

Workshop Ontologie-basiertes Wissensmanagement (WOW 2003)

Workshop auf der 2. Konferenz Professionelles Wissensmanagement -
Erfahrungen und Visionen (WM2003), Luzern (Schweiz), 2.-4.4.2003

York Sure
Institut AIFB, Universität Karlsruhe
sure@aifb.uni-karlsruhe.de

Hans-Peter Schnurr
Ontoprise GmbH
schnurr@ontoprise.de

Thema

Wissensmanagement hat sich in den vergangenen Jahren zu einem kritischen Erfolgsfaktor für Unternehmen entwickelt. Die Globalisierung der Märkte, das Entstehen virtueller Unternehmen, die stärkere Kundenorientierung oder die zunehmende Komplexität von Produkten sind einige der Gründe, weshalb das systematische und gezielte Managen von Wissen immer mehr an Bedeutung gewinnt. Wissensmanagement ist offensichtlich inhärent interdisziplinär: Personalmanagement, Unternehmensorganisation sowie Unternehmenskultur sind wesentliche Aspekte des Wissensmanagement, gleichwohl spielt die Informationstechnologie (IT) eine wichtige Rolle, als "Enabler" für eine Vielzahl von Funktionalitäten einer Wissensmanagement-Lösung.

IT-basierte Wissensmanagement-Lösungen beinhalten typischerweise ein Unternehmensgedächtnis (Organizational Memory), das unformales, semiformales und formales Wissen umfasst, welches vom Wissensarbeiter zur Bearbeitung seiner Aufgaben benötigt wird. Um dem Wissensarbeiter einen flexiblen und personalisierten Zugang zum Wissen anbieten zu können, muss das im Unternehmensgedächtnis bereitgestellte Wissen entsprechend modelliert, strukturiert und vernetzt werden. Ontologien haben sich hierzu als die Lösung herauskristallisiert, da sie eine konzeptuelle Strukturierung und Modellierung einer Domäne zur Verfügung stellen, die von einer Gruppe von Personen, z.B. einem Unternehmensbereich, gemeinsam getragen wird.

Der Workshop soll Vertretern aus Forschung und Praxis eine gemeinsame Basis zum Austausch über Erfahrungen und Visionen bieten. Folgenden Themenfelder sind relevant für den Workshop:

- Methoden und Techniken
- Anwendungen
- Erfahrungen
- Anforderungen

- Kosten/Nutzen Analysen
- Szenarien und Visionen

Der Workshop richtet sich an Teilnehmer, die sich einen Überblick über den aktuellen Stand der Forschung und Erfahrungen aus der Praxis verschaffen wollen und die bereit sind, den Workshops durch Einreichung von Beiträgen oder durch aktive Teilnahme zu gestalten.

Ziel des Workshops ist es dabei, sowohl Vertreter aus Forschung als auch aus der betrieblichen Praxis anzusprechen. Beiträge werden u.a. erwartet von (i) Forschern aus den Bereichen Wissensmanagement und Semantic Web, welche z.B. an state-of-the-art Methoden und Technologien arbeiten, (ii) Praktikern aus Unternehmen, die z.B. Ontologie-basierte Wissensmanagement-Anwendungen entwerfen, implementieren, installieren und warten, sowie (iii) Kunden, die z.B. Anforderungen an Ontologie-basierte Wissensmanagement Methoden, Techniken und insbesondere Anwendungen formulieren.

Organisations-Komitee

- York Sure (Kontakt Person), Institut AIFB der Universität Karlsruhe (D)
- Hans-Peter Schnurr, Ontoprise GmbH (D)

Programm-Komitee

- Andreas Abecker, DFKI Kaiserslautern
- Robert Engels, CognIT a.s
- Dieter Fensel, VU Amsterdam
- Christian Fillies, Semtation GmbH
- Hermann Kaindl, Siemens AG
- Jürgen Klenk, definiens AG
- Alexander Mädche, FZI Karlsruhe
- Ralf Molitor, Swiss Life
- Christian Ohlms, McKinsey
- Josef Schneeberger, SCHEMA GmbH
- Peter Smolle, Net Dynamics
- Heiner Stuckenschmid, VU Amsterdam

Ontologies in Cross-Language Information Retrieval

Martin Volk
Eurospider Information Technology AG
Schaffhauserstrasse 18
CH-8006 Zürich, Switzerland
volk@eurospider.com

Špela Vintar, Paul Buitelaar
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
{vintar, paulb}@dfki.de

Abstract: We present an approach to using ontologies as interlingua in cross-language information retrieval in the medical domain. Our approach is based on using the Unified Medical Language System (UMLS) as the primary ontology. Documents and queries are annotated with multiple layers of linguistic information (part-of-speech tags, lemmas, phrase chunks). Based on this we identify medical terms and semantic relations between them and map them to their position in the ontology.

The paper describes experiments in monolingual and cross-language document retrieval, performed on a corpus of medical abstracts. Results show that semantic information, specifically the combined use of concepts and relations, increases the precision in monolingual retrieval. In cross-language retrieval the semantic annotation outperforms machine translation of the queries, but the best results are achieved by combining a similarity thesaurus with the semantic codes.

1 Introduction

The task of finding relevant information from large, multilingual and domain-specific text collections is a field of active research within the information retrieval and natural language processing communities. Methods of Cross-Language Information Retrieval (CLIR) are typically divided into: approaches based on bilingual dictionary look-up or Machine Translation (MT); corpus-based approaches utilizing a range of IR-specific statistical measures; and concept-driven approaches, which exploit multilingual ontologies or thesauri to bridge the gap between surface linguistic form and meaning. The latter seem particularly appropriate for domains (and languages) for which extensive multilingual ontologies are available, such as UMLS (Unified Medical Language System) in the medical domain.

The experiments reported in this paper were performed within the MUCHMORE project¹, which aims at systematically comparing concept-based and corpus-based methods in cross-language medical information retrieval.

