborexperiment, dass das Gebotsverhalten vom Erfolg vorheriger Gebote abhängt. Allerdings gehen [DEHS05] nicht darauf ein, welche Annahmen zu einem Modell führen könnten, das das wahre Verhalten der Käufer besser modelliert.

Die neuste Arbeit von [SpTe06] zeigt anhand von realen NYOP-Datensätzen, dass Käufer sich nur in einem sehr beschränkten Maße rational verhalten. Dazu nutzen die Autoren zwei verschiedene Datensätze und kommen zum Schluss, dass das Internet nicht geholfen hat, irrationales Verhalten zu eliminieren wie es sich viele erhofft hatten. Allerdings gehen [SpTe06] nicht darauf ein, ob es eventuell Modelle unter anderen Annahmen gibt, die die Abweichung vom untersuchten rationalen Modell erklären könnten.

Insgesamt kann festgestellt werden, dass bisher keine Arbeit existiert, die die Annahme einer exogen gegebenen Preisschwelle aufhebt. Einige Arbeiten wie [HaTe03] und [SpSS04] haben allerdings Modelle für rationales Verhalten unter Annahme einer exogen gegebenen

Preisschwelle entwickelt. Diese Modelle bilden die Grundlage für unsere normative Modellierung des Konsumentenrente-maximierenden Verhaltens im folgenden Kapitel. Im Anschluss werden wir unser Modell vorstellen, das in Anlehnung an Verhandlungsspiele aus der Spieltheorie die Überlegungen des Verkäufers beim Setzen der Preisschwelle mitberücksichtigt und somit die Annahme einer exogen gegebenen Preisschwelle aufhebt.

3 Normative Modellierung

3.1 Maximierung der Konsumentenrente unter Annahme einer exogen gegebenen Preisschwelle

Das folgende Modell beruht auf den Erkenntnissen von [HaTe03] und [SpSS04] und trifft die Annahme eines risikoneutralen Käufers mit einer Zahlungsbereitschaft WTP, der seine erwartete Konsumentenrente maximiert. Weitere Annahmen sind, dass der Käufer von einer exogen festgelegten Preisschwelle ausgeht, die uniform im Intervall [LB, UB] verteilt ist. Das abgeleitete Verhalten lässt sich zwar auch für andere Verteilungsannahmen nachweisen, ist aber der Argumentation von [Stig61], [HaTe03] und [DEHS05] folgend für die Ergebnisse nachrangig.

Die erwartete Konsumentenrente ECS kann also unter Berücksichtigung von Suchkosten c durch folgenden Ausdruck dargestellt werden:

$$ECS = \int_{LB}^b (WTP - b) \cdot g(p_T) dp_T - c = (WTP - b) \cdot \frac{b - LB}{UB - LB} - c$$

Mit einem höheren Gebot b wird zwar die Wahrscheinlichkeit gesteigert, dass das Gebot die Preisschwelle trifft oder überbietet, aber auf der anderen Seite führt ein höheres Gebot natürlich zu einer geringeren Konsumentenrente ($WTP - b$).

Um nun das optimale Gebot zu maximieren, leiten wir den Term nach b ab und setzen gleich 0.

$$\Rightarrow \max ECS = \frac{dECS}{db} = \frac{1}{UB - LB} \cdot [(-1) \cdot (b) - LB + (WTP - b) \cdot 1] \stackrel{!}{=} 0$$

Durch eine Umformung nach b kommen wir zu folgendem Ergebnis:

$$\Leftrightarrow b^* = \frac{WTP + LB}{2}$$

Demnach ist das optimale Gebot eine Funktion der Zahlungsbereitschaft und der Vermutung über das untere Ende der Verteilung der Preisschwelle und beide Faktoren wirken in gleicher Stärke auf die Gebotshöhe.

Sowohl [HaTe03] als auch [SpSS04] benutzen ähnliche Modelle, um auf die Zahlungsbereitschaft und die Suchkosten c zurückzurechnen. Allerdings wird experimentell in diesen Arbeiten nicht überprüft, ob sich das tatsächliche Verhalten wie oben dargestellt, modellieren lässt, was demnach zu suboptimalen Rückschlüssen führen könnte. [SpTe06] zeigen nämlich, dass sich Abweichungen zwischen einem ähnlichen rationalen Modell und dem tatsächlichen Verhalten ergeben.

3.2 NYOP als Verhandlungsspiel

Das vorangegangene normative Modell basiert auf der starken Annahme, dass die Preisschwelle exogen gegeben ist. In der Realität wird aber der Verkäufer beim Setzen der Preisschwelle das Verhalten der Käufer in seine Überlegungen mit einbeziehen. Diese wiederum werden das antizipieren. Um solche Probleme zu modellieren, eignet sich die Spieltheorie hervorragend.

Daher modellieren wir das Setzen der Preisschwelle und die Gebotsabgabe als Spiel in Normalform, d.h. beide Spieler (Käufer und Verkäufer) legen ihre Strategien ohne Kenntnis der Wahl des anderen Spielers fest. Das Spiel kann demnach als statisches Bargaining-Spiel modelliert werden.

Da es sich um ein Spiel mit zwei Spielern, $I = \{\text{Käufer, Verkäufer}\}$, handelt, sind die Strategiemengen endlich und überschaubar und man kann dieses Spiel in Normalform auch als Tabelle darstellen. Wir gehen davon aus, dass jeder Spieler drei Strategien zur Verfügung hat, um so das Prinzip zu demonstrieren, da sich anschließend der Strategieraum in die Unendlichkeit erweitern lässt.

Der Verkäufer hat die Möglichkeit die Preisschwelle auf einen Preis nahe seiner variablen Kosten, einen mittleren Preis und einen hohen Preis zu setzen.

Bietet der Käufer über oder genau auf der Preisschwelle, findet eine Transaktion statt. Der Käufer erhält als Nutzen die Differenz zwischen seiner Zahlungsbereitschaft und seinem Gebot, also die Konsumentenrente. Die Produzentenrente stellt in diesem Fall den Nutzen der Transaktion für den Verkäufer dar. Bietet der Käufer unterhalb der Preisschwelle, so ist die Auszahlung für beide Marktteilnehmer gleich 0.

Der Verkäufer kann sich zwischen drei Strategien entscheiden: Der Wahl einer sehr niedrigen Preisschwelle (L_V), einer mittleren Preisschwelle (M_V) und einer hohen Preisschwelle (H_V),

während sich der Käufer für ein niedriges Gebot, das bei Erfolg zu einer hohen Konsumentenrente (H_K) führen würde, ein mittleres Gebot (M_K) oder ein sehr hohes Gebot, das in einer niedrigen Konsumentenrente (L_K) resultieren würde, entscheiden kann. Insgesamt ist nur der Fall zu betrachten, dass die Zahlungsbereitschaft des Käufers größer ist als die Kosten des Verkäufers. Andernfalls würde kein Verkauf zustande kommen. Insgesamt ist zwischen den beiden Parteien auch nur der Kuchen (Zahlungsbereitschaft-Kosten), also Informationsrente+Basisrente+Konsumentenrente, zu verteilen. Den Anteil an dieser ökonomischen Rente, den die Parteien jeweils erhalten, hängt vom Verhandlungsergebnis ab. Bietet der Käufer sehr hoch, so erhält der Verkäufer den größten Teil, bietet der Käufer dagegen sehr niedrig und ist erfolgreich, so erhält er einen größeren Anteil. Die Gebote müssen jedoch die Preisschwelle treffen oder übertreffen, ansonsten erhalten beide Parteien eine Auszahlung von 0.

Betrachten wir zum Beispiel den Fall, dass der Verkäufer variable Kosten von 50€ hat und ein Käufer eine Zahlungsbereitschaft von 110€. Dann ist die verhandelbare Rente $110€ - 50€ = 60€$. Je nach Verhandlungserfolg bekommt der Verkäufer oder der Käufer davon mehr oder weniger. Tab. 1 stellt die entsprechende Auszahlungsmatrix in absoluten Werten dar.

Käufer/Verkäufer	H_V Preisschwelle=100€	M_V Preisschwelle=80€	L_V Preisschwelle=60€
H_K Gebot=60€	0; 0	0; 0	50€; 10€
M_K Gebot=80€	0; 0	30€; 30€	30€; 30€
L_K Gebot=100€	10€; 50€	10€; 50€	10€; 50€

Tab. 1: Beispielhafte Auszahlungsmatrix für einen Käufer mit der Zahlungsbereitschaft 110€ und einem Verkäufer mit Kosten 50€

In allgemeiner Form stellt sich die Auszahlungsmatrix wie in Tab. 2 dar. K sei dabei die verhandelbare Rente (=Zahlungsbereitschaft-Kosten) und zusätzlich muss gelten, dass $1 > A_H > A_M > A_L > 0$ und $K > 0$.

Käufer/Verkäufer	H_V Hohe Preisschwelle	M_V mittlere Preisschwelle	L_V Niedrige Preisschwelle
H_K niedriges Gebot	0; 0	0; 0	$K^*(A_H; 1-A_H)$
M_K mittleres Gebot	0; 0	$K^*(A_M; 1-A_M)$	$K^*(A_M; 1-A_M)$

L_K hohes Gebot	$K^*(A_L; 1-A_L)$	$K^*(A_L; 1-A_L)$	$K^*(A_L; 1-A_L)$
-------------------	-------------------	-------------------	-------------------

Tab. 2: Auszahlungsmatrix in allgemeiner Form

Wir führen nun eine Eliminierung schwach dominierter Strategien durch. Dazu verwenden wir Tab. 3 mit typischen Werten ($A_L=0,1$; $A_M=0,5$; $A_H=0,7$ und $K=1$) zur besseren Anschaulichkeit.

Käufer/Verkäufer	H_V hohe Preisschwelle	M_V mittlere Preisschwelle	L_V niedrige Preisschwelle
H_K niedriges Gebot	0; 0	0; 0	70%; 30%
M_K mittleres Gebot	0; 0	50%; 50%	50%; 50%
L_K hohes Gebot	10%; 90%	10%; 90%	10%; 90%

Tab. 3: Auszahlungsmatrix des untersuchten Spiels

Mit Hilfe der Spieltheorie kann gezeigt werden, dass für den Verkäufer die Strategie H_V von der Strategie M_V schwach dominiert wird, da so eine Auszahlung von 50% gegenüber 0 bei der Strategiewahl M_K durch den Käufer möglich wird. Die anderen Alternativen bleiben in allen anderen Fällen genauso gut. Der Verkäufer würde also nie Strategie H_V (eine hohe Preisschwelle) wählen. Tab. 4 zeigt die Auszahlungsmatrix nach der Eliminierung der Strategie H_V .

Käufer/Verkäufer	M_V mittlere Preisschwelle	L_V niedrige Preisschwelle
H_K niedriges Gebot	0; 0	70%; 30%
M_K mittleres Gebot	50%; 50%	50%; 50%
L_K hohes Gebot	10%; 90%	10%; 90%

Tab. 4: Nach Eliminierung der Strategie H_V

Antizipiert der Käufer diese Überlegung, ist die Wahl von Strategie L_K (hohes Gebot) nicht mehr attraktiv, da Strategie M_K (Auszahlung 50% oder 50% gegenüber 10% oder 10% bei Strategie L_K) die Strategie L_K dominiert. Daher eliminieren wir Strategie L_K und erhalten Tab. 5.

Käufer/Verkäufer	M_V mittlere Preisschwelle	L_V niedrige Preisschwelle

H_K niedriges Gebot	0; 0	70%; 30%
M_K mittleres Gebot	50%; 50%	50%; 50%

Tab. 5: Nach Eliminierung der Strategie L_K

Damit wird die Strategie L_V für den Verkäufer schwach dominant gegenüber M_V, da er 30% als Auszahlung erhält, falls der Käufer die Strategie H_K wählt, während sich keine Veränderung (50%) der Auszahlung ergibt, falls der Käufer M_K wählt.

Käufer/Verkäufer	L_V niedrige Preisschwelle
H_K niedriges Gebot	70%; 30%
M_K mittleres Gebot	50%; 50%

Tab. 6: Nach Eliminierung der Strategie M_V

Schließlich wird durch eine weitere Eliminierung schwach dominierter Strategien ein iteratives Gleichgewicht (schwach) dominanter Strategien in der Strategiekombination (H_K; L_V) erreicht, da der Käufer mit der Wahl H_K 70% statt 50% Anteil erhält.

Die Anzahl der Strategien kann sowohl für den Käufer als auch den Verkäufer beliebig erweitert werden (Beweis durch Induktion möglich); das Gleichgewicht wird immer in der obersten, rechten Zelle liegen, d.h. also, dass der Verkäufer immer eine Preisschwelle nahe seiner variablen Kosten wählen wird, und ein strategischer Käufer wird unabhängig von seiner Zahlungsbereitschaft nur sehr niedrige Gebote abgeben und dabei weitgehend von seiner Vermutung über die Preisschwelle geleitet werden. Dieses Modell führt also zu einem anderen Ergebnis als das Modell aus Kapitel 3.1 und deutet darauf hin, dass der NYOP-Mechanismus den Käufer zu einem Schnäppchenjäger macht.

In einem Modell mit n Gebotsmöglichkeiten und bei Vernachlässigung von Biet- und Suchkosten, kann ein Gleichgewicht hergeleitet werden, in dem der Käufer zunächst (n-1) unakzeptable Gebote abgibt, um dann im n-ten Gebot ein Take-it-or-leave-it-offer zu unterbreiten (Prinzip der Rückwärtsinduktion). Dies führt letztlich zurück zum oben präsentierten Modell des Ein-Gebots-Falles.

Während also das erste Modell voraussagt, dass die Zahlungsbereitschaft einen starken Einfluss auf das Gebotsverhalten hat, sehen wir in diesem zweiten Modell, dass ausschließlich die

Vermutung über die Preisschwelle das Gebotsverhalten beeinflussen sollte. Um nun zu testen, welches Modell das reale Gebotsverhalten besser beschreibt, testen wir die Modelle empirisch.

4 Empirische Validierung der Modelle

Um die vorgestellten Modelle vergleichen zu können, greifen wir auf die Daten aus einem Laborexperiment mit dem induced-values-Paradigma nach [Smit76] und [Smit82] zurück. Dabei werden den Teilnehmern Zahlungsbereitschaften induziert, indem ihnen für die zu ersteigernden Produkte Wiederverkaufswerte genannt werden. Wenn es den Teilnehmern gelingt, das Produkt zu ersteigern, so ist die Differenz zwischen einem erfolgreichen Gebot und dem Wiederverkaufswert der Gewinn, der ihnen später ausgezahlt wird. Auf diese Art und Weise wird erreicht, dass sich die Teilnehmer ökonomisch verhalten und tatsächliche Anreize haben, sich vergleichbar zu wirklichen Kaufsituationen zu verhalten, in denen Käufer versuchen, ein Produkt möglichst günstig zu erwerben.

Neben induzierten Wertschätzungen müssen wir die Verteilungsannahme der Teilnehmer über die Preisschwelle kontrollieren. Wir greifen daher auf die Erkenntnisse von [HiSp06] zurück, die in ihrer Arbeit zeigen, dass durch Informationen wie „Ein Gebot in Höhe von 130 EUR wurde abgelehnt“ bei den Experimentteilnehmern eine Trunkierung der Verteilungsannahme erreicht werden kann. Auf diese Weise wird sichergestellt, dass neben der Zahlungsbereitschaft auch die untere Grenze der Zahlungsbereitschaft induziert wird. Durch dieses kontrollierte Experiment können wir also mittels einer Regressionsanalyse feststellen, welchen Einfluss sowohl die induzierte LB als auch die induzierte Zahlungsbereitschaft WTP auf das Gebotsverhalten haben.

Das Experiment wurde in sechs Sitzungen durchgeführt und insgesamt nahmen 112 Teilnehmer daran teil.¹ Die Teilnehmer des Experiments waren größtenteils Studenten (97,3%) und männlich (59,8%).

Insgesamt durften die Teilnehmer auf zehn verschiedene Produkte bieten, so dass wir insgesamt 1120 Gebote erhielten. Für die hier vorliegende Frage benutzen wir allerdings lediglich die Fälle, bei denen den Teilnehmern eine Information über das untere Ende der

¹ Die Daten wurden für das Projekt von [HiSp06] erhoben, aber eignen sich perfekt auch für die vorliegende Fragestellung.

Preisschwellenverteilung vorgelegt wurde und somit eine Induzierung der LB vorliegt. Daher reduziert sich die Anzahl der relevanten Fälle auf n=224.

Nach dem Modell aus Kapitel 3.1 erwarten wir sowohl einen deutlichen Einfluss der Zahlungsbereitschaft als auch von LB, während das Modell aus Kapitel 3.2 lediglich einen starken Einfluss von LB voraussagt.

Wir führen eine lineare Regression mit der Gebotshöhe als abhängige Variable und WTP und LB als unabhängige Variable durch.

	B	Standardfehler	Beta	T	Signifikanz
(Konstante)	-9,344	5,702		-1,639	0,103
WTP	0,764	0,048	0,808	15,861	0,000
LB	0,188	0,063	0,152	2,982	0,003

Tab. 7: Regressionsanalyse mit der Gebotshöhe als abhängige Variable (n=224)

Das Modell kann mit einem R^2 von 89,9% für Querschnittsdaten als sehr gut bezeichnet werden. Wie aus Tab. 7 schnell deutlich wird, haben sowohl WTP als auch LB einen hoch signifikanten positiven Einfluss auf die Gebotshöhe. Allerdings signalisiert sowohl der B-Wert der Koeffizienten als auch insbesondere der Beta-Wert, dass die Zahlungsbereitschaft WTP einen wesentlich stärkeren Einfluss auf die Gebotshöhe hat als LB.

Diese Ergebnisse stützen die These, dass Käufer sich vorwiegend an der eigenen Zahlungsbereitschaft orientieren und Verhalten wie wir es aus Bargaining-Spielen kennen eher nicht in der Laborsituation beobachtet werden kann. Allerdings ist in der Laborsituation die Preisschwelle tatsächlich eher exogen gegeben als in der wirklichen Marktsituation. Des Weiteren muss einschränkend gesagt werden, dass sich Equilibria in der Regel erst nach mehrfach wiederholtem Spiel einstellen und daher die Wiederholungen im Laborexperiment eventuell nicht ausreichend waren. Von daher kann an dieser Stelle weder die Überlegenheit des einen noch des anderen Modells mit Sicherheit festgestellt werden. Weitere Experimente sind daher notwendig.

Insgesamt bleibt allerdings festzuhalten, dass die Experimententeilnehmer eher sehr vorsichtig geboten haben, da sie sehr nahe an der eigenen Zahlungsbereitschaft geboten haben, um auf jeden Fall zu gewinnen. Diese Beobachtung könnte die Annahme der Risikoneutralität der Käufer angreifen, die beiden Modellen zugrunde liegt. Eventuell sollten daher die

ökonomischen Modelle um Faktoren aus der Verhaltenswissenschaft erweitert werden und somit der Argumentation von [TvKa91] folgen.

5 Zusammenfassung und Implikationen

In dieser Arbeit untersuchen wir, welche Modelle das Verhalten von Käufer in NYOP-Auktionen gut beschreiben. Dazu wird das Verhalten unter unterschiedlichen Annahmen mikroökonomisch modelliert. Je nach Annahme kann unterschiedliches Verhalten erwartet werden. Daher prüfen wir die Modelle empirisch in einem Laborexperiment mit induzierten Zahlungsbereitschaften und induzierten Vermutungen über die Preisschwelle.

Die Ergebnisse sind interessant und deuten darauf hin, dass sich Käufer - zumindest in der Laborsituation - eher an der eigenen Zahlungsbereitschaft orientieren. Diese Erkenntnisse haben erhebliche Konsequenzen für das Market Engineering. Damit stellt NYOP nämlich nicht nur ein Preismechanismus zum gleichzeitigen Abverkauf mehrerer Einheiten dar, sondern scheint besonders gut dazu geeignet, die Konsumentenrente abzuschöpfen. Ferner können Käufer anhand der Gebote auf die Zahlungsbereitschaft der Käufer zurück schließen, da erste Experimente gezeigt haben, dass „bid shading“ in NYOP kein Problem darstellt. Diese Erkenntnisse könnten besonders für die Marktanalyse von Interesse sein und es könnten Entscheidungsstützungssysteme entworfen werden, die anhand der Gebotsanalyse den Markt analysieren und aus den Ergebnissen bessere Designvorschläge ableiten.

Alternativ könnten eingehenden Gebote „on-the-fly“ ausgewertet werden und das System könnte versuchen, Käufer in bestimmte Segmente einzuordnen. Anhand dieser Käufersegmente könnten Verhandlungssysteme versuchen, die Konsumentenrente noch besser abzuschöpfen oder aber die Allokationseffizienz des Marktes zu verbessern. Ob ein solches Verhandlungssystem allerdings auf Akzeptanz auf Käuferseite stößt, bleibt eine offene Forschungsfrage.

Mit Hilfe der normativen Modelle können Märkte adäquat modelliert werden und es können Vorraussagen über die Marktdynamik abgeleitet werden. Marktineffizienzen können durch regulatorische Maßnahmen oder durch die Einführung neuer Anreize beseitigt oder gemindert werden. In dem vorliegenden Fall von NYOP empfehlen wir allerdings eine verhaltenswissenschaftliche Analyse des Gebotsverhaltens, um noch genauere Modelle entwickeln zu können.

Ferner fehlt in der Literatur ein Vergleich interaktiver Preismechanismen (z. B. verschiedene Auktionsformate und elektronischer Verhandlungen) mit NYOP. Dieser Vergleich könnte eine interessante Forschungsfrage für die Zukunft sein, die unter anderem die Einsatzgebiete von NYOP besser ausleuchten könnte.

Literaturverzeichnis

- [Bako97] Bakos, Y.: The Emerging Role of Electronic Marketplaces on the Internet. In: Communications of the ACM, 41 (1998) 8, S. 35-42.
- [BeHi05] Bernhardt, M.; Hinz, O.: RPXML - Standardisierung von Reverse-Pricing-Mechanismen, In: Ferstl, O. K.; Sinz, E. J.; Eckert, S.; Isselhorst, T. (Hrsg.): Wirtschaftsinformatik 2005, Heidelberg 2005, S. 309-329.
- [Bern04] Bernhardt, M.: Classification of Design Options in Reverse Pricing Mechanisms. In: Bichler, M.; Holtmann, C.; Kirn, S.; Müller, J. P.; Weinhardt, C. (Hrsg.): Coordination and Agent Technology in Value Networks. GITO: Berlin 2004, S. 29-43.
- [Cher03] Chernev, A.: Reverse Pricing and Online Price Elicitation Strategies in Consumer Choice. In: Journal of Consumer Psychology, 13 (2003) 1/2, S. 51-62.
- [DEHS05] Ding, M.; Eliashberg, J.; Huber, J.; Saini, R.: Emotional Bidders – An Analytical and Experimental Examination of Consumers' Behavior in Reverse Auctions. In: Management Science, 51 (2005) 3, S. 352-364.
- [Fay04] Fay, S.: Partial Repeat Bidding in the Name-Your-Own-Price Channel. In: Marketing Science 23 (2004) 3, S. 407-418.
- [HaTe03] Hann, I.-H.; Terwiesch, C.: Measuring the Frictional Costs of Online Transactions: The Case of a Name-Your-Own-Price Channel. In: Management Science 49 (2003) 11, S. 1563-1579.

- [HiSp06] Hinz, O.; Spann, M.: The Impact of Information Diffusion on Bidding Behavior and Seller Profit in Name-Your-Own-Price Markets, Arbeitspapier, Goethe-Universität Frankfurt.
- [Kris02] Krishna, V.: Auction Theory. Academic Press, San Diego 2002.
- [Pric05] Priceline.com: Reports Financial Results for 4th Quarter and Full-Year 2005, http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=pcln&script=410&layout=-6&item_id=818428, Abruf am: 24.06.2006.
- [Smit76] Smith, V. L.: Experimental Economics: Induced Value Theory. In: American Economic Review, 66 (1976) 2, S. 274-279.
- [Smit82] Smith, V. L.: Microeconomic Systems as an Experimental Science. In: American Economic Review, 72 (1982) 5, S. 923-955.
- [SpSS04] Spann, M., Skiera, B.; Schäfers, B.: Measuring Individual Frictional Costs and Willingness-to-Pay via Name-Your-Own-Price Mechanisms. In: Journal of Interactive Marketing, 18 (2004) 4, S. 22-36.
- [SpTe06] Spann, M.; Tellis, G. J.: Does the Internet Promote Better Consumer Decisions? The Case of Name-Your-Own-Price Auctions. In: Journal of Marketing, 70 (2006) 1, S. 65-78.
- [Stig61] Stigler, G. J.: The Economics of Information. In: Journal of Political Economy, 69 (1961) 3, S. 213–225.
- [TeSH05] Terwiesch, C.; Savin, S.; Hann, I.-H.: Online Haggling and Price-Discrimination in a Name-Your-Own Price Channel. In: Management Science, 51 (2005) 3, S. 339-351.
- [TvKa91] Tversky, A.; Kahneman, D.: Loss Aversion in Riskless Choice: A Reference-Dependent Model. In: Quarterly Journal of Economics, 4 (1991), S. 1039-1061.
- [WeHN03] Weinhardt, C.; Holtmann, C., Neumann, D.: Market Engineering (Schlagwort). In: Wirtschaftsinformatik, 45 (2003) 6, S. 635-640.

Design von Reputationssystemen in Online-Auktionen

Eine vergleichende Marktanalyse

Kai Riemer, Dorothee Korn

European Research Center for Information Systems (ERCIS)

Universität Münster

48149 Münster

{kai.riemer, dorotheekorn}@uni-muenster.de

Abstract

Ein Reputationssystem stellt Mechanismen zum Vertrauensaufbau zwischen den Transaktionspartnern in Online-Auktionen zur Verfügung. In diesem Beitrag werden die Reputationssysteme der sechs größten Online-Auktionshäuser im deutschen Markt evaluiert; hierzu wurde ein Kriterienkatalog mit Gestaltungsoptionen für Reputationssysteme erarbeitet. Da eBay.de für das eigene Reputationssystem vielfach kritisiert wird, wurde davon ausgegangen, dass die Wettbewerber bestrebt sind, sich in diesem Bereich zu profilieren und vom Marktführer zu differenzieren. Die Ergebnisse deuten jedoch im Gegenteil auf ein dominantes Design mit nur marginalen Abweichungen im Detail hin. Die Gründe hierfür, die Unterschiede zwischen den Reputationssystemen, sowie einige allgemeine Grenzen und Risiken werden diskutiert.

1 Motivation und Hintergrund

1.1 Online-Auktionen als Form des Distanzhandels

Online-Auktionen haben sich zu einem ernst zu nehmenden Vertriebskanal für den Handel zwischen und mit Konsumenten (C2C/B2C) entwickelt. Sie bieten einen einfachen Marktzugang, zeichnen sich durch eine schnelle Transaktionsabwicklung aus und ermöglichen den Handel unter räumlicher Distanz zwischen weitgehend anonymen Marktparteien [Dell03].

Die Abwicklung der Transaktionen leidet jedoch unter der Asynchronität: Der Auktionsgewinner lässt dem Verkäufer nach Auktionsende das Geld zukommen, ohne die Ware vorher phy-

sisch kontrollieren zu können; erst dann verschickt der Verkäufer die Ware. Online-Auktionshäuser treten dabei lediglich als Betreiber der Plattform auf. Die Handelsparteien (insbesondere die Käufer) tragen alle mit der Transaktion einhergehende Risiken [RZFK00]. Damit trotz der bestehenden Risiken ein Handel zwischen anonymen Parteien entstehen kann, muss zwischen diesen ein Vertrauensverhältnis entstehen. Betreiber von Online-Auktionen nutzen zu diesem Zweck so genannte Reputationssysteme [Dell03].

1.2 Zur Rolle von Reputationssysteme für den Vertrauensaufbau

In herkömmlichen Geschäftsbeziehungen entsteht Vertrauen durch wiederkehrende, persönliche Kontakte. Auf Online-Handelsplattformen treffen jedoch anonyme Käufer und Verkäufer aufeinander, die in der Regel keine wiederkehrenden Transaktionen abwickeln. Ein Reputationssystem fungiert hierbei als Vermittlungsinstanz, indem es das zum Zustandekommen von Geschäftsabschlüssen notwendige Vertrauen begünstigt [RZFK00, Dell00]. Reputationen spiegeln das vergangene Verhalten eines Käufers oder Verkäufers wieder und dienen als Indikator für das zu erwartende zukünftige Verhalten in Interaktionen mit anderen Nutzern („shadow of the future“) [FrRe99a]. Sie entstehen aus der Sammlung von Bewertungen über den Ablauf vergangener Transaktionen und werden als verdichtete Größe im Bewertungsprofil des jeweiligen Nutzers zusammengefasst. Reputationssysteme schaffen eine künstliche Form der Mundpropaganda aus traditionellen Märkten, indem Erfahrungen der Vergangenheit unter allen Nutzern im Sinne eines öffentlichen Gutes geteilt werden [Dell04, Dell01]. Dabei ist die Reputation des Verkäufers sowohl als Maß für dessen Vertrauenswürdigkeit zu sehen [MeAl02, ShTa05], als auch als Servicequalität, die den Preis der Ware beeinflusst [MeAl02, LRBR00].

Werden die Verkäufer durch das Reputationssystem zu einem zuverlässigen Handeln angeleitet, wirkt sich das rückwirkend positiv auf die Reputation der Plattform aus. Diese erhält einen guten Ruf und zieht weitere Nutzer an. Dadurch wiederum steigt die Konkurrenz unter den Verkäufern, so dass diese ebenfalls verstärkt auf ihren guten Ruf achten [Koll99]. Dabei besteht eine positive Korrelation zwischen der bestehenden Reputation und dem Anreiz sich weiterhin eine positive Reputation aufbauen zu wollen [CaHo05].

Zusammenfassend lässt sich sagen, dass Reputationssysteme aus Sicht des Betreibers der Gewinnung von Nutzern dienen, die auf der Online-Handelsplattform ihre Geschäfte abwickeln sollen. Dazu reduziert das Reputationssystem die Unsicherheit, indem es Mechanismen bereitstellt, die betrügerisches Verhalten negativ und kooperatives Verhalten positiv markieren. Somit

lenken die Mechanismen das Verhalten der Nutzer und prägen den „Umgangston“ auf der Plattform. Ein Verkäufer mit einer positiven Historie ist zudem an die Handelsplattform gebunden, da er sein Profil nicht auf eine andere Plattform übertragen kann. Dieses stellt somit Wechselkosten dar und erhöht so die Kundenbindung.

1.3 Motivation der vorliegenden Studie

Gerade das Reputationssystem des Marktführers eBay.de wird für eine Reihe von Unzulänglichkeiten kritisiert. So ist es mit vertretbarem Aufwand möglich, eine positive Reputation selbst zu erstellen, z.B. durch den Handel vieler sehr preiswerter Artikel. Darüber hinaus erlaubt das eBay-System, eine so genannte Rachebewertung abzugeben. Bewertet der Käufer den Verkäufer gerechtfertigt negativ, so kann dieser im Gegenzug auch negativ bewerten. Ein solches System erlaubt dem Verkäufer, Druck auf den Käufer auszuüben. Es wird deutlich, dass das Reputationssystem von eBay keineswegs als optimal gelten kann. Es stellt sich daher die Frage, wie die Wettbewerber ihre Reputationssysteme gestalten, d.h. wie sie sich vom Marktführer abheben und welche Schlüsse man für die Verbesserung auf Seiten von eBay hieraus ziehen könnte. Vor diesem Hintergrund präsentiert der vorliegende Beitrag eine im deutschen Markt für Online-Auktionen durchgeführte Vergleichsstudie. Die grundlegende Forschungsfrage lautet: „Wie nutzen die Wettbewerber im Markt für Online-Auktionen ihre Reputationssysteme um sich vom Marktführer eBay zu differenzieren?“ Hinter der Frage steht die Annahme, dass Wettbewerber ein Interesse daran haben, ihre Reputationssysteme zu nutzen, potenziellen Kunden und wechselwilligen Nutzern ein vertrauenswürdiges Handelsumfeld zu bieten. Um dieser Frage nachzugehen, wurden zunächst typische Gestaltungsoptionen von Reputationssystemen herausgearbeitet. Basierend auf einem Katalog von Fragen wurde dann eine Evaluation der Systeme von eBay sowie der fünf größten Wettbewerber im deutschen Markt vorgenommen.

2 Anforderungen an die Gestaltung von Reputationssystemen

2.1 Übersicht über den Reputationsprozess

Ein Reputationssystem muss sich mehreren Herausforderungen stellen. Es soll ein faires Bewerten seitens der Käufer sicherstellen und gleichzeitig die Verkäufer zu einem Einhalten der ange-

prieseenen Qualität anhalten [Dell00]. Den Käufer interessieren aussagekräftige Informationen, die ein Identifizieren von vertrauenswürdigen Verkäufern erlauben [ReRe01]. In Anlehnung an Resnick et al. lassen sich zwei Phasen im Reputationsprozess auf Online-Handelsplattformen unterscheiden. In der ersten Phase wird das Feedback ermittelt, es erfolgt mithin die Bewertung durch die Nutzer. In der zweiten Phase wird das Feedback zu allen vergangenen Transaktionen eines Nutzers verdichtet und einem Interessenten aufbereitet als Bewertungsprofil präsentiert; es steht also die Nutzung der Reputation für die Kaufentscheidung im Vordergrund. Zu beiden Phasen ergeben sich nun unterschiedliche Optionen für das Design eines Reputationssystems. Die im Weiteren genannten Aspekte bilden die Grundlage für die Evaluation der Reputationssysteme der sechs ausgewählten Betreiber.

2.2 Gestaltung der Feedback-Ermittlung

Nach Ende einer Transaktion gilt es, die Benutzer zur Abgabe einer ehrlichen Bewertung zu animieren. Plattformbetreiber haben etliche Freiheitsgrade bei der Gestaltung der Feedback-Vergabe in der Bewertungsphase. Folgende Aspekte sollten hierbei berücksichtigt werden:

- Wer ist zur Abgabe einer Bewertung berechtigt [Koll99]? Bei einer bidirektionalen Kommunikation haben beide Parteien die Möglichkeit, den Ablauf der Transaktion zu beurteilen, dabei kann es jedoch zum Problem der Rachebewertung kommen.
- Liefert das Bewertungssystem Anreize zur Abgabe einer Bewertung [ReZe01]? Hier gilt es das Problem des „free-riding“ zu vermeiden, bei dem Nutzer zwar von positiven Bewertungen anderer Nutzer profitieren, selbst aber nicht bewerten. Was tut der Plattform-Anbieter um das Pflichtbewusstsein der Nutzer zu stärken und die Wichtigkeit der Bewertungsabgabe zu verdeutlichen?
- Ist die Bewertung verpflichtend, d.h. drohen andernfalls Sanktionen?
- Wie ist die Bewertung strukturiert [Koll99], d.h. auf welche Weise wird bewertet, durch Auswählen eines Urteiles aus einer Liste, Vergabe von Punktwerten, Kommentare etc.?
- Daran schließt sich die Frage an, wie ein einzelnes Urteil in das Profil des Bewerteten einfließt, werden Prozentwerte, Punktwerte etc. errechnet?
- Kann eine abgegebene Bewertung modifiziert werden oder ist das Urteil endgültig [Koll99]? Bietet der Anbieter die Möglichkeit im Konfliktfall Bewertungen zu löschen?

- Besteht die Möglichkeit auf eine erfolgte Bewertung eine Stellungnahme abzugeben? Dies kann hilfreich sein, um einen Disput offen und für Dritte sichtbar zu dokumentieren und es einem potenziellen Käufer zu überlassen, sich selbst ein Urteil zu bilden.
- Wie wird die Abgabe ehrlicher Bewertungen durch den Plattformbetreiber gefördert [ReZe01, Dell03]?
- Besteht die Möglichkeit für den Bewerteten (einzelne) Bewertungen auszublenden? Hier besteht die Gefahr, das System zu unterlaufen, wenn unerwünschte Bewertungen versteckt werden können.

2.3 Gestaltung der Bewertungsnutzung

Der Gestaltung der Feedbacknutzung kommt entscheidende Bedeutung zu, da der Nutzen eines Reputationssystems erst aus Sicht eines potenziellen Käufers deutlich wird. Hier geht es also insbesondere um eine übersichtliche und an den Bedürfnissen des Käufers orientierte Präsentation des Bewertungsprofils eines Verkäufers. Im Einzelnen stellen sich die folgenden Fragen:

- Wie werden das Bewertungsprofil und die Bewertungsanzahl visualisiert? Wird dem Käufer ermöglicht, die Reputation des Verkäufers auf der Auktionsseite unmittelbar abzuschätzen?
- (Wie) Kann der Nutzer weitere Informationen zur Reputation des Verkäufers erlangen?
- Wie wird die Bewertungs-Historie des Nutzers dargestellt [Koll99]? Das bloße Anzeigen eines Wertes als Differenz aller positiven und negativen Bewertungen vernachlässigt die Feinheiten einer Online-Transaktion. Deshalb ist es wichtig, ausführlichere Informationen zur Bewertungshistorie zu bekommen.
- Wichtig in diesem Zusammenhang ist außerdem, ob es einen Filter gibt, anhand derer der Käufer im ausführlichen Profil des Verkäufers gezielt suchen oder die Übersicht ändern kann. Kann man sich gezielt nur die negativen Bewertungen anzeigen lassen?
- Zur Vervollständigung des Reputationsprofils kann es sinnvoll sein, weitere Daten anzuzeigen, wie z.B. seit wann ein Nutzer registriert oder ob er gewerblich tätig ist.

- Darüber hinaus kann der Plattform-Anbieter den Nutzern die Möglichkeit geben, sich bei der Registrierung einer Identifikation zu unterziehen, um so ihre Existenz zu bestätigen und das Problem der Anonymität abzuschwächen.
- Zu guter letzt stellt sich die Frage, ob das Bewertungsprofil eines Nutzer immer neben dem Namen angezeigt wird, oder nur wenn der Nutzer als Verkäufer tätig ist [ReZe01].

2.4 Weitere anbieterbezogene Kriterien

Neben den Gestaltungsoptionen, die sich direkt auf die Phasen des Reputationsprozesses beziehen, können einige weitere Kriterien zur Charakterisierung eines Reputationssystems identifiziert werden:

- Ist das System eher positiv ausgerichtet oder negativ [Koll99]? Werden z.B. die nicht-vertrauenswürdigen oder die vertrauenswürdigen Verkäufer herausgestellt?
- Wie kommuniziert der Betreiber sein Bewertungssystem? Wo gibt es Informationen und wie ausführlich dokumentiert der Betreiber den Umgang mit Bewertungen?
- Welche sonstigen Besonderheiten bietet das Reputationssystem?

3 Reputationssysteme ausgewählter Online-Auktionsplattformen

Basierend auf den obigen Kriterien wurden die Reputationssysteme der Online-Auktionshäuser eBay.de, Hood.de, Auxion.de, BesteAuktion.de, Ricardo.ch und Azubo.de analysiert¹. Eine tabellarische Übersicht dieser Evaluation findet sich im Anhang; hier werden die einzelnen Reputationssysteme kurz beschrieben. Dabei wird auf eBay ausführlicher eingegangen, während für die Wettbewerber lediglich Besonderheiten und Abweichungen herausgestellt werden.

3.1 eBay.de

Verkäufer stellen bei eBay ihre Ware als Auktionen ein; dabei erhält der Höchstbietende nach Ablauf der Auktionszeit den Zuschlag. Bevor ein Kunde auf der eBay-Plattform als Käufer oder

¹ Diese sind laut www.auktionssuche.de [Stand Mai 2006] die sechs größten Anbieter. Dabei ist eBay mit über 15.000.000 Auktionen deutlich Marktführer vor dem zweitgrößten Anbieter Hood.de mit ca. 900.000 Auktionen.

Verkäufer tätig werden darf, muss sich dieser mit seinem Namen und seinen Kontaktinformationen registrieren. Dabei prüft eBay aber lediglich die Gültigkeit der E-Mail-Adresse. Zwar ist es möglich, seine Identität über das PostIdent-Verfahren der Deutschen Post AG offiziell bestätigen zu lassen, hiervon machen jedoch nur wenige Nutzer Gebrauch.

Nach Abschluss einer Transaktion haben beide Auktionspartner 90 Tage Zeit sich gegenseitig zu bewerten, jedoch sind sie hierzu nicht verpflichtet. Jede Bewertung besteht aus einer Kommentarzeile als Freitext zuzüglich einer Einstufung in die Kategorien „positiv“ (+1), „neutral“ (0) oder „negativ“ (-1). Das Bewertungsprofil eines Verkäufers wird immer bei seinen Auktionen angezeigt und ist für jeden Nutzer leicht zu erkennen. Es wird jedoch immer erst auf der eigentlichen Auktionsseite angezeigt, eine Nutzung als Auswahlkriterium in der Suchliste ist nicht möglich. eBay stellt folgende verdichtete Daten in einem Kurzprofil zur Verfügung:

1. Die Differenz aus der Anzahl positiver und negativer Bewertungen als Gesamtzahl hinter dem Nutzernamen, sowie einen farbigen Stern entsprechend der Einstufung in eine Kategorie zur Anzahl der aktuellen Bewertungen. Bei der Berechnung der Gesamtzahl zählen nur Bewertungen unterschiedlicher Mitglieder.
2. Den prozentualen Anteil positiver an den insgesamt erhaltenen Bewertungen.
3. Das Datum, an dem sich der Verkäufer bei eBay registriert hat und ob er privat oder gewerblich handelt, sowie
4. Links zu einer ausführlichen Seite mit Bewertungskommentaren und ggf. zu einer persönlichen Seite oder einem Online-Shop des Verkäufers.

Die vollständige Bewertungshistorie vergangener Bewertungen findet sich also auf einer gesonderten Seite. Hier gliedert eBay die Bewertungen in einer Tabelle nach positiven, neutralen und negativen Bewertungen, ebenso nach Wochen, Monaten und Sechs-Monats-Perioden. Eine Liste der neuesten Bewertungen enthält ausführliche Informationen wie das Urteil, den Bewertungskommentar, das Kurzprofil des Bewertenden und ob es sich um einen Käufer oder Verkäufer handelte, den Zeitpunkt, sowie einen Link zur zugehörigen Transaktion (dieser ist 90 Tage lang aktiv). Die Filterung dieser Bewertungen kann der Nutzer nach Zeiträumen, sowie nach den Kategorien „Von Käufern“ und „Von Verkäufern“ vornehmen. Eine Filterung nach neutralen und negativen Bewertungen versteckt eBay jedoch auf Unterseiten, die man erst erreicht, indem man zuerst nach einem Zeitraum filtert.

Einmal abgegebene Bewertungen können nicht gelöscht werden; sie werden von eBay nur im gegenseitigen Einverständnis nach Durchlaufen eines formularbasierten Prozesses entfernt. Be-

wertungen können jedoch mit einer Textzeile kommentiert werden. Zudem erlaubt es eBay dem Nutzer, seine bisher erhaltenen Bewertungen vor anderen Nutzern zu verstecken. Eine solche Entscheidung bezieht sich aber immer auf alle Bewertungen.

Bezüglich der Kommunikation weist eBay sehr intensiv auf das eigene Bewertungssystem hin: In den Hilfeseiten gibt es ausführliche Informationen zur Rolle und zum Verhalten bei der Bewertung. Auch wird der Nutzer nach Abschluss einer Transaktion daran erinnert, die Bewertung nicht zu vergessen. Dabei ist die Philosophie des Reputationssystems von eBay grundlegend positiv; will der Nutzer eine negative oder neutrale Bewertung abgeben, so erfolgt eine Sicherheitsabfrage und er wird darauf hingewiesen, zunächst eine Einigung mit dem Transaktionspartner zu suchen. eBay ist also sehr daran interessiert, den Nutzer zum Verzicht auf eine negative Bewertung zu bewegen.

3.2 Hood.de

Das Reputationssystem von Hood.de weist eine große Ähnlichkeit zu dem von eBay genutzten auf. Wie auf eBay.de bewerten sich die Handelspartner nach einer abgeschlossenen Transaktion gegenseitig durch eine Bewertung „positiv“, „neutral“ oder „negativ“ und einem zusätzlichen Kommentar. Ebenso wird die Gesamtanzahl der bislang erhaltenen Bewertungen numerisch als auch durch Stern-Symbole ausgewiesen. Im Unterschied zu eBay berechnet sich der Wert aber aus allen erhaltenen Bewertungen, sodass sich ein Nutzer selber einen Überblick verschaffen muss, von wie vielen unterschiedlichen Nutzern der Nutzer Bewertungen erhalten hat. Ein weiterer Unterschied besteht darin, dass der Nutzer die Bewertungen auf der ausführlichen Profildseite nicht geeignet filtern kann. Zudem ist Hood.de sehr sparsam bei der Kommunikation des eigenen Bewertungssystems; hier geht eBay deutlich intensiver auf Verhaltensregeln ein.

3.3 Auxion.de

Das Reputationssystem von Auxion.de ähnelt von der Grundstruktur her sehr stark dem von eBay: Käufer und Verkäufer bewerten sich gegenseitig durch ein Gesamturteil („positiv“, „neutral“ oder „negativ“) und einen Kommentar. Bewertungen werden dabei nicht zu einem Zahlenwert verdichtet, sondern einzeln hinter dem Nutzernamen ausgewiesen; ein Symbol wird nicht angezeigt. Das Profil wird nur angezeigt, wenn der Nutzer als Verkäufer aktiv ist. Daher ist die Reputation eines Bewertenden bei den Kommentaren auf der ausführlichen Profildseite

nicht ersichtlich. Allerdings kann der Nutzer im Gegensatz zu eBay hier sehr leicht nach neutralen und negativen Bewertungen filtern. Auch bei Auxion.de sind die Nutzer nicht zur Bewertung verpflichtet, jedoch verfällt die Bewertung nicht, vielmehr wird durch das System nach Ablauf einer Zeitspanne automatisch ein positives Urteil vergeben. Nutzer können sich bei Auxion.de anhand einer Kopie des Ausweises identifizieren lassen; dies wird durch Anzeige eines Symbols im Profil gekennzeichnet. Die Aufklärung hinsichtlich des Bewertungssystems von Auxion.de ist nicht umfassend und fällt im Vergleich zu den anderen Betreibern dürftig aus.

3.4 BesteAuktion.de

BesteAuktion.de unterscheidet sich von eBay lediglich in Details, dabei fällt das Reputationssystem eher durch das Fehlen von aus eBay bekannten Funktionen auf. Eine Bewertung besteht wieder aus „positiv“, „neutral“ oder „negativ“ zuzüglich eines Kommentars. Diese werden einzeln ausgewiesen; dabei zählen alle Bewertungen, nicht nur die unterschiedlicher Mitglieder. Bei BesteAuktion.de wird im Unterschied zu anderen Auktionshäusern dem Nutzer als Anreiz zur Abgabe einer Bewertung ein positiver Viertelpunkt geboten. Im ausführlichen Bewertungsprofil lassen sich die Bewertungen nach „neutral“ und „negativ“, sowie nach der Aktualität filtern. Es fehlt jedoch ein Link zur Transaktion, auch ist nicht ersichtlich, ob die Bewertung von einem Käufer oder Verkäufer stammt. Die Kommunikation bei BesteAuktion.de fällt ebenfalls sehr dürftig aus. Allerdings verlangt der Betreiber bei der Registrierung eine gültige Telefonnummer zur Übermittlung einer Aktivierungs-PIN.

3.5 Ricardo.ch

Die Registrierung bei Ricardo.ch setzt zwingend eine gültige Postadresse voraus, an die ein Aktivierungscode versendet wird. Ricardo.ch bietet die Möglichkeit der Bewertung mit Hilfe eines Assistenten. Für jede positive Bewertung gibt es einen Punkt (+1), für jede negative einen Punktabzug (-1). Auch hier ist die Darstellung des Bewertungsprofils sehr ähnlich der bei eBay: Es wird ein Prozentwert positiver Bewertungen errechnet und mit einem Symbol angezeigt. Auf der ausführlichen Profildseite ist eine Filterung nach negativen Bewertungen möglich, nicht jedoch nach Zeitraum; auch fehlen Links zu den zugehörigen Transaktionen. Wie eBay versucht Ricardo.ch die Nutzer zum Dialog statt zu negativen Bewertungen zu bewegen; in der Kommunikation ist Ricardo.ch recht aktiv, es finden sich viele Informationen zum Bewertungssystem.

3.6 Azubo.de

Azubo.de unterscheidet sich von den anderen Plattformen durch eine besondere Form der Bewertungsabgabe: Mit Hilfe von drei Fragen wird die Zufriedenheit der Handelspartner ermittelt. So beurteilt der Käufer, ob der Zustand der Ware wie beschrieben war, die Lieferung in einer angemessenen Zeit erfolgte und wie seine allgemeine Zufriedenheit war. Der Verkäufer bewertet die Zahlungsmoral und Kontaktfreudigkeit des Käufers und seine allgemeine Zufriedenheit mit der Transaktion. Die Antworten auf die Fragen haben dabei die Gewichtung „gut-mittel-schlecht“. Darauf aufbauend wird jedoch wieder eine entsprechende Gesamtbewertung der Kategorien „positiv“, „neutral“ und „negativ“ generiert, sowie eine Bewertungszahl und ein Prozentwert. Auf der ausführlichen Bewertungsseite visualisiert ein Balkendiagramm das Verhältnis der positiven, neutralen und negativen Bewertungen untereinander. Eine weitergehende Filterung der Bewertungen ist dagegen nicht möglich. Eine Identifikation des Kunden ist mittels einer Kleinstüberweisung auf sein Girokonto optional möglich.

4 Diskussion der Ergebnisse

4.1 Dominantes Design bei den untersuchten Reputationssysteme

Die Eingangsvermutung dieser Studie, nach der die Wettbewerber von eBay daran interessiert sein sollten, sich im Bereich der Reputationssysteme zu differenzieren, um den von eBay.de bekannten Problemen zu begegnen und sich einen Vorsprung zu schaffen, kann nach der Evaluation der sechs Plattformen nicht bestätigt werden. Es stellt sich im Gegenteil heraus, dass die Bewertungssysteme sehr starke Übereinstimmungen in nahezu allen Bereichen der Gestaltung aufweisen. Es besteht große Übereinstimmung zwischen den Anbietern hinsichtlich der Erhebung der Profile, sowie insbesondere auch bei der Aufbereitung Profile. Dies wird auch bei der visuellen Präsentation deutlich, sowohl der Kurzprofile auf der Auktionsseite, als auch bei der Aufteilung der ausführlichen Bewertungsseite. Abbildung 1 zeigt die als Beispiel die Screenshots aller Kurzprofile bei den sechs Plattformen. Es wird deutlich, dass die Darstellung der Wettbewerber eng an die von eBay angelehnt ist, auch wenn natürlich die Zahlenwerte und Symbole variieren und oftmals einige Informationen weniger angezeigt werden.

Die Abweichungen zwischen den Plattformen beschränken sich lediglich auf Details und bestehen meist darin, dass die fünf Wettbewerber bei der Umsetzung einiger Funktionen hinter den Marktführer zurückfallen. Nur wenige Funktionen lassen dagegen eine eigenständige Entwicklung der Reputationssysteme von Seiten der Wettbewerber erkennen. Vielmehr kann hier bereits von einem dominanten Design für Reputationssysteme in B2C-Online-Auktionen gesprochen werden.



Abbildung 1: Screenshots der Kurzprofile bei allen sechs Plattformen

Ein dominantes Design bei einem Produkt oder einem Service existiert dann, wenn es sich im Markt so durchsetzt, dass es die Wettbewerber im Markt zur Vereinheitlichung und Imitation zwingt [Aber78] bzw. wenn Neueinsteiger in den Markt sich gezwungen sehen, das Design übernehmen zu müssen [Utte94]. Hat sich ein dominantes Design gebildet, finden Variationen nur noch in engen Grenzen statt [Voß04]. Dominante Designs können z.B. durch Nachahmungseffekte auftreten; dies ist meist dann der Fall, wenn es in einem Markt einen dominanten Spieler gibt, der den Markt beherrscht [Voß04]. Dies ist im Markt der Online-Auktionshäuser gegeben²; das Internet macht zudem durch seine Transparenz und Offenheit das Imitieren von Frontend-Funktionen wie dem Bewertungssystem denkbar einfach.

Gründe für die Imitation des Reputationssystems durch die Wettbewerber können einerseits im Reduzieren der Unsicherheit liegen; der Betreiber muss nicht mit neuen Mechanismen experimentieren, wenn der dominante Spieler das Funktionieren bereits erfolgreich demonstriert. An-

² Die Anwendbarkeit des Konstrukts auf die E-Commerce-Domäne wurde von Voß gezeigt [Voß04].

dererseits besteht durch die Dominanz von eBay vermutlich bereits ein konditionierender Effekt auf Seiten der Kunden, d.h. die Bedienung und Interpretation eines Bewertungssystems wurde bereits vom Kunden am Beispiel von eBay gelernt. Weicht der Wettbewerber signifikant im Design ab, riskiert er, nicht anschlussfähig zu sein und so keine Kunden vom Marktführer abwerben zu können. Die bestehenden Abweichungen zwischen den Systemen sind denn auch, der Theorie des dominanten Designs folgend, eher marginal und in den Grenzen eines weitgehend einheitlich interpretierten Systemansatzes zu finden.

4.2 Unterschiede zwischen den Plattformen im Detail

Tabelle 1 macht deutlich, dass eBay in einigen wichtigen Details einen Schritt weiter geht als die Wettbewerber, was als typisches Merkmal eines Marktführers gelten kann. Erst die Kombination von Informationen über den Verkäufer und seine Aktivitäten versetzt den Käufer in die Lage, sich ein umfassendes Bild zu verschaffen und unliebsamen Überraschungen im Rahmen des in dieser Form von Distanz-Handel möglichen aus dem Weg zu gehen. Zu diesen Informationen gehören insbesondere die um mehrfache Einträge der gleichen Handelspartner bereinigte Gesamtzahl positiver Bewertungen, Informationen zu den Produkten und der Wertigkeit der Transaktionen, die eine Bewertung ausgelöst haben, sowie besonders die Profile der Bewerter, um auf diese Weise das künstliche Aufbauen von Profilen besser erkennen zu können. Hier und insbesondere bei Art und Umfang der Kommunikation zum Bewertungssystem zeigt sich bei den Wettbewerbern noch Nachholbedarf.