¹MUCHMORE is a European Union project under grant IST-1999-11438 which also cooperates with two US partners. For details see <http://muchmore.dfki.de>.

2 Related Work

Many authors have experimented with machine translation or dictionary look-up for CLIR (see [KH98]). In a comparison of such methods in both query and document translation, Oard [Oar98] found that dictionary-based query translation seems to work best for short queries while for long queries machine translation of the queries performs better than dictionary look-up. An important problem in the translation of short queries is the lack of context for the disambiguation of words that have more than one meaning and therefore may correspond to more than one translation.

Ambiguity is also of importance to interlingua approaches to CLIR that use multilingual thesauri as resources for a language-independent (semantic) representation of both queries and documents. Domain-specific multilingual thesauri have been used for English-German CLIR by [ERS98] who describes the use of the UMLS MetaThesaurus for French and Spanish queries on the OHSUMED text collection, a subset of MEDLINE. He uses the thesaurus as a source for compiling a bilingual lexicon, which is then used for query translation. Next to domain-specific thesauri more general semantic resources such as EuroWordNet [Vos97] have been used in both mono- and cross-language information retrieval.

The work we describe here is primarily an interlingua approach to CLIR in the medical domain, in which we use both domain-specific (UMLS) and general language semantic resources (EuroWordNet). Central to the approach is the use of linguistic processing for an accurate semantic annotation of relevant terms and relations in both the queries and the documents.

3 Corpus Processing and Annotation

3.1 Linguistic Processing

The main document collection used in the MUCHMORE project is a parallel corpus of English-German scientific medical abstracts obtained from the Springer web site². The corpus consists of approximately 9000 documents with a total of one million tokens for each language. Abstracts are taken from 41 medical journals (e.g. *Der Nervenarzt*, *Der Radiologe*, etc.), each of which constitutes a homogeneous medical sub-domain (e.g. Neurology, Radiology, etc.). Corpus preparation included removing special tags and symbols in order to produce a clean, plain text version of each abstract, consisting of a title, text and keywords. The corpus was then linguistically annotated using standard tools for shallow processing: a tokenizer, a statistical part-of-speech tagger, a morphological analyser and a chunker for phrase recognition.

²<http://link.springer.de>

3.2 Semantic Annotation using UMLS and MeSH

The essential part of any concept-based CLIR system is the identification of terms and their mapping to a language-independent conceptual level. Our basic resource for semantic annotation is UMLS, which is organized in three parts.

The **Specialist Lexicon** provides lexical information: a listing of word forms and their lemmas, part-of-speech and morphological information.

Second, the **Metathesaurus** is the core vocabulary component, which unites several medical thesauri and classifications into a complex database of concepts covering terms from 9 languages. Each term is assigned a unique string identifier, which is then mapped to a unique concept identifier (CUI). A simplified entry for *HIV pneumonia* in the Metathesaurus main termbank (MRCON) looks like this:

```
C0744975 | ENG | P | L1392183 | S1657928 | HIV pneumonia
```

The fields in this entry specify (from left to right), the concept identifier, the language of the term, the term status, the term identifier, the string identifier, and the string itself.

In addition to the mapping of terms to concepts, the Metathesaurus - a true ontology - organizes concepts into a hierarchy by specifying relations between concepts. These are thesaurus-type generic relations like *broader_than*, *narrower_than*, *parent*, *sibling* etc. The UMLS 2001 version includes 1.7 million terms mapped to 797,359 concepts, of which 1.4 million entries are English and only 66,381 German. Only the MeSH (Medical Subject Heading) part of the Metathesaurus covers both German and English, therefore we only use MeSH terms for corpus annotation.

The third part is the **Semantic Network**, which provides a grouping of concepts according to their meaning into 134 semantic types (TUI). The concept above would be assigned to the class *T047, Disease or Syndrome*. The Semantic Network then specifies potential relations between those semantic types. There are 54 hierarchically organized domain-specific relations, such as *affects*, *causes*, *location_of* etc.

3.2.1 Terms and Concepts

The identification of UMLS terms in the documents was based on morphological processing of both the term bank and the document, so that term lemmas were matched rather than word forms. The preparation of the term bank included filtering and normalization procedures, such as case folding, removal of long terms, inversion of term variants with commas (*Virus, Human Immunodeficiency* → *Human Immunodeficiency Virus*), conversion of special characters etc. The annotation tool matches terms of lengths 1 to 3 tokens, based on lemmas if available and word forms otherwise.

The decision to use MeSH codes in addition to concept identifiers (CUIs) was based on our observation that the UMLS Semantic Network, especially the semantic types and relations, does not always adequately represent the domain-specific relationships. MeSH codes on the other hand have a transparent structure, from which both the semantic class

of a concept and its depth in the hierarchy can be inferred. For example, the terms *infarction* (C23.550.717.489) and *myocardial infarction* (C14.907.553.470.500) both belong to the group of diseases, but the node of the first term lies higher in the hierarchy as its code has fewer fields.

3.2.2 Semantic relations

Semantic relations are annotated on the basis of the UMLS Semantic Network, which defines binary relations between semantic types in the form of triplets, for example *T195 - T151 - T042* meaning *Antibiotic - affects - Organ or Tissue Function*. We search for all pairs of semantic types that co-occur within a sentence, which means that we can only annotate relations between items that were previously identified as UMLS terms. According to the Semantic Network relations can be ambiguous, meaning that two concepts may be related in several ways. For example:

| | | |
|----------------------|--------------------|------------|
| Diagnostic Procedure | analyzes | Antibiotic |
| Diagnostic Procedure | assesses_effect_of | Antibiotic |
| Diagnostic Procedure | measures | Antibiotic |

Since the semantic types are rather general (e.g. *Pharmacological Substance*, *Patient or Group*), the relations are often found to be vague or even incorrect when they are mapped to a document. Given the ambiguity of relations and their generic nature, the number of potential relations found in a sentence can be high, which makes their usefulness questionable. A manual evaluation of automatic relation tagging by medical experts showed that only about 17% of relations were correct, of which only 38% were perceived as significant in the context of information retrieval. On the other hand, low term coverage - particularly for German - severely limits the number of relations that we can identify in the described way. Retrieval experiments performed with German queries over English documents showed that an evaluation of semantic relations in this context is almost impossible (cf. the results in section 4.2).

3.3 Semantic Annotation using EuroWordNet

In addition to annotation with UMLS, terms are annotated with EuroWordNet senses [Vos97] to compare domain-specific and general language use. Each language-specific (Euro)WordNet is linked to all others through the so-called Inter-Lingual-Index, which is based on WordNet1.5. The languages are interconnected via this index, so that it is possible to move from a word in one language to similar words in any of the other languages in the EuroWordNet database.

4 Evaluation in Information Retrieval

In order to evaluate whether the semantic annotations result in a performance gain in information retrieval, several experiments have been carried out. We used our own document collection (the set of medical abstracts described above) as well as a set of 25 queries with human relevance assessments provided by the medical expert in the MUCHMORE project. In these assessments the number of relevant documents per query varies between 7 and 104. They add up to a total of 959 relevant documents for the 25 queries.

The queries are short and usually consist of a complex noun phrase extended by attributes (including prepositional phrases) and coordination. Here is a typical example.

- *Arthroskopische Behandlung bei Kreuzbandverletzungen.*
Arthroscopic treatment of cruciate ligament injuries.

4.1 Monolingual Evaluation Runs

MUCHMORE aims first and foremost at cross-language retrieval (CLIR). In order to set the CLIR performance into perspective, monolingual experiments in German and English were conducted acting as baselines for the cross-language experiments.

For the retrieval experiments we used the commercial *relevancy* information retrieval system from Eurospider Information Technology AG. In regular deployment this system extracts word tokens from documents and queries and indexes them using a straight *Inu.ltn* weighting scheme (for the theoretical background of this scheme see [Sch97]).

For the MUCHMORE evaluation runs we adapted the *relevancy* system so that it indexes the information provided by the annotated documents and queries: word forms (tokens) and their base forms (lemmas) for all indexable parts-of-speech. The indexable parts-of-speech encompass all content words, i.e. nouns (including proper names and foreign expressions), adjectives, and verbs (excluding auxiliary verbs). All semantic information was indexed in separate categories each: EuroWordNet terms, UMLS terms, semantic relations, and MeSH terms.

In table 1 we present the results of the monolingual English retrieval experiments. We present the retrieval results in four columns. The first column contains the overall performance, measured as mean average precision (mAvP) as has become customary in the TREC experiments. This figure is computed as the mean of the precision scores after each relevant document retrieved. This value contains both precision and recall oriented aspects and is the most commonly used summary measure. In the second column we present the absolute number of relevant documents retrieved, a pure recall measure. Third, we present the average precision at 0.1 recall (AvP01). Because this number can vary substantially for different queries, we consider also the precision figures for the topmost documents retrieved (in column four). There we focus on the top 10 documents (P10).

In the baseline experiment for English (EN-token) we find 617 relevant documents (out of 956; cf. table 1). The mean average precision (mAvP) is 0.35, and the average precision

in the top ranks is high ($AvP = 0.80$). So, the few documents that are found are often ranked at the top of the list. On average there are 6.16 relevant documents among the 10 top ranked documents (P10).

Linguistic lemmatization (stemming) worsens the precision for English monolingual retrieval. But it does increase the recall when used in combination with tokens (see line EN-token-lemma). This is very different from German monolingual retrieval which clearly improves with lemmatization both for recall and precision. The additional benefit was particularly due to segmentation of German compounds.

The impact of the different types of semantic information was determined one by one, but always in combination with tokens. We wanted to support the hypothesis that semantic information will improve the precision over pure token information. It turns out that MeSH codes are the most useful indexing features among the semantic codes. Using MeSH codes slightly increases recall (from 617 to 637) but most impressively improves average precision (from 0.3455 to 0.3637). The positive impact of the UMLS terms is less visible.

| | mAvP | Rel. Docs Retr. | AvP 0.1 | P10 |
|----------------|--------|-----------------|---------|--------|
| EN-token | 0.3455 | 617 | 0.8077 | 0.6160 |
| EN-lemma | 0.3097 | 600 | 0.6632 | 0.5360 |
| EN-token-lemma | 0.3320 | 635 | 0.7543 | 0.5760 |
| EN-token-EWN | 0.2155 | 604 | 0.5847 | 0.4000 |
| EN-token-UMLS | 0.3455 | 617 | 0.8077 | 0.6160 |
| EN-token-MeSH | 0.3637 | 637 | 0.8259 | 0.6040 |

Table 1: Results of the monolingual English runs

Using the EuroWordNet terms (EWN) in this combination with tokens degrades the overall performance. We investigated this phenomenon and found that EuroWordNet terms in our queries are mostly general language words like *injury*, *complication* or *treatment*. By using these words as additional indexing features we give them more weight than content-bearing specific terms. This leads to a bias towards the general language words and thus to a loss in retrieval precision.

4.2 Cross-Language Evaluation Runs

For the Cross-Language Information Retrieval we assume that we have a document collection (i.e. a corpus) in one language and a query in another language. We used German queries to retrieve English documents.

As a baseline we investigated the use of Machine Translation (MT) for translating the queries. We employed the PC-based system PersonalTranslator (PT2002; linguattec, Munich) to automatically translate all queries from German to English. PersonalTranslator allows to restrict the subject domain of the translation, and we selected the domains medicine and chemistry. Many translations are incomplete or incorrect but still the automatically

translated queries scored well with regard to recall. In table 2, line DE2EN-MT-PT2002, we see that these queries lead to 440 relevant documents at a (rather low) mean average precision of 0.1381.

Now let us compare these results with the results based on the semantic codes annotated in our corpus and queries. This means we are using the semantic annotation of the German queries to match the semantic annotation of the English documents. We are regarding the semantic codes as an interlingua to bridge the gap between German and English.