Neben diesen Unterschieden lassen sich aber auch bei den Wettbewerbern einige gute Ideen finden, die das von eBay gebotene Reputationssystem in wichtigen Details verbessern könnten (siehe Tabelle 2). Dabei ist insbesondere das gezielte Filtern nach negativen Bewertungen zu nennen, das eBay unverständlicherweise tief in der Navigation versteckt, sowie der Zwang zur Identifikation mittels Postadresse, Telefonnummer oder auch Ausweiskopie. Auch das von A-zubo.de gebotene differenzierte Bewertungsurteil könnte dazu beitragen unliebsame Verkäuferpraktiken aufzudecken, die einen Käufer bei einem einfachen Kategoriensystem aber nicht zu einer schlechten Bewertung veranlassen würden.

Mechanismus/Maßnahme	Bemerkung
Bewertungen werden nur von unterschiedlichen Nutzern gezählt	Dies ist eine wichtige Voraussetzung, dem Problem der Freundschaftsbewertungen zum künstlichen Aufbau von Profilen zu begegnen. Für Wettbewerber mit nur wenigen Nutzern in einzelnen Produktgruppen der Plattform besteht jedoch das Problem, dass sich Profile dann deutlich langsamer aufbauen und das Erreichen der kritischen Masse verzögern können.
Transaktion mit Art und Wertigkeit des Produkts ist bei der Bewertungshistorie zugänglich/ersichtlich	Die Wertigkeit und die Art des Produkts gibt wesentlichen Aufschluss über die Aktivitäten des Verkäufers: Wurden viele niedrigpreisige Artikel gekauft/verkauft, um das Profil aufzubauen? Hat der Verkäufer bisher mit ganz anderen Artikeln gehandelt? Dies kann ein Hinweis auf eine geklaute Identität oder unlautere Absichten des Verkäufers sein.
Profil des Bewertenden ist bei der Bewertungshistorie zugänglich/ersichtlich	Auf diese Weise kann erkannt werden, wenn ein Nutzer sehr viele Bewertungen von Neu-Nutzer bekommt, die nur angelegt wurden, um das Profil künstlich zu erhöhen. Zudem kann der Käufer in den Profilen negativ Bewertender ersehen, ob der Verkäufer Rachebewertungen vergibt.
Anzahl zurückgezogener Gebote	Wie verhält sich der Verkäufer als Käufer? Auch dies kann einen Hinweis auf die Persönlichkeit des Verkäufers geben.
Ausführliche Verhaltensregeln zum Bewertungssystem vorhanden	Die Aufklärung der Nutzer ist eine wesentliche Voraussetzung zur Vermeidung von Betrugsfällen; hier sieht sich eBay als Marktführer vermutlich mit den meisten Problem konfrontiert und ist deshalb hier sehr aktiv.

Tabelle 1: Maßnahmen, die eBay den meisten Wettbewerbern voraus hat

Mechanismus/Maßnahme	Bemerkung
Pflicht-Identifikation per Post, Telefon oder Girokonto	Eine sichere Identifikation der Nutzer hilft Mehrfachidentitäten zu unterbinden und unehrliche Nutzer dauerhaft von der Plattform auszuschließen. Allerdings ist das für einen Marktführer wie eBay mit erheblichen Kosten verbunden und könnte kurzfristig wachstumshemmend wirken.
Möglichkeit einfach und schnell nach negativen Bewertungen zu filtern	Dies ist eine Voraussetzung, sich ein umfassendes Bild vom Verkäufer zu machen. Dies in Verbindung mit dem Zugang zum Profil des Bewertenden hilft Rachebewertungen aufzudecken.
Differenzierung des Bewertungsurteils in mehreren Dimensionen (siehe Azubo.de)	Eine Differenzierung nach dem Verhalten des Verkäufers (Kommunikation und Versand), sowie der Produktqualität erlaubt eine bessere Abschätzung des Verkäufers. Es erlaubt zudem die Vergabe kritischer Urteile ohne gleich ein komplett negatives Urteil abzugeben, was viele Nutzer vermeiden wollen.
Anreize zur (schnellen) Abgabe von Bewertungen	Da sich Nutzer mit Bewertung Zeit lassen, kann ein Fehlverhalten lange Zeit unentdeckt bleiben.
Keine Zeitbeschränkung für das Abgeben von Bewertungen	Die künstliche Zeitbeschränkung führt zu Taktiken wie das Warten bis zur letzten Sekunde, um dann negativ zu bewerten ohne ein Rachebewertung fürchten zu müssen.

Tabelle 2: Maßnahmen von (einzelnen) Wettbewerbern, die über die Funktionen von eBay hinausgehen

4.3 Grenzen und Risiken der bestehenden Reputationssysteme

Das Design der untersuchten Reputationssysteme weist gemeinsame Grenzen und Risiken auf, von denen einige bereits oben am Beispiel von eBay erwähnt wurden. Eine abschließende Auflistung aller Probleme kann im Rahmen dieser Arbeit nicht geleistet werden, so dass hier nur einige typische Probleme diskutiert werden.

Die Möglichkeit der Mehrfachidentitäten durch Lücken im Anmeldeprozess wird häufig als Schwäche von eBay genannt [FrRe99], sie trifft aber genauso für die meisten der Plattformen zu, auch wenn diese bei der Identifikation der Nutzer teils einen Schritt weiter gehen. Des Weiteren können sich Mitglieder ein positives Profil erstellen, indem sie sich gegenseitig gute Bewertungen zuschanzen [Dell00, BhGo05]. Daher zählt eBay lediglich solche Bewertungen, die von unterschiedlichen Nutzern abgegeben wurden; dies erhöht zumindest die Kosten eines solchen Vorgehens. Ein weiteres Problem ist, dass das Bewertungsprofil eines Verkäufers unter Umständen seine aktuelle Servicequalität nicht widerspiegelt. Aufgrund vieler paralleler Transaktionen sind nicht zu jeder Zeit alle Bewertungen immer auch berücksichtigt – viele Nutzer lassen sich mit der Abgabe der Bewertung einige Wochen Zeit. Die von den Betreibern geforderte Klärung von Unstimmigkeiten verzögert zudem eventuelle negative Bewertungen, so dass der Verzögerungseffekt für Käufer unangenehme Folgen haben kann.

Ein weiteres Problem der Systeme liegt in der Möglichkeit des Rufmords oder der ungerechtfertigten Rachebewertung. Die Betreiber erlauben zwar das Abgeben einer Stellungnahme, eine Löschung von Bewertungen ist jedoch nur in Ausnahmefällen und nur nach einer langwierigen Prozedur möglich; bei eBay nur nach gegenseitigem Einverständnis der Parteien. So bleibt die Gefahr, dass Käufer unter Druck gesetzt oder Verkäufer von unliebsamen Wettbewerbern in ihrem Ruf geschädigt werden [Dell00]. Dies könnte verhindert werden, wenn z.B. die Urteile der jeweils anderen Partei erst nach Abgabe der eigenen Bewertung gezählt würden.

5 Fazit

Die Ergebnisse der Studie deuten auf ein dominantes Design bei der Gestaltung von Reputationssystemen hin, das darüber hinaus von eBay bestimmt wird. Nicht nur, dass sich die Wettbewerber nicht entscheidend von eBay abheben können, sie bleiben bei wichtigen Zusatzfunktionen hinter eBay zurück, so dass sich hier die Marktführerschaft nicht nur an Umsatzgrößen und Transaktionsvolumina festmacht, sondern auch durch einen Führungsanspruch bei der Gestal-

tung wichtiger Teilbereiche der Plattform untermauert wird. Die Tatsache, dass der Markt auf eBay.de funktioniert, zeigt, dass die Mehrheit der Nutzer Vertrauen in das System und in eBay als Betreiber haben [RZFK00, ReZe01]. Das Bewertungssystem von eBay erfüllt seine Aufgabe, obwohl es nicht optimal ist. Es vermittelt den Nutzern das Gefühl, dass die Plattform ein zuverlässiger Vermittler zwischen Käufern und Verkäufern ist. Dennoch hat das Online-Reputationssystem seine Grenzen, da es die Vertrauensmechanismen aus dem traditionellen Markt nicht vollständig zu kopieren vermag [BoKO04]. Hier bleibt einiges an Raum für die zukünftige Forschung zu Reputationssystemen; Cabral und Hortacsu bewerten das Reputationssystem von eBay denn auch verhalten positiv: „Obviously, the fact the reputation system has bite does not imply that it’s current structure is optimal. In fact, we believe an exciting area for future research is precisely the design of an efficient reputation mechanism.“ [CaHo05]

Literaturverzeichnis

- [Aber78] Abernathy, William J.: The Productivity Dilemma. Baltimore, 1978.
- [BhGo05] Bhattacharjee, Rajat; Goel, Ashish: Avoiding Ballot Stuffing in eBay-like Reputation Systems. <http://www.stanford.edu/~ashishg/papers/ebay.pdf>, Abruf 2006-06-07.
- [BoKO04] Bolton, Gary E.; Katok, Elena; Ockenfels, Axel: How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. In: Management Science 50 (2004) 11, S. 1587-1602.
- [CaHo05] Cabral, Luis; Hortacsu, Ali: The Dynamics of Seller Reputation: Evidence from eBay. <http://www.colorado.edu/Economics/seminars/cabral.pdf>, Abruf am 2006-06-29.
- [Dell00] Dellarocas, Chrysanthos: Immunizing Online Reputation Systems Against Unfair Ratings and Discriminatory Behavior. <http://ccs.mit.edu/dell/ec00reputation.pdf>, 2000- 10-17, Abruf am 2006-06-06.
- [Dell03] Dellarocas, Chrysanthos: The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. In: Management Science 49 (2003) 10, S. 1407-1424.

- [Dell04] Dellarocas, Chrysanthos: Building Trust Online: The Design of Robust Reputation Reporting Mechanisms for Online Trading Communities. <http://ccs.mit.edu/dell/papers/ideabook.pdf>, Abruf am 2006-06-07.
- [FrRe99] Friedman, Eric J.; Resnick, Paul: The Social Cost of Cheap Pseudonyms. <http://www.si.umich.edu/~presnick/papers/identifiers/081199.pdf>, 1999-08-11, Abruf am 2006-06-06.
- [Koll99] Kollock, Peter: The Production of Trust in Online Markets. http://www.sscnet.ucla.edu/soc/faculty/kollock/papers/online_trust.htm, Abruf am 2006-06-07.
- [LRBR00] Lucking-Reily, David; Bryan, Doug; Reeves, Daniel: Pennies from eBay: The Determinants of Price in Online Auctions. <http://www.econometricsociety.org/meetings/wc00/pdf/1736.pdf>, Abruf am 2006-06-06.
- [MeAl02] Melnik, Mikhail I.; Alm, James: Does a Seller's Ecommerce Reputation Matter? Evidence from eBay Auctions. In: *The Journal of Industrial Economics* 3 (2002), S. 337-349.
- [RZFK00] Resnick, Paul; Zeckhauser, Richard; Friedman, Eric; Kuwabara, Ko: Reputation Systems. In: *Communications of the ACM* 43 (2000) 12, S. 45-48.
- [ReZe01] Resnick, Paul; Zeckhauser, Richard: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. <http://www.si.umich.edu/~presnick/papers/ebayNBER/RZNBERBodegaBay.pdf>, 2001-05-02, Abruf am 2006-06-06.
- [ShTa05] Shmatikov, Vitaly; Talcott, Carolyn: Reputation-based trust management. In: *Journal of Computer Security* 13 (2005), S. 167-190.
- [Utter94] Utterback, James M.: *Mastering the Dynamics of Innovation*, Boston, 1994.
- [Voß04] Voß, Andreas: *Dominantes Design im Electronic Commerce – Analysen und Befunde bei Tourismus-Web Sites*, Lohmar, 2004.

Anhang

	eBay.de	hood.de	auxion.de	BesteAuktion	ricardo.ch	azubo.de	
Bewertungsphase	Wer ist zur Abgabe berechtigt?	Käufer und Verkäufer	Käufer und Verkäufer	Käufer und Verkäufer	Käufer und Verkäufer	Käufer und Verkäufer	
	Anreiz / Entschädigung zur Abgabe einer Bewertung	nein	nein	nein	ja, für die Abgabe erhält der Nutzer einen Viertel-Punkt	nein	nein
	Ist die Bewertung verpflichtend	nein	nein	bewertet, erhält der Verkäufer automatisch eine positive Bewertung nach einer	nein	nein	nein
	Zusammensetzung der Bewertung	Drei Kategorien: positiv/neutral/negativ plus Kurzkommentar	Drei Kategorien: positiv/neutral/negativ plus Kurzkommentar	Drei Kategorien: positiv/neutral/negativ plus Kurzkommentar	Drei Kategorien: positiv/neutral/negativ plus Kurzkommentar	Drei Kategorien: positiv/neutral/negativ plus Kurzkommentar	Beantwortung von drei Fragen zum Verhalten des Verkäufers (Antworten gut/mittel/schlecht)
	Wie geht die Bewertung ins Profil ein?	Im Punktwert (Zahl) werden nur Bewertungen unterschiedlicher Mitglieder gezählt, Addition bei positiv, Subtraktion bei negativ	Im Punktwert (Zahl) werden alle Bewertungen gezählt, Addition bei positiv, Subtraktion bei negativ	Punktwerte der Kategorien werden einzeln ausgewiesen, alle Bewertungen werden gezählt	Punktwerte der Kategorien werden einzeln ausgewiesen	Berechnung eines Prozentwertes der positiven Bewertungen	Umrechnung der Beantwortung der drei Fragen in eine Gesamtwertung der Kategorien positiv/neutral/negativ
	Kann eine bereits abgegebene Bewertung modifiziert werden?	nein, Löschung nur in Ausnahmefällen und bei ggs Einverständnis	nein, Löschung nur in Ausnahmefällen (Beleidigungen, Werbung etc.)	nein	nein	nein, Löschung nur in Ausnahmefällen (Beleidigungen, Werbung etc.)	nein
	Möglichkeit einer Antwort	ja	ja	ja	ja	ja	ja
	Bewertungsregeln: Wie wird die Abgabe ehrlicher Urteile gefördert?	Ausführliche Verhaltensregeln	Ausführliche Verhaltensregeln	Sehr zurückhaltend / keine Angaben	Sehr zurückhaltend / keine Angaben	Ausführliche Verhaltensregeln	Ausführliche Verhaltensregeln
Nutzungsphase	Besteht die Möglichkeit, einzelne Bewertungen auszublenden?	Nutzer kann sein Profil als "privat" kennzeichnen und so alle Kommentare ausblenden	nein	nein	nein, wird aber nicht explizit erwähnt	nein	nein
	Visualisierung der Bewertungsanzahl	Zahl plus verschiedenfarbiger Stern (ab 10 positive), sowie Prozentwert, dazu für sehr umsatzstarke Verkäufer ein "Powerseller"-Symbol	Zahl plus bis zu 7 Sterne je nach Anzahl Bewertungen (ab 5 positive), sowie Prozentwert	Anzahl pos/neutr/neg	Anzahl pos/neutr/neg Bewertungen, sowie Sterne und Lorberkränze in Gold/Silber/Bronze	Prozentwert plus Kombination aus Anzahl und Farbe der Sterne (5 Farben und jeweils bis zu 4 Sterne), zus. Diamantensymbole, wenn Prozentsatz > 99%	Zahl plus bis zu fünf Sterne in drei Größen (ab 5 positive)
	Möglichkeiten der Bewertungseinsicht	Per Klick auf Bewertungszahl oder Link zum detaillierten Profil	Per Klick auf Bewertungszahl oder Link zum detaillierten Profil	Per Klick auf "Bewertungen des Verkäufers ansehen" zum detaillierten Profil	Klick auf Nutzernamen führt zu Profil	Klick auf Nutzernamen führt zum detaillierten Profil	Klick auf Nutzernamen führt zu Profil
Wie wird die Bewertungs-Historie dargestellt?	Tab. Aufschlüsselung nach pos/neutr/neg und Zeitraum, Bewertungskommentare mit Urteil, Kurzprofil des Bewertenden, Zeitpunkt und Link zum Artikel, sowie Angabe (K/V), zurückgezogene Gebote als zus. Information	Tab. Aufschlüsselung nach pos/neutr/neg und Zeitraum (zus. Balkendiagramm), Bewertungskommentare mit Urteil, Kurzprofil des Bewertenden, Zeitpunkt und Link zum Artikel, sowie Angabe (K/V)	Aufschlüsselung nach pos/neutr/neg und Liste der Kommentare mit Urteil, Pseudonym des Bewertenden, Zeitpunkt und Link zum Artikel, sowie Angabe (K/V) (Profil des Bewertenden nicht ersichtlich)	Profilseite besteht im Wesentlichen aus der Liste der Kommentare; Bewertungskommentare mit Urteil, Kurzprofil des Bewertenden und Zeitpunkt (Kein Link zum Artikel, keine Angabe (K/V))	Aufschlüsselung nach Zeitraum und pos/neutr/neg, sowie Liste der Kommentare; Bewertungskommentare mit Urteil, Kurzprofil des Bewertenden und Zeitpunkt, sowie Angabe (K/V) (Kein Link zum Artikel)	Visuelle Aufschlüsselung nach pos/neutr/neg (Balkendiagramme) und nach den Bewertungsfragen (s.o.), sowie Liste der Kommentare; Bewertungskommentare mit Urteil, Kurzprofil des Bewertenden und Zeitpunkt, sowie Angabe (K/V) (Kein Link zum Artikel)	

	eBay.de	hood.de	auxion.de	BesteAuktion	ricardo.ch	azubo.de	
Nutzungsphase	Gibt es Bewertungsfilter, anhand derer man in dem Profil suchen kann?	Kriterien: "Von Käufern/Verkäufern", Filterung nach Zeitraum, Filterung nach pos/neg ist versteckt und nur zugänglich nach Filterung auf Zeitraum	nein, lediglich Auflistung	Filterung nach positiven/negativen Bewertungen, sowie nach "erhalten" und "abgegeben"	Filterung nach positiven/neutralen/negativen Bewertungen	Filterung nach positiven/negativen Bewertungen, sowie nach "abgegeben"	Nein, keine
	Zusätzliche Informationen zur Reputation	"Angemeldet seit", "privat/Gewerblich", Seite "über mich", Symbol für verifizierte Benutzer	"Angemeldet seit", "privat/Gewerblich", Seite "über mich"	"Mitglied seit", "Info-Seite des Verkäufers", "Abgegebene Bewertungen" auf der Profelseite, Hammersymbol für verifizierte Benutzer	Keine	"Mitglied seit", aber nur im detaillierten Profil	"privat/gewerblich", Seite "über mich".
	Gibt es verifizierte Benutzer?	optional, über das Postident-Verfahren der Deutschen Post	nein, nur eMail-Verifikation	optional, mit Ausweiskopie	immer, aber nur durch Mitteilung einer PIN über Telefon	immer, Aktivierungscode per Post	optional, Kontocheck mit Kleinstüberweisung
	Wird das Profil immer angezeigt oder nur bei Verkäufern?	immer	immer	Profil wird nur bei einem angebotenen Artikel angezeigt	immer	immer	immer
Allgemeines	Ist das System eher positiv oder eher negativ ausgerichtet?	Sicherheitsabfrage vor negativer/neutraler Bewertung. Eindringliche Hinweise möglichst nicht negativ zu bewerten. Unterliegende Philosophie von eBay geht von positivem Menschenbild aus.	Eher positiv, da Aufforderung der Verhandlung vor Abgabe einer negativen Bewertung.	Eher positiv, da nicht-abgegebene Bewertungen in positive umgewandelt werden	neutral	Ricardo.ch fordert die Mitglieder zum Dialog auf und will so negative Bewertungen verhindern	neutral
	Wo gibt es Informationen zum Bewertungssystem?	Unter dem Link "Hilfe" - Bewertungen (sehr ausführliche Grundsätze und Regeln)	wie eBay, Informationen aber sehr knapp	Unter dem Link "Hilfeseite" - Bewertungen (sehr versteckt), Informationen sehr knapp	Unter dem Link "Hilfe & Tipps" - Sonstiges, sehr knappe Informationen	Unter dem Link "Hilfe zu dieser Seite" - Bewertungssystem (recht ausführlich)	Unter dem Link "Hilfe" - Extras - Bewertungsschema, Informationen sind sehr knapp
	Besonderheiten des Systems	Neue Mitglieder müssen vor ihrer ersten neutralen oder negativen Bewertung eine kurze Lern-Tour absolvieren. Alle Bewertungen von Nutzern die vor Ablauf von 90 Tagen ausgeschlossen werden, gehen nicht in die Profile ein.		Übernimmt aus 3 anderen Auktionshäusern die Bewertungen. Bezahlt ein Bieter nicht wird ein Symbol eingeblendet, bei Wiederholung erfolgt Sperre. Verkäufer, die Ware nicht versenden werden gesperrt, bis Sachlage geklärt ist.		Ohne Aktivierung kann nur bei max. einer Auktion geboten werden.	Möglichkeit der Nichtbewertung ist gegeben, wenn z.B. der Kauf in ggs Einvernehmen nicht zustande kam.

Tabelle 3: Ausführliche Bewertung der sechs Reputationssysteme

Einführung in den Track

eMedia

Prof. Dr. Thomas Hess

LMU München

Prof. Dr. Matthias Schumann

Universität Göttingen

Dr. Konrad Hilbers

Home Shopping Europe

Der Track fokussiert die folgenden Themengebiete:

- Rechteschutzsysteme
- Interaktives Fernsehen
- Tauschbörsen für Musik- und Videoinhalte
- Wikis und nutzergenerierte Inhalte
- Digitalisierung von Produktionsprozessen
- Applikationsarchitekturen für Medienunternehmen

Programmkomitee:

Prof. Dr. Peter Buxmann, Technische Universität Darmstadt

Prof. Dr. Michel Clement, Universität Hamburg

Prof. Dr. Martin Grauer, Universität Siegen

PD Dr. Hans-Dieter Groffmann, Universität Frankfurt am Main

Prof. Dr. Detlef Schoder, Universität zu Köln

Prof. Dr. Gerd Schwabe, Universität Zürich

Intermediation in der TV-Branche: TV-Sender als Auslaufmodell?

Thomas Wilde

Institut für Wirtschaftsinformatik und Neue Medien
Ludwig-Maximilians-Universität München
D-80539 München
wilde@bwl.uni-muenchen.de

Konrad Hilbers

Home Shopping Europe GmbH & Co KG
D-85737 Ismaning
k.hilbers@hse24.de

Thomas Hess

Institut für Wirtschaftsinformatik und Neue Medien
Ludwig-Maximilians-Universität München
D-80539 München
thess@bwl.uni-muenchen.de

Abstract

TV-Sender befinden sich heute in einer strategischen Situation, die derjenigen der Musikindustrie vor einigen Jahren nicht unähnlich ist: Interactive TV, Personal Video Recorder und andere Technologien stellen die Rolle des Intermediärs als Vermittler zwischen Angebot und Nachfrage grundsätzlich infrage – ganz ähnlich wie Tauschbörsen und digitale Endgeräte vor ein paar Jahren in der Musikindustrie. Der vorliegende Artikel analysiert diese Entwicklung mithilfe der Intermediärstheorie und leitet daraus Thesen zur Entwicklung zur Wertschöpfung im deutschen TV-Sektor ab. Im Gegensatz zur Musikindustrie zeichnet sich dabei ein Bedeutungszuwachs bei den Intermediären – namentlich von TV-Sendern und Infrastruktur-anbietern – ab. In Experteninterviews fanden diese Thesen weitgehend Bestätigung.

1 Einleitung: Status-Quo im TV-Sektor

Digitales Fernsehen, interactiveTV, Video-On-Demand und Electronic Program Guides (EPGs) sind konvergente Dienste, die als Folge der Digitalisierung von Inhalten und Übertragungswegen den TV-Sektor verändern sollen – und das seit etwa 10 Jahren. Für die Branche zeichneten sich bis vor ca. 2 Jahren jedoch kaum nennenswerte Änderungen ab. Pilotprojekte von Microsoft, Intel oder AOL/TimeWarner scheiterten an zu schwacher Infrastruktur, fehlenden Standards oder nicht kundengerechten Diensten. Die Umstellung auf digitale Verbreitungswege wurde in Deutschland nur zögerlich vorangetrieben. Mit der Diffusion von Festplattenrekordern, Triple-Play (Telefon, Fernsehen und Internet aus einer Hand), IPTV oder HDTV scheint der Konvergenzprozess nun Fahrt aufzunehmen [CISA05]. Infrastruktur- und Internetprovider wie Kabel Deutschland, Unity Media oder T-Online und Arcor bieten entsprechende Produkte an. Welche Bedeutung hat diese Beschleunigung des Konvergenzprozesses nun für TV-Sender? Auf den ersten Blick lassen sich Bündelung und Distribution durch die konsequente Digitalisierung einfacher abwickeln, was zur Entstehung neuer und zur Substitution vorhandener Dienste führen kann. In einer solchen Situation befand sich die Musikbranche etwa um 2000 [BeCS06]. Die Digitalisierung und Kompression von Content im MP3-Format sowie neue, digitale Übertragungswege im Internet (Peer2Peer-Tauschbörsen, Musikportale) beeinträchtigen das traditionelle Geschäftsmodell von Musiclabels massiv. Musiclabels begegneten dem Problem lange nur auf rechtlicher und technischer Ebene [Scho00], um ihre angestammte Position zu verteidigen – nur mit mäßigem Erfolg. Vor diesem Hintergrund stellt sich die Frage, ob der TV-Branche ähnliche Strukturänderungen bevorstehen. Der Fokus dieses Beitrags liegt auf auf kommerziellen TV-Content-Anbietern, der Sachverhalt wird beispielhaft für den deutschen Markt analysiert.

2 Problemstellung und Analyserahmen

Bezüglich dieser Fragestellung sind drei Teilfragestellungen zu unterscheiden: Welche konvergenzinduzierten Dienste entstehen im TV-Sektor? Wie verändert sich die Wertschöpfungsstruktur der TV-Branche dadurch? Und welche Bedeutung haben diese Veränderungen für TV-Sender als Vermittler zwischen Contentproduzenten und Contentrezipienten? Zu den Themenbereichen finden sich sowohl segmentübergreifende als

auch medienspezifische Analysen, sowohl aus ökonomischer als auch technologieorientierter Perspektive. Einen Überblick gibt Abbildung 1.

	Segmentunabhängig	spezifisch für die Medienbranche
Ökonomische Perspektive	<ul style="list-style-type: none"> • Intermediation [Fing97;AlSa98] • Branchenkonvergenz [Stie04; Yoff97] 	<ul style="list-style-type: none"> • Intermediation in der Medienbranche [Pica02, 34-49; ScHe06, 211-214] • Medientypspezifische Konvergenzanalysen [LöFa02; Thie00]
Technologische Perspektive	<ul style="list-style-type: none"> • Änderung der Intermediationsstruktur durch Digitalisierung, z.B. [ChKa99] 	<ul style="list-style-type: none"> • Implikationen von technologischer Konvergenz in der Medienbranche [Schm05; Clem04] • Technologiegetriebene Anwendungs-konzeption, z.B. [QiSX05]

Abb. 1: Stand der Forschung

Der vorliegende Beitrag adressiert das im rechten unteren Quadranten adressierte Themenfeld. In der aufgeführten Literatur wurde das grundsätzliche Verständnis von Medienunternehmen als Intermediär herausgearbeitet und das Aufgabenspektrum für spezifische Kontexte (z. B. für die Musikindustrie) beschrieben. Neben den Konsequenzen der Digitalisierung für Intermediäre im Allgemeinen werden unter dem Schlagwort „Konvergenz“ technologisch getriebene Transformationsprozesse konzeptionell analysiert. Bei der Wahl des Analyserahmens ist im vorliegenden Fall zunächst das Problem zu lösen, dass viele bewährte Konzepte, wie das klassische Porter'sche Instrumentarium (Wertschöpfungsketten, Branchenstrukturanalyse, etc.), bei konvergierenden Branchen und Produkten zum Teil an ihre Grenzen stoßen [WaHe05, 14-16]. Da Medienunternehmen generell, wie auch TV-Sender im Speziellen, neben ihrer Rolle als Contentproduzenten vor allem als Vermittler bzw. Intermediär auf Medienmärkten agieren, scheint mit der Contentintermediation (CI) [WaHe05] ein vergleichsweise junges Konzept passend, welches sich bereits bei der Beschreibung und Erklärung der Wettbewerbsveränderungen in der Musikbranche bewährt hat [HeWa06].

Das Intermediationskonzept hat seine Wurzeln in der Mikroökonomie und postuliert die Einschaltung von Intermediären, falls die so anfallenden Kosten geringer sind als bei direktem Handel zwischen Produzenten und Nachfragern [Fing97]. Im Handel und in der Finanzindustrie wird es seit längerem erfolgreich als Analyserahmen eingesetzt. Auch zur Analyse der Wirkung digitaler Technologien auf die Intermediation hat es sich bereits bewährt [Scho00]. Das Konzept der Contentintermediation [HeWa06] knüpft an diese beiden Überlegungen an und passt das Konzept auf die Medienbranche an, insbesondere auf deren Digitalisierung. In diesem Konzept werden Contentintermediäre als Träger von Intermediationsfunktionen aufgefasst, die Koordinationsprobleme auf definitionsgemäß imperfekten Märkten zwischen Produzenten von

Content (z. B. Filmstudios) und Nutzern von Content (z. B. privaten Filmkonsumenten) am transaktionskosteneffizientesten lösen.

Das aus der Mikroökonomie stammende Intermediärskonzept sieht originär ein recht abstraktes Kostenkalkül vor, welches aber auf Kataloge mit einzelnen Intermediärfunktionen heruntergebrochen werden kann (z.B. [AlSa98]). So existiert auch für die Medienindustrie ein erster Vorschlag für einen solchen Funktionskatalog [HeWa06, 5-7]:

Funktion	Beitrag der Funktion zur Bedürfnisbefriedigung
<i>Identifikation</i>	Die Identifikation umfasst das Finden der Grundgesamtheiten von Contentangebot und Contentnachfrage und ermöglicht einen Überblick.
<i>Selektion</i>	Die Selektion umfasst die Auswahl aus den Grundgesamtheiten und stellt die Eignung des Contentangebots für die Contentnachfrage sicher.
<i>Transformation</i>	Die Transformation ist die Funktion, die die Verwendbarkeit des Content gewährleistet.
<i>Aggregation</i>	Die Aggregation stillt das Bedürfnis nach Anordnung der einzelnen Inhalte.
<i>Reproduktion</i>	Die Reproduktion umfasst die Vervielfältigung des Content nach der „First-Copy“ und sichert so die Verfügbarkeit des Content.
<i>Distribution</i>	Die Distribution ist die Funktion, die die Erreichbarkeit des Content für die Nachfrage sicherstellt.
<i>Präsentation</i>	Die Präsentation ist die Darstellung des Content gegenüber der Nachfrage und ermöglicht so den Zugang zum Content.

Abb. 2: Intermediärfunktionen in der Medienbranche

Zur Anwendung des Konzepts auf die TV-Branche wird in Abschnitt drei zunächst der Ist-Zustand in der TV-Branche aus Sicht der Contentintermediation dargestellt. Abschnitt vier systematisiert die Ergebnisse bisheriger Arbeiten zu konvergenzinduzierten Innovationen im TV-Bereich und diskutiert deren Bedeutung für die Intermediation von TV-Content. Durch die Zusammenführung dieser Ergebnisse können in Kapitel fünf Implikationen für die Wettbewerbsstruktur durch eine relative Analyse der Wirkung neuer Technologien auf die Transaktionskosten abgeleitet werden. Die Ergebnisse werden in Abschnitt sechs mit Experteninterviews validiert.

3 Intermediation in der TV-Branche

Das Konzept der Contentintermediation, das zur Analyse der zukünftigen Rolle von TV-Sendern verwendet werden soll, ist neu und daher in der Literatur noch nicht etabliert. Dagegen hat sich ein recht einheitliches Verständnis der typischen Aufgaben eines TV-Senders und unterstützenden Dienstleistern herausgebildet [Wirt05, 351-368; Sjur02, 138; Born04, 7-8;

Herm02, 30-44]. Nachfolgend wird diese Literatur aufgearbeitet, um einen speziellen Katalog an Intermediärsfunktionen für die TV-Branche aus dem allgemeinen Katalog der Intermediärsfunktionen in der Medienbranche zu entwickeln.

Gegenstand der *Identifikationsfunktion* ist das Screening der Beschaffungsmärkte für TV-Contentmodule. TV-Sender erbringen diese Leistung im Rahmen der Contentbeschaffung und werden hierbei von Rechthändlern unterstützt. Output dieser Funktion ist eine Übersicht über am Markt verfügbaren TV-Contentmodule, die grundlegenden Qualitätsstandards entsprechen. Damit sind sie Input für den weiteren Prozess, für jedes Modul existiert also prinzipiell eine Zielgruppe. Hauptkostenblöcke sind klassische Suchkosten, in diesem Fall Kosten für Messebesuche, das Beobachten der Konkurrenz und die Qualitätsbewertung, die aufgrund des Erfahrungsgutcharakters von TV-Content [LöFa02] relativ aufwändig ist. Transaktionskosten entstehen bei der Kommunikation der Marktübersicht, genauer an dem Punkt, an dem die Qualität und grundsätzliche Eignung des Content dem Selektierer zu versichern ist. Diese Eigenschaft ist schwer objektiv beschreibbar, aber im Rahmen der Selektion von essentieller Bedeutung.

Die *Selektionsfunktion* wählt Contentmodule aus dem bereits identifizierten Bestand zielgruppengerichtet aus. Dies wird ebenfalls von der TV-Senderfunktion Contentbeschaffung abgedeckt und durch senderspezifische Programmportfoliostrategien gesteuert. Die Aggregationsstrategie stellt hierbei eine Nebenbedingung dar, da Contentmodule teilweise nur für bestimmte Sendezeiten geeignet sind. Output ist eine entsprechende Auswahl von Contentmodulen. Kostentreiber dieses Schritts ist einerseits der hohe Aufwand zur Ermittlung der Rezipientenpräferenzen bezüglich des TV-Content [RuNi97, 220-227] und andererseits die entsprechende Bewertung und Klassifikation der identifizierten Contentmodule. Transaktionskosten entstehen in geringem Umfang bei der Übermittlung der Bezugsdaten der ausgewählten Module an den Akteur, der die Transformation durchführt. Der Abstimmungsaufwand mit dem Aggregator hingegen ist enorm hoch: da die Programmportfolioplanung sowohl Timing- als auch Anordnungsaspekte umfasst, entstehen für die Entscheidungen in Selektions- und Aggregationsfunktion jeweils wechselseitig geltende Restriktionen.

Die *Transformation* stellt einen technischen Aufbereitungsschritt dar, der Formatanpassungen (z. B. Kompression, Bildseitenverhältnis, Tonkanäle), Einblendungen, Kürzungen und ähnliches umfasst. In der TV-Branche wird hierunter die Postproduktion verstanden, die entweder durch den TV-Sender selbst, oft aber durch technische Dienstleister abgewickelt wird.

Output sind technisch aufbereitete Contentmodule, Kosten entstehen bei digitaler Verarbeitungstechnik primär hinsichtlich IT. Transaktionskosten entstehen bis auf den Austausch des Content nahezu keine, da die Aufgabe stark standardisierbar ist.

Gegenstand der *Aggregation* ist die Anordnung der aufbereiteten Contentmodule und damit die Produktion des linearen TV-Programmstroms. Diese Erfüllung dieser Funktion ist komplex und umfasst einen hohen redaktionellen Aufwand, da mit Zielen wie der Maximierung der Rezipientenverweildauer, Mehrfachverwendung, Berücksichtigung von Konkurrenzverhalten und ggf. Werbeplatzierung zahlreiche Nebenbedingungen bestehen. Sie stellt die wichtigste Funktion der TV-Sender dar und ist durch den Umfang des notwendigen Know-how nur schwer imitierbar. Die Sender leisten diese Funktion im Rahmen der Contentaggregation und erzeugen als Output einen linearen Programmstrom. Kosten entstehen analog zur Selektion durch die Ermittlung der Rezipientenpräferenzen bezüglich der Anordnung und der entsprechenden Bewertung von Contentmodulen.

Infrastrukturanbieter erfüllen die *Reproduktion* im Rahmen der Contentdistribution [Schr95]: Die Versionierung des Signals für analoge und digitale Verbreitung stellt einen ersten Reproduktionsschritt dar. Anschließend werden die derart aufbereiteten Daten aller zu übertragenden Programmströme im so genannten Multiplexing-Verfahren zu einem Datenstrom zusammengefasst, der in dieser Form übertragen werden kann. Die Kostenstruktur ist ähnlich den Transformationskosten durch IT-Investitionen und entsprechendes Know-how geprägt. Bei der Verteilung von Reproduktion und Distribution auf zwei Akteure würden Transaktionskosten durch die Abstimmung der technischen Spezifika und der Übertragung des Datenmaterials entstehen.

Eng an die Reproduktion gebunden, betrifft die *Distribution* den physikalischen Transport des Signals über Satellit, Kabel, Terrestrik oder neuerdings auch DSL zu den Rezipientenhaushalten. Verstärkung und Konservierung der Signalqualität sind zentrale Aufgaben. Output dieser Funktion ist ein endgerätadäquates Signal. Den enormen Fixkosten, die hier auf Infrastrukturseite entstehen, sind variable Verbreitungskosten nahe Null gegenüberzustellen. Transaktionskosten zu nachgelagerten Funktionen entstehen nur durch die Abstimmung mit Endgerätherstellern, die Vermarktung von Endgeräten und eventuelle Entschlüsselungstechnik.

Die *Präsentation* findet auf zwei Ebenen statt: Zum einen wird das ankommende Signal durch das Endgerät des Rezipienten in rezeptionsfähigen Content umgewandelt (Teil der

Contentdistribution), zum anderen wird der Rezipient beim Konsum des Content durch Aktivitäten des Contentmarketings unterstützt. Hierunter fallen Maßnahmen der Kommunikationspolitik, vor allem die Verbreitung von Programminformationen, typischerweise durch TV-Zeitschriften. Ein zweiter Distributionskanal besteht für diese Informationen, vorwiegend im digitalen Fernsehen, mit EPGs. Kosten entstehen vor allem durch Endgeräte und die Herstellung der Programminformationsprodukte.

Abbildung 3 ordnet die oben herausgearbeiteten Intermediärfunktionen Unternehmen als deren Träger zu. Zu beachten ist, dass die Wertschöpfungskette der TV-Branche auch Filmstudios und andere originäre Contentproduzenten sowie den Rezipienten umfasst. Sie sind jedoch nicht aufgeführt, da sie nicht Träger von Intermediärfunktionen sind.

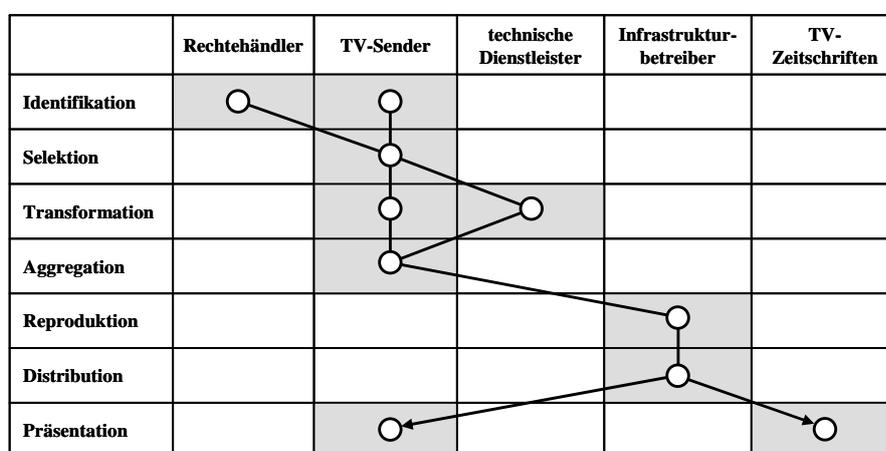


Abb. 3: Heutige Allokation der Intermediärfunktionen in der TV-Branche

So ist – folgt man dem Pfeil und blendet man mögliche regulatorische Eingriffe aus – der „transaktionskostenminimale Weg“ des TV-Content von Produzent zu Rezipient im heutigen technologischen Umfeld abgebildet. Abbildung 3 zeigt auch, dass Selektion, Aggregation, Reproduktion und Distribution nur von einer Klasse von Akteuren abgewickelt werden, während die anderen Funktionen sich auf mehrere Klassen von Akteuren aufteilen lassen. Hintergrund sind technische und inhaltliche Abhängigkeiten zwischen diesen Funktionen, die eine Teilung deutlich erschweren. Dieser Aspekt wird später nochmals aufgegriffen.

4 Technologische Innovationen in der TV-Branche

TV-Sender werden mit einer kaum noch überschaubaren Zahl von neuen digitalen Technologien konfrontiert. Um die angestrebte Analyse der zukünftigen Rolle von Content-

intermediären durchführen zu können, wird als zweiter Baustein ein systematischer Überblick über die anstehenden Technologieschübe ermittelt. Die Systematisierung orientiert sich an der für die Medienbranche zweckmäßigen Unterscheidung zwischen Infrastruktur-, Dienst-, Content- und Endgeräteebene [Thie00, 22-26]. Abbildung 4 zeigt die wichtigsten Technologien im Überblick. Im nachfolgenden Text finden sich Erklärungen dazu.

Endgeräteebene	Personal Video Recorder
Contentebene	Mehrsprachigkeit, Multiperspektiv-TV, interaktive Werbung, Quizzes, Spiele, Wetten, Shopping, weitere Informationsdienste (Wetter etc.)
Dienstebene	Timeshifting, Personalisierung, Video-on-Demand, Electronic Program Guide Mehrkanalton/-video, Informationsdienste, interaktive Applikationen
Infrastrukturebene	DVB, HDTV, Multicasting, IPTV, IP-Rückkanal

Abb. 4: Technologische Innovationen in der TV-Branche

Auf *Infrastrukturebene* sind Innovationen zur Digitalisierung, Adressierbarkeit und Bidirektionalität der Übertragungswege angesiedelt. Über die bestehenden Übertragungswege (Kabel, Satellit, Terrestrik) werden seit Ende der 90er Jahre zunehmend digitale Signale im DVB-Standard (Digital Video Broadcasting) verbreitet [CISA05]. Damit verbunden ist die Erhöhung der maximal übertragbaren Anzahl an TV-Kanälen und die Steigerung der Bildqualität (z. B. HDTV). Möchte man nun über dieses Medium unterschiedliche Informationen an verschiedene Rezipienten übertragen, ist das Problem der Adressierbarkeit zu lösen [Herm02, 34]: Punkt-zu-Punkt-Verbindungen für jeden Rezipienten aufzubauen (Unicasting), würde die Kapazität der vorhanden Netze um ein Vielfaches übersteigen und zudem hohe Verzögerungszeiten nach sich ziehen. Als ein Lösungsansatz verspricht das Multicastingverfahren, Rezipienten gruppenindividuell anzusprechen, indem für jede Gruppe nur ein Datenstrom verschickt und dieser an den entsprechenden Verteilerstellen intelligent vervielfältigt wird. Bidirektionalität wird in Form eines integrierten IP-Rückkanals derzeit in den Kabelnetzen ergänzt. Satellitenbetreiber arbeiten hier mit einem Ergänzungsmedium wie z. B. dem Handy (Zusammenspiel von Set-Top-Box, Bluetooth und Handy im Fall von Astra Bluecom). Neben diesen Neuerungen etabliert sich mit dem DSL-Netz und dem IPTV-Standard eine alternative Distributionsvariante. Der Glasfaserbackbone soll entsprechend ausgebaut werden, um eine hinreichende Anzahl an Punkt-zu-Punkt- und Multicastverbindungen übertragen zu können [CISA05].

Auf *Dienstebene* lassen sich im Kontext TV programm- und programmmodulbezogene Dienste unterscheiden. Programmbezogene, konvergente Dienste betreffen Auswahl und Anordnung der Programmelemente. Innovationen in diesem Bereich zielen auf den Einbezug des Nutzers und erwirken dadurch die Personalisierung des Programms. Hierbei ist davon auszugehen, dass Rezipienten abhängig vom TV-Contenttyp und der Nutzungssituation unterschiedlich intensiv eingreifen möchten, da in gewissem Umfang auch eine Ausrichtung nach vorgefertigten TV-Programmstrukturen stattfindet und der Konsum vieler Inhalte eher impulsiv als geplant verläuft [Schm05, 278; Stip04, 125]. Mit Timeshifting [Clem04], Personalisierungstechniken und Video-On-Demand [HePS05] spannt sich auf technologischer Ebene ein Kontinuum der Eingriffsmöglichkeiten für den Rezipienten auf, das von zeitversetzter Betrachtung über die weitgehend unterstützte Auswahl und Anordnung hin zu individueller Programmzusammenstellung reicht. Diese Unterstützung bzw. stufenweise Aufbereitung des Programms (z. B. durch feiner werdende Klassifikationen) werden im Kontext von Profilbildungs- und Collaborative-Filtering-Techniken seit längerem diskutiert [HePS05, 34; Herm02, 120-121]. Als Bedienungsoberfläche für solche Dienste käme ein Electronic Program Guide infrage, der sowohl Informations- als auch Profil- und Personalisierungsfunktionen umfasst [HePS05, 42]. Bezogen auf einzelne Programmelemente ist die Erweiterung von Contentmodulen mit neuen Funktionen möglich [Born04, 12-14]: Mehrkanalnutzung ermöglicht alternative Ton- und Videospuren, Informationsdienste liefern Metadaten oder verknüpfen den Videostrom mit interaktiven Applikationen.

Hinsichtlich der *Contentebene* wird über Mehrsprachigkeit, freie Perspektivwahl (z. B. bei Sportereignissen), über interaktive Werbung und Product-Placements mit Ad-hoc-Bestellmöglichkeit, sowie über Quiz-, Spiele-, Wett- und Shoppingapplikationen diskutiert [Born04, 15-19; SiRe97]. Allen gemein ist die Erweiterung einer bestehenden Contentform um Zusatzinformationen, zum Teil interaktiv.

Auch auf *Endgeräteebe* finden sich konvergente Innovationen: Personal Video Recorder (PVR), die das TV-Programm auf Festplatte synchron aufzeichnen und so Timeshifting ermöglichen [Clem04], enthalten meist die Funktionalität einer Set-Top-Box und werden zunehmend mit DVD-Recordern und Netzwerkanschlüssen ausgerüstet um eine integrierte Home-Video-Plattform zu schaffen. Mit dieser Tendenz findet eine Produktbündelung statt, deren preisdiskriminatorischer Effekt sich positiv auf die Diffusionsproblematik von TV-Endgeräten auswirken könnte.

5 Implikationen für die Struktur des TV-Sektors

Aus dem Blickwinkel der Intermediationstheorie hat eine Innovation in zwei Fällen Einfluss auf die Struktur einer Branche. Sie kann (1) die Bedeutung von Intermediationsfunktionen verändern, indem sie den Beitrag einer Funktion zur Reduktion der gesamten Transaktionskosten verringert oder erhöht. In der Musikindustrie sind z. B. Selektion und Aggregation durch Suchmaschinen und Empfehlungssysteme deutlich kostengünstiger geworden. Andererseits kann eine Innovation (2) das Zusammenwirken der Träger von Intermediationsfunktionen verändern, wenn sich die Transaktionskosten zwischen den einzelnen Funktionen verändern. In der Musikbranche ermöglichte in diesem Sinne der Einsatz des MP3-Formats zur Kompression der Audiodaten die technische Weiterverbreitung von Songs (wiederholte Selektion, Aggregation, Reproduktion und Distribution) über unterschiedliche Tauschbörsensysteme hinweg bei deutlich reduzierten Transaktionskosten. In Abbildung 5 sind beide Ansatzpunkte an einem Beispiel dargestellt.

Im Folgenden soll diese Betrachtung für die generischen Intermediationsfunktionen durchgeführt werden, der Referenzpunkt ist hierbei die Ist-Situation wie in Abschnitt 3 skizziert.

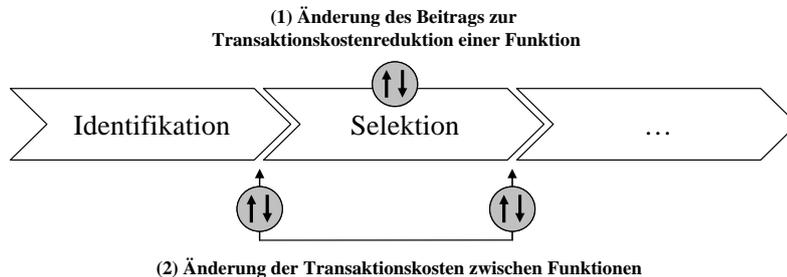


Abb. 5: Einfluss einer technischen Innovation auf die Transaktionskostenstruktur

Die *Identifikationsfunktion*, das Screening der Beschaffungsmärkte gestaltet sich komplexer, da der Markt um Anbieter von interaktiven Applikationen wächst und die Bewertung der featurereicheren Contentmodule (Multichannel, Interaktivität und entsprechender Content) aufwändiger wird. Die Bedeutung der Funktion nimmt zu, die Transaktionskosten in Bezug auf weitere Intermediärsfunktionen ändern sich nicht.

Hinsichtlich der *Selektionsfunktion* wird die zielgruppengerichtete Auswahl zunächst dadurch aufwändiger, dass die zeitliche Beschränkung des Hinkanals durch Personalisierung und Video-On-Demand auf 24 Stunden pro Tag wegfällt. Zudem ist aus dem Angebot an neuen Diensten und Contentformen zu selektieren und diesbezüglich neue Präferenzen der Rezipienten zu

berücksichtigen. Neben der so zunehmenden Bedeutung steigt auch der Abstimmungsaufwand mit dem Aggregator, da neben der Programmportfoliostrategie eine Personalisierungsstrategie zu berücksichtigen ist, die der Aggregator als Profildienstleister maßgeblich bestimmt. Ein weiterer Aspekt ist diesbezüglich die Ausgestaltung der Ausstrahlungsrechte, die bisher auf tradierten Strukturen beruht. Da die tatsächliche Anzahl der Rezipienten eines Contentmoduls zukünftig ermittelbar ist (IPTV oder separater Rückkanal) und neue Reichweitopotenziale via IPTV bestehen, werden sich vermutlich neue Formen von Rechtepaketen herausbilden. Um das optimale Rechtpaket zu ermitteln entstehen ebenfalls neue Interdependenzen zwischen Aggregation und Selektion. Die Effizienz des Intermediationsprozesses würde demnach fallen, sofern Selektion und Aggregation von separaten Akteuren ausgeführt würden.

Die *Transformationsfunktion* fällt kostenintensiver aus, da die technische Aufbereitung von neuen, insbesondere interaktiven Contentformen aufwändiger als bei herkömmlichem Videomaterial ist (z.B. Applikationsportierung für Set-Top-Box-Middleware). Neben dem größeren technischen Beitrag zur Intermediation, vor allem hinsichtlich der neuen Contentformen, bleibt die Transformationsfunktion weitgehend unberührt.

Die *Aggregationsfunktion* mündet neben einem linearen TV-Programmstrom nun zusätzlich in einem Programmportfolio, welches dem Rezipienten über Personalisierungsfunktionen zugänglich ist. Dieser Zugang ist eine maßgebliche Erweiterung des Leistungsangebots von Aggregatoren: Sie können ein Personalisierungssystem ausbilden, welches eine situations- und contenttypadäquate Entscheidungsunterstützung des Rezipienten ermöglicht. Derartige Ansätze werden bereits im Kontext der bereits angesprochenen EPGs diskutiert [HePS05]. Effektivität und Bedeutung der Aggregation nehmen auf diese Weise stark zu, zusätzliche Transaktionskosten entstehen bei der Abstimmung mit der Präsentationsebene, da anstelle der einfachen TV-Programmdaten nun Rezipientendaten und ein komplexes Regelsystem zu übermitteln sind. Im Hinblick auf Reproduktion und Distribution entstehen möglicherweise neue Transaktionskostenblöcke, da sich die Übertragung personalisierter TV-Inhalte wesentlich komplexer gestaltet, als die eines linearen Programmstroms. Je nach Technologie (IPTV oder DVB, Uni- oder Multicasting) entstehen verschiedene Restriktionen, die in der Personalisierungsstrategie berücksichtigt werden müssen.

Die Digitalisierung der Übertragung und der Einsatz verschiedener Standards ermöglichen es Infrastrukturanbietern (TV und Telekommunikation) mit unterschiedlichen technologischen Lösungsansätzen die Funktionen *Reproduktion und Distribution* zu erfüllen. Die Bedeutung im

Rahmen des Intermediationsprozesses nimmt zu, da die notwendigerweise bandbreiten-optimierte Distribution personalisierter Inhalte die Komplexität der Funktion erhöht, Transaktionskostenänderungen entstehen hingegen keine.

Die Anforderungen an die *Präsentationsfunktion*, vor allem an das Contentmarketing steigen mit personalisiertem TV, da ein komplexeres Produkt zu vermarkten ist. Wie bereits geschildert, nimmt mit der Erweiterung von EPGs um Steuerelemente für das TV-Programm auch der Leistungsumfang dieser Funktion zu. Input ist hier das entsprechend aufbereitete Programmportfolio, dessen Darstellung und Vermittlung an die Rezipienten ist Aufgabe der Präsentationsfunktion. Erfüllt wird dieser Aspekt hauptsächlich durch neue Endgeräte, die zudem einfache (nicht vonseiten des Programmanbieters vorbereitete) Funktionen wie Timeshifting bereitstellen. Die Änderungen innerhalb dieser Funktionen ziehen einen starken Bedeutungszuwachs nach sich, neben den für die Aggregation bereits beschriebenen Transaktionskostenänderungen entstehen keine weiteren Änderungen.