The second block in table 2 has all the results. This time the UMLS terms lead to the best results with respect to recall, but MeSH is (slightly) superior regarding precision. EuroWordNet leads to the worst precision and the semantic relations have only a minor impact due to their specificity. If we combine all semantic information, we achieve the best recall (404) and mean average precision (0.1774).

| | mAvP | Rel. Docs Retr. | AvP 0.1 | P10 |
|-------------------------|--------|-----------------|---------|--------|
| DE2EN-MT-PT2002 | 0.1381 | 440 | 0.3747 | 0.2920 |
| DE2EN-EWN | 0.0090 | 111 | 0.0311 | 0.0160 |
| DE2EN-UMLS | 0.1620 | 366 | 0.3724 | 0.2800 |
| DE2EN-MeSH | 0.1699 | 304 | 0.3888 | 0.2600 |
| DE2EN-Semrel | 0.0229 | 23 | 0.0657 | 0.0480 |
| DE2EN-all-combined | 0.1774 | 404 | 0.3872 | 0.2720 |
| DE2EN-SimThes | 0.2290 | 409 | 0.4492 | 0.3640 |
| DE2EN-SimThes+all-comb. | 0.2955 | 518 | 0.5761 | 0.4600 |

Table 2: Results of the cross-language runs: German queries and English documents

For the last two experiments we have built a similarity thesaurus (SimThes) over the parallel corpus. The similarity thesaurus contains words (adjectives, nouns, verbs) from our corpus, each accompanied by a set of words that appear in similar contexts and are thus similar in meaning. In our case we built the similarity thesaurus over the parallel corpus. We were interested in German words and their similar counterparts in English. The similarity thesaurus is thus an automatically constructed bilingual lexicon with a broad translation set (in our case 10 similar English words per German word). For example, for the German word *Myokardinfarkt* the similarity thesaurus contains the following 10 words in decreasing degrees of similarity: *infarction, acute myocardial infarction, myocardial, thrombolytic, acute, thrombolysis, crs, synchronisation, cardiogenic shock, ptca*.

We used these words for cross-language retrieval. Each German word from the queries was substituted by the words of its similarity set. This resulted in a recall of 409 relevant documents found and a relatively good mean average precision of 0.2290 (see DE2EN-SimThes in table 2). Note that unlike in our previous experiments, we have now exploited the parallelism of the documents in our corpus for the construction of the similarity thesaurus. The bilingual similarity thesaurus is only available if we have a parallel or comparable corpus (cf. [BS00]) whereas the semantic annotations will also be applicable for a monolingual document collection.

Finally we checked the combination of all semantic annotations with the similarity thesaurus. Each query is now represented by its EuroWordNet, UMLS, MeSH and semantic relations codes as well as by the words from the similarity thesaurus. This combination leads to the best results for CLIR. We retrieved 518 relevant documents with a mean average precision of 0.2955 (cf. the last line DE2EN-SimThes+all-combined in table 2). And the figures for the high precision area (AvP and P10) are also outstanding.

5 Conclusions

We have explored the use of different kinds of semantic annotation derived from the UMLS ontology for both monolingual and cross-language retrieval. In monolingual retrieval (for both English and German) semantic information from the MeSH codes (Medical Subject Headings) were most reliable and resulted in an increase in recall and precision over token and lemma indexing.

In cross-language retrieval the combination of all semantic information outperformed machine translation. It was only superseded by the use of a similarity thesaurus built over the parallel corpus. The highest overall performance resulted from a combination of the similarity thesaurus with the semantic information.

So far, semantic annotation in our approach was based on the use of existing resources (UMLS and EuroWordNet) without applying disambiguation. In future work we hope to improve the performance by the integration of disambiguation for UMLS and EuroWordNet terms as well as including novel extracted terms and relations for UMLS.

References

- [BS00] Martin Braschler and Peter Schäuble. Using Corpus-Based Approaches in a System for Multilingual Information Retrieval. *Information Retrieval*, (3):273–284, 2000.
- [ERS98] D. Eichmann, M. Ruiz, and P. Srinivasan. Cross-Language Information Retrieval with the UMLS Metathesaurus. In *Proc. Of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, 1998.
- [KH98] W. Kraaij and D. Hiemstra. TREC6 Working Notes: Baseline Tests for Cross Language Retrieval with the Twenty-One System. In *TREC6 Working Notes*, Gaithersburg, MD, 1998. National Institute of Standards and Technology (NIST).
- [Oar98] D. Oard. A Comparative Study of Query and Document Translation for Cross-Lingual Information Retrieval. In *Proc. of AMTA*, Philadelphia, PA, 1998.
- [Sch97] Peter Schäuble. *Multimedia Information Retrieval. Content-based Information Retrieval from Large Text and Audio Databases*. Kluwer Academic Publishers, Boston, 1997.
- [Vos97] Piek Vossen. EuroWordNet: A Multilingual Database for Information Retrieval. In *Proc. Of the DELOS Workshop on Cross-Language Information Retrieval*. Zurich, Switzerland, March, 5-7 1997.

SemanticMinerTM: Ein integratives Ontologie-basiertes Knowledge Retrieval System *

Eddie Mönch

Ontoprise GmbH, Haid-und-Neu-Str. 7, D-76131 Karlsruhe

moench@ontoprise.de

Abstract: Oft stellt sich bei der Analyse von Wissensprozessen in Unternehmen heraus, dass der einfache Zugriff auf das vorhandene Unternehmenswissen in Dokumenten nicht möglich ist. Für den Zugriff auf Dokumenten- und Datenbestände des Unternehmens nehmen die Technologien des *Information Retrieval (IR)* eine zentrale Rolle ein. Im Folgenden beschreiben wir die Theorie des SemanticMinerTM-Systems, das heißt Methoden und Technologien sowie weiterführende Ansätze, um mithilfe semantischer Technologien aus dem Information Retrieval ein *Knowledge Retrieval (KR)* zu erreichen.

1 Einführung in Ontologie-basiertes Wissensmanagement

Bereits Aristoteles versuchte in seiner Kategorienlehre, die Dinge der Welt nach bestimmten Kriterien zu untersuchen und zu ordnen. Daraus entstand über Jahrhunderte eine philosophische Subwissenschaft namens Ontologie. Diese vergleichsweise neue Bezeichnung, die sich aus dem Griechischen zusammengesetzt — „ontos“ für Sein und „logos“ für Wort [Sow00] — wird benutzt, um die Lehre vom Sein zu unterscheiden von der Lehre des Seienden in den Naturwissenschaften.

Die Informatik entlehnte den Begriff der Ontologie zum Zwecke der Repräsentation und Nutzung von Wissen. Seit Anfang der neunziger Jahre wurden Ontologien zu einem beliebten Forschungsthema in Teilgebieten der Künstlichen-Intelligenz-Forschung. In letzter Zeit breitet sich die Idee der Ontologie auf immer mehr Bereiche aus, wie Intelligent Information Integration, Cooperative Information Systems, *Information Retrieval*, Electronic Commerce und Knowledge Management (für weitere Beispiele sei auf [Sow00] verwiesen). Der Grund für die stetig wachsende Popularität von Ontologien, liegt größtenteils an dem, was sie versprechen: Ein geteiltes und gemeinsames Verstehen einer Domäne, das zwischen Personen und Anwendungssystemen kommuniziert werden kann (vgl. [Fen01]).

Bedeutung: Ontologien werden entwickelt um eine maschinen-verarbeitbare Semantik an Informationsressourcen, die zwischen verschiedenen Agenten (Software und Menschen) kommuniziert werden kann, bereitzustellen.

* Ausführliche Version unter <http://www.ontoprise.de/documents/SemanticMinerKR.pdf>

Definition und Eigenschaften. Die am häufigsten zitierte Definition für Ontologie ist die von Gruber: „Eine Ontologie ist eine formale, explizite Spezifikation einer gemeinsamen Konzeptualisierung“ [Gru93].

Durch eine explizite Spezifikation der Entitäten (Konzepte), die mit anderen Entitäten über Axiome (Relationen) verknüpft, oder mit Attributen detaillierter beschrieben werden entsteht daraus eine Ontologie. Üblicherweise sind Ontologien in Taxonomien mit mehrfacher Vererbung und disjunkten Unterkategorien organisiert. Neben dieser Kategorisierung beschreiben sie für einen Wissensbereich ebenfalls Regeln, die die Konzepte durch Constraints oder Inferenzregeln in Beziehung setzen. Diese werden typischerweise in logischen Formalismen repräsentiert, die auf der Prädikatenlogik basieren.

F-Logic. Für das SemanticMiner-System verwenden wir die Sprache *Frame-Logic* (F-Logic). F-Logic entspricht syntaktisch gesehen einer Obermenge der Prädikatenlogik erster Stufe (*FOL, first order logic*), wobei die Ausdrucksmächtigkeit beider Sprachen allerdings äquivalent ist. F-Logic ist eine logik- und objektorientierte Sprache, die 1995 von Kifer et al. [KLW95] entwickelt wurde. Sie verbindet die Ausdrucksstärke von Normallogik (Horn-Logik mit Negationen) mit den Datenmodellierungsmöglichkeiten des objektorientierten Ansatzes. Da die grundlegenden Prinzipien der Vererbung, Kapselung, Klassenbildung, Polymorphie und Typüberprüfung durch die Ausdrucksstärke und die Inferenzmöglichkeiten von Logik ergänzt werden, ist sie besonders für die Modellierung von Ontologien geeignet.

Das Allwissenden-Paradigma. Mit der Verwendung einer Ontologie akzeptiert man automatisch das „Allwissenden“-Paradigma, das einem traditionellen Ansatz der Kognition in sozialen Systemen entstammt. Wissen wird dabei in einer einzigen, von allen geteilten kohärenten Struktur repräsentiert und organisiert, völlig unabhängig von wem, wie, wo und warum dieses Wissen ursprünglich geschaffen wurde. Der heute aufstrebende Ansatz der „Verteilten Intelligenz“ basiert hingegen auf der Annahme, dass Wissen immer und unteilbar mit verschiedenen sogenannten Kontexten verknüpft ist, wie beispielsweise Individuen, Gruppen, Zeiträumen und Orten und daher nicht generell zentral organisiert werden kann: Wissen ist demnach immer kontextspezifisch [NSB00]. Zu erwähnen ist auch, dass sich die spätere Nutzergruppe des angestrebten wissensbasierten Systems auf die Ontologie geeinigt haben muss [Gru95]. Durch diese Formalisierung wird jedoch Mehrdeutigkeit vermieden.

Weitere Ansätze existieren um Wissensmodelle aufzubauen. Eine ebenfalls verbreitete Methode ist die Verwendung von TopicMaps für die Einordnung und Kategorisierung von Begriffen. Hierbei werden vorhandene Themen (Topics) miteinander verbunden, ein semantisches Netz entsteht. TopicMaps eignen sich insbesondere zur Navigation vorhandener Begrifflichkeiten. Ontologien stellen zusätzlich zur Navigationsunterstützung mächtigere Modellierungsmöglichkeiten zur Verfügung, welche zusätzliche Funktionen des Wissensmodells ermöglichen [SM01].

Im Gegensatz zu allen anderen Technologien bestehen weitere Zusatznutzen von Onto-

logien darin, dass sie Ableitungen erlauben und Auswertungen der oben beschriebenen regelbasierten Zusammenhänge mittels einer Inferenzmaschine (z.B. OntoBrokerTM) erlauben. Implizites Wissen wird dadurch ebenfalls abgefragt und dargestellt — explizit gemacht.