Abbildung 6 fasst diese Überlegungen zum Zusammenwirken von Intermediärfunktionen (Ansatzpunkt (2) aus Abbildung 5) zusammen, indem sie die vier Stellen hervorhebt, an denen sich die Transaktionskosten erhöhen.

	Infrastrukturebene	Dienstebene
Identifikation		
Selektion	IPTV, Rückkanal	Personalisierung Video-On-Demand
Transformation		
Aggregation		Personalisierung Video-On-Demand Electronic Program Guides
Reproduktion	DVB, Multicasting	
Distribution	Rückkanal, IPTV	
Präsentation		Electronic Program Guides Timeshifting

Abb. 6: Einfluss neuer Technologien auf die Transaktionskosten zwischen Intermediärfunktionen

Bezüglich der Änderung des Funktionsbeitrags zum Intermediationsprozess (Ansatzpunkt 1 aus Abbildung 5) bleibt festzuhalten, dass die Bedeutung aller Intermediärfunktionen aufgrund der Anforderungen durch neue Dienst- und Contentformen zunimmt. Um (1,2) Selektion und Aggregation, (3) Aggregation, Reproduktion und Distribution sowie (4) Aggregation und Präsentation transaktionskostenminimal abzuwickeln, ist es nahe liegend, diese Funktionsgruppen jeweils in einem Akteur gebündelt abzuwickeln. Für einen TV-Sender bedeutet dies einerseits die Festigung der angestammten Position (1,2), da der hohe

Abstimmungsbedarf eine Markteintrittsbarriere für reine Selektionsdienstleister oder Aggregatoren darstellt. Zum anderen besteht die Notwendigkeit einer stärkeren Kooperation mit Infrastrukturanbietern (3), vor allem wenn die verschiedenen Diversifikationsmöglichkeiten hinsichtlich Personalisierung genutzt werden sollen (4). Von der anderen Seite betrachtet entsteht ebenfalls Diversifikationspotenzial für Infrastrukturanbieter (3), die Aggregation zumindest in Teilen anbieten könnten.

Abbildung 7 stellt die somit zu erwartenden Allokationsänderungen im Überblick dar.

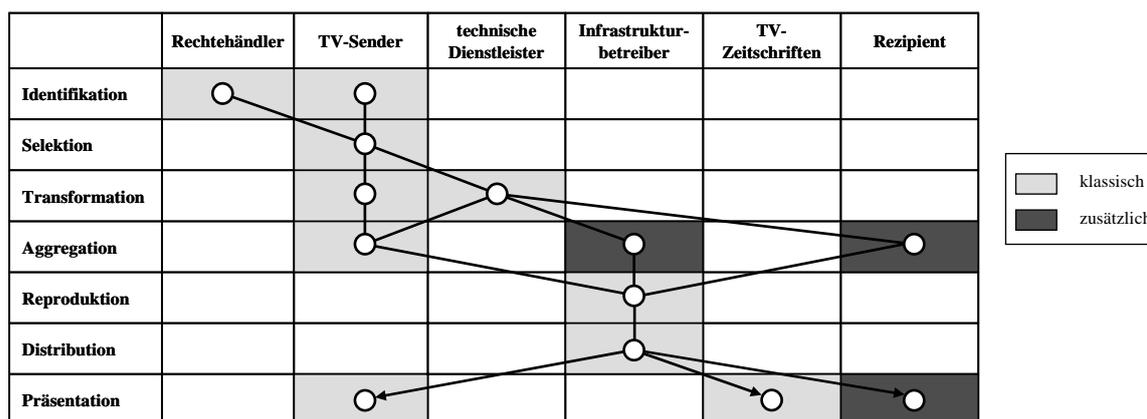


Abb. 7: Zukünftige Allokation von Intermediärfunktionen in der TV-Branche

Während die Digitalisierung in der Musikindustrie die Transaktionskosten zwischen den generischen Funktionen stark gesenkt hat [HeWa06], so dass Reproduktion, Distribution und Präsentation heute mit minimalen Transaktionskosten über Webportale abgewickelt werden können, lässt sich für die TV-Branche gerade das Gegenteil erwarten. Durch die Erhöhung bestehender und das Entstehen neuer Transaktionskostenblöcke wird die Intermediationsstruktur in ihrem Kern weitgehend stabil bleiben.

6 Add-on: Eine erste Validierung der Ergebnisse

Das Ergebnis der Analyse ist zunächst eher contra-intuitiv. Zur grundsätzlichen Validierung der Ergebnisse wurden daher Interviews mit fünf Experten der deutschen TV-Branche geführt. Befragt wurden Matthias Büchs (Direktor Operations SuperRTL, RTL Disney Fernsehen GmbH & Co KG), Andrej Henkler (Geschäftsführer, inLive Interactive GmbH), Malte Hildebrandt (Leiter Marketing, ProSieben Television GmbH), Bernd Schlötterer (Geschäftsführer Tele München Fernseh GmbH & Co Produktionsgesellschaft) und Beate Uhrmeister

(Programmeinkauf, VOX Film & Fernseh GmbH & Co KG). Die Interviews wurden telefonisch in jeweils ca. 20 Minuten durchgeführt. Die Ergebnisse wurden hierfür zu fünf Thesen verdichtet.

Hinsichtlich der Bewertung von Innovationen für die TV-Branche (vgl. Abschnitt 4) ergab sich hinsichtlich Nachfrage und Adaption *Personalisierungsthese I*: „Je nach Contenttyp und Benutzersituation wird ein unterschiedlicher Grad an Personalisierung hinsichtlich Auswahl und Anordnung von Inhalten nachgefragt.“, sowie hinsichtlich der möglichen Implementierung *Personalisierungsthese II*: „Bedingt durch die Komplexität des Produkts TV-Programm muss der Einbezug des Nutzers in den Prozess der Selektion und Aggregation aktiv unterstützt werden.“ Zur technischen Realisierbarkeit wurde die *Diversifikationsthese* vorgestellt: „Mit IPTV, Multicastingverfahren und PVRs bestehen die technischen Möglichkeiten, das bisherige TV-Angebot um verschiedene Arten von Nutzereingriffsmöglichkeiten zu erweitern.“ Aufbauend auf die Analyseergebnisse in Abschnitt 5 wurden zwei Thesen zu den Implikationen der Konvergenz für Intermediäre in der TV-Branche vorgestellt. Einerseits *Intermediationsthese I, die an die Festigung der Position von TV-Sendern anknüpft*: „Ein unmittelbares Eingreifen von Rezipienten in die Selektion und Aggregation von Content – analog der Musikbranche – würde eine lose, marktliche Koordination der beiden Funktionen voraussetzen. Aufgrund prohibitiv hoher Transaktionskosten ist eine solche Lösung in der TV-Branche ökonomisch nicht sinnvoll.“ Und abschließend *Intermediationsthese II, die das entstehende Diversifikationspotenzial umschreibt*: „Durch die zunehmend notwendige Abstimmung von Aggregatoren und Infrastrukturanbietern bei der Distribution von personalisierten TV-Inhalten entsteht Synergiepotenzial durch eine stärkere Integration der beiden Funktionen. So werden Infrastrukturanbieter zunehmend auch einzelne Teile der TV-Senderfunktionen übernehmen.“ Zur Auswertung wurden die Aussagen der Experten klassifiziert in „volle Zustimmung“, „teilweise Zustimmung“ und „keine Zustimmung“ und die weiteren Anmerkungen schlagwortartig verdichtet. So stimmen drei Experten der Personalisierungsthese I vollkommen, zwei Experten nur teilweise zu. Sie wenden zum einen ein, dass die individuelle Nutzensituation nur am Rande durch Personalisierung geprägt ist, maßgeblich seien weiterhin Exklusivität und Qualität der Inhalte. Zum anderen wird auf das gelernte Mediennutzungsverhalten verwiesen, was z. B. dazu führen würde, dass ältere Rezipienten gar keine Personalisierungsfunktionen nachfragen. Personalisierungsthese II wird von allen Experten als zutreffend erachtet. Hinsichtlich der Diversifikationsthese stimmen vier Experten

vollkommen zu, ein Experte verweist auf die seit Jahren geringe Nachfrage von Endgerätfunktionen wie Timeshifting und die Vorteile der Zeitgebundenheit von TV-Inhalten, beispielsweise bei Kultserien, über die sich Zuschauer nach der Rezeption austauschen wollen. Er stimmt daher nicht zu. Intermediationsthese I wird von drei Experten vollkommen zugestimmt. Mit der Einschätzung, dass auch neue Akteure „im Kleinen“ Markenführung und Programmauswahl wie TV-Sender leisten können und sich damit Nischenangebote etablieren können, stimmen zwei Experten der These nur teilweise zu. Der Intermediationsthese II wird von drei Experten nur bedingt zugestimmt: Infrastrukturanbieter versuchen zwar derartige Angebote zu etablieren, jedoch sei das Erlernen der Programmkompetenz sehr aufwändig, weshalb mit einer Spezialisierung der neuen Programmanbieter zu rechnen ist und somit hier keine unmittelbare Konkurrenz zum FreeTV entstünde. Zwei Experten stimmen der These vollkommen zu.

Mit nur einer „Nicht-Zustimmung“ können die ersten drei Thesen als bestätigt angesehen werden. Die in Teilen nur bedingte Zustimmung zu Intermediationsthese I und II unter Angabe von Ausnahmen, kann als grundsätzliche Bestätigung interpretiert werden, wenn auch das Bild offenbar noch nachzuschärfen ist, um diese Fälle erklären zu können.

7 Fazit

Ziel des Beitrags war die Analyse des Einflusses von Innovationen im TV-Sektor auf die Wertschöpfungsstruktur der Branche. Es ließen sich drei Schnittstellen zwischen Intermediationsfunktionen identifizieren, an denen zukünftig Transaktionskosten entstehen oder steigen. Der Effekt dieser Änderung ist die Festigung (!) der bisherigen Position von TV-Sendern als Selektions- und Aggregationsakteure sowie das Entstehen von Diversifikationspotenzial sowohl für einen Sender hinsichtlich Personalisierungsleistungen, als auch für Infrastrukturanbieter hinsichtlich Contentaggregation. In Summe entspricht dies genau dem Gegenteil der bereits angesprochenen Erfahrungen aus der Musikindustrie: Anstelle der Disintermediation nimmt die Bedeutung der TV-Intermediäre im deutschen Markt eher zu. In Großbritannien und den USA, die typischerweise eine Vorreiterrolle bei technischen Innovation im Mediensektor einnehmen, zeigt sich übrigens ein ganz ähnliches Bild [Colw05].

Literaturverzeichnis

- [AlSa98] Allen, F. ; Santomero, A. M.: The Theory of Financial Intermediation. In: Journal of Banking and Finance 21. Jg. (1998) Nr. 11/12, S. 1461-1485.
- [BeCS06] Becker, J. U.; Clement, M.; Schusser, O.: Peer-to-Peer-Netzwerke und die Generation Napster. In: Clement, M. and Schusser, O. (Hrsg.): Ökonomie der Musikindustrie. Deutscher Universitäts-Verlag, Wiesbaden 2005, S. 201-214.
- [Born04] Bornemann, J.-U.: Das digitale Programmangebot werbefinanzierter Fernsehveranstalter. Arbeitspapiere des Instituts für Rundfunkökonomie an der Universität zu Köln, Köln 2004.
- [ChKa99] Chircu, A. M.; Kauffman, Robert J.: Strategies for Internet Middlemen in the Intermediation/Disintermediation/Reintermediation Cycle. In: Electronic Markets 9. Jg. (1999) Nr. 1/2, S. 109-117.
- [Clem04] Clement, M.: Fernsehen im Zeitalter von Networked Personal Video Recordern. In: Zeitschrift für Betriebswirtschaftliche Forschung 12. Jg. (2004) Nr. 12, S. 760-779.
- [CISA05] Clement, Michel; Schneider, Holger; Albers, Sönke: Status Quo des Digitalen Fernsehens in Deutschland. In: Medienwirtschaft 2. Jg. (2005) Nr. 1, S. 1-12.
- [Colw05] Colwell, T.: Interaktives TV: Als Ergänzung zum traditionellen Fernsehen akzeptiert. In: Media Perspektiven 3 (2005) S. 125-133.
- [Eisn00] Eisner, Hubert: Möglichkeiten und Probleme von Electronic Program Guides. In: Kruse, J. (Hrsg.): Ökonomische Perspektiven des Fernsehens in Deutschland. Fischer, München 2000, S. 149-162.
- [Fing97] Fingleton, John: Competition between intermediated and direct trade and the timing of disintermediation. In: Oxford Economic Papers 49. Jg. (1997) Nr. 4, S. 543-556.

- [Herm02] Hermann, Michael: Vom Broadcast zum Personalcast: ökonomische Potenziale der Individualisierung audiovisueller Medienprodukte. Deutscher Universitäts-Verlag, Wiesbaden 2002.
- [HePS05] Hess, Thomas; Picot, Arnold; Schmid, Martin: Intermediation durch interaktives Fernsehen: eine Zwischenbilanz aus ökonomischer Sicht. In: zu Salm, C. (Hrsg.): Zaubermaschine interaktives Fernsehen? TV-Zukunft zwischen Blütenträumen und Businessmodellen. Gabler, Wiesbaden 2005, S. 17-50.
- [HeWa06] Hess, Thomas; Von Walter, Benedikt: Toward Content Intermediation: Shedding New Light on the Media Sector. In: The International Journal on Media Management 8. Jg. (2006) Nr. 1, S. 2-8.
- [LöFa02] Löbbecke, Claudia; Falkenberg, Marcia: A Framework for Assessing Market Entry Opportunities for Internet-Based TV. In: International Journal on Media Management 4. Jg. (2002) Nr. 2, S. 95-104.
- [Pica02] Picard, Robert G.: The Economics and Financing of Media Companies. Fordham University Press, New York 2002.
- [QiSX05] Qiang, Xie; Shibao, Zheng; Xiaojing, Yu: A Smart-Card-based Conditional Access Subsystem Separation Scheme for Digital TV Broadcasting. In: IEEE Transactions on Consumer Electronics 51. Jg. (2005) Nr. 3, S. 925-932.
- [RuNi97] Ruhrmann, Georg; Nieland, Jörg-Uwe: Interaktives Fernsehen: Entwicklung, Dimensionen, Fragen, Thesen. Westdeutscher Verlag, Opladen 1997.
- [Scho00] Schoder, D.: Die ökonomische Bedeutung von Intermediären im Electronic Commerce. Albert-Ludwigs-Universität, Freiburg 2000.
- [Schm05] Schmid, Martin: Die Abkehr vom synchronen Broadcast - Die Zersplitterung der Distributionskanäle im TV-Bereich und ihre Bedeutung für das Wertschöpfungs-system des Fernsehens. In: Rebensburg, K. (Hrsg.): NMI 2005 - Neue Medien der Informationsgesellschaft. Shaker Verlag, Aachen 2005, S. 259-281.

- [ScHe06] Schumann, Matthias; Hess, Thomas; Schumann, Hess: Grundfragen der Medienwirtschaft : eine betriebswirtschaftliche Einführung. 3., akt. und überarb. Auflage, Springer-Verl., Berlin u.a. 2006.
- [Schr95] Schrape, Klaus: Digitales Fernsehen: Marktchancen und ordnungspolitischer Regelungsbedarf. Fischer, München 1995.
- [SiRe97] Silberer, Günter; Rengelshausen, Oliver: Interaktive Werbung: Marketingkommunikation auf dem Weg ins digitale Zeitalter. Schäffer-Poeschel, Stuttgart 1997.
- [Sjur02] Sjurts, Insa: Strategien in der Medienbranche : Grundlagen und Fallbeispiele. 2. Auflage, Gabler, Wiesbaden 2002.
- [Stie04] Stieglitz, Nils: Strategie und Wettbewerb in konvergierenden Märkten. Dt. Univ.-Verl., Wiesbaden 2004.
- [Stip04] Stipp, Horst: Der Zuschauer und das Bedürfnis nach Interaktivität: ein Blick in die USA. In: zu Salm, C. (Hrsg.): Zaubermaschine interaktives Fernsehen? : TV-Zukunft zwischen Blümenträumen und Businessmodellen. Gabler, Wiesbaden 2004, S. 121-127.
- [Thie00] Thielmann, Bodo: Strategisches Innovations-Management in konvergierenden Märkten : Medien- und Telekommunikationsunternehmen in Online-Diensten und im digitalen Fernsehen. Dt. Univ.-Verl., Wiesbaden 2000.
- [WaHe05] Von Walter, Benedikt; Hess, Thomas: Content-Intermediation - Konzeption und Anwendungsgebiet. Arbeitsbericht 5/2005 des Instituts für Wirtschaftsinformatik und Neue Medien der LMU München, München 2005.
- [Wirt05] Wirtz, Bernd W.: Medien- und Internetmanagement. 4. Auflage, Gabler, Wiesbaden 2005.
- [Yoff97] Yoffie, David B.: Chess and Competing in the Age of Digital Convergence. In: Yoffie, D. B. (Hrsg.): Competing in the Age of Digital Convergence. Harvard Business School Press, Harvard 1997, S. 1-36.

Einführung in den Track

eServices: Anwendungen, Technologien und Management

Prof. Dr. Martin Bichler

Technische Universität München

Prof. Dr. Lutz Heuser

SAP Research

Prof. Dr. Peter Buxmann

Technische Universität Darmstadt

Der Track fokussiert die folgenden Themengebiete:

- Dienstmodellierung, -komposition und -choreographie (BPEL, WSCI, WS-Security, etc.)
- Geschäftsprozessintegration (Enterprise Service Bus, EAI, XML/EDI)
- Service-orientierte Architekturen in der Softwareentwicklung
- Prozesse und Methoden des IT-Service-Management
- Verrechnung von eServices (Utility Computing, etc.)
- Anwendungsszenarios, Geschäftsmodelle und betriebswirtschaftliche Bewertung

Programmkomitee:

Prof. Dr. Manfred Grauer, Universität Siegen

Prof. Dr. Stefan Kirn, Universität Hohenheim

Prof. Dr. Winfried Lamersdorf, Universität Hamburg

Dr. Heiko Ludwig (IBM Research)

Prof. Dr. Gustaf Neumann, WU Wien

Prof. Dr. Dirk Stelzer, Technische Universität Ilmenau

Dr. Tim Weitzel, Universität Bamberg

Optimized Dynamic Allocation Management for ERP Systems and Enterprise Services

Valentin Nicolescu, Martin Wimmer, Raphael Geissler, Daniel Gmach, Matthias Mohr,

Alfons Kemper and Helmut Krömer

Technische Universität München,
85748 Garching b. München, Germany

{nicolesc|wimmerma|geisslra|gmach|mohrm|kemper|krcmar}@in.tum.de

Abstract

To ensure the operability and reliability of large scale Enterprise Resource Planning Systems (ERP) and enterprise services, a peak-load oriented hardware sizing is often used, which results in low average utilization. The evaluation of historical load data revealed that many applications show cyclical resource consumption. The identification of load patterns can be used for static as well as dynamic allocation optimization.

In this paper we show the extraction of load patterns and present self-organizing service allocation concepts. This practical evaluation of theoretical adaptive computing concepts is of particular importance for the configuration of emerging service oriented architectures (SOA).

1 Introduction

Enterprise resource planning systems (ERP) undergo significant changes with regard to the hardware and software architecture: Clusters of commodity servers are displacing classical mainframe architectures and monolithic software systems are decomposed into smaller modular enterprise services. These new computing paradigms provide higher levels of flexibility, but also demand for new administration principles and more powerful resources resulting in an increased administration overhead [Erl05, 67]. Thereby, ERP systems have to respond with consistently low latency as requests are posted in a dialog mode. Thus, an integrated Quality of Service (QoS) management is an important issue for ERP systems and their services. User requests can vary according to the frequency they occur and the load they induce. Nevertheless,

using an aggregated view on the number of served users, ERP applications often show statistically periodic load characteristics.

In order to justify this theory, we analyzed the time dependent workload of monolithic ERP training systems (SAP R/3 Enterprise edition). Our contribution is an approach for the extraction of load patterns, which we evaluated on the basis of real-world monitoring data that was provided by the SAP Hochschul Competence Center (SAP HCC) at Technische Universität München (TUM). Furthermore we achieved to extract central monitoring data for realtime analysis and pattern identification purposes. Thus we focused on the runtime behaviour of a large ERP landscape. The cyclical behaviour of applications allows us to optimize the static application-to-server allocation and to supervise the deployment at runtime. Moreover, predictable critical situations like overload on servers can be prevented. Knowing the load characteristics of applications, those with complementary behavior can reliably be deployed onto the same servers.

2 Related Work

The analysis of monitoring data and the detection of application specific load characteristics are beneficial for the prediction of future system states. Such load patterns can be used to predict future resource requirements. A flexible architecture supporting the statistical evaluation of log data is presented in [BFBL05]. In [DiOH; SmFT98], CPU usage and application runtime have been evaluated and in [VaSh03] the data requirements of applications were examined. In this paper we present our pattern extraction algorithms by means of CPU monitoring data. Furthermore, we evaluated various log data like memory usage and quality of service parameters like the number of served users and application response times. These evaluations show that in particular business services oftentimes reveal quite regular behaviours. As shown in [HeZS01; RZAA04], load traces of such applications in many cases demonstrate clear daily and weekly patterns. In [ShHe00], also the trend and residual process of load traces was analyzed. Corresponding to the results presented in [CSCD05] and [CCDS05], our experience is that business applications oftentimes show future load demands that are similar to their requirements in the past, meaning that the trend process has minor impact on short-term forecasts. As we will show in this paper, load forecasts allow warning of potential critical situations, like service crashes and overload situations. Thus, sophisticated feedback and feedforward control mechanisms can

be realized [HDPT04]. For instance, in case of a server overload situation, servers that can satisfy the applications' resource requirements can be determined reliably. Future overload situations can be anticipated and avoided through a dynamic allocation management. The authors of [CZGS05] show how system state patterns can be extracted automatically at runtime. Failure diagnosis can be leveraged through the evaluation of such system "signatures". Case studies on the application of prediction schemes were presented in [VAHM02] and [XZSW06]. The forecasts of failures or anomalies have for example been analyzed in [CKFF02; KiFo04] and [BFBL05]. In [CKFF02; KiFo04], anomaly detection methods are employed to determine problems of internet services. In [BFBL05], the authors show, how HTTP logs can be analyzed and, e.g., certain site failures can be inferred from changes in the user behaviour. Monitoring and analysis of log data are crucial for the realization of self-organizing computing principles – also referred to as adaptive or autonomic computing [Horn01]. Astrolabe [ReBV03] is an example of a self-organizing computing infrastructure. It provides self-management computing concepts for distributed environments. Our main focus is rather on vast distributed applications than on classical intra-organizational enterprise resource planning (ERP) systems and enterprise grids. In [SGKK06] we presented our self-organizing computing infrastructure called AutoGlobe. AutoGlobe is a research prototype, used to evaluate upcoming adaptive computing trends, mainly by means of simulations. In the context of a long-term collaboration with the SAP HCC [MWNK06] – which is an application service provider for the academia (see Section 3) – self-organizing techniques examined in the AutoGlobe project are also step by step evaluated on the basis of productive systems. First results of this collaboration were presented in [WNGM06]. In this paper we present further technical details about the extraction of application load patterns and the current state of the collaboration. In particular, the evaluation of application load patterns, which we present in this paper, can constitute the basis for a sophisticated admission control [SeBH06].

The demand for flexible, self-organizing infrastructures stems from the requirement of enterprises to sooner react on changes of the market. Therefore, leading hardware and software vendors provide self-organizing products in their portfolio, like IBM Director [MeYa04], Sun N1 Grid [Sun06] or Fujitsu-Siemens FlexFrame [FuSi06]. The SAP HCC provides classical ERP (training) applications. But the evaluation results gained in this environment are supposed to apply to a large extent also to the emerging service oriented architectures [BeCT05; Erl05].

These provide higher levels of service flexibility, thus, supporting dynamic allocation schemes even more comprehensive than classical ERP systems.

3 Target Infrastructure

We examined the CPU load caused by the ERP training applications hosted by the SAP HCC at TUM. The SAP HCC acts as an application service provider (ASP) for academia [BIJa04]. As shown in Table 1, it provides support for 61 academic customers with an estimated number of about 18,400 users. An ERP training system in terms of the SAP HCC is typically an SAP R/3 Enterprise edition, which includes the classical enterprise functionalities. Some of these basic configurations are enhanced by business analytics covered by SAP's Business Warehouse (BW). We identified three different effects that determine the load curve of such an ERP training system [MoSK05].

- The basic characteristics can be described as a *long-term component*. For example the number of user requests is noticeably reduced during weekends and holidays. This effect can also be recognized in productive enterprise environments, with the load depending on the work time of the employees.
- *Short-term influences* can be identified by the kind of performed tasks that vary during a day depending on the users' work rhythm, e.g., the course structure. In an enterprise environment this effect can for example be seen, if sales representatives work in-house or a number of batch jobs are worked off.
- A third component describes *unusual workloads* that emerge when certain tasks are performed by all course participants simultaneously. In general, this does not apply to ERP production systems but for training environments like the one regarded in this work.

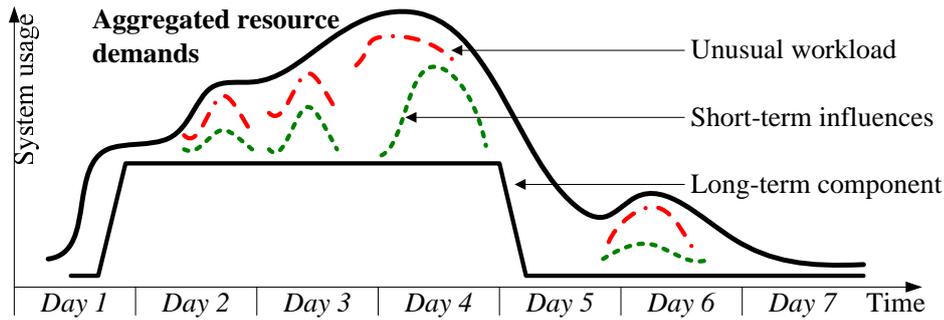


Figure 1: Schematic composition of a 7-day load pattern of an ERP training application

If several of these characteristics occur at the same time, a so-called *whiplash effect* can be observed, meaning that high load peaks appear after short build-up times (see Figure 1).

The SAP HCC employs a Blade server architecture that allows to be adaptively extended. Normally, each of the 55 SAP systems consists of several dialog instances (DI) and one central instance (CI). Initially, only one instance runs on each server. While the DIs can be allocated dynamically on the available hardware, the CI is by default statically assigned to one of the database servers. However, the static deployment can be changed easily due to virtual host names and an SAN infrastructure. The CI provides basic functionality, like dispatching user requests onto the available instances. The assignment of requests to instances can be parameterized. For example, the percentage of user requests for each instance can be defined.

Institutions (partially more than one per system)	Universities	13
	Schools for applied sciences	27
	Vocational schools	17
	University of cooperative education	3
	Total	61
	SAP systems	55
Users	Students	18,390
Application servers	Blade Servers (Sun B100s)	80
	Others (Sun T2000, V40z, X2100, X4200)	17
DB servers	Sun Fire V210, V240, V880, V890	50
Storage	RAID capacity	42 TB

Table 1: SAP HCC customers and infrastructure (June 2006)

Initially, the hardware allocation was done in a traditional, peak-load oriented way. Better, cost-effective results can be achieved by considering the characteristics of courses. In this regard, applications with statistically complementary load patterns are supposed to be allocated onto the same servers. Therefore, we analyzed the historical load information and extracted application specific load patterns. The protocolled load represents 15 min average-aggregates of CPU load

that was measured every 5 seconds. The basis for our experimental evaluation constitutes the monitoring data for 38 ERP training systems that was generated during a German summer semester (February to July). As the predominantly homogeneous HCC landscape is running in a separate subnet, allocation constraints regarding different subnets and performance of servers were not taken into account.

4 Extraction of Application Load Patterns and Application Classification

We developed a three-staged approach for the extraction and evaluation of load patterns. Thereby, we kept the parametrization at a minimum level – a prerequisite for its integration into a self-organizing infrastructure. During the *preprocessing phase* the historical load information is transformed into time series of equidistant sampling rate. Hypothetical load patterns are determined for each time series during the *analysis phase*. Finally, in the *classification phase*, the quality of the extracted patterns is estimated.

In our experiments we analyzed the CPU load induced by ERP systems of the SAP HCC, but our approach is not restricted to a certain notion of load. For example, memory usage or Quality-of-Service (QoS)-relevant parameters like response time, system throughput, or the number of served users can be handled as well. In the following paragraphs we therefore look at time series referring to generic historical load information.

4.1 The Preprocessing Phase

At runtime, all application instances are monitored and the load induced by them is logged. Let S be the set of all ERP training systems. In order to determine the characteristics of a certain ERP training system $s \in S$, the aggregated historical load information of all instances of s has to be determined. Thereby, in heterogeneous computing environments, the load induced on host machines of varying capability is considered through performance normalization. Consequently, for each system s an equidistantly sampled time series $(l^{(s)}(t_i))_{1 \leq i \leq N}$ is calculated, representing its entire load influence over a monitored period of length $T = t_N - t_1$.

4.2 The Analysis Phase

The extraction of a load pattern proceeds under the assumption of the time series being cyclic. Whether this assumption holds is evaluated in the classification phase. The stages of the analy-

sis phase are exemplified for two time series in Figure 2. The left part of the figure represents an ERP training system with periodic characteristics, while the right part represents a system with irregular load characteristics. According to the classical additive component model, a time series consists of a trend component, a cyclical component, and a remainder, e.g., characterizing the influence of noise. The trend is a monotonic function, modelling an overall upwards or downwards development. We focus on the cyclical component that describes the periodic load characteristics of a system.

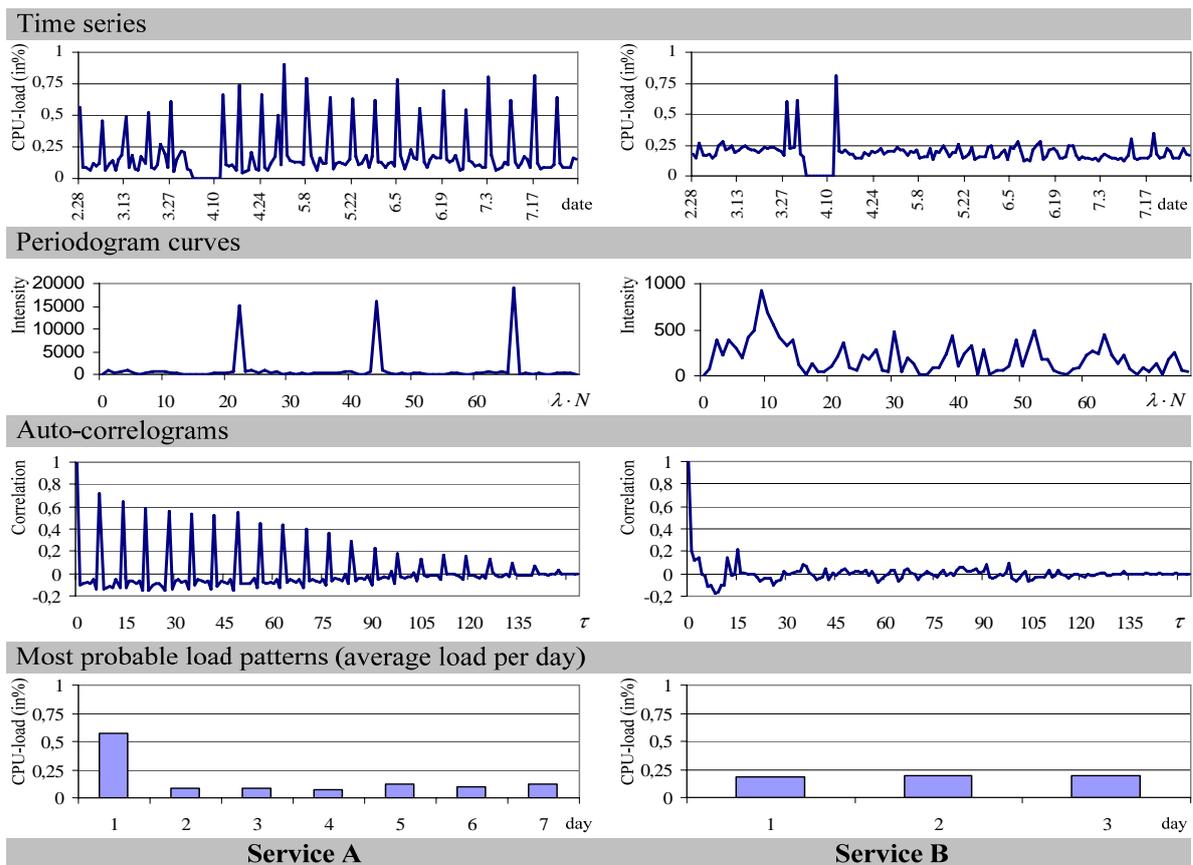


Figure 2: Illustration of the analysis phase – left: system with statistically period, right: system with irregular load characteristics

First, we determine the yet unknown duration of the load pattern, starting with the *evaluation of the periodogram function*: Via Fourier Transformation the time series can be represented as an overlay of harmonics. The periodogram function

$$I(\lambda) = \frac{1}{N} \cdot \left[\sum_{1 \leq i \leq N} (l^{(s)}(t_i) - \bar{l}) \cos 2\pi\lambda i \right]^2 + \frac{1}{N} \cdot \left[\sum_{1 \leq i \leq N} (l^{(s)}(t_i) - \bar{l}) \sin 2\pi\lambda i \right]^2 \quad (1)$$

defines the intensity, with which the harmonic of frequency $\lambda > 0$ is present in the time series that is normalized by its average value \bar{l} . The most dominant frequencies provide information

about the duration of the potential pattern. Intuitively, if I has a local maximum at λ_m , then it is likely that there exists a pattern of length $T \cdot \lambda_m$. In general, taking the position of the global maximum (named \max_I) is error-prone. Thus, we determine the set $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ of local maxima positions, with $I(\lambda_j) > \max I/2$ for every $1 \leq j \leq n$. The local maxima positions are computed through evaluating the periodogram at i/N with $1 \leq i \leq N/2$ (using Fast Fourier Transformation) and subsequent iterative approximation.

We then derive, which $\lambda_j \in \Lambda$ determines the pattern duration. For this purpose we *evaluate the auto-correlation* (r_τ) that is defined as

$$r_\tau = c_\tau / c_0, \text{ with } \tau \in \mathbb{Z}, -N < \tau < N \text{ and } c_\tau = \frac{1}{N} \sum_{1 \leq i \leq N} (I^{(s)}(t_i) - \bar{l})(I^{(s)}(t_{i+\tau}) - \bar{l}) \quad (2)$$

The auto-correlation describes linear dependencies within the time series. If the auto-correlation shows local extrema at a quite regular lag τ_e (see system A in Figure 2), it is a sign that there exists a temporal dependency of length $\tau_e \cdot T / N$. In a quite similar way as for the evaluation of the periodogram, we determine the set $\Pi = \{\tau_1, \dots, \tau_m\}$ of significant local extreme positions.

As a next step, the compliance of the elements of Λ and Π is evaluated and the best matching $\lambda^{(s)}$ is determined. Intuitively, $\lambda^{(s)}$ is the best matching frequency, if the elements of Π exhibit a lag that is (approx.) equal to $\lambda^{(s)} N$, i.e., $\tau_j = j \cdot \lambda^{(s)} N$. We estimate the relationship of $\lambda \in \Lambda$ and Π through:

$$\delta(\lambda, \Pi) \stackrel{def}{=} \frac{d_{cal}(\lambda)}{card(\Pi)} \cdot \sum_{\tau \in \Pi} \left| \tau - rnd\left(\frac{\tau}{\lambda N}\right) \cdot \lambda N \right| \quad (3)$$

In the above definition, rnd is the rounding operation and $card$ determines the cardinality of a finite set. $d_{cal}(\lambda)$ is used to take calendar specific time periods into account, what will be discussed in more detail below. Let $\lambda^{(s)}$ be the frequency that minimizes δ , i.e.,

$\delta^{(s)} \stackrel{def}{=} \delta(\lambda^{(s)}, \Pi) = \min\{\delta(\lambda, \Pi) \mid \lambda \in \Lambda\}$. Assuming that s shows periodic characteristics,

$T_1^{(s)} \stackrel{def}{=} T \cdot \lambda^{(s)}$ is considered to be the most probable duration of the load pattern.

ERP training systems and business processes typically show a periodicity which is a multiple of hours, days, weeks and so forth. Due to unavoidable computational inaccuracies and the influence of noise, $T_1^{(s)}$ can diverge from the actual period. Thus, we perform a *comparison to cal-*

endar specific periods. Therefore, a rounding operation is required that determines the nearest calendar interval for $T_1^{(s)}$. As core intervals, hours, days, and weeks are used. When calculating the best matching calendar dependent interval, longer core intervals are prioritized, e.g., the approximation of $13\frac{1}{2}$ days by two weeks is favored compared to 14 days and especially compared to 324 hours. Let $T_2^{(s)}$ denote the best matching period regarding $T_1^{(s)}$. The enforcement of calendar related periods is also expressed in the above definition of δ by the weight $d_{cal}(\lambda)$. There, $d_{cal}(\lambda)$ describes the difference of $\lambda \cdot T$ and the best matching calendar interval relative to the respective core interval.

Finally, two *patterns are extracted* from the original time series with respect to the supposed pattern durations $T_1^{(s)}$ and $T_2^{(s)}$. One possible approach is to take extracts of the respective lengths out of the original time series. The approach we follow, is to take the average over the pattern occurrences within the time series. In order to reduce the influence of noise and computational inaccuracies, the starting points of pattern recurrences have to be determined reliably. Candidates for distinctive starting points mark significant changes of the derivation. Figure 3 illustrates the proceeding for selecting the k^{th} starting point. Let (t, l) be the supposed starting point according to the estimated pattern duration, i.e., $t = t_1 + k \cdot T_1^{(s)}$ or $t = t_1 + k \cdot T_2^{(s)}$. Instead of approximating the derivation function, we use a more descriptive geometric approach: The distinctive starting point (t', l') in the figure is determined as the point with the maximum distance ε to the line, which is determined by (t, \bar{l}) and the local extremum (t_e, l_e) (with $l_e > \bar{l}$ in case of a local maximum and $l_e < \bar{l}$ in case of a local minimum). Having determined the starting points, patterns $P_1^{(s)}$ and $P_2^{(s)}$ that refer to $T_1^{(s)}$ and $T_2^{(s)}$ are extracted as the average over the respective occurrences. Thus, patterns are time series as well, but (e.g., through interpolation) they can be seen as functions $\mathbb{N} \rightarrow \mathbb{R}_0^+$, with timestamps as input parameters (e.g., milliseconds and thus $\in \mathbb{N}$). The quality of a pattern $P = (l_p(t))$ is defined as the normalized distance to the original time series, estimated as $\Delta(P) = \frac{1}{N} \sum_{1 \leq i \leq N} |l^{(s)}(t_i) - l_p(t_i)|$. The best matching pattern $P^{(s)}$ is the one with

$$\Delta^{(s)} \stackrel{def}{=} \Delta(P^{(s)}) = \min\{\Delta(P_1^{(s)}), \Delta(P_2^{(s)})\}. \quad (4)$$

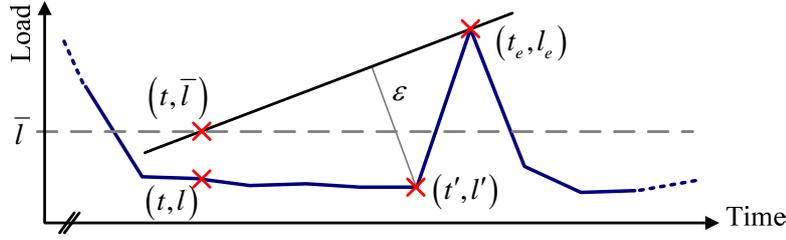


Figure 3: Detection of significant starting points

4.3 The Classification Phase

Separating time series with cyclical characteristics from those without is done in the classification phase. The likelihood for a time series (denoting the load induced by an ERP training system $s \in S$) showing a characteristic cyclical component is estimated by

$$\Omega^{(s)} \stackrel{\text{def}}{=} \omega \cdot \overbrace{\delta^{(s)} / \max\{\delta^{(s')} \mid s' \in S\}}^{\text{compliance of periodogram and auto-correlation}} + (1 - \omega) \cdot \overbrace{\Delta^{(s)} / \max\{\Delta^{(s')} \mid s' \in S\}}^{\text{pattern quality}}. \quad (5)$$

$\omega \in [0,1]$ is a weighting factor and in fact the only parameter that has to be chosen for pattern extraction. In our experiments, we used $\omega = 0.5$.

The lower $\Omega^{(s)}$, the higher the probability of s inducing statistically periodic load. Using a classical Lloyd-Max quantizer, the services are clustered.

4.4 Experimental Results

We evaluated the characteristics of 38 ERP training systems based on monitoring data that was generated during a German summer semester. As mentioned in Section 3, the examined time series represent the aggregated CPU load that was measured on the application servers. Our evaluation showed that CPU load is more suitable for detecting cyclical characteristics than other monitoring data like the number of registered users. We calculated three application clusters: Regarding their likelihood of showing periodic characteristics, we classified the ERP training systems as *periodic*, *fuzzy* and *non-periodic*. Considering Figure 2, system A is classified as *periodic*, while system B is *non-periodic*. The periodogram and auto-correlogram diagrams support this decision: Considering system A the graphs reveal a regular behavior in contrast to the diagrams for system B. Table 2 gives an overview of the results we gained based on the evaluation of CPU monitoring data.

The ranking of applications is useful for the static deployment as discussed in the next section. The table below illustrates that for 23 of the 25 ERP systems that were classified as *periodic* or *fuzzy* 7-day-patterns were determined, which corresponds to the nature of courses. For two of the ERP training systems patterns of length 30 days, respectively 33 days were determined. In contrast to more traditional pattern recognition applications, like image processing, wrongly classified patterns do not have such a strong impact on the optimization technique: False negatives, i.e., services that were not classified as periodic though they actually reveal a cyclical characteristics, reduce optimization possibilities, as for them a standard peak-load oriented distribution will be applied, but do not prohibit optimization at all. False positives, i.e., services that were classified as periodic although they show more or less unpredictable characteristics, are tolerable through the dynamic allocation management that is described in Section 5.2.

Cluster sizes	Periodic	17
	Fuzzy	8
	Non-periodic	13
	Total	38
Pattern lengths	7-days	23
	More than 30 days	2

Table 2: Experimental results

5 Self-organizing Allocation Management

We make use of application-specific load patterns to optimize the application-to-server allocation of large scale ERP landscapes. This idea was developed in the AutoGlobe project [SGKK06]. AutoGlobe is a research project focusing on the design and evaluation of self-organizing computing concepts for emerging service oriented architectures.

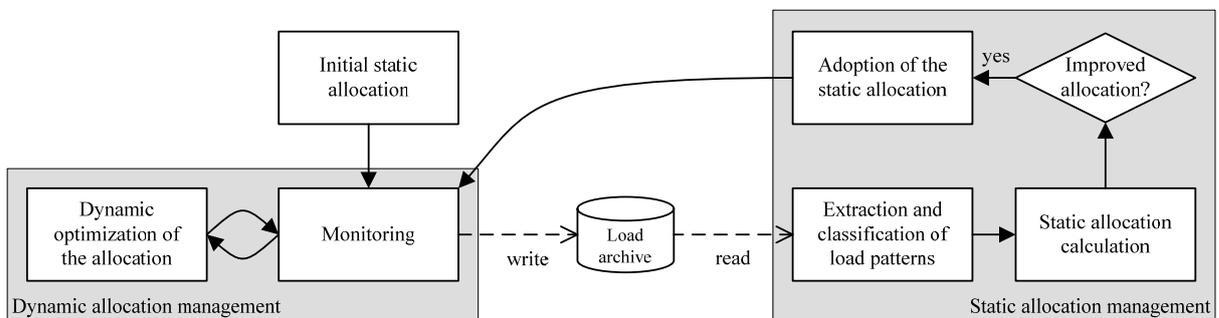


Figure 4: The AutoGlobe framework

Our aim is to successively integrate self-organizing computing techniques into the SAP HCC infrastructure for (semi-) automatic allocation optimization. The goal of bringing AutoGlobe and the SAP HCC together is to reduce hardware and maintenance costs and to evaluate theoretical concepts by means of a real world application scenario. Figure 4 illustrates the basic concepts of the AutoGlobe framework that are grouped into techniques for static and dynamic allocation optimization.

5.1 Static allocation management

Initially, a system is typically set up with a predefined configuration, e.g., designed by the system administrator. Alternatively, if further knowledge about the characteristics of applications is given, AutoGlobe can calculate an initial static allocation. A prerequisite for the computation of a static application-to-server allocation is knowledge about the estimated behavior of applications. Therefore, the load induced by the different ERP training systems is monitored at runtime and protocolled in a load archive. The historical load information is then analyzed and course specific load patterns are determined as described in the previous section. If no load pattern can be determined for an ERP training system, either the average, or for more pessimistic estimations the maximum of the time series is used. By use of these load characteristics, an optimized static allocation for a predefined timeframe is calculated. In this regard, optimized means that the workload will be almost balanced and especially critical situations like overloaded systems will be prevented. One further optimization criterion is that the available hardware shall not be lavishly utilized. That is, hardware, which can be used for further applications, has to be determined.

Let C denote the set of available servers and S the set of ERP training systems, i.e., applications. In our setting, the elements of C represent Blade servers as presented in Section 3. Finding the optimal static allocation then is in $O(\text{card}(C)^{\text{card}(S)})$ a complexity that cannot be handled in the general case. Therefore, we employ a greedy approach that successively assigns the systems $s \in S$ to the servers, whereby the systems are ranked according to the quality of their patterns as motivated in Section 4.4. One necessary parameter of the greedy heuristics is the *allocation limit* (AL), which determines the upper bound of resource utilization, e.g., the maximum percentage of processor load. Any resource utilization higher than AL will be considered as overload situation.

Let T be the duration of the time period, for which a static allocation has to be calculated. Suppose that some systems have already been assigned to the available servers and that the next system to be deployed is $s \in S$. The load induced by s is described by the pattern $(l^{(s)}(t))_{1 \leq t \leq T}$. Analogously, $(l^{(c)}(t))_{1 \leq t \leq T}$ represents the aggregated load scheduled for systems that are already allocated on server $c \in C$. As mentioned in the previous section, these time series are normalized. The load increase on server c , when allocating s on it, is estimated by:

$$\Delta(s, c) \stackrel{\text{def}}{=} \max_{1 \leq t \leq T} (l^{(c)}(t) + l^{(s)}(t)) - \max_{1 \leq t \leq T} (l^{(c)}(t)) \quad (6)$$

The so-called *best-match-server* is the one that minimizes cost function (6). In case all servers show a resource utilization higher than AL, the server with minimal exceedance is chosen. The greedy heuristics terminates when all systems have been allocated and, thus, a new static allocation is determined. Figure 5 shows the AutoGlobe user interface and illustrates the optimization calculation. The right part of the image shows two exemplary application patterns. As these reveal somehow complementary characteristics (system A shows a load peak on Monday, while system B has peaks on Tuesday and Friday) they can well be allocated on one host.

AutoGlobe supports optimized allocations to be applied automatically, e.g., at regular maintenance intervals. At the test stage, the allocations calculated by the AutoGlobe framework are used as recommendations. Before adapting the allocation, AutoGlobe evaluates, whether the new allocation provides significant benefit and is thus worth being applied. Therefore, the actual application-to-server assignment is compared with the new alternative considering the expected overload and idle situations and the administrative costs for modifying the system landscape (e.g., application downtimes).

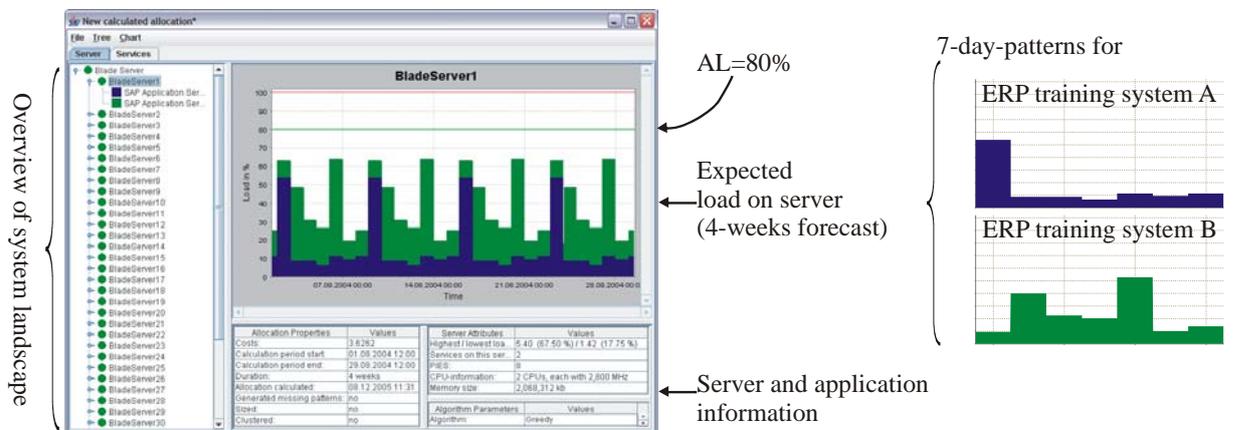


Figure 5: AutoGlobe-GUI: shown is an example for an allocation of two applications onto one server

5.2 Dynamic allocation management

Though a reliable static deployment has been arranged, still unforeseeable situations like crashes of applications, overload situations due to wrongly classified patterns or the start of further systems for additional courses can occur and have to be handled. Therefore, the system is supervised by a controller that reacts on exceptional situations. For example, in case of system breakdowns applications have to be restarted automatically. If an overload situation is detected, i.e., user requests can't be handled sufficiently because a training system s is executed on an overloaded host c , the controller calculates recommendations to remedy the situation. Depending on the flexibility the application provides, it is suggested to either start additional instances of s on other servers or to move instances onto a new host that provides enough free computational resources. For the selection of an appropriate host, load patterns are evaluated for estimating the load on hosts within $[t, t + \Delta t]$ for t being the time the exception occurs and Δt being a predefined look-ahead period. The best matching host is the one, on which an additional instance of s does not cause an exceptional situation during $[t, t + \Delta t]$ and the load increase is at minimum according to cost estimation (6). In case such a host does not exist, the system cannot react and a warning message is created. When an ERP system is split into several instances, the dispatcher assigns user requests to instances of s depending on the performance of the host machines these instances are running on (see Section 3).

The described approach is exception-triggered, i.e., the controller reacts in case of an exceptional situation. Moreover, predictive and proactive capabilities are provided through load-forecasting by regularly estimating the load induced on hosts $c \in C$ in between t and $t + \Delta t$. In case overload situations are predicted, either warning messages are created, or in the sense of an automatism, the system reacts by autonomously migrating applications or starting additional instances.

The first step towards an automatic realtime service allocation, is the access to performance data of the HCC system landscape. Therefore we built a Java based interface to the central monitoring system of the landscape. Within this system all performance data is collected and stored in the SAP Central Performance History (CPH). The interface was built using the library SAP Java Connector, which provides access to the SAP development platform via Java. This interface can be used to extract real-time monitoring data, e.g. CPU performance data, from an ERP system for dynamic as well as for static allocation optimization. The current version of the realtime analyzing tool provides a wizard for the configuration of the connection parameters, the polling

intervall and the type of monitoring data to be transferred, e.g. CPU performance or number of users logged on. After closing the wizard, the monitoring values are extracted from the central monitoring system and the received values are analyzed regarding their periodical load patterns. In a final step these, patterns and analysis parameters are shown in a graphical user interface. Thus the methods that compute the patterns can be verified visually and the practical use can be tested.

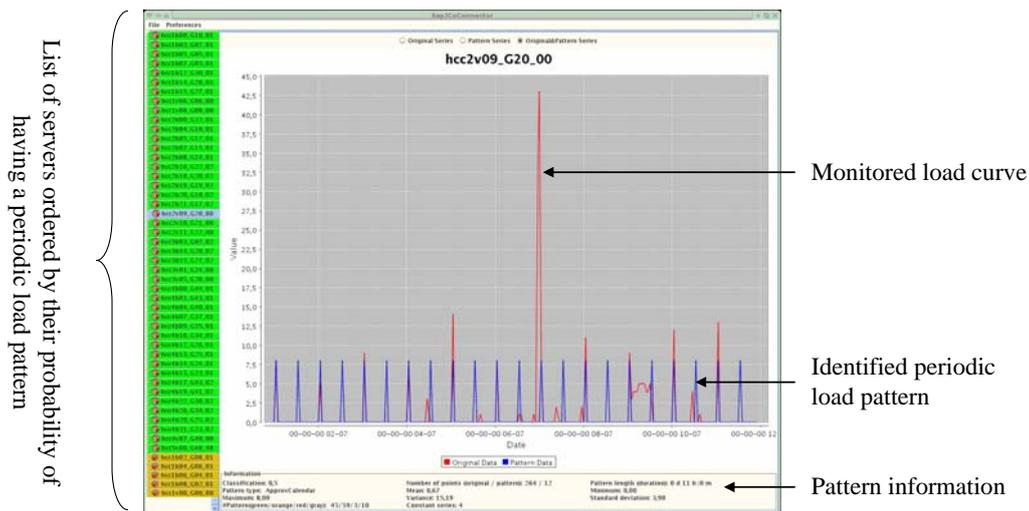


Figure 6: Identified load pattern within the GUI of the realtime analyzing tool

In the next phase, AutoGlobe’s optimization and allocation capabilities will be combined with the realtime analyzing tool to supervise the SAP HCC system in order to create realtime recommendations for allocating systems respectively services. These recommendations based on decision rules will then be further controlled by a system administrator prior to their realization. Later on, the reaction process shall be automated as far as possible.