2 Information Retrieval

Für den Begriff bzw. das Gebiet des Information Retrieval (IR) gibt es keine allgemein akzeptierte Definition oder Abgrenzung. Historisch gesehen wurde IR zum besseren (Wieder)auffinden von wissenschaftlicher Literatur entwickelt. Auch wenn dieses Gebiet nach wie vor einer der Schwerpunkte des IR ist, haben sich sowohl der Bereich der Objekte, mit denen IR umgeht, als auch die Aufgabenstellung erweitert. Eine Beschreibung gibt die Fachgruppe Information Retrieval der Gesellschaft für Informatik [Fuh96]:

„Im Information Retrieval werden Informationssysteme in Bezug auf ihre Rolle im Prozess des Wissenstransfers vom menschlichen Wissensproduzenten zum Informationsnachfragenden betrachtet.“ Ziel des IR ist es also, gespeicherte Daten (Texte, strukturierte Daten, Bilder, Fakten u.a.) so aufzubereiten und anzubieten, dass sie bei einem konkreten Informationsbedarf mit problemgerechten Suchstrategien möglichst präzise und vollständig herausgesucht werden können.

2.1 Qualitätsbewertung von IR-Systemen: Recall und Precision

Die am häufigsten verwendeten Maße zur Beurteilung der Güte eines IR-Systems sind *Recall* und *Precision*. Durch diese beiden Maße wird die Suche mit einem IR-System aufgrund des gelieferten Retrievalergebnisses bewertet. Grundlage bildet der Begriff der Relevanz eines Dokuments.

Eine Reihe von verschiedenen Definitionen des Begriffs Relevanz sind beispielsweise in [Kai93] zu finden. Wir verwenden die Definition *Relevanz* nach [CLvRC98]:

Definition 2.1 (Relevanz) *Wenn der Benutzer ein Dokument zu einer haben will, dann ist dieses relevant zu dieser Anfrage.*

Nun können die beiden Maße Recall und Precision definiert werden [BYRN99]:

Definition 2.2 (Recall) *Recall stellt das Maß für die Vollständigkeit des Retrievalergebnisses dar und ist definiert als das Verhältnis zwischen gefundenen, relevanten Dokumenten und der Gesamtzahl der im Dokumentenbestand vorhandenen relevanten Dokumente.*

Genauer gilt: Gegeben sei ein Informationsbedarf I und eine Anfrage q des Benutzers. Dann berechnet sich der Recall durch

$$req(q, I) = \frac{|\mathcal{R}(q, I)|}{|\mathcal{R}(I)|}, \quad (1)$$

wobei $|\mathcal{R}(I)|$ die Anzahl aller relevanten Dokumente zum Informationsbedarf I und $|\mathcal{R}(q, I)|$ die Anzahl der mit der Anfrage q gefundenen, zum Informationsbedarf I relevanten Dokumente bezeichnet (vgl. Abbildung 1).

Der Wertebereich des Recalls geht von 0 bis 1. Ein Recall von 0 wird für das schlechteste Ergebnis, 1 für das bestmögliche vergeben.

Definition 2.3 (Precision) *Precision dient zum Messen der Genauigkeit des Retrievalergebnisses und als Indikator für die Fähigkeit eines IR-Systems, nicht relevante Dokumente nicht auszugeben. Precision ist definiert als das Verhältnis der gefundenen relevanten Dokumente zur Zahl aller Dokumente.*

Genauer gilt: Gegeben sei ein Informationsbedarf I und eine Anfrage q des Benutzers. Dann berechnet sich Precision durch

$$pres(q, I) = \frac{|\mathcal{R}(q, I)|}{|\mathcal{E}(q)|}, \quad (2)$$

wobei $|\mathcal{R}(q, I)|$ die Anzahl der mit der Anfrage q gefundenen, zum Informationsbedarf I relevanten Dokumente und $|\mathcal{E}(q)|$ die Anzahl aller mit Anfrage q gefundenen Dokumente bezeichnet (vgl. Abbildung 1).

Der Wertebereich von Precision geht ebenfalls von 0 bis 1. Auch bei Precision wird versucht, den Wert zu maximieren.

Sinnvoll ist nur die Betrachtung beider Maße, da Recall die Zahl der irrelevanten, ausgegebenen Dokumente unberücksichtigt lässt und leicht auf das Maximum von 1 gesetzt werden kann, indem alle im Dokumentenbestand vorhandenen Dokumente ausgegeben werden. In diesem Fall wäre dann allerdings der Precisionwert sehr niedrig. Die alleinige Betrachtung von Precision wiederum würde nichts über die Vollständigkeit des Retrievalergebnisses aussagen. Precision allein könnte dadurch maximiert werden, dass nur sehr wenig Dokumente ausgegeben werden.

Bei Suchen mit einem hohen Anspruch auf Vollständigkeit des Suchergebnisses wird ein hoher Recall angestrebt, so dass wir innerhalb des SemanticMiner Systems ein größeres Augenmerk auf dieses Maß legen (siehe hierzu Kapitel 3.1).

3 Knowledge Retrieval — Semantisches Information Retrieval

Ein Indikator für die Retrievalqualität der derzeitigen Ad-hoc-IR-Systeme¹ stellen die Ergebnisse der jährlich stattfindenden TREC-Veranstaltungen dar. TREC bezeichnet eine Veranstaltung, bei der Softwareimplementierungen von derzeitigen Algorithmen im IR auf ihre Qualität getestet werden. In [Har00] werden die Ergebnisse der an TREC teilgenommenen Ad-hoc-IR-Systeme über die letzten Jahre verglichen. Es zeigt sich, dass

¹Unter Ad-hoc-Suche versteht sich die vollautomatische Suche.

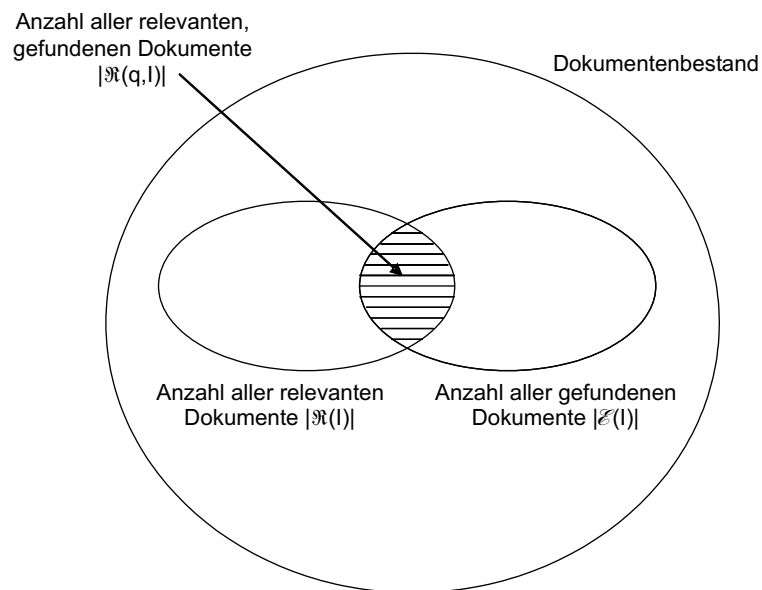


Abbildung 1: Recall und Precision für einen gegebenen Beispiel-Informationsbedarf

seit 1996 bei den Ad-hoc-IR-Systemen eine Stagnation hinsichtlich der Retrievalqualität (Recall/Precision) zu verzeichnen ist. Daraus läßt sich folgern, dass nach dem derzeitigen Kenntnisstand der Wissenschaft die Entwicklung von Retrievalalgorithmen und Indexierungsalgorithmen ausgereizt ist.

3.1 Abhängigkeit der Retrievalqualität von der Anfrage

Die Qualität eines Ad-hocSuchdienstes im Sinne von Recall und Precision ist sehr abhängig von der aktuellen Anfrage. Diese Eigenschaft wurde bei Ad-hoc-IR-Systemen in [Har00] praktisch bestätigt. Es wurden unterschiedliche Ad-hoc-IR-Systeme bei TREC pro Anfrage verglichen und es zeigte sich, dass ein Ad-hoc-IR-System bei der einen Anfrage eine sehr hohe Qualität im Vergleich zu den anderen Ad-hoc-IR-Systemen und bei anderen Anfragen nur eine sehr schlechte Qualität im Vergleich zu den anderen Ad-hoc-IR-Systemen besitzen kann.

3.2 Änderung der Anfrage

Das Ziel der Änderung der Anfrage besteht in der Adaptivität des Wortschatzes des Benutzers an das IR-System. Es ist eine sehr verbreitete Methode. Es existieren sehr viele

automatische Anfragemodifikationsalgorithmen in der Literatur, z.B. [BMS98].

Definition 3.1 (Anfragemodifikation) *Anfragemodifikation entspricht der automatischen Änderung einer Anfrage aufgrund von Zusatzwissen (Thesaurus, Relevanz Feedback, Statistiken, usw.) mit dem Ziel, bessere Retrievalergebnisse zu erzielen. Es besteht dabei die Gefahr des Anfrageabtriebs (engl. Query Drift), also die Gefahr, dass die erweiterte Anfrage nicht mehr den ursprünglichen Informationsbedarf widerspiegelt.*

Teilweise werden für den Begriff Anfragemodifikation auch die Begriffe *Anfrageerweiterung* oder *Anfragereformulation* verwendet.

Unser Verfahren unterscheidet sich von bekannten Anfragemodifikationsalgorithmen dahingehend, dass die Anfrage von der Dokumentenmenge völlig abgekoppelt und die Erweiterung wie in Kapitel 1 beschrieben allgemein gültigen Status besitzt. Ebenfalls wird dadurch der Gefahr des Anfrageabtriebs entgegengesteuert.

3.3 Query-Expansion

Die Abhängigkeit der Retrievalqualität von der Anfrage unterstützt unsere Motivation im SemanticMinerTM-System den Fokus auf die Anfrage an ein Ad-hoc-Suchdienst zu legen. Der dem System zugrunde liegende *Query-Expansion*-Ansatz kann den Anfragen an das Ad-hoc-IR-System automatisch ontologisches Wissen hinzufügen und so die Qualität der Antworten verbessern. Das führt zu einer Verbesserung des Recall-Wertes, da mehr relevante Dokumente durch die qualitative Erhöhung der Suchterme gefunden werden. Über den Precision-Wert kann keine allgemeine Aussage getroffen werden, da die Anzahl der mit der Anfrage q gefundenen, zum Informationsbedarf I relevanten Dokumente — also $|\mathcal{R}(q, I)|$ — als auch die Anzahl aller mit Anfrage q gefundenen Dokumente, also $|\mathcal{E}(q)|$ steigt.

Jedoch schaut sich typischerweise der Suchende nur die ersten 10 bis 20 Dokumente eines Suchergebnisses an. Durch die Ranking-Funktion im SemanticMinerTM-System führt dies, in Kombination mit der Query-Expansion durch ontologisches Wissen, zu einer wesentlichen Steigerung des „subjektiven“, also für den Suchenden relevanten (die ersten 10-20 Dokumente), Precision-Wertes, da die Dokumente mit hoher Termübereinstimmung *aller* Anfrageterme das höchste Ranking erfahren.

Aus [Har00] kann man ebenfalls folgern, dass der Wechsel eines Ad-hoc-Suchdienstes während einer Suche bei gleicher Anfrage durchaus Sinn macht. Die zugrunde liegenden Ad-hoc-IR-Systeme des SemanticMinerTM-Systems sind für das System transparent und können beliebig ausgetauscht oder ergänzt werden.

3.4 Semantik der Anfrage

Ein weiterer Nachteil allgemeiner IR-Ansätze ist, dass eine reine syntaktische Suche nach Begriffen stattfindet, ohne dabei die Bedeutung der Wörter innerhalb der einzelnen Dokumente zu berücksichtigen. Dies führt bei der Suche zu einer hohen Zahl von Treffern, bei denen auch Dokumente gefunden werden, in denen der Begriff in anderer Bedeutung verwendet wird. Eine Suche nach ähnlichen Begriffen, bzw. Eingrenzungen oder Verallgemeinerungen kann von diesen statistischen Ansätzen nicht unterstützt werden.