6 Conclusion and Future Work

The extensive evaluation of the SAP HCC monitoring data supports our assumption that many applications exhibit periodic performance characteristics, which can be used to optimize the application-to-server allocation and, thus, to achieve a better balanced system landscape. Using our pattern classification approach, we are able to automatically identify system clusters that show a reliable cyclical behaviour and to separate them from applications with more or less unpredictable load behavior. This helps to reduce hardware costs as combinations of applications onto the same servers can be determined reliably. Thus, free resources are identified which

can be used for additional applications or economized, e.g., by ordering less resources from an ASP. Regarding the SAP HCC landscape, we were able to automatically classify applications that relate to regularly held courses, which allows a sophisticated analysis and revision of the application deployment. Currently, we are working on a tighter integration of the AutoGlobe system into the HCC landscape, which will help to reduce hardware and maintenance costs. Our intention is to autonomously supervise the ERP training systems, to monitor their “health”, report system breakdowns and to calculate optimization propositions. These can be evaluated and confirmed by the system administrator, who can decide to apply them. This will allow us to prove theoretical adaptive computing techniques in a representative real world ERP environment and to evaluate, how far automatism for system administration can be driven.

The SAP HCC is a service provider for academia, but classical ERP systems are likely to reveal applications with periodic characteristics as well. Concrete examples are classical OLTP applications with regard to financial operating services (FI) or the logistics execution system (LES). Therefore, the results obtained in the academic field of application will also provide vital benefits for load-balancing requirements of classical ERP landscapes in particular hosted ERP applications. Currently, monolithic ERP systems still dominate enterprise system landscapes and restrict a flexible application-to-server allocation. It becomes apparent that a lot of such systems will be replaced by service oriented architectures (SOA), like SAP’s Enterprise Service Architecture (ESA) with its underlying technology platform SAP NetWeaver. Such an enterprise grid architecture provides a higher level of modularity by successively decomposing monolithic applications into fine-grained services that can be allocated in a more flexible way. These services are cleanly partitioned and consistently represented software components [Erl05]. Due to the split-up of applications into a multitude of services the need for intelligent and highly adaptive allocation management will gain in importance.

Our analysis is a step towards evaluating adaptive computing concepts for SOA. We analyzed the characteristics of applications of a large scale academic landscape with a multitude of individual systems. The multitude of systems/services is also characteristic for SOA and it is likely that many enterprise services show cyclical load characteristics as the regarded ERP training systems do. Thus, we assume that the obtained results are also applicable for a pure service oriented approach. Nevertheless, an in-depth analysis of the described concepts for SOA remains as future work.

Bibliography

- [BeCT05] Benatallah, B.; Casati, F.; Traverso, P. (Ed.). (2005). *Proceedings of Third International Conference Service-Oriented Computing – ICSOC 2005, Amsterdam, The Netherlands, December 12-15, Springer 2005*.
- [BFBL05] Bodik, P.; Friedman, G.; Biewald, L.; Levine, H.; Candea, G.; Patel, K.; Tolle, G.; Hui, J.; Fox, A.; Jordan, M.I.; Patterson, D. (2005). *Combining Visualization and Statistical Analysis to Improve Operator Confidence and Efficiency for Failure Detection and Localization*. In: *Proceedings of Second International Conference on Autonomic Computing 2005 (ICAC 2005)*, 89-100.
- [BlJa04] Bleek, W.-G.; Jackewitz, I. (2004). *Providing an E-Learning Plattform in a University Context - Balancing the Organisational Frame for Application Service Providing*. In: *Proceedings of 37th Annual Hawaii International Conference on System Sciences, Los Alamitos, CA*
- [CCDS05] Castellanos, M.; Casati, F.; Dayal, U.; Shan, M.-C. (2005). *A Platform for Business Operation Management*. In: *Proceedings of 21st International Conference on Data Engineering 2005 (ICDE 2005)*, 5.-8.04.2005 1084-1095
- [CKFF02] Chen, M.; Kiciman, E.; Fratkin, E.; Fox, A.; Brewer, E. (2002). *Pinpoint: Problem determination in large, dynamic internet services*. In: *Proceedings of Dynamic Internet Services: International Conference on Dependable Systems and Networks (DSN'02)*, 595-604.
- [CSCD05] Castellanos, M.; Salazar, N.; Casati, F.; Dayal, U.; Shan, M.-C. (2005). *Predictive Business Operations Management*. In: *Proceedings of Databases in Networked Information Systems (DNIS 2005)*, 4th International Workshop, Aizu-Wakamatsu, Japan, 28.-30.03.2005, 1-14.
- [CZGS05] Cohen, I.; Zhang, S.; Goldszmidt, M.; Symons, J.; Kelly, T.; Fox, A. (2005). *Capturing, indexing, clustering, and retrieving system history*. In: *Proceedings of 20th ACM symposium on Operating systems principles, Brighton, United Kingdom*, 105-118.
- [DiOH] Dinda, P.A.; O'Hallaron, D.R. (1999). *An Evaluation of Linear Models for Host Load Prediction*. In: *Proceedings of HPDC '99: The 8th IEEE International Symposium on High Performance Distributed Computing*, 87-96.
- [Erl05] Erl, T. (2005). *Service-Oriented Architecture - Concepts, Technology, and Design*, New York: Prentice Hall.
- [FuSi06] Fujitsu-Siemens (2006). FlexFrame for Oracle, http://www.fujitsu-siemens.com/campaigns/flexframe_oracle/index.html, Accessed at: 2006-07-12.
- [HDPT04] Hellerstein, J.L.; Diao, Y.; Parekh, S.; Tilbury, D.M. (2004). *Feedback Control of Computing System*: Wiley-IEEE Press.
- [HeZS01] Hellerstein, J.L.; Zhang, F.; Shahabuddin, P. (2001). A Statistical Approach to Predictive Detection. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 35(1), 77-95.
- [Horn01] Horn, P. (2001). Autonomic computing: IBM's perspective on the state of information technology, *IBM Corporation*, http://www.research.ibm.com/autonomic/manifesto/autonomic_computing.pdf, Accessed at: 2006-07-10.
- [KiFo04] Kiciman, E.; Fox, A. (2004). Detecting application-level failures in component-based internet services. *IEEE Trans Neural Netw.*, 16(5), 1024-1041.

- [MeYa04] Melenovsky, M.; Yang, J. (2004). IBM Director : Driving Efficiencies in Scale-Out Computing, IDC, http://www-8.ibm.com/my/express/IDC_Sys_Mgt_WhitePaper_0324.pdf, Accessed at: 2006-07-12.
- [MoSK05] Mohr, M.; Simon, T.; Krcmar, H. (2005). *Building an Adaptive Infrastructure for Education Service Providing*. In: *Proceedings of Wirtschaftsinformatik 2005*. eEconomy, eGovernment, eSociety, Bamberg, 847-859.
- [MWNK06] Mohr, M.; Wittges, H.; Nicolescu, V.; Krcmar, H.; Schrader, H. (2006). Einbindung und Motivation informeller Multiplikatoren im IT-Training am Beispiel Education Service Providing. *Wirtschaftsinformatik-Ausbildung mit SAP®-Software - Publikation zum Track der Multikonferenz Wirtschaftsinformatik 2006 in Passau*, 1-24. Lohmar: Joseph Eul Verlag.
- [ReBV03] Van Renesse, R.; Birman, K.P.; Vogels, W. (2003). Astrolabe: A robust and scalable technology for distributed system monitoring, management, and data mining. *ACM Trans. Comput. Syst.*, 21(2), 164-206.
- [RZAA04] Rolia, J.; Zhu, X.; Arlitt, M.; Andrzejak, A. (2004). Statistical service assurances for applications in utility grid environments. *Performance Evaluation*, 58(2+3), 319-339.
- [SeBH06] T.Setzer; Bichler, M.; Hühn, O. (2006). Adaptive Zugriffskontrollverfahren - Ein Entscheidungsmodell für die Kontrolle des Zugriffs auf gemeinsam genutzte IT-Infrastrukturen. *Wirtschaftsinformatik*, 48(4).
- [SGKK06] Seltzsam, S.; Gmach, D.; Krompass, S.; Kemper, A. (2006). *AutoGlobe: An Automatic Administration Concept for Service-Oriented Database Applications*. In: *Proceedings of 22nd International Conference on Data Engineering, ICDE 2006*, Atlanta, GA, USA, 3.-8. April, 90.
- [ShHe00] Sheng, D.; Hellerstein, J.L. (2000). Predictive Models for Proactive Network Management: Application to a Production Web Server. *NOMS 2000 IEEE/IFIP Network Operations and Management Symposium*.
- [SmFT98] Smith, W.; Foster, I.; Taylor, V. (1998). *Predicting Application Run Times Using Historical Information*. In: *Proceedings of IPPS/SPDP '98: Workshop on Job Scheduling Strategies for Parallel Processing*, 122-142.
- [Sun06] Sun Microsystems (2006). Sun N1 Software, Sun, <http://www.sun.com/software/n1gridsystem/>, Accessed at: 2006-07-12.
- [VAHM02] Vilalta, R.; Apte, C.V.; Hellerstein, J.L.; Ma, S.; Weiss, S.M. (2002). Predictive algorithms in the management of computer systems. *IBM Systems Journal*, 41(3), 461-474.
- [VaSh03] Vazhkudai, S.; Schopf, J.M. (2003). Using Regression Techniques to Predict Large Data Transfers. *International Journal of High Performance Computing Applications*, 17(3), 249-268.
- [WNGM06] Wimmer, M.; Nicolescu, V.; Gmach, D.; Mohr, M.; Kemper, A.; Krcmar, H. (2006). *Evaluation of Adaptive Computing Concepts for Classical ERP Systems and Enterprise Services*. In: *Proceedings of IEEE Joint Conference on E-Commerce Technology and Enterprise Computing, E-Commerce and E-Services (CEC'06 and EEE'06)*, San Francisco, California, June 26-29, 352-355.
- [XZSW06] Xu, W.; Zhu, X.; Singhal, S.; Wang, Z. (2006). *Predictive control for dynamic resource allocation in enterprise data centers*. In: *Proceedings of 10th IEEE/IFIP Network Operations & Management Symposium (NOMS'06)*, Vancouver, Canada, 3.-7. April 2006

Analyse des Beitrages von Axiomatic Design zum Entwurf Serviceorientierter Architekturen

René Fiege, Dirk Stelzer

Fachgebiet für Informations- und Wissensmanagement
Technische Universität Ilmenau
98693 Ilmenau
rene.fiege@tu-ilmenau.de, dirk.stelzer@tu-ilmenau.de

Abstract

Axiomatic Design (AD) ist eine Methode, die den Entwurf beliebiger Systeme unterstützen kann. AD hilft, Anforderungen klar voneinander abzugrenzen und es unterstützt die Entwicklung von Systemen, deren Komponenten eine überschaubare Komplexität aufweisen und weitgehend unabhängig voneinander sind. Diese Ziele des AD korrespondieren mit wesentlichen Architekturzielen für Serviceorientierte Architekturen (SOA), nämlich „ausgewogene Granularität“, „lose Kopplung“ und „hohe Autonomie“ von Services. In diesem Papier analysieren wir, welchen Beitrag AD zum Entwurf von SOA leisten kann. Anhand eines Fallbeispiels untersuchen wir, inwiefern AD helfen kann, Services zu entwerfen, welche eine ausgewogene Granularität aufweisen und die in sich autonom und untereinander lose gekoppelt sind.

1 Einleitung

Serviceorientierte Architekturen (SOA) sollen es ermöglichen, wandlungsfähige bzw. „agile“ Architekturen für Informationssysteme (IS) zu realisieren, so dass diese leicht an neue Anforderungen angepasst werden können [SiHu2005, 71 ff.; ZiTP2003]. In SOA werden so genannte Services einer Vielzahl von Teilnehmern zur Nutzung bereitgestellt. Services kapseln wiederverwendbare Funktionen. Sie sollen lose gekoppelt sein und je nach Bedarf zu beliebigen Anwendungen zusammengestellt werden können [DJMZ2005, 7 ff.].

Das SOA-Konzept ist relativ jung und befindet sich immer noch in einer Phase der Erprobung und Weiterentwicklung [Erl2005, 72]. Obwohl es bereits einige viel versprechende Modelle zur Unterstützung von Entwurf, Implementierung, Betrieb und Wartung serviceorientierter Systeme gibt [z. B. BuGa2005; EAAC2004, 83 ff.; Erl2005, 359 ff.; KoHB2005; LaMB2005; MaBe2006, 99-149], sind verschiedene Herausforderungen bisher nicht zufrieden stellend gelöst worden. Hierzu gehört unter anderem, wie Services mit einer ausgewogenen Granularität entworfen werden können und wie bereits im Entwurf darauf hingewirkt werden kann, dass Services entstehen, welche in sich möglichst autonom und untereinander lose gekoppelt sind. Die Architekturziele ausgewogene Granularität, lose Kopplung und hohe Autonomie der Services sind wichtig, um die Wiederverwendbarkeit und Komponierbarkeit der Services zu erhöhen [BuGa2005, 602; Erl2005, 290 ff.].

Axiomatic Design (AD) ist eine Methode zur strukturierten Gestaltung von Objekten¹ [Suh2001]. Urheber und Anwender von AD behaupten, dass diese Methode geeignet ist, Systeme zu entwerfen, deren Komponenten (a) eine überschaubare Komplexität aufweisen sowie (b) weitgehend unabhängig voneinander sind und dass (c) Anforderungen an das zu entwerfende System klar voneinander abgegrenzt werden können [Suh2001, 29 ff.]. Diese Ziele des AD korrespondieren mit den Architekturzielen für SOA.

Ziel dieses Beitrages ist es zu überprüfen, inwiefern AD dazu beitragen kann, die oben genannten Architekturziele beim Entwurf von SOA zu erreichen. Hierzu werden zunächst Architekturziele für SOA vorgestellt. Anschließend stellen wir Grundlagen des AD dar und demonstrieren mit Hilfe einer Fallstudie, wie AD im Rahmen des Entwurfs von SOA eingesetzt werden kann. Im Anschluss analysieren wir den Beitrag von AD für den Entwurf von SOA kritisch. Dabei legen wir insbesondere dar, in wie weit AD helfen kann, die oben genannten Architekturziele zu erreichen. Der Beitrag wird mit einer kurzen Zusammenfassung und einem Ausblick abgeschlossen.

2 Architekturziele Serviceorientierter Architekturen

Architekturziele repräsentieren Prinzipien, die eine SOA charakterisieren [Erl2005, 290]. Diese Ziele müssen bereits im Entwurf berücksichtigt werden, damit sie sich in der resultierenden

¹ Objekte können Materialien, beliebige Systeme, Software, Hardware, strategische Geschäftspläne, Organisationen und Prozesse sein.

Architektur widerspiegeln. Typische Architekturziele von SOA sind: Wiederverwendbarkeit und Komponierbarkeit sowie angemessene Granularität, lose Kopplung, hohe Autonomie, Zustandslosigkeit, Auffindbarkeit, Abstraktheit, Interoperabilität, Geschäftsorientiertheit, Nachhaltigkeit, Neutralität und wohldefinierter Servicekontrakt [BuGa2005, 602; DJMZ2005, 9; EKAP2005, 28; EMPR2005, 3-4; Erl2005, 290; MaBe2006, 39 ff.; SiHu2005, 76-77]. Wir beschränken uns in diesem Beitrag auf fünf dieser Architekturziele. Diese Ziele und ihre Zusammenhänge sind in Abb. 1 zusammengefasst.

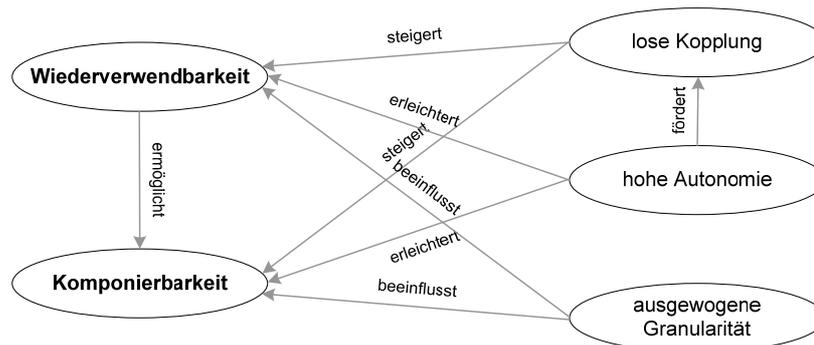


Abb. 1: Wesentliche Architekturziele Serviceorientierter Architekturen

Eines der wichtigsten Architekturziele ist die Wiederverwendbarkeit der Services. Sie sollte möglichst hoch sein, damit die SOA an Änderungen von Anforderungen ohne unangemessen hohen Entwicklungsaufwand angepasst werden kann [Erl2005, 292]. Damit unterstützt die Wiederverwendbarkeit auch die Agilität einer SOA, also die Leichtigkeit, mit der sich Änderungen oder Erweiterungen einer Architektur durchführen lassen [HaSc2006, 277].

Komponierbarkeit bedeutet, dass ein höherwertiger Service aus mehreren generischen Services zusammengesetzt werden kann. Die Logik, die in einem generischen Service gekapselt ist, kann so auf verschiedenen Granularitätsebenen wiederverwendet werden. Auch die Komponierbarkeit der Services einer SOA sollte möglichst groß sein, damit die SOA leicht an zukünftige Änderungen von Anforderungen angepasst werden kann [Erl2005, 301]. Die Wiederverwendbarkeit wirkt daher ebenfalls auf die Agilität der SOA [HaSc2006, 277].

Komponierbarkeit und Wiederverwendbarkeit hängen stark von der gewählten Granularität der Services ab [HaSc2006, 281]. Die Granularität wird bestimmt durch das Abstraktionsniveau der Aufgaben, die durch den Service unterstützt werden [Erl2005, 299]. Sie ist daher vergleichbar mit dem Prinzip der Abstraktion in der Softwareentwicklung [Balz1998, 559]. Granularität beschreibt den Umfang und die Art der Funktionen, welche durch den Service unterstützt werden [Balz1998, 559; Erl2005, 299, 302; MaBe2006, 40, 124]. Je weniger Funktionen und je

konkreter die Funktionen auf einen einzelnen Aufgabenbereich ausgerichtet sind, desto feiner ist die Granularität (und desto niedriger das Abstraktionsniveau). Die Servicegranularität sollte angemessen gewählt werden, um die Kompositions- und Wiederverwendungspotentiale zu erhöhen. Eine zu grobe Servicegranularität kann dazu führen, dass ein Service auf Grund von Performanceproblemen nicht wiederverwendet werden kann. Eine zu feine Granularität kann das Wiederverwendungspotential senken, da der Service auf einen bestimmten Aufgabenbereich zugeschnitten ist und nicht in anderen Aufgabenbereichen verwendet werden kann [MaBe2006, 40].

Lose Kopplung umfasst die Reduzierung von Abhängigkeiten zwischen Services [Balz1998, 474; CaGl1990, 31-41; Kaye2003; Raas1993, 364; VACI2005, 114]. Sie wird unter anderem dadurch erreicht, dass die Services untereinander über genau definierte Schnittstellen interagieren [Erl2005, 314]. Dadurch kann die Implementierung eines Service leicht ausgetauscht werden. Durch Reduzierung der Abhängigkeiten zwischen den Services wird außerdem das Potential zur Komposition und Wiederverwendung erhöht [MaBe2006, 41]. Daher sollte eine möglichst lose gekoppelte SOA angestrebt werden.

Serviceautonomie erfordert, dass die Logik, die innerhalb eines Service gekapselt ist, im Hinblick auf einen definierten Kontext (z. B. eine zu erfüllende Funktion) klar abgegrenzt werden kann [Erl2005, 303]. Autonomie entspricht dem Prinzip der Kohäsion in der Softwareentwicklung [Balz1998, 474; VACI2005, 117]. Eine hohe Autonomie liegt insbesondere dann vor, wenn sich die Servicelogik auf genau einen Kontext bezieht [Erl2005, 304]. Eine hohe Serviceautonomie minimiert Abhängigkeiten zwischen Services und fördert eine lose Kopplung [Balz1998; Erl2005, 303; VACI2005, 118]. Außerdem erleichtert die klare Abgrenzung und Kapselung der Servicelogik die Wiederverwendbarkeit und Komponierbarkeit von Services [Erl2005, 318]. Es sollte daher auch eine hohe Autonomie der Services angestrebt werden.

3 Grundlagen des Axiomatic Design

AD wurde Ende der 70er Jahre von Nam Pyo Suh am Massachusetts Institute of Technology (MIT) entwickelt [Suh1990, 18]. Es handelt sich dabei um eine Methode, mit der man strukturiert beliebige Objekte entwerfen kann. Das Grundprinzip von AD umfasst die strukturierte Suche und Zuordnung geeigneter Lösungen für zuvor festgelegte Anforderungen.

Ein Entwurf ist definiert als das Ergebnis dieses Zuordnungsprozesses [Suh2001, 2 ff.]. Er beschreibt, welche Anforderungen durch welche Lösung erfüllt werden können. AD basiert auf dem Konzept der Domänen sowie auf dem so genannten Unabhängigkeits- und dem Informationsaxiom. Beide Axiome formulieren Richtlinien für den Entwurfsprozess. Aus Platzgründen gehen wir in diesem Beitrag auf das Informationsaxiom nur kurz ein.

3.1 Konzept der Domänen

Das Konzept der Domänen umfasst die Kundendomäne, die funktionale und die physische Domäne sowie die Prozessdomäne (Abb. 2). Der Entwurfsprozess erstreckt sich über alle Domänen. Er beginnt in der Kundendomäne und endet in der Prozessdomäne. Jede Vorgängerdomäne beschreibt Anforderungen, jede Folgedomäne die korrespondierenden Lösungen. Zwischen allen Domänen erfolgt eine Zuordnung von Anforderungen zu korrespondierenden Lösungen [Suh2001, 10-14].

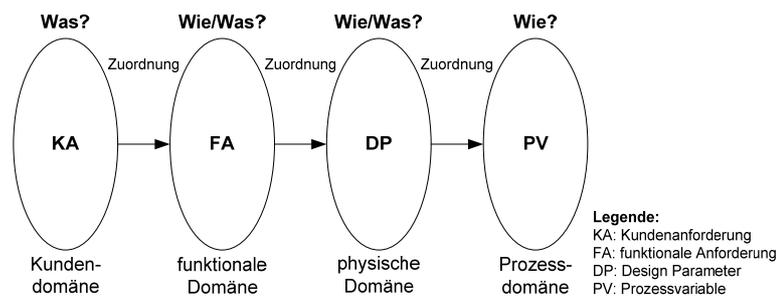


Abb. 2: Konzept der Domänen

Die Kundendomäne beinhaltet die Kundenanforderungen (KA) an das zu gestaltende Objekt. Sie beschreiben relativ grob, welche Eigenschaften das zu entwickelnde Objekt haben soll. Suh empfiehlt, die Kundenanforderungen nach ihrer Wichtigkeit zu ordnen [Suh2001, 14]. Die Kundenanforderungen werden in der funktionalen Domäne weiter zu funktionalen Anforderungen (FA) und Restriktionen (R) verfeinert. Funktionale Anforderungen beschreiben aus Sicht des Designers alle Anforderungen, die geeignet sind, die Bedürfnisse des Kunden abzudecken. Anzustreben ist die minimale Anzahl funktionaler Anforderungen, indem nur die wesentlichen Erfordernisse identifiziert werden. Alle weiteren Anforderungen werden auf tiefer liegenden Entwurfsebenen (vgl. Abschnitt 3.2) bearbeitet, da sie die Komplexität des Entwurfs erhöhen würden. Restriktionen ergänzen die funktionalen Anforderungen, indem sie den Lösungsraum einschränken. Sie repräsentieren zusätzliche Anforderungen und beschreiben Grenzen für korrespondierende Lösungen der funktionalen Anforderungen. Derartige Lösungen werden als Designparameter (DP) bezeichnet. Sie sind Inhalt der physischen Domäne. Im

Idealfall erhält jede funktionale Anforderung genau einen korrespondierenden Designparameter. Bei der Auswahl geeigneter Parameter müssen die Restriktionen berücksichtigt werden. Die Designparameter repräsentieren ihrerseits wieder Anforderungen für die Folgedomäne. Die Prozessdomäne umfasst die Prozessvariablen (PV). Hierbei handelt es sich um alle Hilfsmittel zur Erzeugung der Designparameter. Sie charakterisieren den Herstellungsprozess der Parameter. Idealerweise wird jedem Designparameter genau eine Prozessvariable zugeordnet.

3.2 Unabhängigkeitsaxiom

Während der Zuordnung zwischen den Domänen wird das Unabhängigkeitsaxiom angewendet. Im Folgenden wird dies anhand der Zuordnung zwischen der funktionalen und der physischen Domäne erläutert. Die Zuordnung und Anwendung des Axioms zwischen den anderen Domänen verläuft analog [SuDo2000, 37-38; Suh2001]. Das Unabhängigkeitsaxiom verlangt, dass die Unabhängigkeit der funktionalen Anforderungen nach Zuordnung geeigneter Designparameter gewahrt bleibt. Vollständige Unabhängigkeit liegt vor, wenn der gefundene Designparameter für eine spezifische funktionale Anforderung keine Auswirkungen auf andere funktionale Anforderungen hat. D. h., dass jede funktionale Anforderung durch genau einen Designparameter erfüllt wird. Der Zuordnungsprozess zwischen den Domänen wird hierarchisch im Top-down-Verfahren durchgeführt, um den Entwurf weiter zu verfeinern. Man spricht in diesem Zusammenhang vom Dekompositionsprozess (Abb. 3).

Der Dekompositionsprozess verlangt, dass zwischen den Domänen hin und her gesprungen wird. Wie in Abb. 3 dargestellt, springt man ausgehend von einer funktionalen Anforderung in die physische Domäne, um einen geeigneten Designparameter zuzuordnen. Anschließend erfolgt der Rücksprung in die funktionale Domäne.

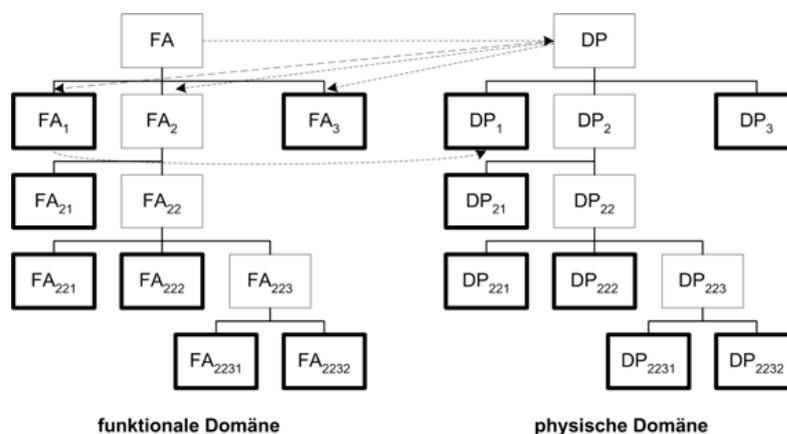


Abb. 3: Dekompositionsprozess

Die funktionale Anforderung wird auf der zweiten Ebene in ihre Teilanforderungen (FA₁, FA₂ und FA₃) zerlegt. Für jede Teilanforderung erfolgt wieder die Zuordnung geeigneter Designparameter (DP₁, DP₂ und DP₃). Dieser Prozess wird solange wiederholt, bis so genannte „elementare FA-DP-Kombinationen“ gefunden wurden. Eine Elementarkombination (diese sind in Abb. 3 fett hervorgehoben) liegt vor, wenn für eine funktionale Anforderung ein Designparameter gefunden wird, der unmittelbar, d. h. ohne weitere Dekomposition, implementierbar ist.

Sobald eine Hierarchieebene im Dekompositionsprozess fertig gestellt wurde, wird das Unabhängigkeitsaxiom zur Prüfung der Unabhängigkeit der funktionalen Anforderungen herangezogen. Hierzu wird die Einflussmatrix [A] gebildet. Sie zeigt die Beziehung zwischen den funktionalen Anforderungen und Designparametern einer Hierarchieebene. Diese Beziehung wird durch folgende Gleichungen zum Ausdruck gebracht.

$$\{FA\} = [A]\{DP\} \text{ bzw. } \begin{cases} FA_1 \\ FA_2 \\ \dots \\ FA_n \end{cases} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{cases} DP_1 \\ DP_2 \\ \dots \\ DP_n \end{cases} \quad (1)$$

Die funktionalen Anforderungen und Designparameter einer Hierarchieebene konstituieren die Vektoren {FA} und {DP}. Die Einflussmatrix [A] ist eine $n \times n$ -Matrix. Die Elemente A_{ij} beinhalten entweder den Wert „0“ oder „X“ (im Falle eines linearen Entwurfs, von dem hier ausgegangen wird [Suh2001, 19]). $A_{ij} = 0$ bedeutet, dass FA_i nicht durch DP_j beeinflusst wird. $A_{ij} = X$ sagt aus, dass DP_j Einfluss auf FA_i hat. Die drei folgenden Ausprägungen der Einflussmatrix [A] werden unterschieden: 1. Wenn für [A] gilt: alle $A_{ij} = 0$ für $i \neq j$, dann nehmen alle Elemente der Einflussmatrix, außer die auf der Hauptdiagonalen, den Wert „0“ an. Eine solche Matrix wird als Diagonalmatrix bezeichnet. 2. Wenn für [A] gilt: entweder oberhalb oder unterhalb der Diagonalen sind alle $A_{ij} = 0$, dann wird [A] als Triangulärmatrix bezeichnet. 3. Wenn für [A] gilt: sowohl oberhalb als auch unterhalb der Diagonalen existieren $A_{ij} = X$, dann wird [A] als Vollmatrix bezeichnet.

1. ungekoppelter Entwurf (Diagonalmatrix)	2. entkoppelter Entwurf (Triangulärmatrix)	3. gekoppelter Entwurf (Vollmatrix)
$\begin{cases} FR_1 \\ FR_2 \\ FR_3 \end{cases} = \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix} \begin{cases} DP_1 \\ DP_2 \\ DP_3 \end{cases}$	$\begin{cases} FR_1 \\ FR_2 \\ FR_3 \end{cases} = \begin{bmatrix} X & 0 & 0 \\ X & X & 0 \\ X & X & X \end{bmatrix} \begin{cases} DP_1 \\ DP_2 \\ DP_3 \end{cases}$	$\begin{cases} FR_1 \\ FR_2 \\ FR_3 \end{cases} = \begin{bmatrix} X & 0 & X \\ X & X & 0 \\ X & X & X \end{bmatrix} \begin{cases} DP_1 \\ DP_2 \\ DP_3 \end{cases}$

Tab. 1: Ausprägungen der Einflussmatrix

Entsprechend der Ausprägungen der Einflussmatrix unterscheidet AD zwischen den drei Entwurfstypen der Tab. 1. Nur im ungekoppelten Entwurf sind alle funktionalen Anforderungen unabhängig voneinander, da jede funktionale Anforderung durch genau einen Designparameter erfüllt wird. Er repräsentiert daher die ideale Erfüllung des Unabhängigkeitsaxioms. Bei den anderen Typen gibt es zusätzliche Abhängigkeiten. Diese treten beim gekoppelten Entwurf am häufigsten auf. Er stellt daher eine Verletzung des Unabhängigkeitsaxioms dar. Ein Designer wendet das Unabhängigkeitsaxiom auf jeder Hierarchieebene des Dekompositionsprozesses an. Stellt er auf einer Ebene fest, dass ein gekoppelter Entwurf entstanden ist, versucht er die ermittelten FA-DP-Kombinationen so zu überarbeiten, dass entweder ein ungekoppelter oder ein entkoppelter Entwurf entsteht. Erst danach wird der Dekompositionsprozess auf tiefer liegenden Ebenen fortgesetzt.

3.3 Informationsaxiom

Liegen alternative FA-DP-Kombinationen vor, die das Unabhängigkeitsaxiom erfüllen, wird das Informationsaxiom angewendet, um den besten Entwurf zu ermitteln. Das Informationsaxiom ermöglicht eine quantitative Bewertung gegebener Entwürfe und ermöglicht eine Reduktion der Komplexität sowie eine Erhöhung der Zuverlässigkeit [ClHi2000, 274; Suh2001, 39 ff.]. In diesem Beitrag gehen wir aus Platzgründen nicht weiter auf das Informationsaxiom ein.

4 Axiomatic Design im Entwurf Serviceorientierter Architekturen

AD wurde für den Maschinenbau entwickelt und zunächst zum Entwurf von Produkten angewendet [Suh2001, 11]. Mittlerweile wurde diese Methode auch in vielen anderen Gebieten erfolgreich eingesetzt [DoSu2000; EnNo2000; Suh1990, 323-352; Suh1998; Suh2001, 341-375]. Tab. 2 gibt einen Überblick über publizierte Anwendungen von AD im Softwareentwurf.

Anwendung des Axiomatic Design zum Entwurf...	Quelle
... von Steuerungssoftware für elektromechanische Systeme	[HiNa1999]
... von Benutzerschnittstellen von Softwaresystemen	[Jams2004]
... einer Bibliotheksverwaltungssoftware	[KiSK1991]
... der objektorientierten Software Acclaro.	[SuDo2000]
... von Software für Programmable Logic Controller	[ScTs2000]

Tab. 2: Anwendung des Axiomatic Design im Softwareentwurf

Die oben aufgeführten Projektbeispiele (Tab. 2) belegen, dass die Vorteile des AD auch im Softwareentwurf erzielt werden können [DoSu1999, 121; HiNa1999, 1-2; Jams2004, 1;

ScTs2000, 270; SuDo2000, 95-96]. Zwischen dem Entwurf SOA und dem Entwurf von Software gibt es viele Parallelen [CeHa2005, 11; Erl2005, 321 ff.; KoHB2005]. Wir stellen die These auf, dass AD auch den Entwurf von SOA verbessern kann. Wir wollen überprüfen, ob AD helfen kann, die Architekturziele „ausgewogene Granularität“, „lose Kopplung“ und „hohe Autonomie“ zu erreichen.

Die Anwendung des AD auf den Entwurf SOA basiert auf Grundlagen, die in den vorhergehenden Abschnitten beschrieben wurden. Der Entwurfsprozess und die Inhalte der Domänen müssen allerdings an die Besonderheiten von SOA angepasst werden. Im Folgenden werden die relevanten Schritte im Entwurf SOA abgegrenzt. Anschließend illustrieren wir die Anwendung des AD beim Entwurf von SOA anhand einer Fallstudie.

4.1 Eingrenzung des relevanten Gegenstandsbereiches

Als Grundlage für die Anwendung des AD dient ein idealtypisches Vorgehensmodell zur Entwicklung SOA (Abb. 4).

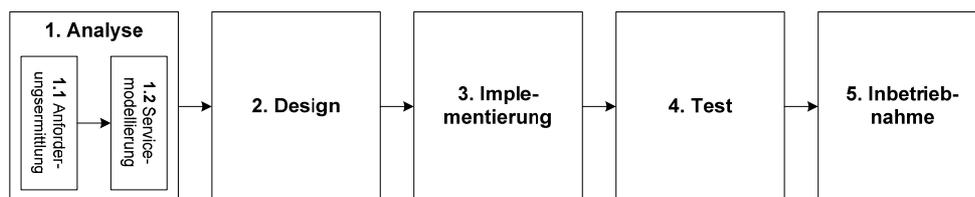


Abb. 4: Entwicklungsprozess Serviceorientierter Architekturen

Dieses Modell haben wir aus einer Synopse der Vorgehensmodelle von [BuGa2005; EAAC2004, 83 ff.; Erl2005, 359 ff.; KoHB2005; LaMB2005; MaBe2006, 99-149] abgeleitet. Es repräsentiert einen reinen Top-Down-Ansatz. Auf Grund der Problemkomplexität werden in diesem Beitrag Bottom-Up-Vorgehensweisen und Mischformen bewusst ausgeblendet. Dieses Modell bezeichnen wir im Folgenden als Entwicklungsprozess für SOA.

4.2 Anwendung des Axiomatic Design im Entwurf Serviceorientierter Architekturen

Die Anwendung des AD auf den Entwurf SOA haben wir auf der Grundlage des so genannten V-Modells des AD (Abb. 5) durchgeführt [DoSu2000, 279 ff.]. Wir haben uns dabei an der Anwendung des AD im Softwareentwurf orientiert. Anpassungen für den Entwurf SOA waren hinsichtlich der Auswahl der zu modellierenden Gegenstände in den einzelnen Domänen des AD notwendig. Außerdem mussten geeignete Vorschriften zur Ableitung von serviceorientierten Konstrukten aus der Gesamteinflussmatrix gefunden werden. Das V-Modell des AD darf nicht mit Vorgehensmodellen der Softwareentwicklung, zum Beispiel dem V-

Modell von Boehm [Boeh1981] oder dem V-Modell[®] XT der Bundesbehörden [KBSt2005], verwechselt werden. Die Intention des V-Modells ist, dass alle methodischen Schritte des linken Astes (Schritte eins bis vier) im Sinne des AD bearbeitet werden. Alle Schritte des rechten Astes (Schritte fünf bis sieben) können mit beliebigen Hilfsmitteln der Softwareentwicklung kombiniert werden [Suh2001, 266]. AD hilft, den Entwurf auf relevante Architekturziele zu fokussieren. Im Folgenden werden die einzelnen Schritte des V-Modells und deren Anwendung auf den Entwurf SOA beschrieben und an einem Fallsbeispiel demonstriert. Dieses Beispiel basiert auf einer Fallstudie von Erl [Erl2005, 430-444]. Grundlage ist ein Prozess zur Vorlage und Prüfung von Arbeitszeitznachweisen der Mitarbeiter, welcher durch eine SOA abgebildet werden soll. Dieser Prozess soll automatisiert werden.

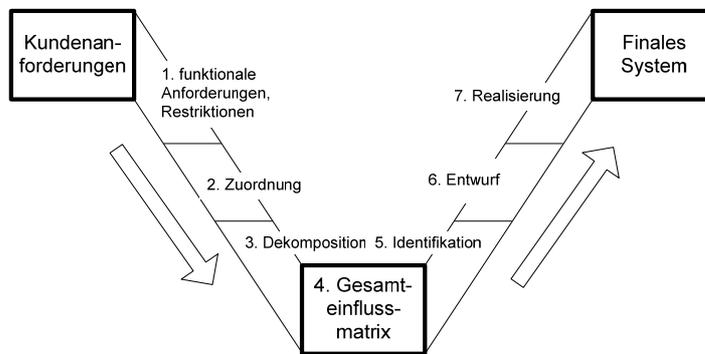


Abb. 5: V-Modell des Axiomatic Design (In Anlehnung an [SuDo2000, 96])

Der erste Schritt betrifft die Kundendomäne und die funktionale Domäne im AD. Er umfasst die Ermittlung von Kundenanforderungen an das zu entwerfende Softwaresystem. Die Kundenanforderungen umreißen relativ grob die Eigenschaften der zu entwickelnden SOA. Sie werden aus den Geschäftsprozessen abgeleitet, die in Phase „1.1 Anforderungsermittlung“ des Entwicklungsprozesses für SOA ermittelt werden [Erl2005, 363-365]. Diese Phase liefert den Informationsinput für den ersten Schritt des V-Modells des AD. Im Rahmen des Fallbeispiels wurde der folgende Prozess ermittelt (Abb. 6).

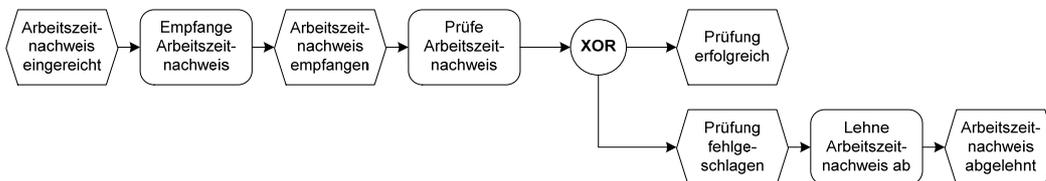


Abb. 6: Prozess zur Vorlage und Prüfung von Arbeitszeitznachweisen (In Anlehnung an [Erl2005, 430-444])

Den Kundenanforderungen werden funktionale Anforderungen und Restriktionen zugeordnet. Die funktionalen Anforderungen bilden dabei logische Kontexte ab (z. B. die betriebswirtschaftliche Funktion „Rechnung buchen“ oder die Entität „Buchhaltungssystem“),

nach denen später Operationen und Services gruppiert werden können. Restriktionen ergänzen die funktionalen Anforderungen, indem sie den Lösungsraum der physischen Domäne einschränken. Die wichtigste Kundenanforderung, die in unserem Beispiel ermittelt wurde, lautet: KA1: „Wir benötigen eine SOA, welche die vollautomatische Bearbeitung des Prozesses zur Vorlage und Prüfung von Arbeitszeitnachweisen ermöglicht“. Dieser Kundenanforderung wird folgende funktionale Anforderung zugeordnet: FA1: „Bilde den Prozess der Vorlage und Prüfung von Arbeitszeitnachweisen ab“. Eine typische Restriktion beim Entwurf SOA ist die Anforderung, die SOA auf Basis von Webservices zu entwickeln. Im zweiten und dritten Schritt werden den funktionalen Anforderungen geeignete Designparameter in der physischen Domäne zugeordnet. Die Designparameter repräsentieren Daten der Serviceverarbeitung [Erl2005, 35-37]. Es muss darauf geachtet werden, dass die Designparameter keine Restriktionen verletzen. Anschließend erfolgt die Dekomposition der funktionalen Anforderungen und Designparameter durch Sprung zwischen der funktionalen und der physischen Domäne. Während der Dekomposition muss auf jeder Hierarchieebene sichergestellt werden, dass das Unabhängigkeitsaxiom erfüllt ist. Dies geschieht durch Prüfung der Zuordnungsbeziehungen der Einflussmatrizen, die auf jeder Ebene gebildet werden. Abhängigkeiten, die auf der Diagonale einer Matrix liegen, werden durch Großbuchstaben, alle sonstigen Abhängigkeiten durch Kleinbuchstaben dargestellt. In diesem Fallbeispiel werden insgesamt fünf Dekompositionsebenen gebildet. Aus Platzgründen stellen wir nur die Matrix der zweiten Ebene dar (Tab. 3). Die Inhalte der Matrix verdeutlichen, dass z. B. zur Erfüllung der funktionalen Anforderung: FA11: „nehme Arbeitszeitnachweis entgegen“ und FA12: „verarbeite Arbeitszeitnachweis“ die Daten des Arbeitszeitnachweises, repräsentiert durch: DP11: „Arbeitszeitnachweis“, benötigt werden.

	DP11: Arbeitszeitnachweis	DP12: Daten zur Verarbeitung des Arbeitszeitnachweises
FA11: nehme Arbeitszeitnachweis entgegen	B	
FA12: verarbeite Arbeitszeitnachweis	A	C

Tab. 3: Einflussmatrix der zweiten Dekompositionsebene

Zur Erfüllung der funktionalen Anforderung FA12: „verarbeite Arbeitszeitnachweis“ werden zusätzlich Daten benötigt, die während der Verarbeitung eines Arbeitszeitnachweises wichtig sind. Diese Daten werden auf der zweiten Dekompositionsebene noch sehr abstrakt als: DP12: „Daten zur Verarbeitung des Arbeitszeitnachweises“ bezeichnet. Auf den tiefer liegenden

Dekompositionsebenen (Tab. 4) werden diese Daten jedoch weiter verfeinert und konkretisiert, z. B. zu: DP1141: „Profildaten“ oder DP11221: „abgerechnete Stunden“.

In Schritt vier wird die Gesamteinflussmatrix (Tab. 4) gebildet. Diese Matrix zeigt die Beziehungen zwischen den funktionalen Anforderungen und Designparametern aller Hierarchieebenen. Sie ergibt sich aus der Zusammenfassung der Einflussmatrizen aller Hierarchiestufen des Dekompositionsprozesses. Die Gesamteinflussmatrix ist eine wichtige Grundlage für die folgenden Schritte. Mit ihrer Hilfe wird überprüft, ob die Designentscheidungen konsistent sind, die während der Dekomposition getroffen wurden.

		DP1: Daten des Prozesses der Arbeitszeitzachweisvorlage											
		DP12: Daten zur Verarbeitung des Arbeitszeitzachweises											
		DP112: Prüfdaten				DP113: Prüfergebnisdaten		DP114: Mitarbeiterdaten	DP115: Mitteilungsdaten				
		DP1121: Arbeitszeitzachweisdaten		DP1122: Rechnungsdaten		DP1131: Kundenabrechnungsübereinstimmung	DP1132: Genehmigungsexistenz	DP1141: Profildaten	DP1151: Mitarbeiterablehnung	DP1152: Vorgesetztenmitteilung			
		DP11211: Stundenachweis	DP11212: Überstundenachweis	DP11213: Genehmigungsachweis	DP11221: abgerechnete Stunden								
FA1: bilde den Prozess der Arbeitszeitzachweisvorlage ab	FA12: verarbeite Arbeitszeitzachweis	FA111: speichere Arbeitszeitzachweis		D: speichere Arbeitszeitzachweis						C			
		FA112: ermitte Profildaten	FA1121: ermitte Arbeitszeitzachweisdaten	FA11211: ermitte Stunden	a: Nachricht D	P: ermitte Stunden					E		
				FA11212: ermitte Überstunden	a: Nachricht D	Q: ermitte Überstunden					I		
				FA11213: ermitte Genehmigungen	a: Nachricht D	R: ermitte Genehmigungen					J		
			FA1122: ermitte Rechnungsdaten	FA11221: ermitte abgerechnete Stunden					S: ermitte abgerechnete Stunden		F		
		FA113: prüfe Arbeitszeitzachweis	FA1131: prüfe Übereinstimmung mit Kundenabrechnung		b: Nachricht P						K: prüfe Übereinstimmung		
			FA1132: prüfe Genehmigungen für Überstunden		b: Nachricht Q		b: Nachricht R						L: prüfe Genehmigungen
		FA114: aktualisiere Mitarbeiterdaten	FA1141: aktualisiere Profildaten						c: Nachricht K	c: Nachricht L	M: aktualisiere Profildaten		H
		FA115: sende Mitteilungen	FA1151: sende Ablehnung an Mitarbeiter						d: Nachricht K	d: Nachricht L	e: Nachricht M		N: sende Ablehnung
			FA1152: sende Mitteilung an Vorgesetzten						d: Nachricht K	d: Nachricht L	e: Nachricht M		O: sende Mitteilung

Tab. 4: Gesamteinflussmatrix²

Der fünfte Schritt dient der Identifikation serviceorientierter Konstrukte. Die Inhalte der Gesamteinflussmatrix werden dabei auf Services und Operationen abgebildet. Tab. 5 zeigt einen Ausschnitt des Ergebnisses. Aus den funktionalen Anforderungen werden vorläufige Services, so genannte Servicekandidaten, abgeleitet. Die Daten, die der Service verarbeitet, ergeben sich aus den Designparametern. Operationen der Services werden durch die Elemente innerhalb der Gesamteinflussmatrix repräsentiert.

² Eine größere Darstellung dieser Matrix kann über die Website <http://www.wirtschaft.tu-ilmeneau.de/im/> abgerufen werden.

Service	FA1 Arbeitzeitsnachweisvorlageprozess	FA11 Arbeitszeitsnachweisentgegennahme	FA12 Arbeitszeitsnachweisverarbeitung	FA112 Prüfdatenermittlung	FA113 Arbeitszeitsnachweisprüfung
Daten	DP11 Arbeitzeitsnachweis DP12 Daten zur Verarbeitung des Arbeitzeitsnachweises	DP111 Arbeitzeitsnachweis	DP112 Prüfdaten DP113 Prüfergebnisdaten DP114 Mitarbeiterdaten DP115 Mitteilungsdaten	DP1121 Arbeitszeitsnachweisdaten DP1122 Rechnungsdaten	DP1131 Kundenabrechnungsübereinstimmung DP1132 Genehmigungs-existenz
Operationen	A Arbeitzeitsnachweisvorlageprozess	B Arbeitszeitsnachweisentgegennahme D speichereArbeitszeitsnachweis	C Arbeitszeitsnachweisverarbeitung	E Prüfdatenermittlung	F Arbeitszeitsnachweisprüfung K prüfeÜbereinstimmung L prüfeGenehmigungen

Tab. 5: Identifikation serviceorientierter Konstrukte

Im sechsten Schritt werden die identifizierten Konstrukte in eine vorläufige SOA überführt. Die Ergebnisse werden anschließend der Phase „2. Design“ des Entwicklungsprozesses für SOA zugeführt. Die Spezifikation der SOA kann mit Hilfsmitteln der UML abgebildet werden [LaPi2005]. Sie beinhaltet die zuvor identifizierten Services, deren Operationen und Daten. Außerdem werden Abhängigkeiten zwischen Services auf Grund einer Servicekomposition oder auf Grund des Nachrichtenaustausches zwischen Services spezifiziert. Diese Abhängigkeiten werden ebenfalls aus der Gesamteinflussmatrix abgeleitet. Servicekompositionen ergeben sich aus der hierarchischen Strukturierung der funktionalen Anforderungen und Designparameter. Der Nachrichtenaustausch zwischen Services wird aus den nichtdiagonalen Inhaltselementen der Gesamteinflussmatrix abgeleitet.

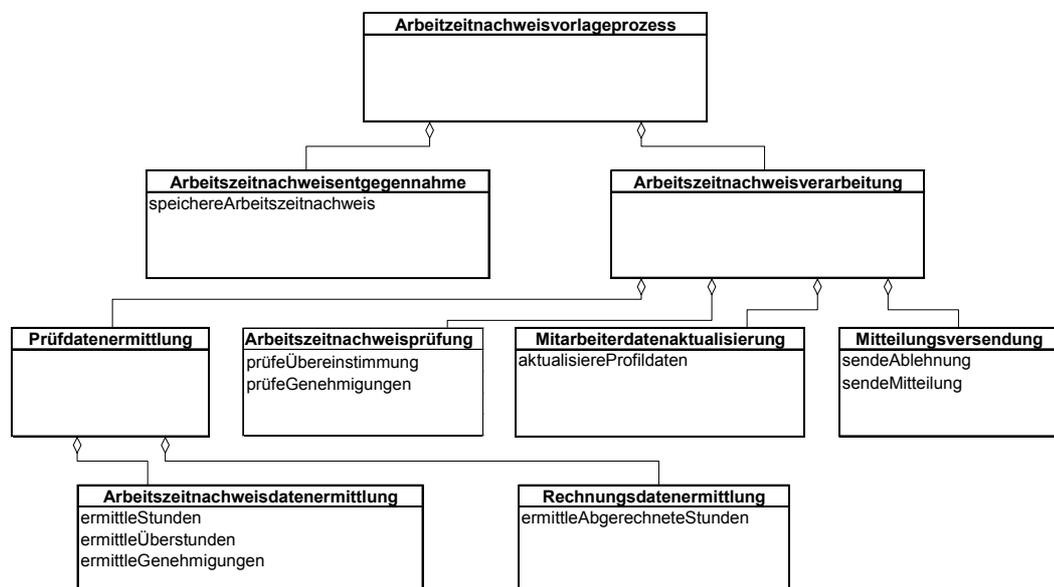


Abb. 7: Servicekomposition

Abb. 7 wurde in Anlehnung an ein UML-Klassendiagramm erstellt. Es beinhaltet alle im letzten Schritt identifizierten Services sowie deren Operationen und Abhängigkeiten.

Kompositionsbeziehungen zwischen Services werden durch Aggregationsbeziehungen im UML-Diagramm ausgedrückt.

Schritt sieben betrifft die Prozessdomäne im AD. Er bezieht sich auf die Realisierung der SOA auf einer technischen Plattform. Da es sich hierbei um Aufgaben außerhalb des Entwurfs handelt, verzichten wir auf eine Beschreibung dieses Schritts.

4.3 Kritische Analyse des Beitrags von Axiomatic Design

Im Folgenden wollen wir darlegen, welchen Beitrag AD zur Erreichung der Architekturziele „hohe Autonomie“, „lose Kopplung“ und „ausgewogene Granularität“ beim Entwurf von SOA leisten kann und welche Nachteile dem gegenüber stehen.

Die Anwendung von AD fördert eine hohe Autonomie der Services, da Entwurfsanforderungen klar voneinander abgegrenzt werden. Im Rahmen der Zuordnung und Dekomposition (Schritt zwei und drei im V-Modell des AD) wird sichergestellt, dass für jede funktionale Anforderung einer Dekompositionsebene auf der nächst tieferen Ebene ausschließlich FA-DP-Kombinationen erarbeitet werden, die einen Beitrag zur Erfüllung dieser funktionalen Anforderung leisten. Im Fallbeispiel (Tab. 4) zielen z. B. die funktionalen Anforderungen FA1131, FA1132 und korrespondierende DP ausschließlich auf die Erfüllung der funktionalen Anforderung FA113: „prüfe Arbeitszeitnachweis“. Dies zeigt, dass auf jeder Dekompositionsebene kohärente Teillösungen gruppiert werden, die auf die Erfüllung einer funktionalen Anforderung abzielen. Auf diese Weise entstehen Services, deren Operationen ebenfalls auf genau einen definierten Kontext (z. B. eine Aufgabe) ausgerichtet sind. Aus der funktionalen Anforderung FA113 wird der Service „Arbeitszeitnachweisprüfung“ abgeleitet (Abb. 7). Er beinhaltet die zwei Operationen „prüfeÜbereinstimmung“ und „prüfeGenehmigung“. Diese Operationen sind kohärent, da sie beide auf genau einen Kontext (die Aufgabe, den Arbeitszeitnachweis zu prüfen) ausgerichtet sind.

Durch AD wird auch das Ziel der losen Kopplung gefördert, da die Unabhängigkeit von Entwurfsbestandteilen angestrebt wird. Dies wird durch das Unabhängigkeitsaxiom erzielt. Es stellt sicher, dass jede funktionale Anforderung durch möglichst wenige Designparameter beeinflusst wird. Abhängigkeiten sind nur auf oder unterhalb der Diagonalen der Gesamteinflussmatrix erlaubt. Im Fallbeispiel (Tab. 4) ist das Unabhängigkeitsaxiom erfüllt, da oberhalb der Diagonalen der Gesamteinflussmatrix alle Felder leer sind. Durch Verminderung der Abhängigkeiten in der Gesamteinflussmatrix wird in der SOA die Anzahl der Beziehungen zwischen den Services reduziert. Im Fallbeispiel erhält z. B. der Service

„Arbeitszeitnachweisprüfung“ einen Dateninput vom Service „Prüfdatenermittlung“. In der Gesamteinflussmatrix wird dies durch die nichtdiagonalen Elemente „b“ verdeutlicht. Dieser Service sendet selbst Daten an die Services „Mitarbeiterdatenaktualisierung“ und „Mitteilungsversendung“. Dies wird durch die nichtdiagonalen Elemente „c“ und „d“ verdeutlicht.

AD trägt auch zu einer ausgewogenen Servicegranularität bei, da Entwurfsbestandteile in überschaubarer Komplexität entstehen. Das Top-Down-Vorgehen bei der Zuordnung und Dekomposition bewirkt, dass ein Gesamtsystem rekursiv in immer feiner granulierten Subsysteme zerlegt wird. Jede Dekompositionsebene enthält nur die FA-DP-Kombinationen, die dem Abstraktionsniveau dieser Ebene entsprechen. In der Gesamteinflussmatrix (Tab. 4) wurde z. B. auf der dritten Dekompositionsebene die noch relativ abstrakte FA112: „ermittle Prüfdaten“ festgelegt. Erst auf der vierten und fünften Ebene erfolgt eine Konkretisierung hinsichtlich der zu ermittelnden Daten – z. B. konkretisiert FA1121, dass Arbeitszeitnachweisdaten ermittelt werden müssen, FA11211 konkretisiert noch stärker, dass es sich dabei u. a. um Stunden handelt. Auf diese Weise wird die Komplexität der gesamten SOA über mehrere Ebenen auf Einheiten überschaubarer Größe verteilt. So wird sichergestellt, dass keine zu grob granulierten Services entstehen. Im geschilderten Beispiel wurde aus FA112 der Service „Prüfdatenermittlung“ abgeleitet (Abb. 7). Er sorgt für die Komposition der feiner granulierten Services „Arbeitszeitnachweisdatenermittlung“ und „Rechnungsdatenermittlung“. Diesen Vorteilen stehen einige negative Aspekte von AD gegenüber. Ein Nachteil der Anwendung von AD ist der hohe Dokumentationsaufwand. Auf jeder Ebene des Zuordnungs- und Dekompositionsprozesses müssen Designgleichungen und -matrizen erstellt werden. Zwar kann dieser Aufwand durch Verwendung der Software Acclaro[®] verringert werden.³ Die FA-DP-Kombinationen müssen aber in jedem Fall manuell eingegeben werden. Die Verfechter des AD führen die starke Strukturierung und Formalisierung von Entwurfsprozessen als Vorteil auf, da positive Effekte, wie die Reduzierung von Entwurfsschritten und eine Erhöhung der Kreativität der Designer, entstehen sollen [Suh2001, 239 ff.]. Allerdings ist die Erstellung der Gesamteinflussmatrix aufwändig. Die benötigte Zeit fehlt evtl. für andere Aufgaben. Wie die Softwareentwicklung zeigt, kann eine zu starke Strukturierung und Formalisierung auch zu nachteiligen Effekten führen, z. B. zur Einschränkung von Kreativität.

³ Informationen zur Software Acclaro[®] sind unter: <http://www.axiomaticdesign.com> abrufbar.

5 Zusammenfassung und Ausblick

Wir haben an einer Fallstudie demonstriert, dass AD dazu beiträgt, die Architekturziele „ausgewogene Granularität“, „lose Kopplung“ und „hohe Autonomie“ für SOA zu fördern. AD hilft auch, den Entwurfsprozess für SOA aus einer fachlichen Perspektive zu strukturieren.

Wir konnten zwar zeigen, dass AD einen positiven Beitrag zum Entwurf SOA leistet. Offen ist allerdings, wie groß dieser Beitrag in realen Entwicklungsprojekten ist. In weiteren Experimenten und Fallstudien bleibt daher zu prüfen, inwieweit die beschriebenen Vorteile in realen Projekten zur Entwicklung umfangreicher SOA erreicht werden können. Außerdem muss untersucht werden, ob die Vorteile der Anwendung von AD den Aufwand rechtfertigen, der mit der Anwendung der Methode verbunden ist.

In Rahmen dieses Beitrages konnten wir aus Platzgründen nur auf einige Aspekte des AD eingehen. Es wäre z. B. interessant, auch den Beitrag des Informationsaxioms zum Entwurf von SOA zu analysieren, da es die quantitative Bewertung der Komplexität eines Entwurfes ermöglicht und somit zur Bewertung und Auswahl alternativer SOA-Spezifikationen herangezogen werden kann. Außerdem bietet AD weitere Hilfsmittel zur Unterstützung der Implementierungsphase an [Suh1998; Suh2001, 196-198]. Wir planen, auch den Beitrag dieser Hilfsmittel für die Entwicklung von SOA zu untersuchen.

Wir haben uns darauf beschränkt, die Auswirkungen von AD auf ausgewählte Architekturziele zu untersuchen. Wir vermuten, dass AD auch einen positiven Einfluss auf weitere Ziele – wie z. B. Geschäftsorientiertheit – hat. Allerdings gilt dies nicht für alle Ziele. Bestimmte Ziele, wie die Zustandslosigkeit oder Abstraktheit repräsentieren grundlegende Prinzipien für SOA, die zwar beim Entwurf berücksichtigt werden müssen, aber nicht direkt durch eine spezifische Entwurfsmethode beeinflusst werden können.

Fraglich ist auch, inwieweit die am Fallbeispiel demonstrierten Erkenntnisse allgemeine Gültigkeit besitzen. Die Konzepte von Axiomatic Design sind in jedem Anwendungsgebiet – beim Entwurf von Produkten, Software, SOA, etc. – dieselben. Daraus schlussfolgern wir, dass auch die Vorteile des Axiomatic Design in jedem Anwendungsgebiet erzielt werden können. Wir vermuten, dass die Erreichung der Architekturziele für SOA, unabhängig von den Besonderheiten eines spezifischen Entwicklungsprojektes, durch den Einsatz von Axiomatic Design gefördert werden kann. Zur Überprüfung dieser Vermutung, haben wir die Durchführung und Evaluierung weiterer Fallstudien und Praxisprojekte geplant.

Literaturverzeichnis

- [Balz1998] *Balzert, H.:* Lehrbuch der Software-Technik: Software-Management, Softwarequalitätssicherung, Unternehmensmodellierung. Spektrum, Heidelberg et al. 1998.
- [Boeh1981] *Boehm, B. W.:* Software Engineering Economics. Prentice-Hall, Englewood Cliffs et al. 1981.
- [BuGa2005] *Buchmann, I.; Gamber, M.:* Methoden zur Unterstützung der Entwicklung einer SOA. In: *Cremers, A. B. et al. (Hrsg.): Informatik 2005: Informatik live!;* Beiträge der 35. Jahrestagung der Gesellschaft für Informatik e.V. GI, Bonn 2005, S. 601-605.
- [CaGl1990] *Card, D. N.; Glass, R. L.:* Measuring Software Design Quality. Prentice Hall, Englewood Cliffs 1990.
- [CeHa2005] *Cervantes, H.; Hall, R. S.:* Technical Concepts of Service Orientation. In: *Stojanovic, Z. et al. (Hrsg.): Service-Oriented Software System Engineering: Challenges and Practices.* IGP, Hershey et al. 2005, S. 1-26.
- [ClHi2000] *Clapsis, P. J.; Hintersteiner, J. D.:* Enhancing Object-Oriented Software Development through Axiomatic Design. In: *First International Conference on Axiomatic Design (ICAD2000).* Cambridge 2000, S. 272-277.
- [DJMZ2005] *Dostal, W.; Jeckel, M.; Melzer, I.; Zengler, B.:* Service-Orientierte Architekturen mit Web Services: Konzepte - Standards - Praxis. Spektrum, München 2005.
- [DoSu2000] *Do, S.-H.; Suh, N. P.:* Object-Oriented Software Design with Axiomatic Design. In: *Proceedings of ICAD2000.* Cambridge 2000, S. 278-284.
- [DoSu1999] *Do, S.-H.; Suh, N. P.:* Systematic OO Programming with Axiomatic Design. In: *IEEE Computer* 32 (1999) 10, S. 121-124.
- [EAAC2004] *Endrei, M.; Ang, J.; Arsanjani, A.; Chua, S. et al.:* Patterns: Service-Oriented Architecture and Web Services. <http://www.redbooks.ibm.com>, 2004, Abruf am 2006-06-27.
- [EKAP2005] *Erradi, A.; Kulkarni, N.; Anand, S.; Padmanabhuni, S.:* Designing Reusable Services: An Experimental Perspective for the Securities Trading Domain. In: *Chung, J.-Y. et al. (Hrsg.): Proceedings of the First International Workshop on Design of Service-Oriented Applications (WDSOA'05).* IBM Research Division, Amsterdam 2005, S. 25-32.
- [EMPR2005] *Eidson, B.; Maron, J.; Pavlik, G.; Raheja, R.:* SOA and the Future of Application Development. In: *Chung, J.-Y. et al. (Hrsg.): Proceedings of the First International Workshop on Design of Service-Oriented Applications (WDSOA'05).* IBM Research Division, Amsterdam 2005, S. 1-8.
- [EnNo2000] *Engelhardt, F.; Nordlund, M.:* Strategic Planning based on Axiomatic Design. In: *Proceedings of ICAD2000.* Cambridge 2000, S. 26-34.
- [Erl2005] *Erl, T.:* Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall PTR, Upper Saddle River et al. 2005.
- [HaSc2006] *Hagen, C.; Schwinn, A.:* Measured Integration - Metriken für die Integrationsarchitektur. In: *Schelp, J. et al. (Hrsg.): Integrationsmanagement. Planung, Bewertung und Steuerung von Applikationslandschaften.* Springer, Berlin et al. 2006, S. 267-292.
- [HiNa1999] *Hintersteiner, J. D.; Nain, A. S.:* Integrating Software into Systems: An Axiomatic Design Approach. In: *Proceedings of the 3rd International Conference on Engineering Design and Automation.* Vancouver 1999, S. 1-7.

- [Jams2004] *Jamshidnezhad, B.*: Towards a Rational Basis for User Interface Design Methods. In: Proceedings of ICAD2004. Seoul 2004, S. 1-4.
- [Kaye2003] *Kaye, D.*: Loosely Coupled: The Missing Pieces of Web Services. RDS Press, Marin County 2003.
- [KBSt2005] *Koord.- und Beratungsstelle der Bundesreg. für IT in der Bundesver: V-Modell XT*, Version 1.2.0. <http://www.v-modell-xt.de>, 2005, Abruf am 2006-04-25.
- [KiSK1991] *Kim, S. J.; Suh, N. P.; Kim, S. G.*: Design of Software Systems based on Axiomatic Design. In: Robotics & Computer-Integrated Manufacturing 8 (1991) 4, S. 243-255.
- [KoHB2005] *Kotonya, G.; Hutchinson, J.; Bloin, B.*: A Method for Formulating and Architecting Component- and Service-Oriented Systems. In: *Stojanovic, Z. et al. (Hrsg.): Service-Oriented Software System Engineering: Challenges and Practices*. IGP, Hershey et al. 2005, S. 155-181.
- [LaMB2005] *Laures, G.; Meyer, H.; Breest, M.*: An Engineering Method for Semantic Service Applications. In: *Chung, J.-Y. et al. (Hrsg.): Proceedings of the First International Workshop on Design of Service-Oriented Applications (WDSOA'05)*. IBM Research Division, Amsterdam 2005, S. 79-86.
- [LaPi2005] *Latchem, S.; Piper, D.*: Service-Oriented Design Process Using UML. In: *Stojanovic, Z. et al. (Hrsg.): Service-Oriented Software System Engineering: Challenges and Practices*. IGP, Hershey et al. 2005, S. 88-108.
- [MaBe2006] *Marks, E. A.; Bell, M.*: Service-Oriented Architecture: A Planning and Implementation Guide for Business and Technology. Wiley, Hoboken et al. 2006.
- [Raas1993] *Raasch, J.*: Systementwicklung mit strukturierten Methoden: ein Leitfaden für Praxis und Studium. 3. Aufl., Hanser, München et al. 1993.
- [ScTs2000] *Schreyer, M.; Tseng, M. M.*: Hierarchical State Decomposition for Design of PLC Software by applying Axiomatic Design. In: Proceedings of ICAD2000. Cambridge 2000, S. 264-271.
- [SiHu2005] *Singh, M. P.; Huhns, M. N.*: Service-Oriented Computing: Semantics, Processes, Agents. Wiley, Chichester et al. 2005.
- [SuDo2000] *Suh, N. P.; Do, S.-H.*: Axiomatic Design of Software Systems. In: Annals of the CIRP 49 (2000) 1, S. 95.
- [Suh1990] *Suh, N. P.*: The Principles of Design. Oxford, New York et al. 1990.
- [Suh1998] *Suh, N. P.*: Axiomatic Design Theory for Systems. In: Research in Engineering Design 10 (1998) 4, S. 189-209.
- [Suh2001] *Suh, N. P.*: Axiomatic Design: Advances and Applications. Oxford, New York 2001.
- [VACI2005] *Vogel, O.; Arnold, I.; Chughtai, A.; Ihler, E. et al.*: Software-Architektur: Grundlagen - Konzepte - Praxis. Spektrum, München 2005.
- [ZiTP2003] *Zimmermann, O.; Tomlinson, M.; Peuser, S.*: Perspectives on Web Services: Applying SOAP, WSDL and UDDI to Real-World Projects. Springer, Berlin et al. 2003.

Services and resource profiles as metrics for the allocation of IT infrastructure costs

Reinhard Brandl

Chair of Internet-based Information Systems (IBIS)
Institute of Informatics
Technische Universität München
85748 Garching, Germany
brandlr@in.tum.de

Abstract

This paper addresses the general problem of allocating costs for shared and distributed IT infrastructures to business entities, like cost centres or processes. In contrast to earlier times, when monolithic mainframes were the central resources in data centres, client/server architectures with distributed heterogeneous components now dominate. In such an environment, resource consumption (e.g. CPU time) usually cannot be fully apportioned to users or other business entities. We therefore propose an approach which is based on predetermined resource profiles as estimates for the mean resource consumption of business services. Instead of metering every component during operation and consolidating the log data afterwards, only service invocations need to be logged. For the determination of the resource profiles we developed a measurement methodology and a software toolkit. The approach was successfully tested and evaluated for OLTP systems at the BMW Group. Furthermore, by combining resource profiles with Queuing Network Theory, we could demonstrate their appropriateness for capacity planning.

1 Introduction

Throughout the literature on IT management and controlling, three general topics are commonly found: firstly, the minimisation of operational and security risks; secondly, the increasing of IT Business Value (effectiveness); and thirdly, the need for a continuous reduction of costs

(efficiency). In the context of the latter two aspects, one of the major challenges, particularly for internal IT managers, is to establish a transparent relationship between IT spending and business entities like cost centres, cost units, processes or activities. Otherwise, an IT unit runs the risk of appearing as black box with high fixed costs and no measurable business value. The lack of cost transparency of internal IT service provision is one of the major motivations for outsourcing activities [DGHJ04].

Allocating costs to business entities is a particularly intricate task when the corresponding resource is shared and an apportionment of its consumption is not practicable. In the IT context, this is usually the case for corporate data centres and infrastructure services. Whereas for desktop- and office-oriented services a known customer exists, the consumption of infrastructure resources usually cannot be fully apportioned to users or other business entities. As a single user activity might involve a multitude of separated components (e.g. servers, network, storage, etc.) which mostly have no access to the original business context of the operation, a consolidation of the distributed log information is elaborate and error-prone. As long as major resources are dedicated, e.g. one server per application, direct costing approaches can be applied. However, with the advances in virtualisation technologies and the move towards shared resource pools, new accounting concepts and metrics are required.

We address this problem in the context of distributed OLTP systems, which are the beating heart of modern enterprise computing environments. We assume first that an OLTP system provides one or more business services to its users and secondly that the invocation of such a service always results in similar resource consumption (e.g. transferred bytes, CPU seconds). The idea is to determine for each service a complete resource profile in a test environment. During regular operations, only service invocations per user or cost centre are logged. If the approach is practicable, the information would be sufficient for a usage-based cost apportionment of IT infrastructures. A distributed metering of resource consumption could be omitted. Furthermore, the resource profiles might be valuable inputs for capacity planning.

We tested the approach, under realistic conditions, in the central IT unit of the BMW Group. In a first step, we developed a methodology and a software tool for the determination of resource profiles in heterogeneous environments. In a second step of the project, we used these profiles for the parameterisation of Queuing Network Models (QNM). The results of these models were compared with the outcomes of real load tests. We had two main reasons for undertaking this

procedure. Firstly, the resource profiles should be validated and, secondly, their appropriateness for capacity planning should be demonstrated. In this paper we present our results.

The remainder of the article is structured as followed. Chapter 2 introduces the problem of IT infrastructure cost allocation and briefly discusses current practiced approaches. In Chapter 3 we present our idea of allocating infrastructure costs via resource profiles. The concept is further illustrated in Chapter 4 by the example of the central IT unit of the BMW Group. The overall results are evaluated in Chapter 5. The paper concludes with an outlook on future areas of research in Chapter 6.

2 Problem statement

Cost accounting is a managerial accounting activity which increases transparency and supports decisions by allocating overhead and direct costs to business entities. Traditionally, business entities are cost centres (e.g. departments, subsidiaries) or cost units (e.g. product or service produced), but more recent forms of cost accounting also consider processes [HoMa89] or activities [KaBr87]. The goal is to measure the economic performance of business entities and the value of the resources consumed in producing goods and services [Harp93]. Cost transparency is a quite general need. However, the purposes the information is used for (e.g. internal chargeback, benchmarks or management evaluation), and to what extent, is strongly organisation-specific.

IT is mostly organised as a service function which supports more than one business entity. The budget is dominated by salaries, consulting fees and infrastructure costs. As a direct breakdown of those costs to business entities is for the most part not feasible, one can either treat them as general overhead or try to identify appropriate cost drivers as accounting objects (IT products or services). They should enable a usage-based cost allocation and serve as bridges between IT and business. On the IT side they integrate different resources and cost types; on the business side they allow usage and cost control. In the context of desktop- and office-oriented IT services, accounting objects could be the provision of a desktop PC, an e-Mail Box or a telephone line. For non-standardised IT services, the identification of accounting objects is more sophisticated. Consider the management and operation of business applications, hosted in data centres. In the example in Figure 1, there is one OLTP system used by five different business functions.

From their point of view, the provision and/or usage of the system and the related support services are reasonable accounting objects. As the percentage rates indicate, a good portion of the total costs can be allocated either to the application or directly to the business functions.

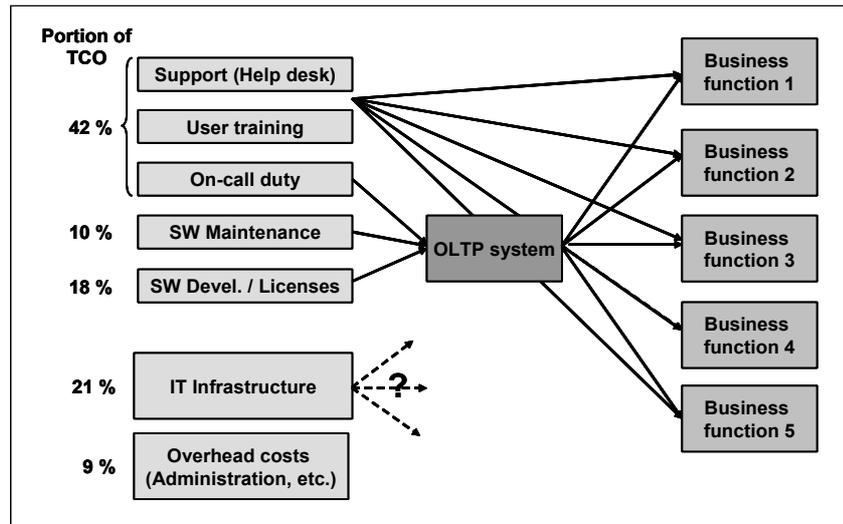


Figure 1: Costs of an OLTP system. (Rates adapted from [Zieh04])

We assume that ratios (e.g. number of logins, number of employees, transactions) for an apportionment of the application-specific costs (development, maintenance, etc.) exist and focus in the following on the allocation of the remaining infrastructure costs.

In the literature (e.g. [Aure97; Bert01; Hein02]), the ITIL library [OoOC01], and in our cooperation with the BMW Group (Chapter 4), we identified several ways how organisations handle these costs.

Direct cost allocation. In case a resource (e.g. a server) is dedicated to a specific application or a business unit, the incurring costs can be directly allocated. This proceeding is transparent and easy to implement. However, this approach is mainly limited to computing infrastructure. Other resources, like network or EAI components, are usually shared. With regard to the average server utilisation of less than 25% in modern data centres and the cost saving potential of virtualisation technologies, dedicated resources will more and more give way to shared resource pools.

Overhead costs. Infrastructure costs which cannot be directly allocated are handled in analogy to other IT overhead costs (administration, buildings, etc.). Particularly for small to medium sized companies, with IT organised as a service cost centre, this approach might be sufficient.

Measured usage. At first glance, measuring resource consumption and allocating it to the originator is the most obvious method for the apportionment of shared infrastructures costs. A measured usage approach is reasonable for certain resources like disk space or telephony as well as for very computing-intensive jobs, like batch processing or simulations. However, in an OLTP context, measuring and apportioning resource consumption to a specific user or even to

an application is not trivial. In contrast to earlier times, when monolithic mainframes were the dominating resources, modern OLTP systems are based on the client/server model and integrate different and distributed components. A single business transaction may cause resource consumptions on multiple components. In a simple three-tier architecture, with a web server, application server and database server, three different computing resources are already involved. Due to performance and security reasons (and the lack of standards), the original business context of a transaction is mostly not available in the backend. For instance, user names are not further transmitted after a successful authorization and for database access connection pools are used. This lack of information complicates the establishment of an end-to-end perspective. So, particularly in heterogeneous environments, the collection and consolidation of log data for accounting purposes quickly becomes tedious.

Per provided application. The provision of an application in a data centre is an “IT product” to which cost are allocated, either by a single or a tiered flat rate (e.g. classification according to the expected number of concurrent users, required service levels or number of interfaces). The BMW Group uses such flat rates for applications on its J2EE infrastructure. This approach is particularly easy to implement, as no explicit differentiation of resource costs and no metering is required. The accompanying lack of transparency may be accepted, if the applications are roughly of a similar nature (complexity, resource consumption, etc.). However, the notion of an “application” is a little bit fuzzy. Depending on the software architecture, the same business functionality may be implemented in a single or in multiple separate applications (i.e. executables, ear-files, etc.). Furthermore, one user of an application can cause a lot more resource consumption than 100 users of another. If cost accounting is combined with chargeback, this simplification potentially leads to acceptance problems.

In our opinion, none of the approaches presented above really cope with the complexity of modern OLTP systems. It’s not surprising that the lack of cost transparency is one of the major reasons for outsourcing IT activities. In a market environment, accounting metrics can be found which are not directly linked to costs (e.g. “pay-per-business transaction”), but which rely on a risk sharing model between customer and service provider. This establishes cost transparency at the customer side, but does not solve the fundamental problem.

3 Resource profiles for accounting and capacity planning

In search of alternatives, we abstracted from the technical view on “applications” and took a business perspective, with an IT system offering a number of useful services (e.g. „process order“, „update stock“, and „add customer“). An application may implement one or more of those services, but a service can also comprise multiple applications. The invocation of a service causes resource consumption in the infrastructure (CPU time, transferred bytes, etc.). The exact amount is dependent on input factors like parameters passed or the internal states of the components involved.

Nevertheless, for repeated service invocations an average value or, more precisely, a vector with average consumption values for different resources, exists. If these vectors and the number of service invocations per user were known, a consumption-based cost apportionment to services and users would be possible.

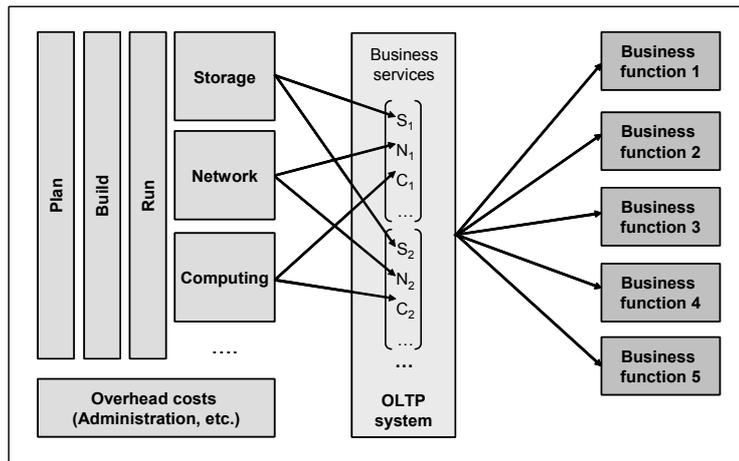


Figure 2: Cost allocation by business services and resource profiles.

The elaborate process of measuring and consolidating log data from different components could be bypassed. Furthermore, these vectors would be valuable inputs for the alignment of business forecasting and IT capacity planning.

Whether such an approach is actually practicable in an industrial context depends on the efforts required for the determination of appropriate services and average consumption vectors. As the problem of allocating infrastructure costs is of a quite general nature, we further followed this idea and analysed its potential in cooperation with our industrial partner, the BMW Group.

4 Results and experiences from the pilot test

4.1 Organisational context

At the BMW Group, management and operation of the considered OLTP systems is organised on two levels. A central IT department provides standardised infrastructure services, such as server computing, network and storage for the whole company. At each Business Unit (e.g.

Production, Sales) a dedicated IT department is responsible for the management and operation of their applications. The Business Units are charged for their consumption of infrastructure resources. If resources are dedicated, direct costing is applied. Dedicated resources are required in particular for business-critical applications with peak demands. However, most applications run on shared infrastructure and we assume that with increasing hardware performance and advances in virtualisation technologies, their share will further grow.

The predominant application platform at the BMW Group is Java/J2EE. The central IT department therefore provides standardised environments on shared and dedicated computers. The Business Units are either charged per application (shared infrastructure) or by the performance characteristics of their dedicated computers. These flat rates cover all other resource costs. This approach is quite efficient and comprehensible. However, it does not consider the different nature of applications and their resource consumption (see Chapter 2). In a pilot test, we evaluated if the approach, presented in the previous chapter, would be a feasible alternative.

4.2 Integration of the approach into existing IT processes

The IT processes of the BMW Group specify that prior to the first deployment or after a major software change, every application has to pass an operational approval. This process consists of several well-defined tests (e.g. installation, load and stress tests, cluster failover) in an isolated environment. In contrast to tests during the development phase, the operational approval is conducted by infrastructure specialists and not by the responsible software engineers. The objective is to verify whether the application meets stability and performance requirements (at the BMW Group also referred to as “operational maturity”). Such approval or acceptance tests are common practice in industrial data centres [OoOC02] and we consider them as a suitable opportunity for the determination of the required resource profiles. So, we designed an additional process step for the operational approval at the BMW Group, to acquire consumption data of typical user behaviour.

4.3 A methodology for the determination of resource profiles

As relevant run-time consumption metrics, we identified CPU time, network and storage I/O. We excluded disk space, as it is usually a priori allocated to a specific application or a database

and elaborate accounting tools are available. Memory is usually also considered as a scarce resource. However, the maximum amount of physical memory a server can allocate on a machine is already determined at startup (e.g. by setting a range for virtual memory) and we propose to take this value as the basis for accounting.

The determination of resource profiles in heterogeneous environments is not trivial. Existing profiling tools focus on diagnosing performance problems, like memory leaks or CPU consumption of methods or transactions. They are technology- or vendor-specific, (e.g. Java/J2EE [JaPeoJ], .NET [Schw06], Intel [IntelJ], different ERP/CRM systems [SymaoJ]) and do not cover all the resources mentioned above. Load generators (e.g. Mercury Loadrunner [MercoJ], SilkPerformer [SeguoJ], The Grinder [TheGoJ], httpperf [MoJi98]), on the other hand, have elaborate means to simulate different user behaviours in various environments. However, their main analysis focus lies on response times experienced by their virtual clients. Although most tools provide consoles for a remote monitoring of hardware and server performance, none of the products we evaluated calculates resource consumption for specific user activities.

To overcome the limitations of existing tools, we developed a new methodology and an appropriate software toolkit. The underlying motivation was to enable a flexible and non-invasive determination of resource profiles for business-oriented services, independent from underlying hardware and software technologies. The basic idea is to install the system under consideration in an isolated environment (as is usual for operational approvals), simulate consecutive service invocations by means of a load generator and correlate start and end times of services with performance log files recorded at the different components. The prerequisite of the approach is that no other users or applications distort the measurements. So, besides inevitable background activities, resource consumption can be fully allocated to service invocations. The corresponding software toolkit consists of three separate packages. Firstly, a collection of log file parsers to transfer the measurements into a database, secondly, an analysis component which correlates the data and computes the resource profiles, and finally, a tool for the visualisation of the results.

4.4 Experimental setup

We tested the approach with several applications. For the experiments we set up an infrastructure which is similar to the standards of the BMW Group (see Figure 3). To increase

the empirical coverage, we combined different Operating Systems (Linux, Windows and Unix) and servers (Apache HTTP, Bea Weblogic and Oracle Database).

As example scenario for this publication we use the J2EE reference implementation Java PetStore [SunoJ]. We chose this example for three major reasons. Firstly, it is publicly available and well documented, secondly, as reference implementation, it covers most J2EE technologies (EJBs, Servlets, JSPs, Web Services and XML, JMS, CMP, etc.), and thirdly, the software architecture, with several interacting applications in the front- and back-end, is an appropriate representation of the structure of modern enterprise systems. For the simulation of service invocations, we chose Mercury Loadrunner [MercoJ], as it had been already used during the operational approval at the BMW Group. To monitor the resource consumption at the different computers, we relied on

standard OS tools (Unix/Linux: sar, Winodws: perfmon). Thus, we were independent from server technologies and could avoid the installation of additional software.

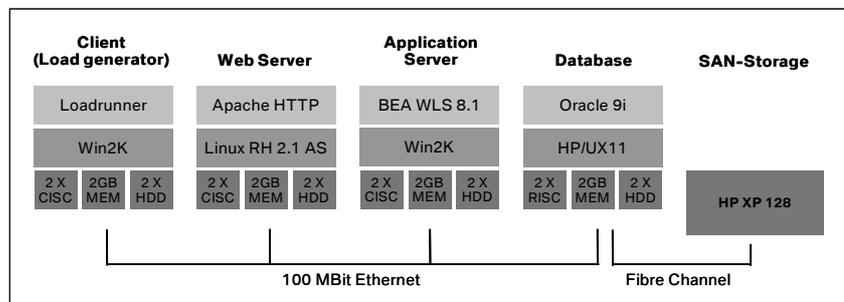


Figure 3: Infrastructure used in the experiments

4.5 Definition of services

We assume that, from a business perspective, an IT system provides a number of useful services and that we can determine resource profiles as estimates for the mean resource consumption of service invocations. To minimise the gap between estimated and actual resource consumption one could define every possible user activity within the system as a separate service. Nevertheless, in our opinion a large number of fine-granular services would lead to an unreasonable complexity for cost accounting. We propose instead to consider whole or at least parts of business processes as services. For instance, in the PetStore case, we defined the shopping process of an average user as single service. The logging of service invocations could be easily combined with central authentication/authorisation mechanisms. However, as visitors have different possibilities of using the store, i.e. the service, information about the average user behaviour must be available or estimated. This might be a problem, in particular for new systems. In that case, we recommend reviewing the assumptions on services and usage after an observation period. For the operational approval at the BMW Group, software projects were

already obligated to specify the expected usage and provide appropriate load test scripts for the different use cases. We proposed to take this information as the basis for the definition of services.

4.6 Example: Determination of the resource profile for a PetStore shopper

For the determination of user paths through web applications, several tools (e.g. WebTrends Analytics from WebTrend Inc. [WebToJ]) and scientific approaches exist (e.g. Customer Behaviour Model from Menascé and Almeida [MeAl00]) exist. For the PetStore example, we assume that such information is available and we recorded the path of an average shopper in a load test script. As mentioned above, we use standard OS tools for the performance monitoring at the different computers. The major drawback of those tools is that they only provide rough measurements, which are furthermore distorted by background activities. For instance, the minimal interval between two data points is one second, which is much higher than the usual response time of a three-tier web application, like the Java PetStore. Accordingly, the resource consumption of a single PetStore shopper would be hardly measurable. To solve this problem, and to separate the resource consumption of services from background activities, we

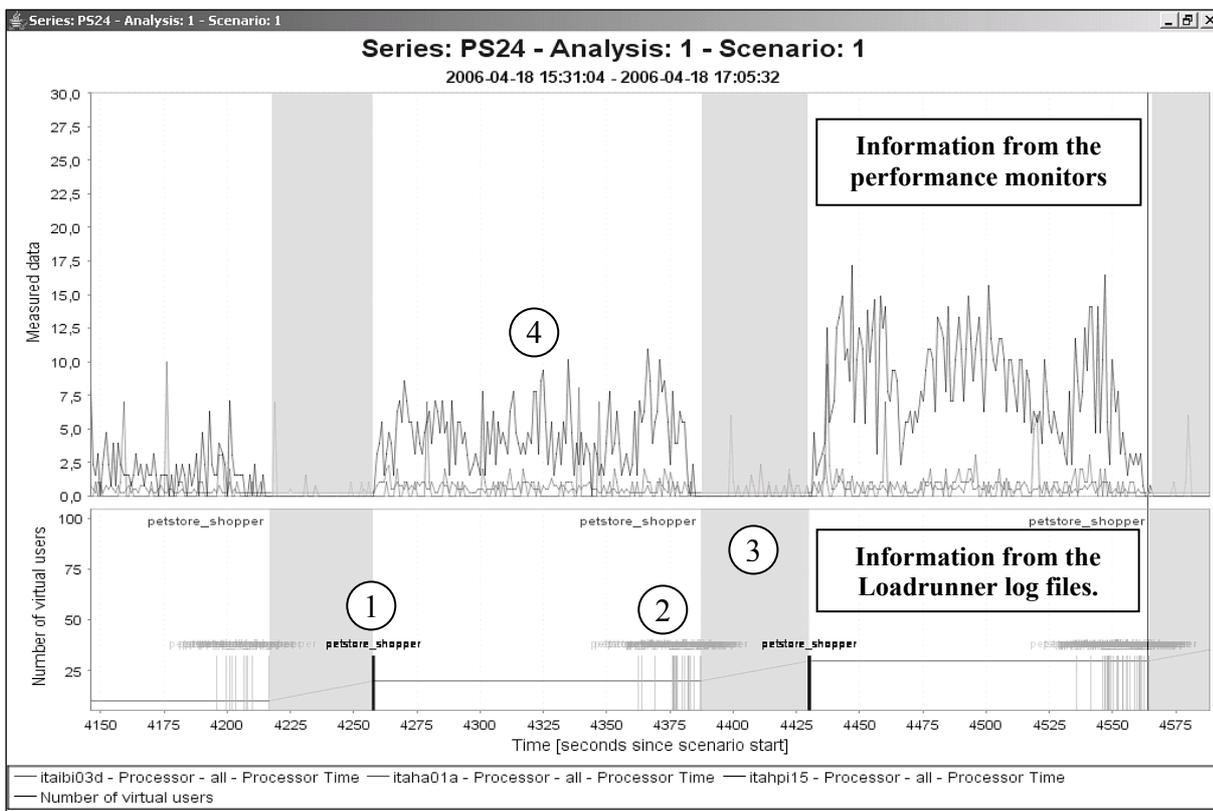


Figure 4: Determination of the resource profile for a PetStore shopper.

successively raise the number of parallel service invocations and determine by means of a linear regression the increase of resource consumption. In the PetStore example, we started with 10 concurrent users and increased the number in steps of 10 until we reached 70 users and repeated this process several times. Figure 4 illustrates this concept with a screenshot from our analysis tool. At label (1), 20 concurrent users start browsing through the PetStore. To avoid peaks, we build in arbitrary think times between the requests. About two minutes later all users have finished their shopping tour (2). After an idle phase of 45 seconds (3), the procedure is restarted with 30 concurrent users. During the whole test, the performance monitors at the Web, Application and Database Servers record the system behaviour in log files on their local disks. The upper diagram (4) shows for instance the overall CPU utilisation ((user + system time) / 2 processors) at the three different computers. The most utilised component is the Application Server. Data from network and storage resources is analysed in the same manner.

After the test, log file parsers consolidate the data and copy it into a database. Then the analysing component correlates start and end times of transactions with selected performance data and computes the average resource consumption for every measurement cycle. In the following, we apply a linear regression to these values and compute the resource consumption per service invocation (i.e. slope of the regression line). We use the correlation coefficient ($r \in [-1;1]$) as a first quality indicator for the results. Figure 5 shows the results of the analysis of CPU times. Accordingly, a single PetStore shopper consumes during a 2 minute visit 0.207 seconds of CPU time¹ at the Application Server ($r=0.825$), 0.012 seconds at the Web Server ($r=0.968$) and 0.025 seconds at the Database ($r=0.875$). For the other resources in the profile – network (transferred bytes) and storage (transferred blocks) – we got similar results. Nevertheless, the coefficient of correlation only evaluates the fitting of the regression line and not the resource consumption as such. So, we were looking out for a further sanity check for the resource profiles.

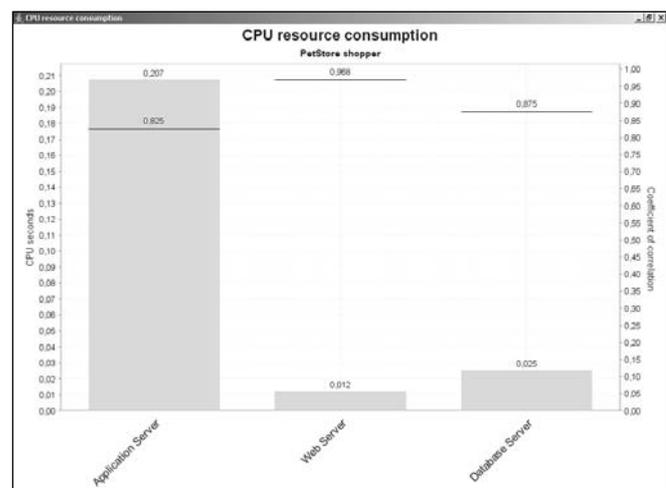


Figure 5: CPU resource consumption of a PetStore shopper

¹ These are raw values, dependent on the architecture and performance characteristics of the hardware used during the tests. For accounting purposes, they must be normalised. We therefore propose the usage of standard industry benchmarks.

4.7 Resource profiles as input parameters for analytical capacity planning

Queuing Network Theory is a well-studied analytical approach for performance modelling and capacity planning of distributed computer systems. However, is rarely used in practice. One reason for this is that the input parameters, like service time for jobs, are not directly available from common performance monitors. The resource profiles presented above might fill this gap. We further evaluated this idea, for two major reasons. By comparing performance predictions with measured values, we could, firstly, validate the resource profiles (see previous section) and, secondly, test their appropriateness for capacity planning. In the following section we briefly present the results.

As the first step, we developed an appropriate load test set-up. We use the same scripts as for the determination of the resource profiles. However, instead of simultaneous starts and stops, we put the users in an endless loop. Every five minutes, we added ten more users until the first component reaches its bottleneck. During the load test, performance monitors record the system behaviour in log files. The data is afterwards copied into a database, but instead of analysing resource consumption, we determine average values for utilisation and response times.

In a second step, we developed a general performance model for the three-tier infrastructure used during the experiments (see Figure 3). We separately analysed each of the above-mentioned five minute slots. As the number of users remains constant within each slot, we applied a Closed Queuing Network Model (see Figure 6). As the solution algorithm, we implemented the Mean Value Analysis algorithm [ReLa80]. For further information on Queuing Network Theory, we refer the interested reader to the work of Menascé *et. al.* [MeAl00; MeAD04] and Bolch *et al.* [BoGM06].

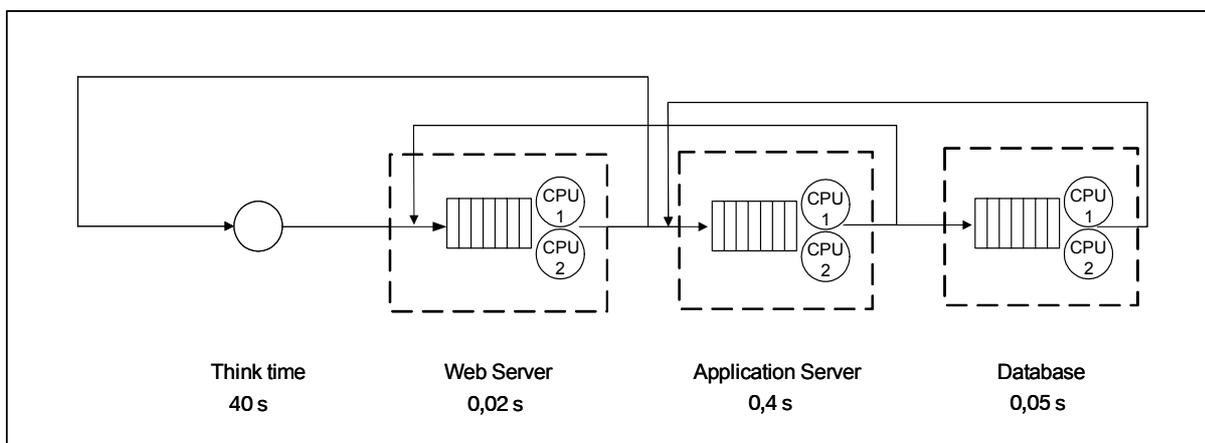


Figure 6: QN-Model of the infrastructure used in the experiments.

In the PetStore case we solely modelled the processors, as the time required for disk accesses was negligible. Due to the dual-processor architecture of the considered hardware (see Figure 3), we modelled each server as a queue with two separate stations. The service demands were taken from the resource profiles. As the values in Figure 5 are based on the average CPU utilisation of both processors, we had to double them for the performance model.

We conducted a load test according to the above-mentioned process and compared the outcome to the predictions of the Queuing Network model. As the length of the think time between two requests must not have an effect on the resource consumption, we intentionally changed it for this load test (in total 40 seconds instead of 120 seconds). The bottleneck in the PetStore example was the Application Server. Figure 7 compares its measured utilisation (average value of the time interval with constant number of users) to the computed values from the Queuing Network Model. The maximum absolute deviation is about 10%, which we consider to be tolerable. Nevertheless, we are still looking for improvements.

Analytical capacity planning has a number of advantages over common approaches. Once a validated model exists, the effect of parameter changes, concerning the hardware performance, the expected load and the user behaviour can be evaluated. These “what-if”-analyses are particularly helpful in situations where the hardware from the test environment differs from the production environment. The results are usually more precise and substantiated than “rule of thumb” estimates or general sizing guidelines from hard- and software vendors.

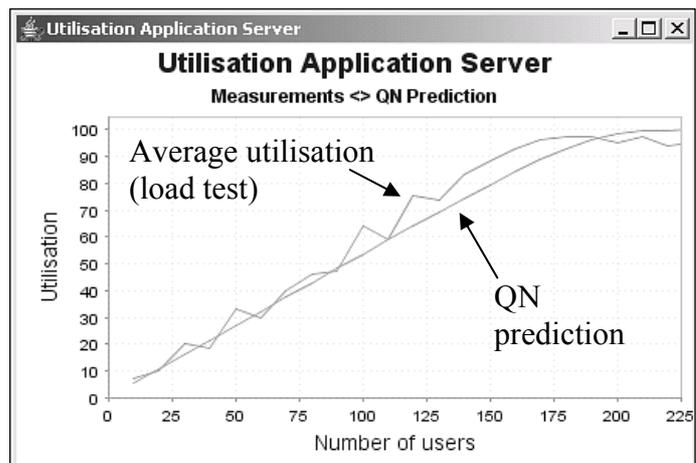


Figure 7: Comparison of average CPU utilisation during a load test with QN predictions.

4.8 Contribution and related approaches

The contribution of our concept for the determination of resource profiles is threefold. Firstly, by utilising the capabilities of a professional load generator, we can determine resource profiles not only for single transactions, but for whole user sessions or business processes. Until now we tested it only for applications with web-based interfaces. However, this should not be a

restriction. Tools like the Mercury Loadrunner can simulate load in an analogous manner on a multitude of different front- and backend interfaces. Secondly, as we fully rely on standard performance monitors of Linux/Unix and Windows Operating Systems, new components can be easily integrated as the case arises. Despite the limitations of these tools (e.g. minimal measurement intervals of one second), we could demonstrate that the concept is also appropriate for very small values of resource consumption. Thirdly, through the combination of Queuing Network Theory and load tests we provide means for a quick sanity check of the resource profiles.

As illustrated in the Section 4.3, the determination of overall resource profiles for services on distributed systems is not yet supported by tools available on the software market. This problem had been also addressed by Nagaprabhanjan and Apte who recently presented a tool [NaAp05] for automated resource consumption profiling of distributed transactions. Analogous to our approach, they combine load generation with OS performance monitors for the determination of resource profiles. However, their focus is not accounting, but the determination of input parameters for performance analysis and capacity planning. The major strength of Nagaprabhanjan and Apte is the run-time coordination of load generation and resource profiling. This enables more precise measurements. Accordingly, they require fewer measurement cycles and generate less data for the analysis. Our approach is more flexible concerning the definition of services and the integration of additional resources. However, due to the above mentioned aspects it takes more time for measurement and analysis.

5 Evaluation of results

We could demonstrate during the project that it is technically feasible to determine comprehensive resource profile for business-oriented services in a distributed and heterogeneous environment. However, our methodology is subject to the following two major conditions. Firstly, we assume that knowledge about appropriate services and user behaviour is available. This assumption may not hold for new applications. If no historical usage data exists, measurements have to be based on estimations. This can lead to distorted resource profiles. Secondly, our metering concept requires an environment with no external interferences. In organisations with formal IT testing and approval processes, such an infrastructure might be

given. If not, considerable extra efforts and investments are necessary. In that case the efficiency of the approach has to be seriously questioned.

If the above requirements are fulfilled and if there is a need for usage-based cost allocation, the approach is a feasible alternative to “measured usage”. Beyond its operational advantages (a distributed metering of resource consumption can be omitted) and its business orientation (service invocation instead of CPU seconds), we identified the following strengths. Firstly, resource consumption, which cannot be apportioned during normal operation (e.g. network traffic, storage I/O), can also be included. Secondly, if the resource profiles of services are transparent, they can be used as a criterion for the evaluation of software. So, an incentive for economic resource usage for developers and architects may already be created. According to the literature on Performance Engineering (e.g. [Smit90; BMIS03]) and confirmed by experiences of the BMW Group, most potential for savings are in the design and implementation phases. Finally, resource profiles are valuable inputs for capacity planning and performance analysis. In this context analytical approaches, like Queuing Network Theory, have a number of advantages and are quite well studied. However, due to their complexity and the lack of appropriate input data, they are rarely used in practice. We demonstrated that resource profiles partially fill this gap. In combination with usage forecasts, they facilitate “what-if” analyses and the sizing of infrastructures. In the context of cost allocation “per provided application” and tiered flat rates (see Chapter 2), the predicted capacity requirements are a reasonable criterion for the classification of an application.

Legacy or COTS software components are usually not designed to log service invocations for accounting purposes. We therefore propose linking usage metering to central authentication or authorization services. So, typically the login to an application would be considered as a service invocation. However, this is a trade-off between practicability and flexibility. Furthermore, accounting systems which are designed for an enterprise-wide usage may by default not support hundreds or even of thousands of different services with individual prices and must be adapted accordingly.

For the case of the BMW Group we propose, for reasons of practicability, to retain the cost allocation per application. However, instead of one single flat rate, resource profiles should be used for the determination of rates, based on the expected resource consumption of an application. So, even if the actual consumption cannot be apportioned, the cost driver “infrastructure resource requirements” is considered in cost accounting.

6 Conclusion and outlook

During the project at the BMW Group, we could demonstrate the feasibility of our approach. Nevertheless, the root cause of the difficulties with allocating infrastructure costs is that accounting mechanisms are hardly considered in the architectural guidelines for modern IT systems. Hence, current practice is a per-resource-accounting based on technical metrics, such as CPU seconds. What is still missing, are comprehensive and user-oriented approaches. Predetermined resource profiles are an elegant way out, but do not constitute a general solution. Against the background of advancing virtualisation technologies and the ongoing commercialisation of shared resource pools, we predict a growing demand for uniform standards and protocols to facilitate the allocation of infrastructure resource consumption. Without well-defined methodologies for usage accounting the foundation for economic efficiency in a competitive world for service provisioning is not given. We strongly believe that in this domain further scientific work could provide valuable contributions for the industrial practice.

References

- [Aure97] Aurenz, H.: Controlling verteilter Informationssysteme. Client/Server-Architekturen. Bd. 4, Peter Lang Publishing Group, 1997.
- [BMIS03] Balsamo, S.; Marco, A. D.; Inverardi, P.; Simeoni, M.: Software Performance: state of the art and perspectives.
<http://www.di.univaq.it/adimarco/technicalreports/techReport-perf.pdf>, Abruf am 2006-07-13.
- [Bert01] Bertleff, C.: Einführung einer IT-Leistungsverrechnung zur Unterstützung des strategischen IT-Controllings. In: H. Heilmann (Hrsg.): Strategisches IT-Controlling. 2001.
- [BoGM06] Bolch, G.; Greiner, S.; Meer, H. d.: Queueing Networks and Markov Chains. Wiley-Interscience, 2006.

- [DGHJ04] Dibbern, J.; Goles, T.; Hirschheim, R.; Jayatilaka, B.: Information systems outsourcing: a survey and analysis of the literature. In: SIGMIS Database 35 (2004) 4, S. 6-102.
- [Harp93] Harper, W. M.: Cost Accounting. 3rd. Aufl., Pitman Publishing, 1993.
- [Hein02] Heinrich, L. J.: Informationsmanagement. Bd. 7, Oldenburg, 2002.
- [HoMa89] Horváth, P.; Mayer, R.: Prozesskostenrechnung. Der neue Weg zu mehr Kostentransparenz und wirkungsvolleren Unternehmensstrategien. In: Controlling 1 (1989) 4, S. 214-219.
- [Intel0J] Intel Corporation: Intel VTune Performance Analyzer.
<http://www.intel.com/cd/software/products/asm-na/eng/vtune/index.htm>, Abruf am 2006-07-13.
- [JaPeoJ] Java Performance Tuning: Tool reports.
<http://www.javaperformancetuning.com/tools/>, Abruf am 2006-07-13.
- [KaBr87] Kaplan, R. S.; Bruns, W. J.: Accounting and Management: Field Study Perspectives. Harvard Business School Press, 1987.
- [MeAl00] Menascé, D. A.; Almeida, V. A. F.: Scaling for E-Business: Technologies, Models, Performance, and Capacity Planning. Prentice Hall, 2000.
- [MeAD04] Menascé, D. A.; Almeida, V. A. F.; Dowdy, L. W.: Performance by Design: Computer Capacity Planning by Example. Prentice Hall, 2004.
- [MercoJ] Mercury Interactive Corporation: Mercury LoadRunner.
<http://www.mercury.com>, Abruf am 2006-07-13.
- [MoJi98] Mosberger, D.; Jin, T.: httperf - a tool for measuring web server performance. In: SIGMETRICS Performance Evaluation Review 26 (1998) 3, S. 31-37.
- [NaAp05] Nagaprabhanjan, B.; Apte, V.: A Tool for Automated Resource Consumption Profiling of Distributed Transactions. In: G. Chakraborty (Hrsg.): Second

International Conference on Distributed Computing and Internet Technology.
Bhubaneswar, India 2005, S. 154-165

- [OoOC01] Office of Government Commerce: Service Delivery. Stationery Office Books, 2001.
- [OoOC02] Office of Government Commerce: ICT Infrastructure Management. Stationery Office Books, 2002.
- [ReLa80] Reiser, M.; Lavenberg, S. S.: Mean-Value Analysis of Closed Multichain Queuing Networks. In: Journal of the ACM 27 (1980) 2, S. 313-322.
- [Schw06] Schwichtenberg, H.: Tools and Software Components for the.NET Framework. <http://www.dotnetframework.de/dotnet/tools.aspx>, Abruf am 2006-07-13.
- [SeguoJ] Segue Software Inc.: SilkPerformer. <http://www.segue.com/>, Abruf am 2006-07-13.
- [Smit90] Smith, C. U.: Performance Engineering of Software Systems. Addison-Wesley, 1990.
- [Sunoj] Sun Microsystems: Java Pet Store Sample Application. <http://java.sun.com/reference/blueprints/>, Abruf am 2006-07-13.
- [SymaoJ] Symantec Corporation: Application Performance Management. <http://www.symantec.com/Products/enterprise?c=prodc&refId=1021>, Abruf am 2006-07-13.
- [TheGoJ] The Grinder: A Java Load Testing Framework. <http://grinder.sourceforge.net/>, Abruf am 2006-07-13.
- [WebToJ] WebTrends Inc. <http://www.webtrends.com/>, Abruf am 2006-07-13.
- [Zieh04] Zieh, O.: Government on Demand. Neue Wege in der Projektfinanzierung. In: I. Corporation (Hrsg.): 7. Deutscher Verwaltungskongress "Effizienter Staat". Berlin 2004

Einführung in den Track

Informationsmanagement in der Energie- und Umweltinformatik

Prof. Dr. Wolf Fichtner

Universität Cottbus

Prof. Dr. Axel Tuma

Universität Augsburg

Dr. Wolfram Münch

EnBW AG

Die Energie- und Umweltinformatik hat sich seit ihren Anfängen, Mitte der achtziger Jahre, zu einer wichtigen Disziplin im Schnittfeld von Natur-, Ingenieur-, Wirtschaftswissenschaften und Informatik entwickelt. Sie umfasst eine Reihe interessanter Teilgebiete, wie die Modellierung von Stoff- und Energieflusssystemen, "Geo-Informationssysteme", "Organisationsmodelle und Informationssysteme zur Umsetzung nachhaltiger Wertschöpfungsnetzwerke" und Systeme zur effizienten Verkehrssteuerung. Bedingt durch die Situation an den internationalen Energiemärkten sowie der Liberalisierung der Energiewirtschaft kommt der Energieinformatik dabei eine stetig wachsende Bedeutung zu.

Der Track "Energie- und Umweltinformatik" richtet sich gleichermaßen an Wissenschaft und Praxis. Einerseits werden aktuelle Forschungsarbeiten präsentiert sowie entsprechende Konzepte, Methoden und Werkzeuge vorgestellt und diskutiert. Andererseits werden aus Sicht der Praxis Anforderungen an zu entwickelnde Konzepte und Methoden erörtert und bereits realisierte Anwendungen vorgestellt.

Programmkomitee:

Prof. Dr. Wolf Fichtner, Universität Cottbus

Dr. Wolfram Münch, EnBW

Prof. Dr. Claus Rautenstrauch, Universität Magdeburg

Prof. Dr. Guido Siestrup, Fachhochschule Furtwangen

Prof. Dr. Axel Tuma, Universität Augsburg

TOP-Energy - Ein Framework für Softwarelösungen in der Energietechnik

Gregor Wrobel

Gesellschaft zur Förderung angewandter Informatik (GFai)
12489 Berlin
wrobel@gfai.de

Stefan Herbergs

Lehrstuhl für Technische Thermodynamik
RWTH-Aachen
52062 Aachen
herbergs@ltt.rwth-aachen.de

Eckardt Augenstein

perpendo GmbH
52068 Aachen
e.augenstein@perpendo.de

Abstract

Die Entwicklung komplexer Softwaresysteme kann bekanntermaßen nur durch einen hohen Entwicklungsaufwand realisiert werden. Sowohl die fachlichen Anforderungen an die Systeme, als auch hohe Standards in der Bedienbarkeit und Datenverwaltung müssen erfüllt werden. Insbesondere die letztgenannten Anforderungen treiben den Aufwand, existierende Fachlösungen in marktfähige Systeme zu überführen, in die Höhe. In diesem Beitrag wird ein Framework vorgestellt, auf dessen Basis die Entwicklung verschiedener Softwaresysteme der Energie- und Umwelttechnik effizient realisiert werden kann. Die erforderlichen Entwicklungsschritte zur Ableitung neuer Applikationen werden ebenso wie bereits entwickelte Lösungen dargestellt.

1 Herausforderung

Viele Institute von Universitäten und Forschungseinrichtungen implementieren Forschungsergebnisse in umfangreichen Softwaresystemen. Die daraus entstandenen Systeme können häufig nur von deren Entwicklern bedient werden. Das liegt zum einen in der fachlichen Spezialisierung des Aufgabengebietes begründet, oft aber auch in der unzureichenden Softwarearchitektur insbesondere der GUI. Der Aufwand, der bei der Überführung von universitären Lösungen zu einem marktfähigen Produkt notwendig ist, wird dabei oftmals unterschätzt oder kann nicht erbracht werden, so dass viele Ergebnisse ausschließlich der Forschung zur Verfügung stehen.

Am Lehrstuhl für Technische Thermodynamik der RWTH sind in den letzten Jahren verschiedene Systeme entwickelt worden, die zur Analyse, Simulation und Optimierung von Energiesystemen eingesetzt werden können [AWKP04; NN05; NN05a; NN06; BoLu06]. Die Gruppe der Anwender umfasst Energieberater, Energiecontractoren bzw. mit Fragen des Energieeinsatzes betraute Mitarbeiter größerer Industrieunternehmen; oft also Ingenieure. Aufgrund der Komplexität der Aufgabenstellungen reichen die von den Anwendern häufig verwendeten Standardtabellenkalkulationsprogramme nicht aus. Gleichzeitig stellen komplexe Simulationsumgebungen und Analysesoftware keine dem Anwendungsgebiet adäquaten Benutzerführungen zur Verfügung. Insbesondere die Modellierung erfolgt in diesen Systemen sehr abstrakt und erfordert spezielle Systemkenntnisse. Hieraus ergibt sich folgende Herausforderung:

Herausforderung 1: Verschiedene Softwaresysteme zur Analyse, Bewertung und Optimierung von Energiesystemen müssen entwickelt und derart anwenderfreundlich gestaltet werden, dass Ingenieure diese Systeme bedienen können.

Da die zu entwickelnden Lösungen einer gemeinsamen Domäne – der Energiesystemtechnik – zugeordnet werden können, bietet sich die Entwicklung eines Frameworks als Basis für Softwareprodukte für diese Domäne an. Ein solches Framework beinhaltet dann die gemeinsamen Datenmodelle, Benutzerschnittstellen und Editoren etc. Es muss aber auch leicht konfigurierbar sein, um spezielle Applikationen entwickeln zu können, sowie über geeignete Schnittstellen verfügen. Diese Anforderungen an sich stellen schon eine hohe Herausforderung an das Framework dar, die nur durch ein tiefgehendes Domänenengineering ermittelbar sind. Das Framework ermöglicht dann eine effektive Entwicklung neuer Applikationen aus der

betrachteten Domäne. Erstrebenswert wäre jetzt, dieses Applikationsengineering wiederum einer Anwendergruppe zugänglich zu machen. Diese Anwender sind derart gekennzeichnet, dass sie das umfangreiche fachliche Wissen zur Applikationsentwicklung besitzen und darüber hinaus über geeignete Softwarekenntnisse verfügen. Da das fachliche Wissen dieser Anwender im Vordergrund steht, kann nicht davon ausgegangen werden, dass diese Anwendergruppe aus Informatikern besteht. Somit ergeben sich weitere Anforderungen an die Schnittstellen des Frameworks. Die Ableitung neuer Applikationen aus dem Framework muss ohne umfangreiche Programmierung möglich sein.

Herausforderung 2: Es muss ein Framework entwickelt werden, das die Basis für Softwaresysteme der Energietechnik bildet und Applikationsentwicklungen durch Anwender ermöglicht.

2 Strategie

Um die oben beschriebenen Herausforderungen zu erfüllen und das TOP-Energy Framework zu entwickeln, mussten zwei unterschiedliche Softwaresysteme zusammengefasst werden.¹ Am Lehrstuhl für Technische Thermodynamik der RWTH entstanden in den letzten Jahren in verschiedenen Forschungsvorhaben Softwaresysteme zur Analyse und Optimierung von Energieversorgungssystemen. Um die Entwicklung dieser Systeme effizienter zu gestalten, wurde eine Softwarekomponente² (ETL) implementiert, die umfangreiche Berechnungen von erweiterten ingenieurtechnischen Grunddatentypen enthält.

Daneben beschäftigt sich der Forschungsbereich Graphische Ingenieursysteme der GFaI seit mehreren Jahren mit der Entwicklung von Softwaresystemen für komplexe ingenieurtechnische Aufgabenstellungen. Aus diesen Entwicklungen heraus ist ein Framework (CASTool) entstanden, auf dessen Grundlage verschieden graphische Ingenieursysteme abgeleitet werden können [PFPL05, PFE05, BPPS01].

Mit diesen beiden Systemen war die Grundlage für TOP-Energy geschaffen. Die erste Herausforderung, die Entwicklung geeigneter fachlicher Applikationen, konnte durchgeführt

¹ Arbeiten zu diesem Framework wurden in einem Projekt durchgeführt, das aus Haushaltsmitteln des BMWi über die AiF im ZUTECH-Programm unter dem Forschungsvorhaben Nr. 92 ZBG / 1 (Laufzeit 01.01.2003 bis 31.12.2004) über die Mitgliedsvereinigung VEU gefördert wurde. Dem BMWi und der AiF wird für die Förderung gedankt.

² Der Begriff Softwarekomponente beschreibt Teile der Architektur, die bestimmte softwaretechnische oder fachliche Aufgaben erfüllen. Softwarekomponenten sind von Modellkomponenten zu unterscheiden, die die Repräsentation einer realen Anlage oder Randbedingung in einem Energiesystem darstellen. Modulkomponenten oder Module sind fachliche Softwarekomponenten mit bestimmten Eigenschaften bezüglich ihrer Darstellung in TOP-Energy.

werden. Da es sich beim CASTool um ein allgemeines Framework für graphische Ingenieurssysteme handelt, waren Spezialisierungen für TOP-Energy notwendig. Insbesondere die abstrakte Datenhaltung im CASTool erschwerte die Applikationsentwicklung und stand der zweiten Herausforderung im Wege. Aus diesem Grunde wurde ein Objektmodell entwickelt, das die benötigten Objekte der Energiesystemtechnik abbildet. Detailliert wird die Softwarearchitektur des Frameworks im folgenden Kapitel dargestellt.

Damit waren die softwaretechnischen Voraussetzungen für die Applikationsentwicklung geschaffen. Um die zweite Herausforderung, die anwenderorientierte Entwicklung von Applikationen realisierbar zu machen, wurde neben weiteren Softwarekomponenten ein rollen- und vorlagenbasierte Ansatz entwickelt und in das Framework integriert. Das Kernkonzept dieses Ansatzes besteht darin, sämtliche Objektdaten in so genannte Vorlagen zu implementieren. In einer TOP-Energy-Applikation werden diese Vorlagen dann verwendet, um geeignete Objekte zu erstellen. Die Vorlagen werden von der zweiten Anwendergruppe, den so genannten Applikationsentwicklern, erstellt. Die Entwicklung der Vorlagen wird durch geeignete Editoren vom Framework selbst unterstützt, so dass damit die zweite Herausforderung erfüllt wird. Das Erstellen neuer Applikationen wird detailliert in Kapitel 4 beschrieben.

3 Framework

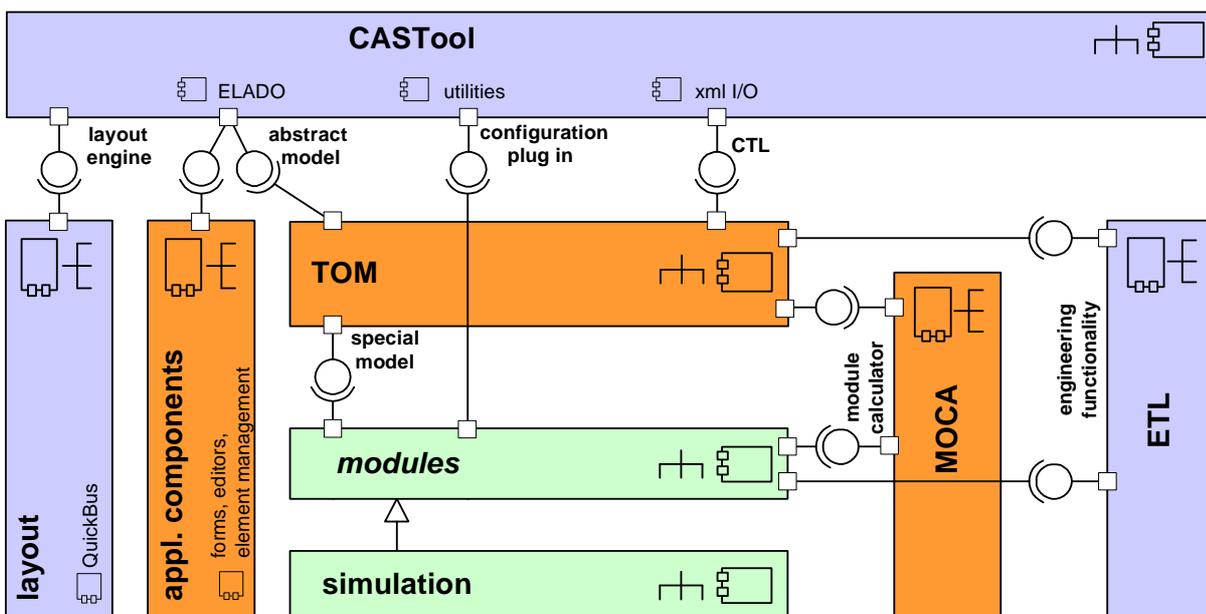


Abb. 1 Architektur des TOP-Energy Frameworks

In diesem Kapitel wird die Systemarchitektur des Frameworks vorgestellt und auf die einzelnen Softwarekomponenten detailliert eingegangen. Abb. 1 stellt die Grundarchitektur des Frameworks dar und hebt wichtige Schnittstellen hervor. Die einzelnen Komponenten werden im Folgenden beschrieben.

3.1 CASTool

Ursprünglich zur Entwicklung von graphischen Ingenieursystemen entworfen bietet CASTool (Computer Aided Schematics Toolbox) Mechanismen, netzartig und hierarchisch strukturierte Applikationsdaten aufzunehmen und schematisch darzustellen, anwendungsspezifische Funktionalität hinzuzufügen sowie unterschiedliches Objekt- und Systemverhalten zu definieren.

Datenmodell

Zentraler Bestandteil des Frameworks ist eine Modellbibliothek (ELADO - Erweitertes Layoutdatenmodell), welche das zugrunde liegende Netzwerkmodell implementiert und applikationsspezifische Parameter aufnehmen kann. Das Modell bildet sowohl strukturelle Informationen als auch zusätzliche Parameter (z.B. spezielle grafische Symbole, technologische Daten usw.) ab. ELADO setzt sich aus drei wesentlichen Teilen zusammen: dem Domänenmodell, dem Applikationsmodell und dem Repräsentationsmodell (vgl. Abbildung 2).

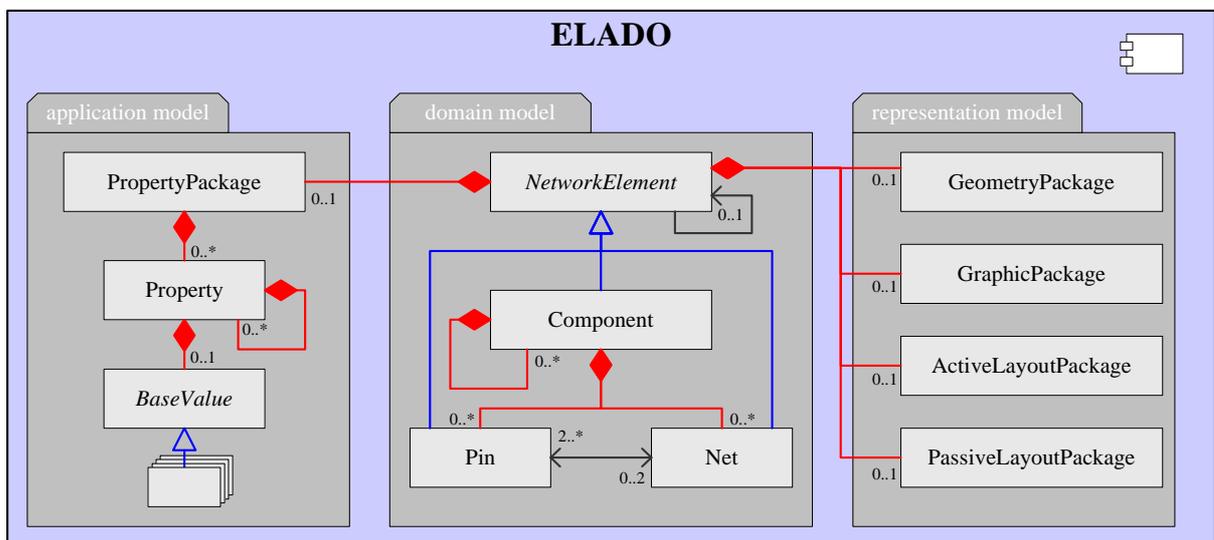


Abb. 2 ELADO Kernpaket und -klassen

Das Domänenmodell, auch als Strukturmodell bezeichnet, bildet Netzwerke ab. Netzwerke sind ähnlich wie mathematische Graphen strukturiert und bestehen im Wesentlichen aus

Komponenten, Netzen und Anschlusspunkten, so genannten Pins. In Erweiterung zu den Graphen können Komponenten, die den Graphknoten entsprechen, wiederum Netzwerke enthalten und Pins definierte Positionen auf diesen Komponenten haben.

Um verschiedene Applikationen entwickeln und insbesondere deren applikationsspezifische Daten (z.B. technische Daten, Geschäftsdaten) im Modell aufnehmen zu können, wurde ELADO um ein Applikationsmodell erweitert. Im Gegensatz zum Domänenmodell, in dem jedes relevante Element durch eine Klasse modelliert ist, ist das Applikationsmodell abstrakt aufgebaut. Applikationsspezifische Daten werden in einem Propertybaum modelliert. Jeder Knoten dieses Baumes kann ein Grunddatenobjekt und untergeordnete Properties besitzen. Ein komplexer Datensatz setzt sich dann aus mehreren Properties zusammen.

Darüber hinaus besitzt ELADO ein Repräsentationsmodell. Dieser Modellteil speichert z.B. Informationen über die grafische Darstellung von Netzwerkelementen, deren Shape und Positionen sowie layoutspezifisches Verhalten.

Erweiterbarkeit

Um verschiedene ingenieurtechnischen Lösungen entwickeln zu können, muss die Systemarchitektur so gestaltet sein, dass neue Funktionalität hinzugefügt und bestehende Funktionalität erweitert bzw. modifiziert werden kann. Dazu kann das Framework um anwendungsspezifische Dienste ergänzt werden; neben speziellen Layoutverfahren auch um ingenieurtechnische Modulkomponenten wie Simulatoren etc.

Das Hinzufügen von Komponenten erfolgt dynamisch zur Laufzeit über eine definierte Schnittstelle. In das System zu integrierende Modulkomponenten betten dazu ihre Funktionalität vollständig in Kommandos ein. Der Datenaustausch erfolgt durch eine abstrakte Parameterstruktur direkt über die Kommandoschnittstelle. Auf diese Art und Weise ist eine elegante Erweiterungsmöglichkeit für das System gegeben, da ein Entwickler ohne Eingriff in das System weitere applikationsspezifische Funktionalitäten hinzufügen bzw. vorhandene modifizieren kann. Die zugrunde liegenden Mechanismen basieren auf entsprechenden Entwurfsmustern [GHJV96].

Über den selben Mechanismus sind neben den Applikationskommandos auch im Framework implementierte Standardkommandos integriert. Dazu gehören z.B. einfache Layoutverfahren, Input/Output-Verfahren, Standardinteraktionen und Kommandos für die Projektverwaltungen. Dadurch erfolgt eine in sich geschlossene Kommandoabarbeitung, so dass für den Benutzer eine Anwendung als ein einheitliches, in sich geschlossenes System erscheint.

Konfigurierbarkeit

Neben der Erweiterbarkeit des Systems durch neue Funktionalitäten ist deren Systemintegration ein wesentlicher Aspekt zur Applikationsentwicklung. Dazu bietet CASTool verschiedene Konfigurationsmechanismen, welche vorrangig für die Gestaltung der grafischen Benutzeroberfläche einer konkreten Applikation genutzt werden. Dies kann einerseits durch die Modifikation von Konfigurationsdateien geschehen, andererseits ist durch dynamische Scriptauswertung auch eine Veränderung des Applikationsverhaltens zur Laufzeit des Systems möglich. Mit Hilfe der integrierten Scriptinterpreter können darüber hinaus einfache Funktionalitäten und Systemverhalten in Form von Scripten zur Systemlaufzeit implementiert werden.

3.2 Layout

In vielen technischen Anwendungen werden graphische Darstellungen von Netzwerken benötigt. Sie sollen die strukturellen Eigenschaften der Netzwerke hervorheben und müssen dabei ästhetisch ansprechend und übersichtlich sein. In der GFaI werden seit vielen Jahren Layoutverfahren für graphische Ingenieursysteme entwickelt.

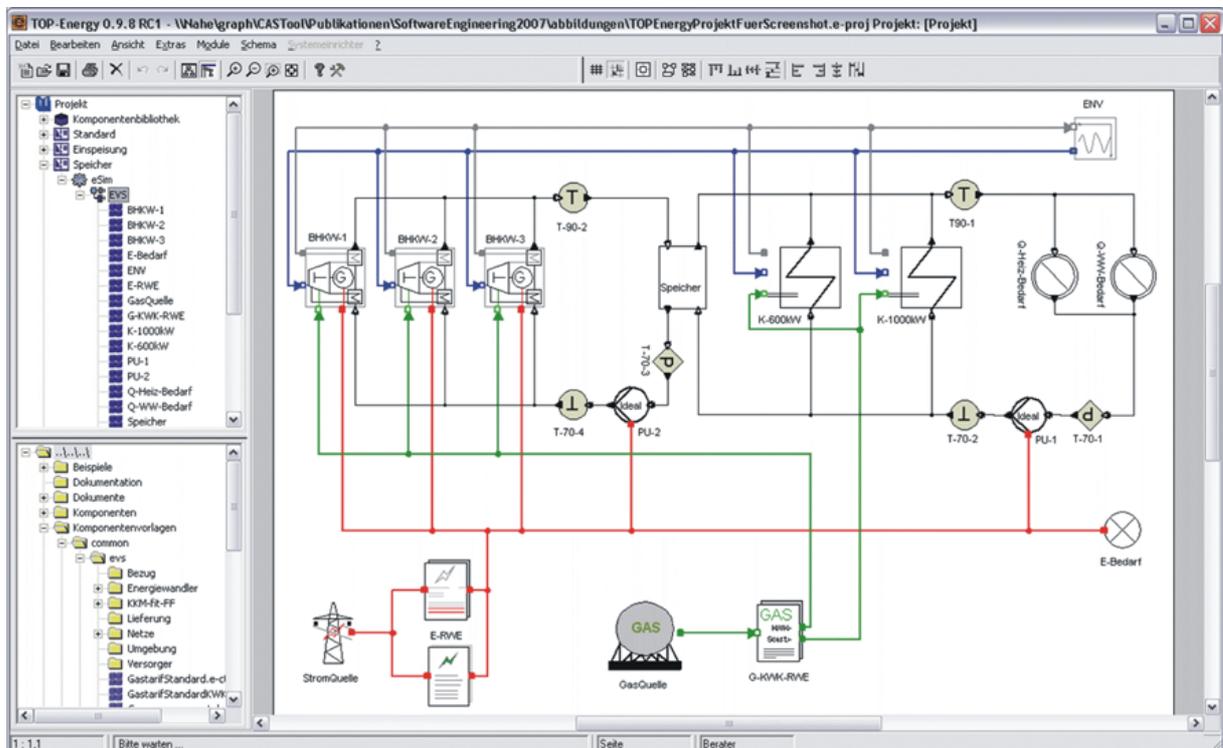


Abb. 3 Flusschemaeditor

Für den Flussschemaeditor von TOP-Energy (vgl. Abbildung 3, Beschreibung siehe Abschnitt 3.5) kommt neben Überlappungstests und Ausrichtfunktionen für Modellkomponenten ein so genanntes Busrouting [GKNP06] zum Einsatz. Dieses unterstützt den interaktiven grafischen Entwurf von Netzwerken, indem es auf interaktive Komponentenverbindungen und Veränderungen der Komponentenplatzierung mit lokalen Anpassungen des Routings der Netze reagieren. Das Busrouting sorgt prinzipiell für ein achsenparalleles Routing zwischen Komponenten, wobei Überlappungsfreiheiten zwischen Netzen und Komponenten sowie zwischen parallel verlaufenden Netzen berücksichtigt werden. Zusätzlich werden für Hyperkanten so genannte Busse ermittelt, die eine möglichst durchgängige geradlinige Struktur aufweisen. Das Busrouting bearbeitet dabei zunächst die betroffenen Netze autark. Danach beseitigt ein Postprocessing durch minimale Veränderungen des Routings Kollisionen mit Komponenten und anderen Netzen.

3.3 Engineering Type Library

Datentypen, wie Zeitreihen, Zahlen mit Einheiten, Kennfelder und Zeitpunkte sind in gängigen Programmiersprachen standardmäßig nicht vorgesehen. Daher bindet das TOP-Energy Framework die Basisbibliothek ETL (Engineering Type Library) ein, die diese Datentypen zur Verfügung stellt. Die ETL stellt sicher, dass in allen Operationen die physikalischen Einheiten der Zahlen berücksichtigt und gegebenenfalls konvertiert werden. Fast alle Datentypen, die auf einem Formular dargestellt werden, können in einen ETL-Datentyp konvertiert, modifiziert und wieder in das Formular geschrieben werden. Darüber hinaus bietet die ETL Funktionen, die das Bearbeiten von Strings oder Tabellen vereinfachen, und stellt einen Parser zur Verfügung, der es ermöglicht, komplexe Ausdrücke zu berechnen, z.B. $12[\text{kW}] * 645[\text{h}] + 72[\text{MJ}] = 7760[\text{kWh}]$

3.4 TOP-Energy Object Model

Das CASTool-Datenmodell ELADO ist, wie oben beschrieben sehr abstrakt, insbesondere in dessen Applikationsmodell. Um Applikationsentwicklern den Modellzugriff zu erleichtern, wurde ein Objektmodell implementiert, deren Objekte Vertreter von ELADO-Objekten darstellen und die sich in ETL-Objekte umwandeln lassen. Die Datenhaltung erfolgt ausschließlich im ELADO während die Vertreterobjekte geeignete Funktionalitäten implementieren.

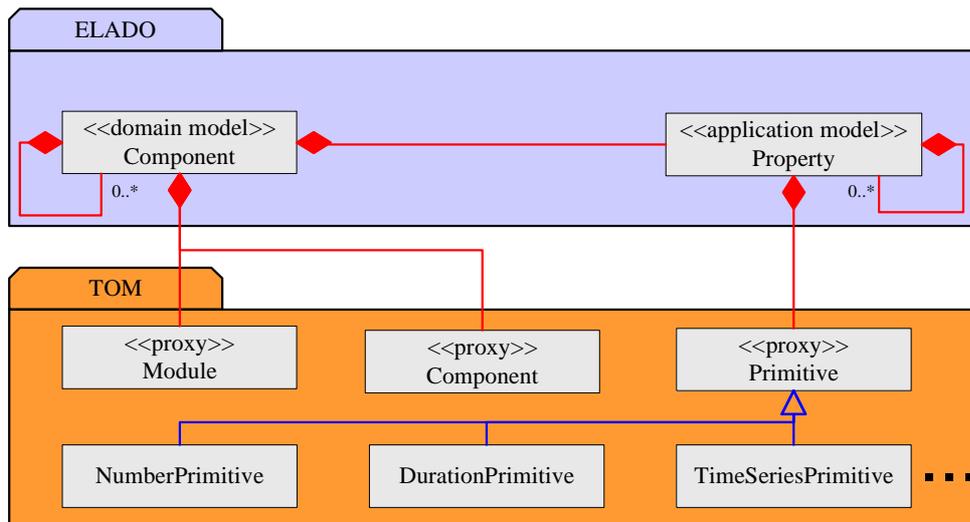


Abb. 4 TOP-Energy Object Model (Auszug)

In diesem Objektmodell können nun die für Softwareapplikationen der Energietechnik wesentlichen Objekte modelliert werden. Fachliche Funktionalitäten werden in so genannten Modulkomponenten gekapselt. Diese beinhalten Modellkomponenten (über die Komposition im ELADO), welche im Allgemeinen Modelle real existierender energietechnischer Komponenten darstellen. Sie beinhalten Daten, die zur Parametrisierung dieser Komponenten erforderlich sind. Die Daten werden durch so genannte Primitive implementiert. Primitive erscheinen nach außen wie unterschiedliche Datentypen (reelle Zahlen, Zeitdauern, Zeitreihen) die zudem über Zusatzinformationen verfügen (Grenzen, Default-Werten, Definitionen, Datenqualität etc.). Die Datenhaltung erfolgt wiederum im ELADO, in diesem Fall durch Propertybäume im Applikationsmodell.

3.5 Applikationskomponenten

Für das TOP-Energy Framework wurden spezielle Applikationskomponenten³, insbesondere ein Flussschema- und Formulareditor entwickelt. Diese Applikationskomponenten machen intensiven Gebrauch von den Basisbibliotheken CASTool und ETL.

Flussschemaeditor

Zur Unterstützung des Anwenders beim Beschreiben der Struktur von Energieversorgungssystemen dient der Flussschemaeditor (vgl. Abbildung 3). In diesem können Modell-

³ Applikationskomponenten sind spezielle Modulkomponenten, die von dem Framework zur Verfügung gestellt werden und von jeder angebundener Modulkomponente genutzt werden können. Sie stellen Dienste zur Verfügung, die sich in ihrer Funktionalität in das Framework eingliedern.

komponenten platziert und verbunden werden. Zur Unterstützung eines effizienten Entwurfs werden Layoutverfahren bereitgestellt, deren Intention es ist, eine optimale grafische Darstellung des Entwurfs zu erzeugen.

Folgende Aufgaben werden von den Layoutverfahren innerhalb des Flussschemaeditors übernommen: *Automatisches Routing*, *Sicherung von Überlappungsfreiheit*, *Ausrichtungsfunktionen*.

Zur Einhaltung von Verbindungsconstraints konnte auf einen allgemeinen Ansatz des CASTool zurückgegriffen werden. Dieser stellt sicher, dass lediglich gültige Netzwerke erstellt werden können. Dazu können die Struktur-Eigenschaften von Netzen definiert werden. Für ein gültiges Stromnetz wurde bspw. festgelegt, dass es immer mindestens zwei bidirektionale Strompins verbindet, während ein Dampfnetz genau zwei Pins verbindet. Damit sind Hyperstromnetze erlaubt jedoch keine Hyperdampfnetze. Darüber hinaus muss ein Dampfnetz immer genau ein Eingangspin und ein Ausgangspin verbinden, womit gerichtete Netze ermöglicht werden. Weiterhin implizieren diese Netzbeschreibungen, dass ein Netz zwischen einem Strompin und einem Dampfpin nicht gültig ist.

Formulare

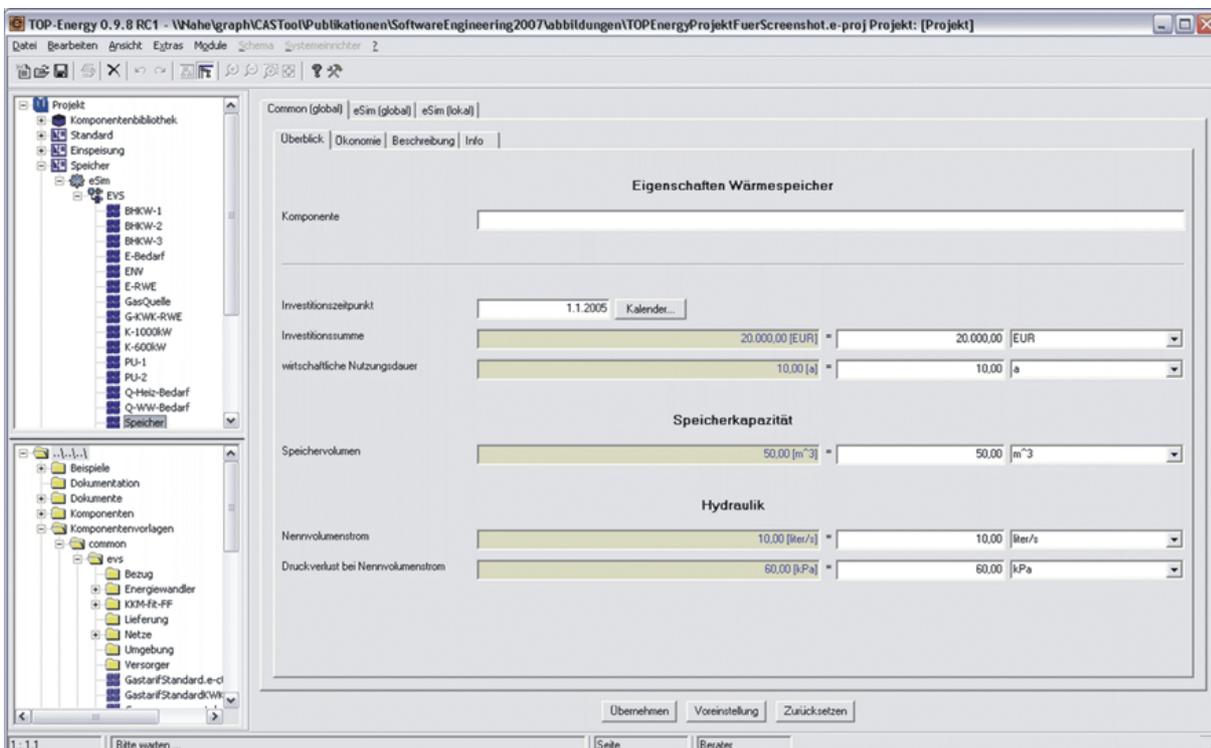


Abb. 5 Screenshot einer Formularansicht

Formulare (vgl. Abbildung 5) stellen in TOP-Energy eine bedeutende Benutzerschnittstelle dar. Sämtliche Daten verschiedener Modellkomponenten (Tarife, Anlagen, Personen, etc.) und Modulkomponenten (Simulator, Erstanalysemodul, etc.) werden über Formulare eingepflegt. Darüber hinaus dienen Formulare zur Visualisierung von Ergebnisdaten.

Die einzelnen Formularseiten sind zeilenbasiert aufgebaut. Jede Zeile repräsentiert ein Datenobjekt und wird durch eine Menge von Eingabeelementen realisiert. Der Aufbau der Formulare, also welche Daten wo dargestellt werden, ist nicht in Quellcode implementiert, sondern wird durch eine Formularbeschreibung in XML definiert. Zur Systemlaufzeit wird diese Beschreibung ausgewertet und die jeweiligen Formulare generiert. Dadurch wird dem Applikationsentwickler das Erstellen von Eingabemasken ohne Programmierung ermöglicht. Individuelle Formulare deren Aufbau von diesem Ansatz abweicht sind derzeit nicht möglich. Diese müssten im Quelltext bzw. in der Formularbeschreibung (nach Erweiterung) kodiert werden.

Neben den beiden beschriebenen Editoren sind weitere Applikationskomponenten entwickelt worden. Erwähnenswert ist insbesondere ein Berichtsgenerator, der ähnlich wie der Formular-editor mit Hilfe einer Berichtsdefinition den Inhalt eines Projektes in einem Bericht ausgibt. Die Berichtsdefinition wird wiederum vom Applikationsentwickler erstellt.

3.6 Module Calculator

Der Module-Calculator (MOCA) bildet eine Schnittstelle zwischen ETL, TOM und dem Entwickler. Er stellt eine verbesserte Fehlerbehandlung zur Verfügung, die es ermöglicht, einfache Berechnungen wesentlich komfortabler zu implementieren. Zudem erhält der Benutzer ein Feedback, warum und an welcher Stelle ein Fehler aufgetreten ist. Die aufgetretenen Fehler werden automatisch mit den zugehörigen Daten gespeichert. So kann der Benutzer überprüfen, warum ein bestimmtes Datum nicht ermittelt werden konnte und wie dieser Fehler behoben werden kann. Diese Schnittstelle eignet sich nicht zur Implementierung rechenzeitintensiver Operationen, da die Fehlerbehandlung viel Rechenzeit verbraucht. Sie ist zur schnellen Berechnung einfacher Kennzahlen gedacht. Durch die erweiterte Fehlerbehandlung wird eine wesentlich effizientere Modulentwicklung erreicht. Erste Tests mit der Softwarekomponente ergeben eine Einsparung von ca. 80% Quellcode, der vorher für die Fehlerbehandlung verwendet wurde.

4 Entwicklung neuer Applikationen

Die oben beschriebene Systemarchitektur stellt die Voraussetzung zur Entwicklung von TOP-Energy-Applikationen dar. Eine solche Applikation setzt sich aus verschiedenen Modulen zusammen. Abb. 1 zeigt, wie sich ein Modul in die Gesamtarchitektur einbettet. Damit ist ein Framework zur Entwicklung von energietechnischen Softwareapplikationen gegeben und somit die erste Herausforderung erfüllt.

Die zweite Herausforderung besteht darin, auf TOP-Energy basierende Applikationen durch *Fachingenieure*, also durch Nichtinformatiker, implementierbar zu machen. Die Konfiguration der Applikation und die Integration der Modulfunktionalität erfolgt dabei über die beschriebenen Mechanismen des CASTool und besteht im Wesentlichen darin, Konfigurationsdateien zu erstellen bzw. zu editieren. Der Zugriff auf die Modelldaten erfolgt wie beschrieben über das TOP-Energy Object Modell, welches das abstrakte Systemmodell ELADO kapselt. Der Entwickler von Applikationen steht darüber hinaus vor der Aufgabe, geeignete Objekte für die jeweilige Applikation sowie die fachlichen Funktionen zu implementieren.

4.1 Vorlagen erstellen

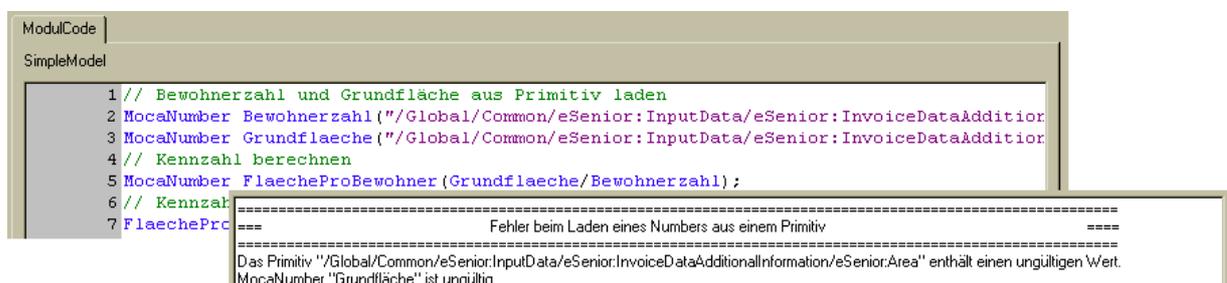
Um eine leichte Entwicklung von Applikationen basierend auf TOP-Energy zu gewährleisten, wurde als Kernansatz ein *rollen- und vorlagenbasierter Ansatz* in das Framework integriert. Der Grundgedanke dieses Ansatzes besteht darin, dass dem Anwender vom Applikationsentwickler eine Konfiguration von TOP-Energy zur Verfügung gestellt wird, mit der er Projekte durchführen kann. Wesentliche Bestandteile dieser Konfiguration sind Vorlagen. Alle für eine Applikation benötigten Modul- und Modellkomponenten werden durch Vorlagen beschrieben. Vorlagen stellen komplexe Objektdefinitionen dar. Sie beinhalten Strukturen zur Aufnahme von allgemeinen und technologischen Daten, grafische Ausprägungen (Symbole), Identifikatoren und Klassifikatoren sowie Dialogbeschreibungen (Formulare) und Berichtsdefinitionen zur Dateneingabe und -präsentation. Ihre Instanzen werden von den Anwendern als die eigentlichen Objekte ihrer Applikation verwendet. Da Applikationsentwickler gemäß Anforderungen keine Informatiker sind, mussten geeignete Werkzeuge zur Erstellung von Vorlagen (Vorlageneditoren) bereitgestellt werden.

4.2 Modulcode implementieren

Neben der Einbindung von Modulcode auf herkömmliche Weise mit einem Compiler und einem Linker kann die Funktionalität der Module auch mit Hilfe einer Skriptsprache in ein Modul integriert werden. Dies bietet dem Applikationsentwickler die Möglichkeit, die Arbeitsweise eines Moduls einzusehen, zu kontrollieren und gegebenenfalls anzupassen.

Die Implementierung von Modulcode mit dem Skriptinterpreter richtet sich an Ingenieure mit wenigen Programmierkenntnissen und benutzt daher das bereits beschriebene MOCA-Paket. Fehlerbehandlung und Benutzermeldungen bleiben bei der Programmierung also außen vor.

Ein Beispiel für eine Berechnung mit einem gemeldeten Fehler ist in Abb. 6 dargestellt.



```
ModulCode
SimpleModel
1 // Bewohnerzahl und Grundfläche aus Primitiv laden
2 MocaNumber Bewohnerzahl ("/Global/Common/eSenior:InputData/eSenior:InvoiceDataAddition
3 MocaNumber Grundflaeche ("/Global/Common/eSenior:InputData/eSenior:InvoiceDataAddition
4 // Kennzahl berechnen
5 MocaNumber FlaecheProBewohner (Grundflaeche/Bewohnerzahl);
6 // Kennzahl
7 FlaecheProBewohner (Grundflaeche/Bewohnerzahl);
=====
Fehler beim Laden eines Numbers aus einem Primitiv
=====
Das Primitiv "/Global/Common/eSenior:InputData/eSenior:InvoiceDataAdditionalInformation/eSenior:Area" enthält einen ungültigen Wert.
MocaNumber "Grundfläche" ist ungültig.
```

Abb. 6 Codebeispiel und zugehörige Fehlermeldung mit Moca

5 Vorhandene Module

Das TOP-Energy Framework bietet die Möglichkeit, neue Module zu erstellen, die das Spektrum der fachlichen Möglichkeiten erweitern. Die Module haben Zugriff auf alle Funktionalitäten des Frameworks und profitieren somit von Weiterentwicklungen des Frameworks. Im Folgenden werden der Simulator und weitere Module vorgestellt.

5.1 TOP-Energy Simulator

Ziele

Aufgabe des Simulationsmodul eSim ist die Berechnung ökonomischer und ökologischer Kennzahlen eines Energieversorgungssystems. Dabei kommen sowohl industrielle [AHK05], als auch kommunale [BoLu06] und Gebäudeenergieversorgungssysteme [AHK06] in Betracht. Ausgangsdaten zur Erstellung dieser Kennzahlen sind zeitlich aufgelöste Lastprofile, ökonomische Randbedingungen, technische Daten der eingesetzten Anlagen und Informationen über die Umgebung.

Anwender dieses Moduls sind Energieberater, Contractoren und kleinere Energieversorger, die mit einem vertretbaren zeitlichen Aufwand eine Entscheidung über die Güte eines Energiesystems treffen müssen. Die aufwendige Implementierung von Steuerung und Regelung soll dabei vermieden werden. Um eine optimale Steuerung zu gewährleisten, führt der Simulator eine Einsatzoptimierung durch.

Des Weiteren soll der Anwender ohne Kenntnis des Simulationsalgorithmus in der Lage sein, eine Simulation durchzuführen. Es müssen also stabile mathematische Verfahren angewendet werden, um ein aufwendiges Anpassen des Modells an den Simulationsalgorithmus zu vermeiden.

Modulansatz

Statt einer dynamischen Simulation, wie sie in vielen herkömmlichen Tools verwendet wird (Modelica, gProms), wurde in TOP-Energy ein quasi-statischer Simulator integriert, der mit einer Auflösung von typischerweise einer Stunde simuliert. Dieser ist im Allgemeinen ausreichend, da die Daten, die zur Vorhersage benutzt werden, zumeist aus historischen Daten stammen und darüber hinaus selten mit ausreichender Genauigkeit bekannt sind. Für eine tendenzielle Aussage ist die quasi-statische Betrachtung aber ausreichend.

Da in einem Energiesystem Anlagen mit diskreten Zuständen vorhanden sind, muss ein gemischt-ganzzahliges Optimierungsverfahren gewählt werden. Als besonders stabil stellte sich dabei die gemischt-ganzzahlig lineare Programmierung heraus. Dieses Optimierungsverfahren musste allerdings erweitert werden, da temperaturabhängige Modellkomponenten quadratische Gleichungen implizieren.

Die Modellierung erfolgt mit Hilfe des Flusschemaeditors auf der Basis einzelner Modellkomponenten. Diese Komponenten stellen entweder real existierende Anlagen dar oder können Randbedingungen in das System implizieren. Typische Komponenten sind BHKWs, Kessel, Wärmepumpen, aber auch Gastarife, Stromtarife, Außenluftzufuhr sowie die thermischen und elektrischen Lasten.

Umsetzung

Um quadratische Gleichungen mit einem linearen Optimierer behandeln zu können wurde um den linearen Löser ein evolutionärer Algorithmus geschaltet, der die Variablen, die in quadratischen Gleichungen vorkommen, vor der linearen Optimierung wählt und optimiert. Dieser Ablauf ist im Folgenden dargestellt.

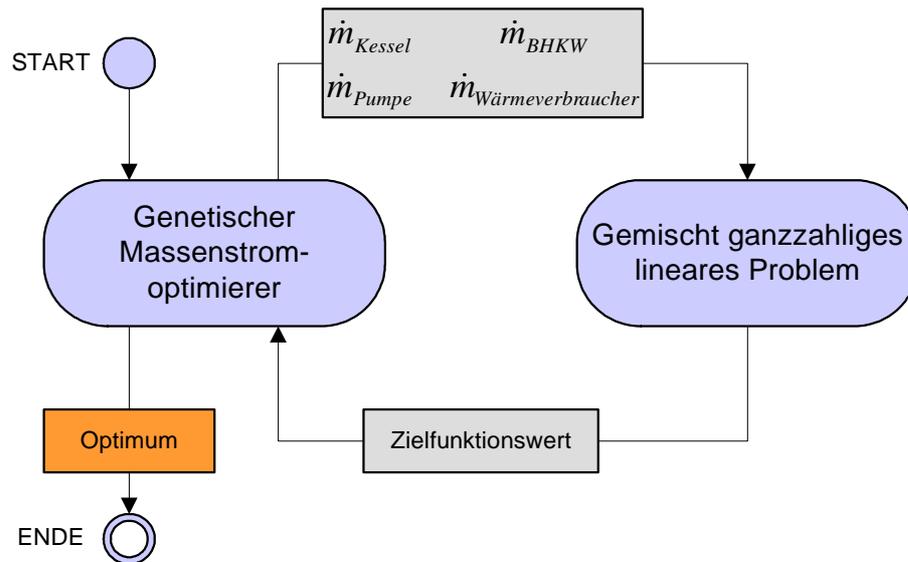


Abb. 7 Ablauf der quadratischen Optimierung

Da sich die Variablen in den quadratischen Gleichungen auf Massenströme zurückführen lassen, wird der evolutionäre Optimierer in der Grafik als Massenstromoptimierer bezeichnet.

Um die Erweiterbarkeit des Simulators um neue Anlagen zu gewährleisten, werden die einzelnen Modellkomponenten mit Hilfe einer Modellbeschreibungssprache modelliert. Es ist also möglich, Anlagen, die auf neuen Technologien beruhen, nachträglich dem Simulator hinzuzufügen.

5.2 Strukturoptimierer

Ziele

Bei der Planung neuer Energiesysteme wird ein erster Vorschlag für eine Systemstruktur typischerweise von einem Fachingenieur gemacht, der auf der Basis seiner Erfahrungen eine Systemstruktur vorschlägt. Zur genauen Auslegung der Anlagen in dieser Struktur werden in einem späteren Arbeitsschritt Simulationstools eingesetzt. Das Auffinden einer optimalen Systemstruktur ist sehr zeitaufwendig, da alle Strukturen erstellt und miteinander verglichen werden müssen. Das Strukturoptimierungsmodul soll automatisch eine Vorauswahl sinnvoller Strukturen treffen und dem Benutzer anbieten.

Umsetzung

Mit Hilfe eines evolutionären Algorithmus werden auf der Basis einer Überstruktur konkrete Strukturen erstellt und evaluiert. Die Evaluation wird von dem in Kapitel 5.1 beschriebenen

Simulator eSim vorgenommen, der Kennzahlen, wie Primärenergiebedarf, CO₂-Emissionen und Kosten berechnet, die als Zielfunktionen für die Strukturoptimierung fungieren. Die Strukturen werden nach einer oder mehreren Zielfunktionskriterien bewertet und sortiert. Ergebnis der Strukturoptimierung ist eine Pareto-optimale Menge aus Strukturen, mit deren Hilfe eine individuellen Gewichtung der Zielfunktionen vorgenommen werden kann.

6 Zusammenfassung

Das vorgestellte Framework bietet geeignete Datenstrukturen, Editoren und Schnittstellen an, um Softwaresysteme für die Energietechnik zu entwickeln. Damit ist eine umfassende Softwarebasis für Weiterentwicklungen geschaffen worden. Insbesondere in einem Umfeld, in dem häufig kleinere Anwendungen für spezifische Aufgaben geschrieben werden (Universitäten, Ingenieurbüros), ist das TOP-Energy Framework gut einsetzbar, da es eine schnelle Entwicklung ermöglicht und die Anwendungen wartbar hält. Die bereits entwickelten Module gehen weit über ein Proof of Concept hinaus. Einige dieser Module stehen in einer kostenlosen Applikation (Beta-Version) zur Verfügung [TOPE06].

Die Möglichkeit, Applikationen durch Fachingenieure und Wissenschaftler entwickeln zu lassen, führte dazu, dass sehr schnell neue Lösungen entstanden und entstehen. Der hohe Abstraktionsgrad des Basisframeworks CASTool wurde durch geeignete Wrapper und Editoren domänenspezifisch verringert. In Zukunft sollen weitere Editoren entwickelt werden, die insbesondere die Systemkonfigurationen zur Modulintegration vereinfachen. Derzeit muss dazu noch intensiv mit der Anwenderdokumentation [TOWI06] gearbeitet werden.

Die Weiterentwicklung des Frameworks und der Module wird von allen Beteiligten angestrebt.

7 Ausblick

Detaillierte Analysen bieten die Module der TOP-Energy Software in der aktuellen Version nur für Energieversorgungssysteme an. Die Betrachtung der industriellen Produktion und des Energiebedarfs bleibt dabei außen vor. Derzeit wird ein Modul entwickelt, das eine genauere Betrachtung der industriellen Produktion ermöglicht. Da für dieses Modul verschiedene Sichten auf einen Produktionsprozess (Planungssicht für den Produktionsablauf, Energiesicht,

Kostensicht) erforderlich sein werden ergibt sich die Anforderung an das Framework, neben dem Flusschemaeditor weitere Schemaeditoren zur Verfügung stellen zu können.

Weiterhin hat sich gezeigt, dass die Optimierungs- und Simulationsmodule sehr aufwendige Rechenoperationen durchführen müssen. Daher wird bei größeren Systemen die Rechenzeit der begrenzende Faktor für die Komplexität der Systeme. Es ist geplant, eine Parallelisierung der Rechenschritte vorzunehmen und das Framework an das D-Grid [DGRI06] anzubinden. Eine Anbindung an das D-Grid bzw. der Aufbau kleiner lokaler Grids soll insbesondere die Ausführung der evolutionären Algorithmen im Simulator und im Strukturoptimierer beschleunigen.

Weitere Arbeiten beschäftigen sich mit der Bearbeitung von Zeitreihen und der Integration einer Versionsverwaltung.

Literaturverzeichnis

- [AHK06] Augenstein E.; Herbergs S.; Kuperjans, I.: TOP-Energy – Ein Werkzeug zur Optimierung der Gebäudeenergieversorgung. KI-Luft und Kältetechnik, 5/2006
- [AHK05] Augenstein, E.; Herbergs, S.; Kuperjans, I.: TOP-Energy - Softwaregestützte Analyse und Optimierung industrieller Energieversorgungssysteme. VDI-Bericht 1908, 2005
- [AWKP04] Augenstein, E.; Wrobel, G.; Kuperjans, I.; Pleßow, M.: TOP-Energy – Computational Support For Energy System Engineering Processes. 1st International Conference „From Scientific Computing to Computational Engineering, IC-SCCE, Paper ID Nr.189, Athen, 2004, <http://ic-scce.upatras.gr>.
- [BoLu06] Bouvy, C.; Lucas, K.: Multicriterial optimisation of communal energy supply concepts. ECOS 2006, Crete.
- [BPPS01] Bartsch, Th.; Pleßow, M.; Pocher, M.; Schmidt, H.-W.: Ein universelles System für die Projektierung von Prozeßleitsystemen. ZwF Zeitschrift für wirtschaftlichen Fabrikbetrieb, ZwF 96 (2001) 4, Carl Hanser Verlag, München, pp. 205 - 211.

- [DGRI06] <http://www.d-grid.de>
- [GHJV96] Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J.: u.a.: Entwurfsmuster: Elemente wiederverwendbarer objektorientierter Software. Addison-Wesley, Bonn, 1996.
- [GKNP06] Goetze, B.; Kurth, R.; Nitz, S.; Pleßow, M.: NetViews - Verfahren zur automatischen Generierung prozessorientierter Sichten auf Kommunikationsnetzwerke, Abschlußbericht, GFaI-NetViews-11-06, Berlin, Mai 2006.
- [PFPL05] Pleßow, M.; Fröhling, R.; Pocher, M.; Lisounkin, A.: Tools for Knowledge Acquisition, Modeling and Visualisation applied to Process Supervision. in Krüger, J. (ed.): Proceedings of the 3. Industrial Simulation Conference 2005 (ISC-2005, Jun .9-11), Berlin, Germany, pp. 358 - 362.
- [PFE05] Pleßow, M.; Fröhling, R.; Ebert, R.-E.: Grafisch-interaktive Modellierung von Versorgungsnetzwerken. in A&D Kompendium 2005, publish industry Verlag GmbH, München, 2005, Seite 80-82. (in PDF-Form über www.aud24.net Code AD6A0305).
- [NN05] N.N.: Effizienter Energieeinsatz – Projekt zeigt neue Wege auf. Altenheim 10/2005, Seite 9.
- [NN05a] N.N.: Branchenenergiekonzept für die Recyclingindustrie. Euwid Nr. 32, 08/2005, Seite 8.
- [NN06] N.N.: Arbeitsgemeinschaft Branchenenergiekonzept Papier gegründet; APR-Allgemeine Papierrundschau, 04/2006, Seite 18.
- [TOPE06] Toolkit for Optimization of Industrial Energy Systems, <http://top-energy.ltt.rwth-aachen.de>
- [TOWI06] TOP-Energy-Wiki, <http://top-energy.ltt.rwth-aachen.de/wiki>

MDA-basierte Kopplung heterogener Informationssysteme im EVU-Sektor

Mathias Uslar, Niels Streekmann

Abteilung Betriebliches Informationsmanagement
OFFIS – Institut für Informatik
D-26131 Oldenburg
{uslar,streekmann}@offis.de

Sven Abels

FlexaDot – Consulting, Research & Development
D-26131 Oldenburg
sabels@acm.org

Abstract: IT-Systeme von Unternehmen der Energieversorgungsbranche sehen sich in den letzten Jahren einem verstärkten Druck zu ständigen Änderungen bedingt durch gesetzliche und markttechnische Anforderungen gegenüber. Diese Änderungen führen insbesondere zu verschiedenen heterogenen Informationssystemen, die oftmals losgelöst voneinander agieren (sog. Automation Islands). Dieser Artikel beschreibt einen Ansatz zur Kopplung dieser Systeme. Er geht dabei auf den MDA-Ansatz der OMG ein und beschreibt eine Möglichkeit, Systeme miteinander zu verbinden, indem eine Kopplung auf Prozessebene durchgeführt wird.

1 Änderung der Anforderungen an die IT-Systeme durch gesetzliche Vorgaben und andere Restriktionen

Der Wandel an den Energiemärkten verändert die Rahmenbedingungen, denen die IT-Landschaft innerhalb eines Versorgungsunternehmens unterworfen ist. Dabei lassen sich vor allem drei Faktoren identifizieren (vgl. auch [US05b]), die einen Einfluss auf das Zusammenspiel der einzelnen komplexen IT-Systeme besitzen.

Betrachtet man die historische Entwicklung von Systemen wie SCADA, GIS, CSS oder DMS, so waren diese zu Beginn innerhalb eines einzigen monolithischen Systems realisiert. Durch Spezialisierung und Komponentenbildung kam es dazu, dass Versorger über die Jahre nur noch einzelne Komponenten verschiedener Hersteller einsetzten und Kommunikationsverbindungen zwischen diesen Komponenten etablieren mussten. Wegen der extrem hohen Einsatz- und Lebensdauer dieser Systeme kommen über die Zeit der Nutzung immer wieder neue Adapter für Verbindungen zwischen den Systemen der verschiedenen Hersteller hinzu. Das führt zu einem Punkt-zu-Punkt Dilemma [Lu05].

Weiterhin hat die in jüngster Zeit zunehmende Verbreitung der dezentralen und regenerativen Energieerzeuger wie Windenergieanlagen, Brennstoffzellen oder Biogasanlagen dazu geführt, dass völlig neue Daten in komplett neuen Kommunikationsprozessen ausgetauscht werden müssen. Dies führt zu einem weiteren zahlenmäßigen Anstieg an Datenformaten und Verbindungen zwischen Systemen, die durch den Versorger mit seiner IT-Infrastruktur bedient werden müssen.

Letzter wichtiger Einflussfaktor ist das so genannte Legal Unbundling, welches zu einer Liberalisierung des Energiemarktes beitragen soll. Das Legal Unbundling bezeichnet die rechtliche Abtrennung zwischen Netz und Vertrieb und soll damit der Diskriminierung von Drittanbietern auf dem liberalisierten Strommarkt entgegenwirken. Es ist nach dem neuen Energiewirtschaftsgesetz (EnWG) für Unternehmen mit mehr als 100.000 Kunden zwingend erforderlich [Bu05]. Das Legal Unbundling ist eng verknüpft mit dem Informativischen Unbundling, welches die Trennung zwischen verschiedenen Teilen der Datenhaltung beschreibt. Ziel des Informativischen Unbundling bei einem Versorger im Strombereich ist beispielsweise die zukünftige Aufspaltung der Daten und Informationssysteme, welche aktuell noch gleichermaßen von den Bereichen Erzeugung und Netz genutzt werden. Das Informativische Unbundling soll für eine diskriminierungsfreie Bereitstellung von Daten für alle Marktteilnehmer sorgen.

Das Unbundling führt dazu, dass auch die Formate und Schnittstellen aller neuen Teilnehmer bedient werden müssen. So wurde beispielsweise von der Bundesnetzagentur am 11.07.2006 in einem aktualisierten Beschluss festgelegt, welche Formate zur Anbahnung und zur Abwicklung der Netznutzung bei der Belieferung von Letztverbrauchern mit Elektrizität zu nutzen sind, wenn zwei Marktpartner Kunden- bzw. Vertragsinformationen austauschen. Es wurden dabei Unterformate des EDIFACT-Standards festgelegt, wobei die folgenden Geschäftsprozesse

betroffen sind: Lieferantenwechsel, Lieferbeginn, Lieferende, Ersatzversorgung, Zählerstands- und Zählwerteübermittlung, Stammdatenänderung, Geschäftsdatenanfrage sowie Netznutzungsabrechnung. Details hierzu finden sich in [Bu06].

Die in den letzten Jahren erlassenen gesetzlichen Rahmenbedingungen schreiben eine Trennung bestimmter Systeme ebenso vor wie die Formate, mit denen eine Kommunikation gegenüber Dritten zu erfolgen hat. Beides führt zu entsprechenden Restriktionen, die bei einer Kopplung von Systemen zu beachten sind. Eine detaillierte rechtliche Betrachtung liegt jedoch außerhalb des Themenbereiches dieses Beitrages. Der geneigte Leser sei deshalb auf die erwähnten Dokumente und auf die gesetzlichen Vorgaben verwiesen.

2 Gängige Standards und Formate im EVU-Bereich

Betrachtet man den Nutzen von Standards, sind folgende Zahlen als erste Hinweise auf die Auswirkungen von Standards zu sehen. [DIN00] untersucht die Auswirkungen von Standards in Relation zum Gesamtumsatz der Branche: Ungefähr 1 Prozent kann durchschnittlich branchenübergreifend an Kosteneinsparungen realisiert werden; weltweit führt dies zu potenziellen Einsparungen in Höhe von 20 Millionen Euro allein im EMS/SCADA Markt und 24 Milliarden Euro in der gesamten Branche der Energiewirtschaft [Mae03]. Damit wäre die Auswirkung von Standardisierung höher als der Nutzen von Patenten oder Lizenzen.

Um den Nutzen von Standards in der Energiebranche genauer zu verdeutlichen, ist es nötig, verschiedene Seiten der entsprechenden IT-Infrastruktur zu beleuchten. Zunächst ist die technische Ebene zu nennen, die aus Steuerung, Kommunikation und dem Betrieb von Generatoren, Transformatoren, Leitungen in der Feldebene sowie dem Betrieb des Leitsystems besteht. Diese Ebene setzt selbstverständlich auf industrielle Kommunikations- und Datenmodellstandards wie dem Common Information Model CIM (IEC 61970) [Ie03] oder dem Feldebenekommunikationsstandard IEC 61850 [BW02] auf, da es sonst kaum möglich wäre, Geräte verschiedener Hersteller miteinander zu verbinden oder parallel zu einander einzusetzen, ohne eine sehr große Anzahl an Protokollen und Schnittstellen aufeinander abzustimmen.

Durch die Insolvenz eines Herstellers oder Diskontinuität in der Herstellung einzelner Produkte könnte ohne Standardisierung der Betrieb einer nötigen Infrastruktur ernsthaft gefährdet

werden. Im Bereich der technischen Anlagen ist die Standardisierung und Interoperabilität von Geräten bis auf der Ebene von Interchangeability für EVU daher bereits seit geraumer Zeit implementiert [Sc04].

Neben der (netz-)technischen Ebene existiert weiterhin eine kaufmännische IT. Die kaufmännische Seite wurde ähnlich wie auch die technische IT durch die Liberalisierung und Deregulierung der Branche stark verändert. Zur Koordination von Energiebeschaffung, -verteilung und -abrechnung ist mittlerweile ein erhöhter Informationsaustausch zwischen den Marktteilnehmern nötig. Lieferant, ÜNB, BKV, Händler etc. müssen sich über standardisierte Nachrichten austauschen können. In Deutschland wird seit langer Zeit auf den Standard EDIFACT gesetzt. Für einige Nachrichten hat der VDEW Vorarbeiten geleistet und Datenformate definiert, die auch im Internet auf der Verbandsseite frei erhältlich sind. Anders als beispielsweise in den Niederlanden existiert neben den Formaten keine weitere Unterstützung durch festgelegte Prozesse oder Referenzmodelle. Datenformate wie MSCONS, ESS (KISS) oder UTILMD sind definiert, ihre Verwendung jedoch erst spät von der Regulierungsbehörde vorgeschrieben worden. Es wird geschätzt, dass die Anzahl der Formate zum Austausch von Zeitreihen und Zählerdaten in Deutschland im dreistelligen Bereich liegt [Wei05]. Lieferanten, die Energie in eine große Anzahl von Netzen liefern, müssen daher nicht selten viel Aufwand betreiben, um alle entsprechenden Formate für ihre Kunden verarbeiten zu können. Zwar setzen mittlerweile fast 35 Prozent aller Nutzer auf MSCONS [Wei05], jedoch reicht auch ein solcher de-facto Standard noch nicht aus, es sind Standards für ein Gesamt-EVU von Nöten [Sc04].

Die IT in der Energiebranche ist historisch durch ihre heterogenen Systeme mit einer Vielzahl an Schnittstellen und Verknüpfungen geprägt. Systeme sind historisch gewachsen, wurden nach und nach miteinander verbunden und müssen Daten austauschen. Dadurch wurde meist eine hohe Anzahl an Punkt-zu-Punkt Verbindungen aufgebaut. Diese Verbindungen bestehen nicht nur zwischen den technischen, sondern auch zu den kaufmännischen Systemen, so dass eine Datenintegration zwischen diesen Systemen nötig ist. Je höher der Grad der Abhängigkeit zwischen den Systemen, d.h. die Intensität der Kopplung, desto teurer sind Anpassungen bei Änderungen an proprietären Schnittstellen bzw. die Integration neuer Systeme [Lu05]. Je mehr Adaptoren und Verbindungen, seien sie inner-betrieblich oder zu externen Systemen, gewartet werden müssen, desto höher die Kosten.

Abhilfe kann hier auf einer ersten Ebene die Einführung einer so genannten Common Language wie der IEC 61970 CIM auf einem Enterprise Message Bus schaffen. Dadurch muss jedes System nur noch einen Adapter zu dem Format auf dem Bus schaffen, sowohl für den Import als auch für den Export von Daten (vgl. auch Abschnitt 5). Die absolute Anzahl von Verbindungen und die Kosten für Erstellung und Pflege können so gesenkt werden. Eine Einschränkung besteht darin, dass natürlich beim Austausch ein Datenformat gewählt werden muss, was alle Elemente der Teilnachrichten vereint, d.h. es wird u. U. ein Format gewählt, welches für einen Teil der Nachrichten und Daten Overhead verursacht. Trotzdem sind die Kosten durch diese Art von Standardisierung zu senken und eine Erweiterung der aktuellen IT um neue Systeme sowie eine Verknüpfung mit den Daten der Altsysteme wird erleichtert.

3 Informationssysteme im EVU-Bereich

In EVU-Unternehmen sind in der Regel viele verschiedene Informationssysteme im Einsatz. Im Folgenden werden drei dieser Systeme beschrieben. Diese Systeme dienen in den nachstehenden Abschnitten zur Bildung eines Szenarios, anhand dessen der Ansatz der MDA-basierten Integration im EVU-Umfeld beschrieben wird. Es handelt sich bei den Systemen um reale und im Einsatz befindliche exemplarische Systeme, die jedoch umbenannt wurden sowie leicht vereinfacht beschrieben werden.

3.1 EMS (Energiemanagement für dezentrale Erzeuger)

Das erste betrachtete System ist ein System zum Management dezentraler Energieerzeugung, welches verschiedene Komponenten wie beispielsweise wetterabhängige Lastprognose, Windleistungsvorhersage, Netzzustandsvorhersage (State Estimation), Flickerüberwachung, Knotenlastprognose oder Dispositionsmanagementsfunktionen umfasst. Diese Einzelkomponenten bieten ihre Dienste über Schnittstellen an, die als WebServices umgesetzt wurden. Neben den reinen Services werden noch Legacy-Daten über eine Einkopplung in eine zentrale EAI-Plattform zur Verfügung gestellt und können durch die Komponenten genutzt werden, was eine dezentrale Datenhaltung mit mehreren zu synchronisierenden Systemen unnötig werden lässt. In einer weiteren Ausbaustufe des Systems wird eine Kopplung an SCADA-Funktionen angestrebt.

Die Daten zwischen den Komponenten werden nachrichtenbasiert mittels SOAP ausgetauscht. Dabei wird eine Microsoft BizTalk Plattform eingesetzt. Auf ihr können Prozesse eines EVU orchestriert werden, die im späteren Verlauf nachrichten-getriggert über die Plattform gesteuert ablaufen. Das EMS-System befindet sich bereits zu Teilen im Produktivbetrieb. Erste Prozesse befinden sich im Einsatz und tragen zu einem besseren Umgang mit den neuen fachlichen Anforderungen des dezentralen Energiemanagements bei.

3.2 CSS (Customer Support und Billing-System)

Das CSS-System ist ein integriertes Customer-Care und Billing-System mit dessen Entwicklung 1994 begonnen wurde und welches sich seit 1997 im produktiven Einsatz befindet. Das System ist bei verschiedenen Energieversorgungsunternehmen im Einsatz und verwaltet derzeit eine Größenordnung zwischen 4000 und 1.7 Mio. Zählern. Das System ist dabei anhand der Funktionalität in die Aspekte *Stammdaten, Verkauf, Vertrieb & Marketing, Kundenverkehr, Inkasso, Outputmanagement* und *Informationssysteme* unterteilt. CSS wurde eigens für den liberalisierten Markt entwickelt und komplett in das SAP R/3 System integriert. Eine der Hauptanforderungen an das System war es, die Vertriebs-, Marketing- und Abrechnungsfunktionalitäten zu vereinen und es den Bedürfnissen des liberalisierten Strommarktes anzupassen.

CSS stellt einerseits eine Customer-Care-Lösung dar, die eine schnelle und umfangreiche Kundenbetreuung ermöglicht und dabei Kontakte, Vorgänge und Dokumente verwaltet, sowie technische Daten zu Kunden, Verträgen und Versorgungsobjekten vorhält.

Auf der anderen Seite stellt CSS eine Billing-Software dar, die das Abrechnen von Verbrauchsarten inklusive sämtlicher Nebenkosten ermöglicht. Aufgaben des Systems sind beispielsweise die Erstellung von Meldungen für den Bilanzkreiskoordinator, die Mehr- und Mindermengenberechnung und die Bilanzierung und Visualisierung der Durchleitung bzw. Netznutzung.

3.3 SAS (Störungsannahmesystem)

Bei dem SAS-System handelt es sich um ein derzeit in der Einführungsphase befindliches Informationssystem zur Störungsannahme und –auskunft. Ziele des Systems sind:

- Die Erfassung von Störungsmeldungen im EVU-Bereich wie z.B. die Erfassung von Ausfällen.

- Die dauerhafte Dokumentation von Störungen und insbesondere von Störungsursachen.
- Die softwaregestützte Benachrichtigung von Außendienstmitarbeitern zur Behebung der Störung.
- Die Auskunft im Krisenfall.
- Die Erstellung von Unternehmensstatistiken zur Auswertung der Ausfallsituation im Unternehmen.

Die Software wurde als komplette Neuentwicklung erstellt und ist nicht ausschließlich auf die Erfassung von Störungen in Stromnetzen ausgerichtet, sondern wird auch für Störungen im Gas- und Wasserbereich eingesetzt. Bei der Entwicklung des Systems wurde ein modellgetriebener Ansatz verfolgt bei dem die Modellierung auf UML-Basis und die automatische Generierung von Quellcode einen zentralen Punkt in der Entwicklung des Systems einnahmen. Neben der Schnittstelle zur Annahme von Störungen mittels einer grafischen, webbasierten Oberfläche bestehen Webservice-Schnittstellen zur Erweiterung und Abfrage des Systems.

Die folgende Grafik verdeutlicht das Vorgehen Fall einer eingehenden Störungsmeldung.

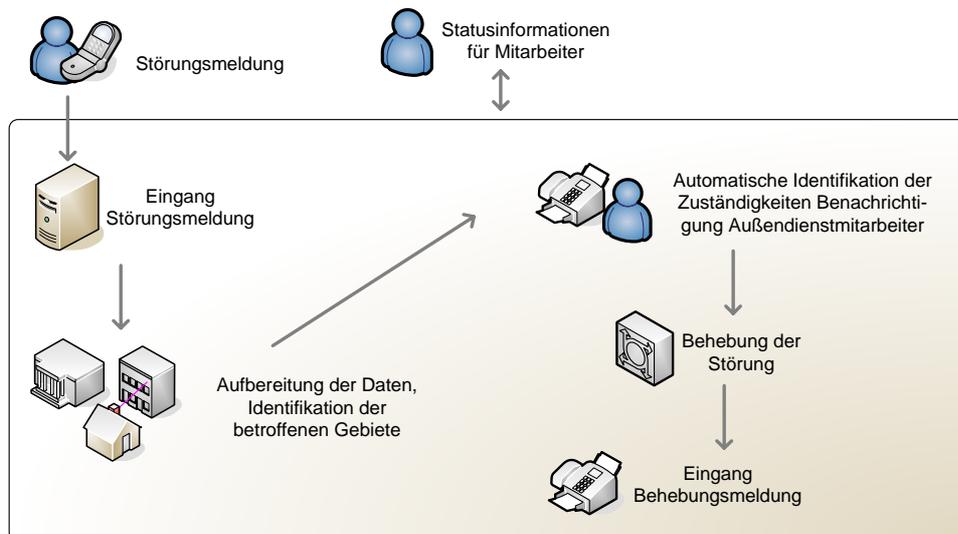


Abbildung 1: Ablauf einer Störungsmeldung im SAS

4 MDA-basierte Kopplung

Für die Etablierung einer ganzheitlichen und in sich homogenen IT-Landschaft spielt die Interoperabilität von Informationssystemen eine wichtige Rolle. Als Interoperabilität wird dabei die nahtlose Zusammenarbeit von Systemen in Bezug auf ihre Daten und Prozesse verstanden

(vgl. [IE90]). In diesem Abschnitt wird ein Ansatz der Kopplung zwischen Systemen im EVU-Bereich vorgestellt, der die Interoperabilität zwischen den skizzierten Systemen gewährleisten soll. Dieser Ansatz basiert dabei auf dem MDA-Ansatz [OMG03] der Object Management Group (OMG).

4.1 Szenario

Die im letzten Abschnitt beschriebenen Systeme sind derzeit als heterogene Informationssysteme in zeitlich und personell unterschiedlichen Projekten entstanden. Sie sind daher bislang weder verknüpft, noch in derselben Programmiersprache erstellt oder nach ähnlichen Programmierparadigmen entwickelt worden. Während EMS aus verschiedenen Systemen besteht, die mittels einer SOA verbunden sind und mittels eines BizTalk-Servers koordiniert werden, ist das CSS-System komplett in die SAP-Systemlandschaft integriert und daher in ABAP bzw. ABAP Objects als SAP-Modul entwickelt worden. Das dritte System (SAS) hingegen wurde als modellgetriebene Entwicklung erstellt und in C# umgesetzt.

Trotz der unterschiedlichen Funktionalitäten und Anwendungsgebiete der Systeme hat sich gezeigt, dass eine Verbindung in einigen Anwendungsfällen zielführend und zeitsparend wäre. Der folgende Anwendungsfall beschreibt ein derartiges Szenario und wird durch Abbildung 2 visualisiert.

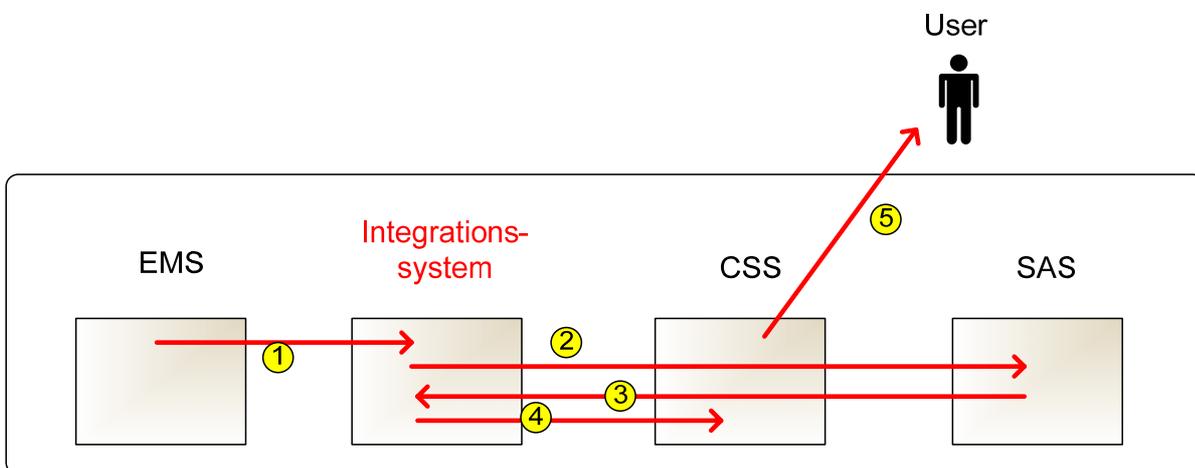


Abbildung 2: Prozessablauf des Beispielszenarios

Die Abbildung zeigt die beteiligten Systeme und deren Kommunikation (Pfeilrichtung) in einer bestimmten Reihenfolge (Nummerierung). Innerhalb des Szenarios wird angenommen, dass im EMS-Prozess eine bestimmte Information zum unerwarteten Ausfall eines Stromerzeugers erkannt wird. Diese Information wird an ein weiter unten beschriebenes Integrationssystem

propagiert, welches das Störungs-Annahme-System (SAS) über den Ausfall informiert. Hier wird der Ort, auf den der Ausfall Einfluss hat (z.B. Straßennamen, Ortsteile, etc.), identifiziert und eine entsprechende Störungsmeldung ausgelöst. Hat eine Identifizierung stattgefunden, so wird im Schritt drei eine Rückmeldung an das Integrationssystem geliefert. Dieses leitet die Information anschließend an das CSS-System weiter. Hier werden diejenigen Zählpunkte und Kunden identifiziert, die direkt von der Störung betroffen sind. Eine entsprechende Meldung ist so durch den Sachbearbeiter direkt erkennbar.

Durch diese Integration kann ein Sachbearbeiter einen Ausfall ohne merkliche Verzögerung erkennen. Andernfalls wäre es ihm durch die getrennte Datenhaltung der Systeme nicht möglich, entsprechende Informationen zu erhalten. Er könnte lediglich die entsprechenden Informationssysteme (EMS und SAS) manuell abfragen. Die oben beschriebene lose Kopplung der Systeme erlaubt eine nahtlose Integration aller Informationen und beschleunigt dadurch den Informationsfluss deutlich.

4.2 Durchführung der Kopplung der Systeme

Die beschriebene Kopplung der heterogenen Systeme kann nach verschiedenen Verfahren und in verschiedenen Ebenen der Systeme erfolgen. Denkbar sind dabei folgende Ebenen [CHK05]:

Persistenzebene

Eine Kopplung der Systeme auf Persistenzebene kann durch die Nutzung einer gemeinsamen Datenbank realisiert werden, auf die alle beteiligten Systeme zugreifen. Dabei müssen alle Systeme so verändert werden, dass Sie ein gemeinsames (relationales) Datenbanksystem verwenden. Der Nachteil der Kopplung auf dieser Ebene besteht zum einen in dem hohen Aufwand zur Vermeidung von Konflikten (Gleichzeitiges Ändern eines Datensatzes etc.) und zum anderen in der Notwendigkeit zur Modifizierung aller Systeme und ggf. der zugrunde liegenden Datenmodelle.

Funktionslogik

Bei einer Kopplung auf Funktionsebene werden die logischen Funktionen der Systeme miteinander verbunden. Die Systeme arbeiten dabei mit getrennten Datenbanksystemen und eigener Funktionslogik, die vom jeweils anderen System im Falle einer Änderung aufgerufen wird. In diesem Fall ist eine starke Änderung der bestehenden Funktionslogik notwendig, was unter Umständen zu hohen Kosten führen kann.

1. Analyse der betroffenen Altsysteme auf die Bereiche, die zur Kopplung benötigt werden.
2. Identifikation und logische Extraktion von WebServices innerhalb der Altsysteme.
3. Technische Umsetzung der WebServices durch die Schaffung einer Serviceschicht und durch Kapselung der vorhandenen Geschäftslogik.
4. Generierung eines Integrationssystems zur Kopplung und Orchestrierung.

Das in der Abbildung als Integrationssystem beschriebene System wird im folgenden Abschnitt beschrieben.

4.3 Umsetzung des Integrationssystems

Als Alternative zur Nutzung von WebServices und dem im folgenden Abschnitt beschriebenen Integrationssystem käme – insbesondere im betrieblichen Kontext – eine Middleware-Komponente wie beispielsweise SAP XI im SAP Netweaver-Umfeld in Frage¹. Die Nutzung von derartigen Produkten hat den Vorteil, dass keine eigenständige Komponente zur Integration entwickelt werden muss und dass das entsprechende System (wie z.B. SAP XI) i.d.R. bereits vielfach im Praxiseinsatz erprobt und daher entsprechend ausgereift ist. Im skizzierten Anwendungsfall wurde sich dennoch aus mehreren Gründen für ein eigenes auf WebServices basierendes Integrationssystem entschieden. Zum einen bietet eine solche Entwicklung die größtmögliche Flexibilität im Erstellungsprozess. Zum anderen kann durch die Nutzung des MDA-Ansatzes der OMG ein großer Teil des Systems ausgehend von der fachlichen Beschreibung automatisiert erstellt werden. Dies ermöglicht die Beteiligung von Domänenexperten im direkten Entwicklungsprozess. Domänenwissen ist insbesondere im EVU-Bereich ein wichtiger Faktor, dessen Nutzung durch dieses Vorgehen begünstigt wird.

Abbildung 4 zeigt eine Übersicht des MDA-Ansatzes, der einen modellgetriebenen Entwicklungsprozess definiert. Dabei werden verschiedene Modelle beschrieben, die mittels Modelltransformation automatisch ineinander überführt werden können und jeweils andere Aspekte eines Systems beschreiben. Das *Computation Independent Model* (CIM) ist eine fachliche Beschreibung des Systems, die zusammen mit Domänenexperten erstellt wird. Dieses Modell wird in ein *Platform Independent Model* (PIM) überführt, das technische Aspekte des Systems unabhängig von der zur Umsetzung verwendeten Technologie beschreibt. Die

¹ <http://www.sap.com/platform/netweaver>

technologie-abhängige Beschreibung liefert schließlich das *Platform Specific Model (PSM)*, das aus dem PIM unter Hinzufügung von Plattforminformationen abgeleitet werden kann. Aus dem PSM kann dann der Quellcode des Systems generiert werden.

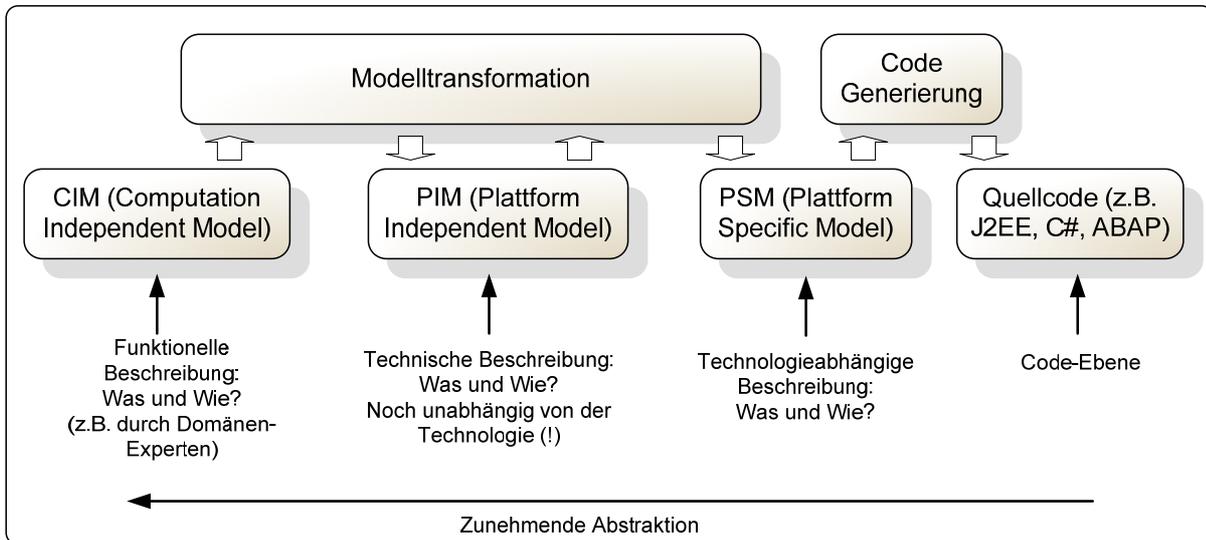


Abbildung 4: Übersicht des MDA-Ansatzes

Um das Integrationssystem mit Hilfe des MDA-Ansatzes umsetzen zu können, muss dieser um Vorgehensweisen und Methoden zur Integration erweitert werden. Ein allgemeiner Ansatz zur Integration von Altsystemen ist die BALES-Methode [HHP00]. Diese beschreibt die Kopplung eines top-down erstellten Referenzmodells des Zielsystems und eines bottom-up erstellten Modells des Altsystems. Die BALES-Methode kann auf den MDA-Ansatz übertragen werden. Dabei wird das zu erstellende System fachlich in Form von Geschäftsprozessen beschrieben, die technisch in Form von WebServices umgesetzt werden (top-down). Aus dem Altsystem werden ebenfalls WebServices extrahiert (bottom-up). Diese werden auf die Services aus dem fachlichen Modell abgebildet.

Für die konkrete Umsetzung des MDA-Ansatzes bedeutet dies, dass im MDA-CIM der integrierende Geschäftsprozess aus Abbildung 1 modelliert wird. Im PIM werden die zu integrierenden WebServices, d.h. die beteiligten bereits bestehenden Services aus dem EMS und dem SAS-System, sowie die aus CSS extrahierten Services, die für den Geschäftsprozess nötig sind, technisch beschrieben. Das PSM stellt die plattformabhängige Beschreibung des Integrationssystems dar, aus dem letztendlich der Quellcode mit Hilfe entsprechender Generatoren erstellt wird. Eine Generierung von Quellcode für die anderen Systeme und damit eine Änderung an diesen Systemen ist nicht nötig, da die Webservice-Infrastruktur im Fall von EMS und SAS bereits besteht und im Fall vom CSS in einem Vorschrift der Integration

bereitgestellt wird. Wichtig für die Umsetzung der Integration ist, dass auf allen Ebenen Informationen über die zu integrierenden Systeme in die entsprechenden Modelle einfließen. Für die Integration auf Geschäftsprozessebene für das oben genannte Integrationssystem ergeben sich damit verallgemeinert folgende Schritte, die für die Erstellung des Systems mittels eines MDA-basierten Integrationsansatzes durchgeführt werden müssen:

- Fachliche Modellierung des integrativen Geschäftsprozesses im MDA-CIM.
- Technische Modellierung der entsprechenden WebServices aller beteiligten Systeme und ihrer Kopplung im PIM.
- Plattformabhängige Modellierung des Integrationssystems im PSM. Um das PIM in ein PSM transformieren zu können ist ein zusätzliches Plattformmodell notwendig, das die Plattform beschreibt, auf der das System implementiert werden soll.
- Generierung des Quellcodes für das Integrationssystem aus dem PSM.

5 Mappinggestützte Integration von ausgetauschten Datenobjekten

Bei der Verwendung des vorgestellten MDA-Ansatzes kommt es zur Notwendigkeit des Austausches der fachlichen Datenobjekte. Es sind deshalb Mappings zwischen den Elementen der ausgetauschten Datenobjekte vorzunehmen. Dabei werden innerhalb des Integrationssystems die einzelnen Nachrichtenformate ineinander oder in eine so genannte Common Language (beispielsweise das Common Information Model CIM) überführt. Dieses Format dient als Austausch- oder Zwischenformat zwischen allen beteiligten Formaten.

Zur Integration werden daher vor einer Kopplung fachliche Mappings durch einen Domänenexperten erstellt, die dann umgesetzt werden. Bei einer Verwendung des Common Information Model CIM lässt sich auch für die Erstellung der XML-Nachrichtenschemata ein MDA-basierter Ansatz nutzen, bei dem aus einer in XMI serialisierten Darstellung des gesamten CIM einzelne Objekte samt Attributen ausgewählt werden können und durch eine Umwandlung in OWL, RDFS oder XSD für die nachrichtenbasierte Integration nötige Schemata erzeugt werden können [KS03].

Im Gegensatz zu der früher üblichen und aufwändigen Erstellung von Adaptern für Punkt-zu-Punkt-Verbindungen zwischen Systemen muss bei der Verwendung einer Common Language nur jeder Standard auf das Zwischenformat abgebildet werden statt wie sonst üblich im schlimmsten Falle paarweise zwischen den Standards aller Systeme. Dies setzt voraus, dass sich

alle fachlichen Objekte in das Zwischenformat (z.B. CIM) abbilden lassen bzw. das Zwischenformat für nötige proprietäre Erweiterungen entsprechende Mechanismen zur Verfügung stellt. Daher ist eine nachrichtenbasierte Integration nicht für jeden Anwendungsfall sinnvoll durchführbar.

Eine Integration von Standards, die direkt miteinander verknüpft werden sollen, ohne die Verwendung einer gemeinsamen Zwischensprache kann beispielsweise über die semantische Ebene erfolgen. Bei einer ontologiebasierten Integration werden gemeinsame Konzepte ermittelt, die dann in einem späteren Schritt mittels einer Upper-Ontologie [KHR05] aufeinander abgebildet werden. Hierfür stehen verschiedene Methodiken des Ontology-Matchings und Ontology-Mappings bereit. Eine detaillierte Beschreibung eines entsprechenden Mappings und eine Marktübersicht über gängige Methoden finden sich in [AHH05]. Durch die Abbildung der einzelnen Ontologien müssen nicht die einzelnen Datenmodelle angepasst werden, was wegen existierender Implementierungen durch Hersteller oftmals nicht möglich ist. Die semantische Verknüpfung über gleiche Konzepte ist in der Energiebranche jedoch noch eine offene Herausforderung, erste Ansätze existieren für die Integration des CIM IEC 61970 sowie der Feldebenekommunikation IEC 61850 [Sc02], [KPF03].

6 Techniken MDA-basierter Integration

Aufgrund des frühen Stadiums der Erforschung MDA-basierter Integrationsmethoden und bisher kaum vorhandenen praktischen Erfahrungen in diesem Bereich können an dieser Stelle noch keine aussagekräftigen Evaluationsergebnisse präsentiert werden, sieht man von Schätzungen und ersten Kopplungen ab, die jedoch keine wissenschaftliche Bewertung des Gesamtverfahrens ermöglichen. Es werden stattdessen an dieser Stelle Metriken festgelegt, die für die Evaluation MDA-basierter Integration als relevant angesehen und zur späteren Evaluation herangezogen werden. Basis dieser Kriterien sind allgemein diskutierte Vor- und Nachteile der MDA sowie Punkte, die sich aus der Integration ergeben. Die folgenden Kriterien wurden dabei identifiziert:

Entwicklungskosten: Einer der erwarteten Vorteile des Einsatzes des MDA-Ansatzes ist die Verkürzung der Entwicklungszeit und damit auch die Senkung der Entwicklungskosten. Das soll vor allem durch Verringerung der manuellen Implementierung durch den Einsatz von Generortechnologie geschehen. Eine Voraussetzung für diese Annahme ist aber, dass der

Komplexitätsgrad der verwendeten Modelle deutlich geringer ist als die Komplexität von Programmcode. Eine weitere Voraussetzung ist, dass ein hohes Maß an Wiederverwendung, vor allem in Bezug auf die Generatoren und die Werkzeuge zur Modelltransformation, möglich ist. Auch die Verwendung von Standards, wie z.B. das Common Information Model, unterstützt dieses Ziel.

Codequalität: Es wird davon ausgegangen, dass durch den Einsatz des MDA-Ansatzes die Qualität des erzeugten Quellcodes steigt. Dazu tragen zu einem großen Teil die Wiederverwendung von Generatoren und Templates bei. Durch häufige Wiederverwendung sinkt die Wahrscheinlichkeit von Fehlern im Gegensatz zu manuell erzeugtem Quellcode. Durch die Beachtung von Programmierkonventionen kann auch die Lesbarkeit des Codes auf einem gleichmäßig hohen Niveau gehalten werden.

Wartbarkeit: Dadurch dass Änderungen am System nur an den entsprechenden Modellen und nicht direkt am Quellcode vorgenommen werden müssen, steigt auch die Wartbarkeit von MDA-basierten Systemen im Gegensatz zu konventionell erstellten Systemen. Dieser Punkt ist besonders wichtig für Integrationssysteme, da alle integrierten Systeme jederzeit Änderungen erfahren können, an die das Integrationssystem möglichst schnell und flexibel angepasst werden muss.

Performanz und Skalierbarkeit: Weitere nicht-funktionale Anforderungen, die an ein Informationssystem gestellt werden sind Performanz und Skalierbarkeit, da sie die Arbeit mit dem erstellten System erheblich beeinflussen können. Auch hier werden dem MDA-Ansatz aufgrund der strukturierten Herangehensweise und der Möglichkeit, diese Faktoren bereits im Stadium der Modellierung berücksichtigen zu können, Vorteile zugesprochen. Im Falle der beschriebenen Integration ist das von hoher Relevanz, da das nicht-funktionale Verhalten heterogener Systeme berücksichtigt werden muss.

Die Untersuchung dieser Kriterien in einem praxisnahen Szenario ist Inhalt zukünftiger Forschung. Eine Möglichkeit dafür bietet das vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Projekt MINT [MINT06]. In diesem Projekt werden Vorgehensweisen zur modellbasierten Integration entwickelt. Ein Teil des Projektes ist zudem die Durchführung einer Fallstudie, in der die Vorgehensweisen im konkreten Anwendungskontext evaluiert werden. Ein mögliches Szenario ist dabei das in Abschnitt 5 erläuterte Beispiel.

7 Ausblick

Dieser Beitrag hat einen Überblick über die aktuelle IT-Situation von Energieversorgungsunternehmen in Deutschland gegeben und die Notwendigkeit zur Kopplung von Informationssystemen in diesem Bereich anhand eines Beispiels verdeutlicht. Es wurde ein Ansatz zur losen Kopplung von Informationssystemen durch eine MDA-basierte Integration mittels eines Integrationssystems gegeben. Der Ansatz basiert dabei auf einer Kopplung aller Systeme auf Prozessebene. Dies ermöglicht es, die Systeme zu großen Teilen lediglich um eine Service-Schicht zu erweitern (sollte diese nicht bereits existieren), anstatt die Funktionslogik der Systeme verändern zu müssen.

Der beschriebene Ansatz ist Gegenstand des Forschungsprojektes MINT, das im Rahmen der „Software Engineering 2006“-Initiative des BMBF durchgeführt wird. Die in dem Projekt gemachten Praxiserfahrungen werden von den Projektpartnern projektbegleitend zur Verfügung gestellt. Interessante und offene Fragen betreffen dabei insbesondere die praktische Umsetzung der Webservice-Extraktion (siehe Kapitel 4) aus Systemen des produktiven Betriebes sowie die Abbildung der verschiedenen Systemnachrichten innerhalb des Integrationssystems (vgl. Kapitel 5).

Literaturverzeichnis

- [AHH05] Abels, S.; Haak, L.; Hahn, A.: Identification of Common Methods Used for Ontology Integration Tasks. In: Proceedings of the first International ACM Workshop on Interoperability of Heterogeneous Information Systems (IHIS05), CIKM conference proceedings. ACM, Sheridan Publishing, 2005.
- [Bu05] Bundesministerium der Justiz: Gesetz über die Elektrizitäts- und Gasversorgung, neues Energiewirtschaftsgesetz (EnWG), http://bundesrecht.juris.de/enwg_2005, 2005.
- [Bu06] Bundesnetzagentur, Beschlusskammer 6, AZ BK6-06-009, Festlegung einheitlicher Geschäftsprozesse und Datenformate zur Abwicklung und Belieferung von Kunden mit Elektrizität, 2006

- [BW02] Brand; Wimmer: Der Standard IEC 61850 Offene Kommunikation in Schaltanlagen im deregulierten Strommarkt. Bulletin SEV/VSE 1/02, 2002; 9-13.
- [CHK05] Conrad, S.; Hasselbring, W.; Koschel, A.: Enterprise Application Integration. Grundlagen - Konzepte - Entwurfsmuster – Praxisbeispiele, Spektrum Akademischer Verlag , 2005
- [DIN00] Deutsches Institut für Normung e.V.: Gesamtwirtschaftlicher Nutzen der Normung: Zusammenfassung der Ergebnisse, wissenschaftlicher Endbericht mit praktischen Beispielen, Berlin, Wien, Zürich, Beuth Verlag, 2000
- [HHP00] van den Heuvel, Willem-Jan; Hasselbring, Wilhelm; Papazoglou, Mike: Top-Down Enterprise Application Integration with Reference Models. In: Australian Journal of Information Systems 8 (2000) 1, S. 126-136.
- [IE90] IEEE, IEEE Standard Computer Dictionary: A compilation of IEEE standard computer glossaries, Institute of Electrical and Electronics Engineers, 1990.
- [In03] International Electrotechnical Commission: INTERNATIONAL STANDARD IEC 61970-301: Energy management system application program interface (EMS-API) – Part 301: Common Information Model (CIM) Base. International Electrotechnical Commission, 2003.
- [KHR05] Kalfoglou, Y.; Hu, B.; Reynolds, D. & Shadbolt, N.: Semantic Integration Technologies Survey; 2005, <http://eprints.ecs.soton.ac.uk/10842/>
- [KPF03] Kostic, Tanja; Preiss, Otto; Frei, Christian: Towards the Formal Integration of Two Upcoming Standards. IEC 61970 and IEC 61850. In (El-Hawary; Little (Hrsg.)): Proceedings for the 2003 LESCOPE Conference, Montreal, 7-9 May, IEEE Publishing, 2003.
- [KS03] Kalfoglou, Y. and Schorlemmer, M.: Ontology mapping: the state of the art. The Knowledge Engineering Review 18(1) pp. 1-31.

- [Lu05] Luhmann, Till: Ein service-orientierter Ansatz zur nachhaltigen Anwendungsintegration in Unternehmen. GI Jahrestagung (2) 2005: 402-406
- [Mae03] März, Wolfgang: Importance of Standardisation – The Business Case, Vortrag auf dem DKE 952 Treffen Dortmund 2003, 2003
- [MINT06] Projektseite des MINT-Projekts. <http://www.mint-projekt.de>. (Abruf: 20.07.06)
- [OMG03] OMG, MDA Guide 1.0.1. <http://www.omg.org/docs/omg/03-06-01.pdf>. (Abruf: 20.07.06), 2003
- [Sc04] Schwarz, Karlheinz: IEC 61850, IEC 61400-25, and IEC 61970: Information models and information exchange for electric power systems. Proceedings of the Distributech 2004, Orlando, Florida, 2004.
- [Us05]. Uslar, M.; Schmedes, T.; Luhmann, T.; Appelrath, H.-J.: Eine serviceorientierte Architektur für das dezentrale Energiemanagement. In: Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005: Informatik LIVE!, Band 2, Beiträge der 35 Jahrestagung der Gesellschaft für Informatik e.V. (GI) 19. bis 22. September 2005 in Bonn, Gesellschaft für Informatik, Bonn, Köllen Verlag, Bonn, S.622-626, 2005
- [Us05b]. Uslar, M.; Schmedes, T.; Luhmann, T.: Rahmenbedingungen und Lösungen für Enterprise Application Integration bei EVU. In: Softwaretechnik-Trends, 25(2), Gesellschaft für Informatik, Bonn, S.74-75, 2005
- [Wei05] Weisshaar, Rainer: Datenaustausch in einem liberalisierten Energiemarkt, emw, Heft 2/05, 2005

Einführung in den Track

Rechtsfragen der Informationsgesellschaft

Prof. Dr. Thomas Dreier

M.C.J., Universität Karlsruhe (TH)

Prof. Dr. Andreas Wiebe

LL.M., WU Wien

Prof. Dr. Rupert Vogel

Bartsch und Partner, Karlsruhe

Die rasante Entwicklung der Informationsgesellschaft führt zu immer größerer Automatisierung und Mobilität. Zugleich werden ganze Lebensbereiche, in denen Informationstechnologien bislang nur punktuell zum Einsatz kommen, zunehmend umfassend technologisch umgestaltet. Zu nennen sind hier neben neuen Geschäftsmodellen der Internetökonomie vor allem die Bereiche e-Health, e-learning und e-government.

Technikgesteuerte und zugleich sozialverträgliche Anwendungen lassen sich jedoch nur dann realisieren, wenn sie auf rechtliche Rahmenbedingungen angemessen Rücksicht nehmen. Denn Ziel dieser rechtlichen Rahmenbedingungen ist es, für die Durchsetzung rechtspolitischer Wertevorgaben zu sorgen, um die Freiheit sowohl der Anwender wie auch der Nutzer zu sichern. Dabei geht zum einen in einer Vielzahl von Fragen darum, welche Auswirkungen die rechtlichen Vorgaben auf die Ausgestaltung der Technologien und ihrer Nutzung haben. Zum anderen ist die Frage zu beantworten, wie das Recht beschaffen sein muss, um den neuen technologischen Möglichkeiten nach Möglichkeit zum Durchbruch zu verhelfen.

IT-Systeme in der Rechnungslegung und entsprechende Prüfungsanforderungen

Rechtliche Rahmenbedingungen: Aktueller Stand und Perspektiven

Elizaveta Kozlova, Ulrich Hasenkamp

Philipps-Universität Marburg
Institut für Wirtschaftsinformatik
35032 Marburg

{kozlova|hasenkamp}@wiwi.uni-marburg.de

Abstract

Für die Korrektheit der in der Buchhaltung eingesetzten Software und für die Richtigkeit der IT-gestützten Buchführung sind die Buchführungspflichtigen verantwortlich und nicht, wie irrtümlich oft angenommen wird, die Softwarehersteller. Der zunehmende Einsatz von IT in der Rechnungslegung findet auch Berücksichtigung in der Wirtschaftsprüfung gemäß §§ 316-324 HGB. Der vorliegende Beitrag gibt einen Überblick über die aktuelle Rechtslage des Einsatzes der IT in der Rechnungslegung sowie dessen Auswirkungen auf die Wirtschaftsprüfung und zeigt die Entwicklungstrends in diesem Bereich auf.

1 Einführung

Aus der stetig gewachsenen Bedeutung von Finanzbuchhaltungsprogrammen für die Rechnungslegung der Unternehmen resultiert der Bedarf an rechtlichen Rahmenbedingungen in diesem Bereich und deren kontinuierlicher Fortentwicklung. In Kapitel 2 dieses Beitrages werden die aktuellen rechtlichen Normen und Entwicklungen im Bereich der IT-gestützten Buchführungssysteme aufgezeigt. Diese Normen haben auch deswegen eine besondere Relevanz, weil sie auf alle Konzepte, die im Zusammenhang mit rechnungslegungsrelevanten Geschäftsprozessen stehen (z. B. Electronic Data Interchange (EDI), Betriebsdatenerfassung (BDE), Materialwirtschaftssysteme, Dokumenten-Management-Systeme, Workflow-Management-Systeme), anzuwenden sind. Im Rahmen der Jahresabschlussprüfung beurteilt der Wirtschaftsprüfer unter

anderem die Ordnungsmäßigkeit der Buchführung. Mit dem zunehmenden Einsatz der IT in der Rechnungslegung entstehen neue Herausforderungen für die Prüfer. In Kapitel 3 werden die Vorschriften und Verlautbarungen diskutiert, die für die Wirtschaftsprüfer im Rahmen der IT-Systemprüfungen relevant sind. Schlussbemerkung und Ausblick folgen in Kapitel 4.

2 Rechtliche Rahmenbedingungen für IT-gestützte Buchführungssysteme

2.1 Überblick

Die wichtigste rechtliche Grundlage für eine Buchführung, die mittels Informationstechnik (IT) erledigt wird, schaffen Handelsgesetzbuch (HGB) und Abgabenordnung (AO). Nach § 239 Abs. 4 HGB und § 146 Abs. 5 AO können Bücher und die sonst erforderlichen Aufzeichnungen auch auf Datenträgern geführt werden, soweit diese Form der Buchführung einschließlich des dabei angewandten Verfahrens den Grundsätzen ordnungsmäßiger Buchführung (GOB) entsprechen. Neben den handelsrechtlichen Vorschriften und den GOB sind für die IT-gestützten Buchführungssysteme weitere Rechtsvorschriften zu beachten, insbesondere auch steuerlicher Art, wie die Einkommensteuer-Richtlinien (EStR).

<p>HGB (Handelsgesetzbuch) AO (Abgabenordnung)</p>	<ul style="list-style-type: none"> • § 239 Abs. 4 HGB und § 146 Abs. 5 AO: Grundsätzliche Möglichkeit, Bücher und sonst erforderliche Aufzeichnungen auch auf Datenträgern zu führen • Die Neufassung von §§ 146, 147 AO zum 1. Januar 2002: Möglichkeit des direkten Zugriffs durch die Prüfer der Finanzämter auf die IT-Systeme
<p>BMF (Bundesministerium der Finanzen)</p>	<ul style="list-style-type: none"> • BMF-Schreiben vom 7. November 1995: „Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme“ (GoBS) • BMF-Schreiben vom 16. Juli 2001: „Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen“ mit der Wirkung zum 1. Januar 2002

Tab. 1: IT-gestützte Buchführungssysteme: Rechtliche Rahmenbedingungen

Neben den handelsrechtlichen rechnungslegungsspezifischen Vorschriften und den Grundsätzen ordnungsmäßiger Buchführung sind bei der Gestaltung und beim Einsatz IT-gestützter Buchführungssysteme weitere gesetzliche Vorschriften zu beachten, die speziell im Hinblick auf den IT-Einsatz entwickelt wurden. Diese Anforderungen wurden in der Tabelle 1 zusammengefasst.

In den nächsten Teilabschnitten werden die „Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme“ und „Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen“ etwas ausführlicher betrachtet.

2.2 Vorstellung der Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme

Im Jahre 1978 wurden die Grundsätze ordnungsmäßiger Speicherbuchführung (GoS) erstmalig veröffentlicht. Die Technik hat sich seitdem enorm weiterentwickelt; heutzutage benutzen praktisch alle Unternehmen für die Buchhaltung Finanzbuchhaltungsprogramme („FiBu-Programme“). FiBu-Software nimmt eine besondere Stellung im Bereich der betrieblichen Anwendungssysteme ein, weil die Programmfunktionen wesentlich durch rechtliche Vorschriften reglementiert sind. Das Zertifikat eines Wirtschaftsprüfers bestätigt die Einhaltung von GoB (Grundsätze ordnungsmäßiger Buchführung) und der daraus abgeleiteten Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme (GoBS) durch die Software.

Die GoBS wurden von der Arbeitsgemeinschaft für wirtschaftliche Verwaltung e.V., Eschborn, (AWV) ausgearbeitet und als Schreiben des Bundesministeriums der Finanzen am 7. November 1995 verabschiedet. Die GoS und die GoBS unterscheiden sich im Wesentlichen in folgenden Punkten [Phil98, 313; Zepf96, 1259]:

- Die Buchhaltung in den GoBS wird als integrierte und nicht als isolierte und eindeutig abgrenzbare Unternehmensfunktion betrachtet. Ein DV-gestütztes System umfasst erheblich mehr als die Finanzbuchhaltung mit ihren Nebenbuchhaltungen Kreditoren-, Debitoren- und Anlagebuchhaltung. Die GoBS sind auf alle Verfahren, die rechnungslegungsrelevante Geschäftsprozesse verarbeiten (z. B. EDI, BDE, Materialwirtschaftssysteme, Dokumenten-Management-Systeme, Workflow-Management-Systeme) anzuwenden.
- Der Anwendungsbereich der GoBS schließt sowohl moderne und zukünftige Informationssysteme als auch moderne Verfahren zur Informationssystementwicklung und -pflege sowie Produktdokumentation ein.
- Durch die Ausführungen zu den Fragen „was, wogegen, wie lange und wie ist zu sichern“ wird der Inhalt des erforderlichen Datensicherungskonzepts deutlicher als in den GoS dargestellt.

- Der Begriff Internes Kontrollsystem (IKS) erhält eine zentrale Bedeutung. Die Anwendung des systemorientierten Prüfungsansatzes wird hervorhoben. Die Bedeutung des IKS wird in Abschnitt 3.1 erläutert.

Die GoBS sind den Grundsätzen ordnungsmäßiger Buchführung untergeordnet. Nach § 238 Abs. 1 S. 1 HGB sind alle Kaufleute verpflichtet, Bücher zu führen und in diesen ihre Handelsgeschäfte nach den Grundsätzen ordnungsmäßiger Buchführung aufzuzeichnen. Nach § 239 (2) HGB müssen die Aufzeichnungen vollständig, richtig, zeitgerecht und geordnet vorgenommen werden. Die GoBS präzisieren die GoB im Hinblick auf den Einsatz der IT und sollen sicherstellen, dass die Buchungen und sonst erforderlichen Aufzeichnungen vollständig, richtig und geordnet vorgenommen werden [StHa05, 340]. Die GoBS stellen keine detaillierten Handlungsanweisungen zur Organisation und Führung einer Buchführung mit GoB-Garantie dar [Zepf96, 1259]. Sie stellen lediglich die aus den GoB abgeleiteten Anforderungen an DV-gestützte Buchführungssysteme dar. Die Anforderungen an die Ordnungsmäßigkeit bleiben im Wesentlichen dieselben wie bei einer manuell erstellten Buchführung. Ein Buchführungssystem ist ordnungsmäßig, wenn es bei objektiver Betrachtung den GoB-Anforderungen gerecht wird. Diese Feststellung beruht auf dem Einführungserlass zur HGB-Änderung von 1977: Im HGB sollen keine bestimmten Buchführungsmethoden und -maßnahmen im Einzelnen vorgeschrieben werden. Laut GoBS sollen die erforderlichen Aufzeichnungen jederzeit innerhalb einer angemessenen Frist verfügbar und lesbar gemacht werden. Die Nachvollziehbarkeit der einzelnen Geschäftsvorfälle muss ergänzend durch eine aussagekräftige Verfahrensdokumentation dargestellt werden. Das dient vor allem der Sicherung der GoB im Hinblick auf die Beleg-, Journal- und Kontenfunktion. Ein wichtiger Aspekt ist die bildliche Übereinstimmung von Handels- oder Geschäftsbriefen und Buchungsbelegen mit dem Original. Alle auf der Originalunterlage enthaltenen Angaben müssen auf dem Bild erkennbar sein, damit die Aussage- und Beweiskraft des Geschäftsvorfalles nicht gefährdet wird. Aus dem Verzicht auf den Buchungsbeleg aus Papier folgt kein Verzicht auf den zwingenden Nachweis des Zusammenhangs zwischen dem buchhaltungspflichtigen Vorgang in der Realität und seiner Abbildung in der Buchführung. Nach § 257 Abs. 3 Ziff. 1 HGB müssen empfangene Handelsbriefe und Buchungsbelege bildlich wiedergegeben werden im Unterschied zu abgesandten Handelsbriefen, Konten, Journalen und sonstigen Organisationsunterlagen, die nur inhaltlich wiederzugeben sind. Die Anforderung ist erfüllt, wenn alle auf der ursprünglichen Originalunterlage enthaltenen Angaben zur Aussage- und Beweiskraft des Geschäftsvorfalles originaltreu wiedergegeben sind. Wenn z.B. die

Schriftart auf dem Original nichts aussagt und beweist, dann muss sie nicht wiedergegeben werden. Wenn ein buchführungspflichtiger Vorgang ausgelöst wird, ohne dass dabei ein originärer Papierbeleg die Basis ist, erfüllt das jeweilige Verfahren die Beweisfunktion (Belegfunktion).

Bei der elektronischen Buchführung spielt die IT-Sicherheit eine besondere Rolle. Der Inhalt des erforderlichen Datensicherungskonzepts wird in GoBS sehr ausführlich diskutiert. Die GoBS betonen, dass die Datensicherungsanforderungen des Unternehmens über die GoBS-Anforderungen hinausgehen. Die relevanten Daten sind gegen Verlust zu sichern und vor unberechtigter Veränderung zu schützen. Funktionierende Zugriffskontrollen sind bei den Buchführungssystemen die entscheidende Voraussetzung dafür, dass eine erfolgte Aufzeichnung (Buchung) nicht in einer Weise verändert werden kann, dass der ursprüngliche Inhalt nicht mehr feststellbar ist. Die Wirksamkeit der Zugriffsberechtigungskontrollen hängt davon ab, wie gut das System der Benutzeridentifizierung funktioniert und wie zuverlässig die Vergabe und Einrichtung von Berechtigungen stattfinden. Die handels- und steuerrechtlichen Aufbewahrungspflichten sollen streng beachtet werden. Die Buchführung gilt nicht mehr als ordnungsmäßig, wenn aufbewahrungspflichtige Unterlagen fehlen.

Ein wichtiges Kriterium für die Ordnungsmäßigkeit eines Buchführungssystems stellt die Verfahrensdokumentation dar. Es ist erforderlich so zu handeln, dass eine aktuelle Verfahrensdokumentation auf Dauer verfügbar ist. Die Verfahrensdokumentation soll den Anwendern die Möglichkeit geben, das Verfahren sachgemäß anzuwenden und notwendige Änderungen schnell und korrekt vorzunehmen. Für die Prüfer ist es allerdings nur wichtig, dass das Vorhandensein der Verfahrensdokumentation nachgewiesen ist. Die Qualität der Verfahrensdokumentation muss für das Prüfungsergebnis nicht ausschlaggebend sein.

Für die Einhaltung der GoB ist auch bei Einsatz der IT-gestützten Buchführungssysteme der Buchführungspflichtige verantwortlich. Wenn eine DB-Buchführung an einen externen Dienstleister ausgelagert wird, trägt ebenso der Buchführungspflichtige die Verantwortung für die Einhaltung der GoBS. Seine Aufgabe besteht darin, aus den allgemein gültigen GoBS-Anforderungen für sein individuell gestaltetes Buchführungssystem die richtigen detaillierten Regelungen und Kontrollen zur Sicherung der Ordnungsmäßigkeit festzustellen.

2.3 Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen

Mit der Neufassung von § 147 AO zum 1. Januar 2002 wurde die digital auswertbare Speicherung für originär digitale und steuerlich relevante Unterlagen vorgeschrieben, die zur Wieder-

verarbeitung in einem IT-gestützten Buchführungssystem geeignet sind. Im Rahmen von Steuerprüfungen macht die Finanzverwaltung zunehmend davon Gebrauch, sich einen digitalen Zugriff auf DV-gestützte Buchführungssysteme zu verschaffen. Hierzu hat das Bundesfinanzministerium am 16. Juli 2001 ein Anwendungsschreiben „Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen“ (GDPDU) erlassen (BMF 2001). Das Finanzamt hat im Rahmen einer Außenprüfung unmittelbaren Zugriff auf alle steuerlich relevanten Daten eines Unternehmens, d.h. die Daten der Finanz-, Anlagen- und Lohnbuchhaltung, einschließlich des Rechts, die Übergabe von Datenträgern zu fordern. Originär digitale Unterlagen sind maschinell verwertbar zu archivieren. Sie dürfen nicht mehr nur in ausgedruckter Form oder auf Mikrofilm aufbewahrt werden. Wenn kryptografische Verfahren und elektronische Signaturen eingesetzt werden, müssen die entsprechenden Schlüssel aufbewahrt werden [StHa05, 341]. Diese Regelungen sind seit Januar 2002 bei der Gestaltung und Anwendung von IT-gestützten Buchführungssystemen zu berücksichtigen.

3 IT-Systemprüfungen

3.1 Zum Verhältnis der Wirtschaftsprüfung und Wirtschaftsinformatik

Die Veränderungen in Unternehmen haben auch Einfluss auf die Art und Durchführung der modernen Prüfung. Nicht zuletzt sind diese Veränderungen auf den zunehmenden Einsatz der IT im Unternehmen im Allgemeinen und in der Rechnungslegung insbesondere zurückzuführen. In Tabelle 2 sind die wichtigsten davon zusammengefasst.

Es gibt zwei Richtungen, wie IT und Wirtschaftsprüfung sich gegenseitig beeinflussen. Zum einen bedeutet der zunehmende Einsatz der Informationstechnik in der Rechnungslegung eine neue Sichtweise auf den Beruf eines Wirtschaftsprüfers. Die Wirtschaftsprüfer müssen sich mit den Informationssystemen, die in Unternehmen eingesetzt werden, gut auskennen, um die entsprechenden IT-Systemprüfungen durchführen zu können. Grundzüge und Prüfung der Informationstechnologie sind nach § 4 der Prüfungsverordnung für Wirtschaftsprüfer vom 20. Juli 2004 (BGBl. I S. 1707) Bestandteile des Wirtschaftsprüferexamens [oV04, 6]. Zum anderen beeinflussen die neuen Entwicklungen im IT-Bereich die Prüfetechniken. Die wachsende Abhängigkeit der Unternehmen vom informationstechnischen Umfeld verlangt die hinreichende Anpassung der Prüfmethoden und die ständige Überwachung und Prüfung der im Un-

ternehmen eingesetzten IT-gestützten Systeme [MaQu03, 541]. In diesem Artikel interessiert primär der erste Aspekt.

Neue Prüfgebiete	<ul style="list-style-type: none"> • Corporate Governance (einschließlich IT-Compliance) • E-Business • Wissensmanagement • Outsourcing • ...
Neue Technologien	<ul style="list-style-type: none"> • Data Warehouse • Objektorientierte Softwareentwicklung • Wissensbasierte Systeme • ...
Neue Testate	<ul style="list-style-type: none"> • Real-Time-Financial-Reporting • Internet Business Reporting • Rating • ...
Neue Prüftechniken	<ul style="list-style-type: none"> • Web-basierte Prüfung • Analytische Prüfung • Kontinuierliche Prüfung • ...

Tab. 2: Wirtschaftsprüfung und Wirtschaftsinformatik: Neue Herausforderungen
Quelle: In Anlehnung an [Brun03].

3.2 IT-Einsatz in der Rechnungslegung und deren Einfluss auf den Prüfungsprozess

Der ansteigende Einsatz der IT in der Rechnungslegung hat einen unmittelbaren Einfluss auf den Prüfungsprozess im Rahmen der Wirtschaftsprüfung gemäß §§ 316-324 HGB. Die Informationstechnik beeinflusst Abläufe, Systeme, Kontrollen und Dokumente des zu prüfenden Unternehmens. Informatik-Risiken gehören zum inhärenten Risiko und zum Kontrollrisiko und haben Einfluss auf das Aufdeckungsrisiko des Prüfers.

Das Institut der Wirtschaftsprüfer (IDW) hat einen Prüfungsstandard IDW PS 330 „Abschlussprüfung bei Einsatz der Informationstechnologie“ am 24. September 2002 veröffentlicht. Laut diesem Standard ist die IT-Prüfung ein Bestandteil der Prüfung des bereits erwähnten Internen Kontrollsystems. Interne Kontrollsysteme gewinnen in Unternehmen in den letzten Jahren immer mehr an Bedeutung. Der Begriff „IKS“ wurde bereits in den GoBS aufgeführt und erläutert: „Als IKS wird grundsätzlich die Gesamtheit aller aufeinander abgestimmten und miteinander verbundenen Kontrollen, Maßnahmen und Regelungen bezeichnet. ... Dabei reichen wegen komplexer Abläufen und Strukturen ... einzelne, voneinander isolierte Kontrollmaßnahmen keinesfalls aus. Vielmehr bedarf es einer planvollen und lückenlosen

Vorgehensweise, um ein effizientes Kontrollsystem im Unternehmen zu installieren.“. Aus den GoBS geht eindeutig hervor, dass es ohne IKS kein ordnungsmäßiges Buchführungssystem geben kann. Seit der Einführung des Gesetzes zur Kontrolle und Transparenz (KonTraG) im Jahre 1998 muss der Prüfer das Risikomanagement eines Unternehmens gesondert betrachten und beurteilen. International finden sich ähnliche rechtliche Anforderungen beispielsweise im Sarbanes-Oxley Act (SOX), einer Rechnungslegungsvorschrift für Unternehmen, die an US-Börsen gelistet sind. Interne Kontrollen haben die Aufgabe, Risiken, aber auch Chancen in einem Unternehmen frühzeitig zu erkennen. Ziele der internen Kontrollen kann man wie folgt zusammenfassen:

- Erreichung der Unternehmensziele,
- Einhaltung von Gesetzen und Vorschriften (Compliance),
- Schutz des Geschäftsvermögens,
- Verhinderung/Aufdeckung von Fehlern und Unregelmäßigkeiten,
- Sicherstellung der Zuverlässigkeit und Vollständigkeit der Buchführung und Berichterstattung.

Durch § 317 Abs. 4 HGB wurde das Überwachungssystem eines Unternehmens in die gesetzliche Abschlussprüfung aufgenommen: „Bei einer Aktiengesellschaft, die Aktien mit amtlicher Notierung ausgegeben hat, ist außerdem im Rahmen der Prüfung zu beurteilen, ob der Vorstand die ihm nach § 91 Abs. 2 AktG obliegenden Maßnahmen in einer geeigneten Form getroffen hat und ob das danach einzurichtende Überwachungssystem seine Aufgaben erfüllen kann.“ Die Überwachung betrifft die Risiken, die den Fortbestand der Gesellschaft gefährden können.

Der Einsatz der IT im Unternehmen ist nicht nur mit Chancen, sondern auch mit Risiken verbunden. Die Risiken sind auf allen Ebenen des IT-Systems vorhanden. Folgende Risikoaspekte sind in diesem Zusammenhang zu erwähnen:

- Abhängigkeit eines Unternehmens von IT,
- Änderungsgeschwindigkeit und deren Bewältigung,
- verfügbares Fachwissen und verfügbare Ressourcen,
- adäquate Ausrichtung an Geschäftserfordernissen.

Die Prüffelder gemäß IDW PS 330 sind IT-Strategie, IT-Umfeld, IT-Organisation, IT-Infrastruktur, IT-Anwendungen, IT-gestützte Geschäftsprozesse, IT-Überwachungssystem und IT-Outsourcing (s. Tabelle 3). Das System ist somit in seiner Gesamtheit zu prüfen.

Risikofeld	Abhängigkeit	Änderung	Knowhow/ Ressourcen	Geschäftliche Ausrichtung
IT-Umfeld	Dominanz der IT-Bereiche	Barrieren, Festhalten	Bewusstsein, Verständnis	Strategie, Konzeption
IT-Organisation	Betrieb, Verfügbarkeit	Projektmanagement	Aufgabenabwicklung	Richtlinien, Kompetenzen
IT-Geschäftsprozesse	Automatisierung, Komplexität	Anwenderakzeptanz	Anwenderunterstützung	Aufgabenabwicklung
IT-Anwendungen	Ausfälle, Geschäftsbetrieb	Funktionalität, Fehler	Entwicklung, Betreuung	Markt-, Benutzeranforderungen
IT-Infrastruktur	Outsourcing, Provider	Technologie, Sicherheitslücken	Strukturen, Betreuung	Inhomogenität, Sicherheit
Daten/Informationen	Umlauf, Inhalt, Aktualität	Migration, Archivierung	Auswertungen, Analysen	Entscheidungsrelevanz

Tab. 3: Risikolandschaft beim Einsatz von IT-Systemen in der Rechnungslegung
Quelle: In Anlehnung an [Hees02, 6].

So gewann die Durchführung von IT-Systemprüfungen in den vergangenen Jahren in den Unternehmen immer mehr an Bedeutung. Die IT-Systemprüfung soll Risiken aufzeigen, die eine fehlerhafte Rechnungslegung verursachen kann. Wenn ein IT-System falsch eingeführt wird, kann es sehr schnell zur Bestandsgefährdung von Daten führen. Weitere Risiken liegen im Bereich der funktionalen Datensicherung sowie im Bereich des Datenschutzes. Im Zusammenhang mit der zunehmenden Bedeutung von IT wurden mehrere Verlautbarungen vom Institut der Wirtschaftsprüfer (IDW) bzw. von dessen Fachausschuss für Informationstechnologie (FAIT) herausgegeben.

3.3 Verlautbarungen des IDW

Im Umfeld der Prüfung von IT-gestützten Systemen müssen viele spezifische Aspekte betrachtet werden [Hees02, 8]. Der Fachausschuss für Informationstechnologie (FAIT) des IDW hat folgende Verlautbarungen in diesem Zusammenhang veröffentlicht:

- Grundsätze ordnungsmäßiger Buchführung bei Einsatz von Informationstechnologie (IDW RS FAIT 1),
- Grundsätze ordnungsmäßiger Buchführung bei Einsatz von Electronic Commerce (IDW RS FAIT 2),

- Grundsätze ordnungsmäßiger Buchführung bei Einsatz elektronischer Archivierungsverfahren (IDW ERS FAIT 3), zurzeit im Entwurfsstadium.

Diese Verlautbarungen sollen eine einheitliche Prüfungsqualität gewährleisten. Sie ersetzen nicht die GoBS, die vor allem für Buchführungspflichtige und für Softwareentwickler gedacht sind, sondern wurden speziell für die besonderen Aufgaben der Wirtschaftsprüfer entwickelt. Die GoBS und IDW RS FAIT 1 weisen inhaltlich viele Überschneidungen auf. Die Ausführungen von FAIT 1 enthalten zusätzlich allgemeine Informationen bzgl. des Einsatzes von IT-Systemen im Unternehmen und speziell eines solchen Einsatzes mit Rechnungslegungsbezug. Das nach dem Rechnungslegungsstandard FAIT 1 definierte IT-System umfasst die Bereiche Infrastruktur, Anwendungen und IT-gestützte Geschäftsprozesse, somit die gesamte Bandbreite des IT-Einsatzes. Besondere Aufmerksamkeit wird dem IT-Kontrollsystem gewidmet. Dieses System umfasst sämtliche Bereiche und beinhaltet Grundsätze, Verfahren sowie Maßnahmen zur Bewältigung der sich aus dem Einsatz der IT im Unternehmen allgemein ergebenden Risiken. Die im FAIT 1 festgelegten Maßnahmen enthalten Regelungen für den IT-Einsatz selbst sowie für die Überwachung der Einhaltung dieser Regelungen.

Für die Einhaltung dieser Vorschriften sowie die Sicherheit der im Systembetrieb verarbeiteten Daten ist die Geschäftsführung verantwortlich. Aus diesem Grund muss ein geeignetes Sicherheitskonzept bzw. -management festgelegt und regelmäßig auf seine Wirksamkeit hin untersucht werden. Besondere Aufmerksamkeit wird drei Begriffen gewidmet [IDW FAIT, Tz. 23]:

- Autorisierung (Anforderungen an den Zugriffsschutz),
- Authentizität (Anforderungen an Berechtigungen und eindeutige Identifikation),
- Verbindlichkeit (Anforderung an die Herbeiführung bindender Rechtsfolgen).

Die Aufgabe des Sicherheitsmanagements ist die Gewährleistung des gesetzlich vorgeschriebenen Grades an Informationssicherheit. Ein solches Sicherheitskonzept setzt seinerseits ein entsprechendes Risikomanagementsystem voraus, in dem die bestehenden Risiken des IT-Einsatzes realistisch bewertet werden.

Im September 2003 wurde eine ergänzende Stellungnahme zu den aus E-Commerce-Systemen erwachsenden Anforderungen herausgegeben (IDW RS FAIT 2). Diese verdeutlicht die im IDW RS FAIT 1 dargestellten Ordnungsmäßigkeits- und Sicherheitsanforderungen im Bereich von E-Commerce. Darüber hinaus beinhaltet IDW RS FAIT 2 ergänzende Anforde-

rungen im Bezug auf besondere IT-Risiken, die mit dem Einsatz von E-Commerce-Systemen zusammenhängen.

IDW ERS FAIT 3 enthält neben der Auslegung der rechtlichen Vorschriften im Zusammenhang mit digitalen Unterlagen und deren Archivierung (die wichtigste rechtliche Quelle stellen GDPdU dar) die Beschreibung des technischen und organisatorischen Umfelds des elektronischen Archivierungsverfahrens. Es werden ausführliche Anweisungen im Zusammenhang mit der Einrichtung eines Archivierungssystems gegeben. Das Dokument knüpft an die FAIT 1 und andere Richtlinien des IDW an. Es geht detailliert auf die Anforderungen vom HGB, AO, GoBS und auch der GDPdU ein. Dabei werden unterschiedliche Archivierungsverfahren, der Charakter von elektronischen Dokumenten, die technische Infrastruktur und auch das Thema Outsourcing behandelt. Ein weiteres wichtiges Thema des Entwurfs stellt die Migration dar. IDW ERS FAIT 3 ist zurzeit als Entwurf zur Diskussion gestellt. Anzumerken ist, dass die Richtlinie mehrere Überschneidungen mit Anweisungen wie dem Code of Practice "Grundsätze der elektronischen Archivierung" des VOI Verband Organisations- und Informationssysteme e.V. (<http://www.voi.de>), dem Grundschutzhandbuch, Kapitel 9, des BSI Bundesamt für Sicherheit in der Informationstechnik (<http://www.bsi.de>) und anderen Dokumenten aufweist. Die AWW Arbeitsgemeinschaft für wirtschaftliche Verwaltung e.V. überarbeitet zurzeit die GoBS [oV06]. Da auch in den neuen GoBS die gleichen Themen behandelt werden, kann man mit Redundanzen und Divergenzen rechnen. Die Vielfalt der Richtlinien und Codes of Best Practice macht die Orientierung schwierig [oV05a]. Ein einheitlicher, prüfbarer Kriterienkatalog für Archivsysteme ist bis heute nicht vorhanden.

Zurzeit befinden sich in der Diskussion bzw. in der Erarbeitung außerdem noch folgende Verlautbarungen:

- IDW-Prüfungsstandard „Grundsätze ordnungsmäßiger Durchführung von Datenschutzaudits gem. § 9a Bundesdatenschutzgesetz (BDSG)“ (IDW PS 870),
- IDW-Prüfungshinweis „Die Prüfung von IT-gestützten Geschäftsprozessen (IDW PH 9.330.x) in Ergänzung zum IDW-Prüfungsstandard „Abschlussprüfung bei Einsatz von Informationstechnologie“ (IDW PS 330),
- Überarbeitung der HFA-Stellungnahme 4/1997 „Projektbegleitende Prüfung EDV-gestützter Systeme“,
- Diskussionspapier „IT-Risikomanagement“.

Das Ziel projektbegleitender Prüfungen ist sicherzustellen, dass neu entwickelte, geänderte oder erweiterte IT-gestützte Rechnungslegungssysteme den GoB entsprechen. Diese Verlautbarung ist deswegen sehr wichtig, weil Compliance-Anforderungen (SOX, Basel II, Solvency II etc.) und deren Umsetzung in die IT eines Unternehmens an Bedeutung gewinnen. Die Herausforderung für die DV liegt in der Umsetzung allgemeiner Business-Anforderungen in die IT. Viele Regelungen enthalten diesbezüglich wenig konkrete Einleitungen. Für die Wirtschaftsprüfung ergibt sich die Aufgabe, gesetzeskritische DV-Kontrollen zu identifizieren.

4 Aktuelle Herausforderungen

4.1 Weiterentwicklung von GoBS

Die GoBS wurden zuletzt im Jahre 1995 aktualisiert. An der neuen Version wird in der Arbeitsgemeinschaft für wirtschaftliche Verwaltung gearbeitet. Da GoBS und IDW RS FAIT 1 inhaltlich viele Überschneidungen aufweisen, empfiehlt sich die Berücksichtigung von IDW RS FAIT 1 auch für diejenigen, die nicht als Prüfer tätig sind.

In der Arbeitsgemeinschaft für wirtschaftliche Verwaltung wurde außerdem ein neuer Arbeitskreis geschaffen: "Auslegung der GoB beim Einsatz neuer Organisationstechnologien". Im Mittelpunkt der Tätigkeit dieses Arbeitskreises steht die Untersuchung und Bewertung der Vereinbarkeit von neuen Technologieformen mit den GoB.

4.2 Neue Einsatzfelder von Informationssystemen in der Rechnungslegung

In den letzten Jahren kann man einen zunehmenden Einsatz von Informationssystemen für das interne und externe Reporting sowie das Risikomanagement beobachten. Laut einer Studie des Business Application Research Center (BARC) verfügen bereits zwei Drittel der Unternehmen über ein durchgängiges Reporting bis hin zum Vorstand. Eine Schwachstelle stellen allerdings laut Umfragen die langen Berichtswege dar. In den deutschen Unternehmen dauert es durchschnittlich fünf Tage, bis Informationen über wesentliche Risiken beim Vorstand ankommen [oV05b].

Die Studie belegt auch, dass eine erhebliche Heterogenität der in der Rechnungslegung eingesetzten Systeme sowie eine große Anzahl von insgesamt verwendeten Informationssystemen zu beobachten sind. Durchschnittlich sind in den deutschen Firmen acht unterschiedliche Informa-

tionssysteme, die Daten für Reporting, Planung, Konsolidierung und Risikomanagement liefern, im Einsatz. Für eine gute Performance spielt ein großer Integrationsgrad eine wichtige Rolle.

4.3 IT-Systeme in der Rechnungslegung und Prozessgedanke in Unternehmen

Die durch GoBS und GDPdU vorgeschriebene Verfahrensdokumentation fördert Prozessgedanken in Unternehmen. Mit dem Sarbanes-Oxley Act (SOX) wurde diese Richtung noch mehr gestärkt. Der SOX wurde als US-Gesetz im Jahre 2002 mit dem Ziel verabschiedet, durch eine bessere Berichterstattung die Informationsversorgung von Anlegern zu verbessern. Unternehmensprozesse müssen definiert und beschrieben werden und entsprechende Kontrollmechanismen sind zu implementieren. SOX ist für die deutschen Unternehmen relevant, wenn sie in den USA börsennotiert sind. Mit den Vorschriften von SOX sollen sich aber auch die nicht betroffenen Unternehmen auseinandersetzen, weil die 8. EU-Richtlinie in Kürze in Kraft treten soll. Sie wird als Antwort der Europäischen Union auf den amerikanischen SOX gesehen. Deswegen bezeichnet man sie auch als „SOX für Europa“. Konsequenzen der 8. EU-Richtlinie für die europäischen Unternehmen sind ähnlich wie die Konsequenzen von SOX für die amerikanischen oder in Amerika börsennotierten Unternehmen.

Durch diese Entwicklungen gewinnt die Nachvollziehbarkeit der Dokumentenverwendung immer mehr an Bedeutung. Durch eine umfangreichere Protokollierung werden die Informationen zur Verfügung gestellt, die zur Verbesserung betrieblicher Abläufe verwendet werden können. Durch die Protokollierung können Engpässe und doppelte oder nicht genau zugeordnete Tätigkeiten aufgedeckt werden [Böhn06, 32].

Durch diese Entwicklungen gewinnen Dokumentenmanagement-Systeme bzw. Enterprise-Content-Management-Systeme immer mehr an Bedeutung. Der Teilbereich Records Management steuert Archivsysteme eines Unternehmens. Neben der Deklaration des Dokumentes als unveränderbar werden weitergehende Informationen über das Dokument selbst (wie Status und Version) sowie über die damit verbundenen Prozesse (z. B. Weiterleitung) in den meisten Fällen automatisch mitgespeichert.

4.4 IT-Compliance

Neue gesetzliche Anforderungen im Bereich der Rechnungslegung sorgen für neue Sichtweisen auf rechnungslegungsrelevante DV-Systeme. Ab 1. Januar 2005 müssen alle börsennotierten europäischen Unternehmen und ihre Tochtergesellschaften die Abschlüsse nach internationalem Rechnungslegungsstandard IFRS (International Financial Reporting Standards) veröffentlichen.

Auch neue Compliance-Regelungen (SOX, Basel II, Solvency II etc.) sorgen für zunehmende Komplexität der Rechnungslegung eines Unternehmens. Für einige Unternehmen wird das in der Konsequenz zu einer Softwareumstellung führen. Die Anforderungen an die DV bei der Einführung der neuen Bilanzierungsregeln sind vielschichtig. Die Software muss zukünftig auch parallele Konten und Buchungskreise zulassen und verarbeiten können, denn deutsche Unternehmen müssen weiterhin eine lokale Bilanz gemäß HGB vorlegen. Dabei kommt es jedoch zu einer vollkommen unterschiedlichen Bewertung der Daten.

Die Rechnungslegung ist in der Regel auf verschiedenen IT-Komponenten basiert. Die neuen Gesetze erfordern eine Bestandsaufnahme und IT-Systeme in umfangreichem Ausmaß: eine umfassende Konsolidierungsrechnung der verschiedenen Segmente, ein erweiterter Kostenplan aufgrund der erhöhten Offenlegungssegmente etc [ScKr05, 89]. Es soll geklärt werden, inwieweit diese Anforderungen mit den vorhandenen Systemen erfüllt werden können. Der Aufwand zur Rechnungslegung wird in der Zukunft deutlich über dem Aufwand der Vergangenheit liegen. In der neuen Version von GoBS sollen deswegen nicht nur die handelsrechtlichen Vorschriften berücksichtigt werden, sondern auch die Vorschriften der internationalen Rechnungslegung.

Mit dem Inkrafttreten von neuen Gesetzen spielt die Richtigkeit der vorzulegenden Finanzberichte eine noch größere Rolle als bisher. Die Haftungs- und Sorgfaltspflichten des Managements im Bezug auf Bilanzierung und Transparenz sind deutlich strenger geworden. Dadurch steigt die Notwendigkeit, die Angemessenheit und Ordnungsmäßigkeit sowohl der Systementwicklung als auch des Systembetriebs nachzuweisen. Damit unmittelbar verbunden sind die steigenden Anforderungen im Bereich Risikomanagement [JoGo06, 10].

5 Schlussbemerkung und Ausblick

Die Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme befinden sich momentan in einem Überarbeitungsprozess. An einer neuen Version von GoBS wird im Rahmen eines Projektes der Arbeitsgemeinschaft für wirtschaftliche Verwaltung gearbeitet. Bei der Umsetzung der GoBS in der betrieblichen Praxis und deren Kontrolle durch die Wirtschaftsprüfer stellte sich heraus, dass ein besonderer Handlungsbedarf in den Bereichen der Erstellung und Pflege von Verfahrensdokumentation besteht [Phil98, 317]. Eine weitere wichtige Herausforde-

nung besteht in der Berücksichtigung der internationalen Rechnungslegungsvorschriften bei der Ausarbeitung der neuen GoBS-Version.

Wie bereits erläutert, gelten die GoBS für alle direkt oder indirekt rechnungslegungsrelevanten DV-Systeme im Unternehmen. Der zunehmende Einsatz der Informationstechnik in der Rechnungslegung bedeutet auch eine neue Sichtweise auf den Beruf des Wirtschaftsprüfers. Die Wirtschaftsprüfer müssen sich mit den Informationssystemen, die in Unternehmen eingesetzt werden, gut auskennen, um die entsprechenden IT-Systemprüfungen durchführen zu können. Die Wirtschaftsprüfer stehen vor neuen Herausforderungen, da die Systeme noch komplexer werden.

Es gibt immer mehr Unternehmen mit einer geschäftsprozessorientierten Organisation. Diese Tendenz wird nicht zuletzt dadurch unterstützt, dass Informationssysteme in weiteren Unternehmensbereichen eingesetzt werden und die Vernetzung einzelner Unternehmensteile fördern. Die IT-unterstützte Buchführung zeigt, dass Buchführung immer mehr als integrierte Unternehmensfunktion betrachtet wird. Die engere Verknüpfung einzelner Unternehmensfunktionen sorgt für neue Herausforderungen bei den Wirtschaftsprüfern. Die Beurteilung der ordnungsmäßigen Abbildung von Geschäftsprozessen und deren Darstellung in der Rechnungslegung gewinnt weiter an Bedeutung [BeVo03, 1233 ff.].

Literaturverzeichnis

- [BeVo03] Berenz, Bernd; Voit, Franz: Die Geschäftsprozessorientierung in der Abschlussprüfung. In: Die Wirtschaftsprüfung 56 (2003) 22, S. 1233-1243.
- [Böhn06] Böhn, Martin: Compliance schafft Prozesstransparenz. In: is report 10 (2006) 4, S. 30-33.
- [Brun03] Brun, Jürg: Corporate Governance und IT-Revision. Uni Zürich 2003. http://www.ifi.unizh.ch/ifiadmin/staff/maurer/Dokumente/Web_SS03/Brun/C%20Governance.pdf, Abruf am 2006-06-13.
- [EIDW05] Entwurf IDW Stellungnahme zur Rechnungslegung: Grundsätze ordnungsmäßiger Buchführung bei Einsatz elektronischer Archivierungsverfahren (IDW ERS FAIT 3). In: Die Wirtschaftsprüfung 58 (2005) 13, S. 746-754.

- [Grun01] Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen. In: Betrieb und Wirtschaft 55 (2001) 18, S. 773-776.
- [Hees02] Heese, Klaus: IT-Systemprüfungen im Rahmen einer risiko- und prozessorientierten Prüfungsstrategie. PwC. <http://www.pwc.com/extweb/pwcpublishations.nsf/DocID/A478650D4955530585256BC6002D16D2>, 4/2002, Abruf am 2006-06-08.
- [IDWS02] IDW Stellungnahme zur Rechnungslegung: Grundsätze ordnungsmäßiger Buchführung bei Einsatz von Informationstechnologie (IDW RS FAIT 1). In: Die Wirtschaftsprüfung 55 (2002) 21, S. 1157-1167.
- [IDWP02] IDW Prüfungsstandard: Abschlußprüfung bei Einsatz der Informationstechnologie (IDW PS 330). In: Die Wirtschaftsprüfung 55 (2002) 21, S. 1167-1179.
- [IDWS03] IDW Stellungnahme zur Rechnungslegung: Grundsätze ordnungsmäßiger Buchführung bei Einsatz von Electronic Commerce (IDW RS FAIT 2). In: Die Wirtschaftsprüfung 56 (2003) 22, S. 1258-1276.
- [JoGo06] Johanssen, W.; Goeken, M.: IT-Governance – neue Aufgaben des IT-Managements. In: HMD Praxis der Wirtschaftsinformatik 43 (2006) 250, S. 7-20.
- [MaQu03] MacKee, Thomas E.; Quick, Reiner: IT-Kenntnisse der wirtschaftsprüfenden Berufsstände. Eine empirische Untersuchung. In: Die Wirtschaftsprüfung 56 (2003) 10, S. 541-547.
- [oV04] o. V.: Prüfungsverordnung für Wirtschaftsprüfer nach §§ 14 und 131I der Wirtschaftsprüferordnung (Wirtschaftsprüferprüfungsverordnung – WiPrPrüfV) vom 20. Juli 2004 (BGBl. I S. 1707). <http://www.wpk.de/pdf/WiPrPruefV.pdf>, Abruf am 2006-07-05.
- [oV05a] o. V.: IDW FAIT ERS 3 PROJECT CONSULT Kommentar, 2005. <http://partner.coextant.com/website/project.nsf/RSDatumWWWScriptGraphics/B80388E546013DFE412570B000330615>, Abruf am 2006-06-13.

- [oV05b] o. V.: IT treibt die Corporate Governance. In: Computerwoche 18 (2005). <http://www.computerwoche.de/heftarchiv/2005/18/1051394/>, Abruf am 2006-11-07.
- [oV06] o.V: Arbeitsgemeinschaft für wirtschaftliche Verwaltung überarbeitet GoBS. Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. (BITKOM) 2006. http://www.bitkom.org/de/themen_gremien/37105_9754.aspx, Abruf am 2006-06-13.
- [Phil98] Philipp, Mathias: Ordnungsmäßige Informationssysteme im Zeitablauf – Umsetzung der GoBS im Informationssystem-Lebenszyklus. In: Wirtschaftsinformatik 40 (1998) 3, S. 312-317.
- [StHa05] Stahlknecht, Peter; Hasenkamp, Ulrich: Einführung in die Wirtschaftsinformatik. 11. Aufl., Springer, Berlin, Heidelberg 2005.
- [ScKr05] Schön, Dietmar; Kröninger, Linda: Rechnungslegung im Umbruch – Was bei der Umstellung auf IAS/IFRS zu beachten ist. In: Controlling (2005) 2, S. 85-91.
- [Zepf96] Zepf, Günter: Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme. Erläuterungen zu den GoBS für die Praxis. In: DStR (1996) 32, S. 1259-1263.

Autorenverzeichnis

(Band 1 und 2)

Sebastian **Abeck**

abeck@cm-tm.uka.de

Sven **Abels**

sabels@acm.org

Maximilian **Ahrens**

maximilian.ahrens@telekom.de

Stephan **Aier**

stephan.aier@unisg.ch

Andreas **Albers**

andreas.albers@m-lehrstuhl.de

Paul **Alpar**

alpar@staff.uni-marburg.de

Anupriya **Ankolekar**

aan@aifb.uni-karlsruhe.de

Felix-Robinson **Aschoff**

aschoff@ifi.unizh.ch

Joerg **Aßmann**

assmann@ieb.net

Eckardt **Augenstein**

e.augenstein@perpendo.de

Tilmann **Bartels**

bartels@sysedv.tu-berlin.de

Thomas **Barth**

barth@fb5.uni-siegen.de

Bernhard **Beck**

Jörg **Becker**

isjobe@wi.uni-muenster.de

Ralph **Bergmann**

bergmann@uni-trier.de

Matthias **Biggeleben**

biggeleben@wiwi.uni-frankfurt.de

Andreas **Birkhofer**

abirkhof@rumms.uni-mannheim.de

Annette **Bobrik**

bobrik@sysedv.tu-berlin.de

Freimut **Bodendorf**

Bodendorf@wiso.uni-erlangen.de

Oliver **Bohl**

bohl@inf.wirtschaft.uni-kassel.de

Tilo **Böhm**

boehmann@in.tum.de

Reinhard **Brandl**

reinhard@brandl-eitensheim.de

Juergen **Branke**

branke@aifb.uni-karlsruhe.de

Oliver **Braun**

ob@itm.uni-sb.de

Jan vom **Brocke**

jan.vom.brocke@ercis.de

Eva-Maria **Bruch**

ena@de.festo.com

Tobias **Bucher**

tobias.bucher@unisg.ch

Sabine **Buckl**

buckls@in.tum.de

Christian **Buddendick**

christian.buddendick@ercis.de

Ulrich **Buhl**

Hans-

ulrich.buhl@wiwi.uniaugsburg.de

Peter **Buxmann**

buxmann@is.tu-darmstadt.de

Gary Carr-Smith
Gary.Carr-Smith@avonfire.gov.uk

Michele Catalano
catalano@dea.unian.it

Pavlina Chikova
pavlina.chikova@iwi.dfki.de

Achim Dannecker
Achim.Dannecker@unibw.de

Sina Deibert
deibert@uni-mannheim.de

Patrick Delfmann
delfmann@ercis.de

Ben Diamond
Ben.Diamond@wmfs.ne

Kai Dingel
dingel@wiwi.hu-berlin.de

Ulrike Dowie
dowie@wi.uni-stuttgart.de

Michael Durica
mdurica@web.de

Helmuth Elsner
helmuth.elsner@iism.uni-karlsruhe.de

Bartholomäus Ende
Ende@wiwi.uni-frankfurt.de

Alexander M. Ernst
ernst@in.tum.de

Torsten Eymann
eymann@uni-bayreuth.de

Ulrich Faisst
ulrich.faisst@bain.com

Otto K. Ferstl
otto.ferstl@wiai.uni-bamberg.de

René Fiege
rene.fiege@tu-ilmenau.de

Andreas Fink
andreas.fink@hsu-hamburg.de

Ronny Fischer
ronny.fischer@unisg.ch

Elgar Fleisch
elgar.fleisch@unisg.ch

Jochen Frank
jochen.frank@wiai.uni-bamberg.de

Ulrich Frank
ulrich.frank@uni-duisburg-essen.de

Bernd Freisleben
freisleb@informatik.uni-marburg.de

Andrea Freßmann
fressmann@uni-trier.de

Swantje Friedrich
friedrich@wiwiss.fu-berlin.de

Thomas Friese
th.friese@siemens.com

Lothar Fritsch
lothar.fritsch@m-lehrstuhl.de

Armin Fügenschuh
fuegenschuh@mathematik.tu-darmstadt.de

Roland Gabriel
rgabriel@winf.rub.de

Martin Gersch
martin.gersch@rub.de

Khaled Ghédira
khaled.ghedira}@isg.rnu.tn

Gianfranco Giulioni
giulioni@dea.unian.it

Daniel Gmach
gmach@in.tum.de

Martin Gneiser
Martin.Gneiser@wiwi.uni-augsburg.de

Peter Gomber
gomber@wiwi.uni-frankfurt.de

Simone Göttlich
goettlich@mathematik.uni-kl.de

Kai-Uwe Götzelt
kai.goetzelt@wiso.uni-erlangen.de

Manfred Grauer
grauer@fb5.uni-siegen.de

Guido Grohmann
guido.grohmann@im-c.de

Roger Gruetter
rgruetter@rgruetter.ch

Markus Gsell
gsell@wiwi.uni-frankfurt.de

Oleg Gujo
gujo@wiwi.uni-frankfurt.de

Michael J. Gundlach
mjj1052@garnet.acns.fsu.edu

Wolfgang Hackenbroch
wolfgang.hackenbroch@wiwi.uni-augsburg.de

Ulrich Hasenkamp
hasenkamp@wiwi.uni-marburg.de

Matthias Henneberger
matthias.henneberger@wiwi.uni-augsburg.de

Andreas Henrich
andreas.henrich@wiai.uni-bamberg.de

Martin Hepp
mhepp@computer.org

Stefan Herbergs
herbergs@lth.rwth-aachen.de

Michael Herty
herty@mathematik.uni-kl.de

Michael A. Herzog
herzog@fhtw-berlin.de

Georg Herzwurm
herzwurm@wi.uni-stuttgart.de

Thomas Hess
thess@bwl.uni-muenchen.de

Lutz Heuser
lutz.heuser@sap.com

Konrad Hilbers
k.hilbers@hse24.de

Oliver Hinz
ohinz@wiwi.uni-frankfurt.de

Mohamed Hmiden
hmiden_mohamed@yahoo.fr

Holger Hoffmann
holger.hoffmann@in.tum.de

Peter Hruschka
hruschka@b-agile.de

Steffi Jahner
jahner@in.tum.de

Christian Janiesch
ischja@wi.uni-muenster.de

Matthias Jarke
jarke@informatik.rwth-aachen.de

Nicholas R. Jennings
nrj@ecs.soton.ac.uk

Martin Jetter
mjetter@de.ibm.com

Liviu Joita
l.joita@cs.cardiff.ac.uk

Wolfgang König
jfranke@wiwi.uni-frankfurt.de

Timo **Kahl**
timo.kahl@iwi.dfki.de

Carolin **Kaiser**
carolin.kaiser@wiso.uni-erlangen.de

Oliver **Kamin**
kamin@ioeb.de

Alfons **Kemper**
kemper@in.tum.de

Hartmut **Klemens**
hartmut.klemens@klemens.com

Natalia **Kliewer**
kliewer@upb.de

Ralf **Knackstedt**
israkn@wi.uni-muenster.de

Uta Franziska **Knebel**
knebel@in.tum.de

Karl **Knispel**
knispel@ioeb.de

Michael **Koch**
kochm@informatik.tu-muenchen.de

Stefan **Koch**
Stefan.Koch@wu-wien.ac.at

Michael **Kohl**
michael.kohl@wiai.uni-bamberg.de

Dorothee **Korn**
dorotheekorn@uni-muenster.de

Elizaveta **Kozlova**
kozlova@wiwi.uni-marburg.de

Wolfgang **Kraemer**
wolfgang.kraemer@im-c.de

Hanna **Krasnova**
krasnovh@wiwi.hu-berlin.de

Helmut **Krcmar**
Krcmar@in.tum.de

Falk **Kretzschmar**
kretzschmar@wiwi.uni-halle.de

Nina **Kreyer**
Nina.Kreyer@uk.ibm.com

Stefan **Kühne**
kuehne@informatik.uni-leipzig.de

Holger **Kühnl**
hk@synspace.com

Dennis **Kundisch**
dennis.kundisch@vwl.uni-freiburg.de

Stefan **Lacher**
stefan.lacher@sap.com

Steffen **Lamparter**
sla@aifb.uni-karlsruhe.de

Josef **Lankes**

Volker **Lanninger**
lanninger@bisor.de

Christoph **Lattemann**
lattema@uni-potsdam.de

Ulrike **Lechner**
Ulrike.Lechner@unibw.de

Lasse **Lehman**
llehmann@KOM.tu-darmstadt.de

Jan **Lehmann**
lehmann@dvz-mv.de

Johann **Leibig**
johann.leibig@hvbsystems.com

Jan Marco **Leimeister**
leimeister@in.tum.de

Katrina **Leyking**
katrina.leyking@iwi.dfki.de

Andreas **Löber**
loeber@ifi.unizh.ch

Peter Loos
Peter.Loos@iwi.dfki.de

Michael Luther
michael_luther@hotmail.com

Christian Lütke Entrup
luetke-entrup@fb5.uni-siegen.de

Josef Maier
josef.maier@tu-berlin.de

Robert Malec
malec@cm-tm.uka.de

Shakib Manouchehri
manouchehri@wirtschaft.uni-kassel.de

Thomas Matheis
thomas.matheis@iwi.dfki.de

D. Harrison McKnight
mcknight@bus.msu.edu

Markus Meister
Markus.Meister@chiru.med.uni-giessen.de

Jan Mendling
jan.mendling@wu-wien.ac.at

Achim Michel-Backofen
achim.michel-backofen@akad.uniklinikum-giessen.de

Daniel Michelis
daniel.michelis@unisg.ch

Frank Milius
frank.milius@im-c.de

Matthias Mohr
mohrm@in.tum.de

Christof Momm
momm@cm-tm.uka.de

Karlheinz Morgenroth
karlheinz.morgenroth@wiai.uni-bamberg.de

Kathrin Möslein
kathrin.moeslein@hhl.de

Felix R. Müller
felix.mueller@wiwi.uni-regensburg.de

Felix Müller-Wienbergen
felix.mueller-wienbergen@ercis.de

Thomas Myrach
thomas.myrach}@iwi.unibe.ch

Thomas Neubauer
neubauer@ifs.tuwien.ac.at

Valentin Nicolescu
valentin.nicolescu@in.tum.de

Björn Niehaves
bjoern.niehaves@ercis.uni-muenster.de

Jasminko Novak
novak@ifi.unizh.ch

Markus Nüttgens
markus.nuettgens@wiso.uni-hamburg.de

Philipp Offermann
Philipp.Offermann@syledv.tu-berlin.de

Sebastian Olbrich
sebastian.olbrich@arcor.de

Dirk Pawlaszczyk
dirk.pawlaszczyk@tu-ilmenau.de

Martin Pellengahr
martin.pellengahr@wi.uni-muenster.de

Simone Perlmann
simone.perlmann@sap.com

Mathias Petsch
mathias.petsch@tu-ilmenau.de

Volkmar Pipek
volkmar.pipek@uni-siegen.de

Key Pousttchi
key.pousttchi@wiwi.uni-augsburg.de

Jörg Purucker
joerg.purucker@wiso.uni-erlangen.de

Mike Radmacher
mike.radmacher@m-lehrstuhl.de

Omer F. Rana
o.f.rana@cs.cardiff.ac.uk

Kai Rannenber
kai.rannenber@m-lehrstuhl.de

Julian Reichwald
reichwald@fb5.uni-siegen.de

Claudia Reininger
claudia.reininger@t-online.de

Annette Renz

Florian Resatsch
resatsch@ieb.net

Kai Riemer
wikari@wi.uni-muenster.de

Peter Rittgen
peter.rittgen@hb.se

Susanne Robra-Bissantz
robra@wiso.uni-erlangen.de

Rainer Röhrig
rainer.roehrig@chiru.med.uni-giessen.de

Dumitru Roman
dumitru.roman@deri.org

Thomas Rose
thomas.rose@fit.fraunhofer.de

Heiko Roßnagel
heiko.rossnagel@m-lehrstuhl.de

Matthias Rothensee
rothensm@cms.hu-berlin.de

Franz Rothlauf
rothlauf@uni-mannheim.de

Werner Rotzoll
rotzoll@dvz-mv.de

Thorsten Rüffer
thorsten.rueffer@gfd.de

Lamjed Said
lamjed.bensaid@isg.rnu.tn

Michael Samtleben
samtleben@bwl.uni-muenchen.de

Nicola Schackmann

Walter Schäfer

Michael Schermann
michael.schermann@in.tum.de

Tobias Scherner
tobias.scherner@m-lehrstuhl.de

Günter Schicker
guenter.schicker@wiso.uni-erlangen.de

Thomas Schildhauer
schildhauer@ieb.net

Michael Schlicker
michael.schlicker@interactive-software.de

Christian Schmidt
schmidt@is.tu-darmstadt.de

Kathrin Schneider
kathrin.schneider@bmw.de

Hagen Schorcht
hagen.schorcht@tu-ilmenau.de

Nina Schroeder
nina.schroeder@o2.com

Christian Schröpfer
Christian.Schroepfer@sysedv.tu-berlin.de

Gerhard Schwabe
schwabe@ifi.unizh.ch

Frank Schwartz
fs@econ.uni-hamburg.de

Christian M. Schweda
schweda@in.tum.de

Michael Schwind
schwind@wiwi.uni-frankfurt.de

Martin Sedlmayr
martin.sedlmayr@fit.fraunhofer.de

Stefan Seidel
seidel@ercis.de

Eva Peggy Sekatzek
peggy.sekatzek@bmw.de

Carlo Simon
simon@uni-koblenz.de

Matthew Smith
matthew@informatik.uni-marburg.de

Zbynek Sokolovsky
zbynek.sokolovsky@ied-consulting.de

Thomas Spengler
t.spengler@tu-braunschweig.de

Sarah Spiekermann
sspiek@wiwi.hu-berlin.de

Mark Srite
msrite@uwm.edu

Gernot Starke
gs@gernotstarke.de

Ingmar Steinzen
steinzen@dsor.de

Dirk Stelzer
dirk.stelzer@tu-ilmenau.de

Gunnar Stevens
gunnar.stevens@uni-siegen.de
Tim Stockheim
stockheim@bisor.de

Stefan Strecker
stefan.strecker@uni-due.de

Niels Streekmann
streekmann@offis.de

Werner Streitberger
werner.streitberger@uni-bayreuth.de

Tobias Stuber

Christian Stummer
christian.stummer@univie.ac.at

Uwe H. Suhl
suhl@wiwiss.fu-berlin.de

Lena Suhl
suhl@dsor.de

Brian Taylor
btaylor@fireservicecollege.ac.uk

Jason Bennett Thatcher
jthatch@clemsun.edu

Frédéric Georges Thiesse
frederic.thiesse@unisg.ch

Heiko Thimm
heiko.thimm@fh-kiel.de

Oliver Thomas
oliver.thomas@iwi.dfki.de

Maik Thränert
thraenert@informatik.uni-leipzig.de

Bettina Thurnher
bettina.thurnher@wi-mobile.de

Matthias Trier
trier@sypedv.cs.tu-berlin.de

Mathias Uslar
uslar@offis.de

Hal R. Varian
hal@sims.berkeley.edu

Daniel Veit
veit@uni-mannheim.de

Hong Tuan Kiet Vo
kiet.vo@iism.uni-karlsruhe.de

Christian Völcker
cv@synspace.com

Thomas Volling
t.volling@tu-bs.de

Jan vom Brocke
jan.vom.brocke@ercis.de

Stefan Voß
stefan.voss@uni-hamburg.de

Jens Vykoukal
jens.vykoukal@wiwi.uni-frankfurt.de

Konrad Walser
konrad.walser@iwi.unibe.ch

Philipp Walter
philipp.walter@iwi.dfki.de

Veronika Waue
veronika@waue.net

Peter Weber
peter.weber@rub.de

Kilian Weiß

Oliver Wendt

Thomas Wilde
wilde@bwl.uni-muenchen.de

Martin Wimmer
wimmerma@in.tum.de

Udo Winand
winand@wirtschaft.uni-kassel.de

Robert Winter
Robert.Winter@unisg.ch

Markus Withopf

Petra Wolf
petra.wolf@in.tum.de

Adrian Wranik
wranik@wiwi.uni-frankfurt.de

Gregor Wrobel
wrobel@gfai.de

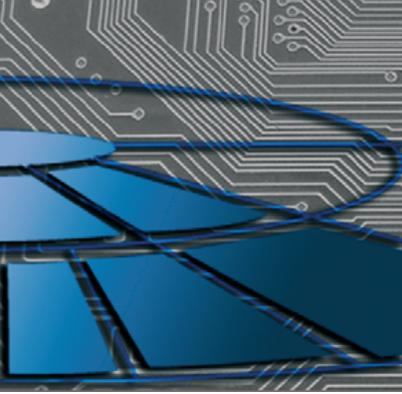
Kim Wuellenweber
wuellenweber@wiwi.uni-frankfurt.de

Volker Wulf
volker.wulf@uni-siegen.de

Jan Zibuschka
jan.zibuschka@m-lehrstuhl.de

Jörg Ziemann
joerg.ziemann@iwi.dfki.de

Volker Zimmermann
volker.zimmermann@im-c.de



“eOrganisation: Service-, Prozess-, Market-Engineering“ ist das Leitthema der 8. internationalen Tagung Wirtschaftsinformatik 2007 (WI 2007). Die Beiträge der WI 2007 greifen folgende Schwerpunkte auf:

Service-Engineering:

- Das Internet der Dinge
- eGovernment
- Information Services in der Logistik
- Mobile Business und Communications
- Outsourcing und IT-Governance
- Product Life Cycle Management in Unternehmenssoftwaresystemen
- Semantic Web
- Service Oriented Computing
- Verkehr und Mobilität

Prozess-Engineering:

- Business Intelligence
- Business Process Engineering
- Collaborative Business
- Customer und Supplier Relationship Management
- eHealth
- eLearning
- IS-Architekturen
- IT-Risikomanagement
- Modellierung als Innovationsmotor
- Softwareprozessverbesserung
- Supply Chain Management
- Operations Research
- Wissensmanagement

Market-Engineering:

- Agenten- und Multiagenten-Technologien für betriebliche Anwendungen
- eFinance
- eLearning Geschäftsmodelle
- Elektronische Märkte und elektronische Verhandlungen
- eMedia
- eServices
- Informationsmanagement in der Energie- und Umweltinformatik
- Internetökonomie
- Rechtsfragen der Informationsgesellschaft

Die Tagungsbände richten sich an Wissenschaftler und Praktiker, und bieten einen Einblick in aktuelle Forschungsfelder der Wirtschaftsinformatik

ISBN: 978-3-86644-095-1 (Band 2)

ISBN: 978-3-86644-093-7 (Set)

www.uvka.de