Während der letzten 30 Jahre fand eine anhaltende Diskussion darüber statt, ob zur Unterstützung natürlicher Sprachverarbeitung (Natural Language Processing, NLP) auf syntaktische oder semantische Technologien fokussiert werden sollte. In beiden Lagern wurden Ansätze diskutiert und vorangetrieben. Immer deutlicher hat sich gezeigt, dass beide Technologien und insbesondere das Zusammenspiel zwischen statistischen Verfahren und semantischer Modellierung die wichtigsten Ansatzpunkte für die Weiterentwicklung der natürlichen Sprachverarbeitung darstellen.

3.5 Integration und Auswertung strukturierter Daten

Durch die Kombination einer Suchanfrage mit (semi)strukturierten Daten (Listen, Datenbanken, Metadaten) und logischen Regelzusammenhängen wird die Mächtigkeit der vorgestellten Ansätze in Abschnitt 3.4 weiter erhöht. Ziel ist es dabei, als Ergebnisliste keine Verweise auf Dokumente, die wiederum den gesuchten Inhalt enthalten, zu liefern, sondern tatsächliche Informationen aus Dokumenten zu lösen und als konkrete Antworten zu präsentieren.

Dies geschieht durch die Bildung von *Kollokationen*, wobei eine Kollokation eine Wortgruppe darstellt, mit der üblicherweise ein Grundbegriff, ein Gegenstand oder eine Handlung bezeichnet wird. Kollokationen wie „maschinelle Übersetzung“ oder „Anwendung schließen“ werden in der Terminologie als eigenständige Termini angesehen. Durch die Bildung von Korrelationslisten aus einer Datenbank oder mittels des Excel2F-Logic-Exports, werden die Kollokationen dem SMS bekannt gemacht.

Dadurch kann man auch aus völlig unstrukturierten Informationsquellen und Textdokumenten zum Beispiel geeignete Mitarbeiter in Unternehmen identifizieren, indem man zu dem gesuchten Begriff eine Kollokationsabfrage über die Mitarbeiter-Datenbank stellt. Ebenfalls können somit Wissenslücken in Unternehmen aufgedeckt oder Wettbewerberübersichten generiert werden.

3.6 Deduktion

Wie in Abschnitt 1 bestehen weitere Zusatznutzen von Ontologien darin, dass sie Ableitungen erlauben und Auswertungen der beschriebenen regelbasierten Zusammenhänge

mittels der Inferenzmaschine des OntoBrokers erlauben. Implizites Wissen wird dadurch ebenfalls abgefragt und dargestellt - explizit gemacht. Das heißt, dass alle Informationen, die durch Regeln ausgewertet wurden, also nur implizit vorlagen, im SemanticMinerTM-System als explizite Informationen dargestellt werden. Der End-Nutzer des Systems ist somit nicht in der Lage zu unterscheiden, ob die Information, die ihm präsentiert wird, explizit vorhanden war, oder durch *Deduktion* anhand von Ableitungsregeln („*inferencing rules*“) ermittelt wurde.

Literatur

- [BMS98] C. Buckley, M.Mandra, and A. Singhal. Improving Automatic Query Expansion. In *21st ACM SIGIR International Conference on Research and Development in Information Retrieval*, pages 206–214, 1998.
- [BYRN99] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, New York, Addison-Wesley, 1999.
- [CLvRC98] F. Crestani, M. Lalmas, C.J. van Rijsbergen, and I. Campbell. Is this Document Relevant? ... Probably - A Survey of Probabilistic Models in Onformation Retrieval. *ACM Computing Surveys*, 30:528–552, December 1998.
- [Fen01] D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, Berlin, 2001.
- [Fuh96] N. Fuhr. Ziele und Aufgaben der Fachgruppe Information Retrieval, January 1996. <http://ls6-www.informatik.uni-dortmund.de/ir/fgir/mitgliedschaft/brochure2.html>.
- [Gru93] T.R. Gruber. A translation approach to portable ontology specifications. *Knowledge Aquisition*, (5):199–220, 1993.
- [Gru95] T.R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, (43):907–928, 1995.
- [Har00] D. Harman. What We Have Learned, and not learned, from TREC. In *BCS-IRSG: 22nd Annual Colloquium on IR Research*, pages 2–20, April 2000. <http://irsg.eu.org/irsg2000online/papers/harman.htm>.
- [Kai93] A. Kaiser. *Computer-unterstütztes Indexieren in Intelligenten Information-Retrieval Systemen. Ein Relevanz Feedback orientierter Ansatz zur Informationserschliessung in unformatierten Datenbanken*. PhD thesis, Wirtschaftsuniversität Wien, 1993.
- [KLW95] M. Kifer, G. Lausen, and J. Wu. Logical Foundations of Object-Oriented and Frame-Based Languages. *Journal of the ACM*, 42:741–843, 1995.
- [NSB00] S. Neumann, L. Schuurmans, and M. Bonifacio. Verteilte Systeme im Wissensmanagement. *Information Management und Consulting*, (15):75–82, 2000.
- [SM01] S. Staab and A. Mädche. Knowledge Portals: Ontologies at Work. *AI Magazine*, 2(21), 2001.
- [Sow00] J. Sowa. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, 2000.

Knowledge Management durch Prozess Mining

Wissen verwerten mit elektronischem „Butler“ FRED

Peter F. Smolle
Net Dynamics Internet Technologies
peter.smolle@netdynamics-tech.com

Abstract

Der Bedarf Wissen, welches sich überreichem Masse weitgehend auch elektronisch vorhanden ist, im Anlassfall verfügbar zu haben ist enorm.

Knowledge Management ist daher als jene Disziplin welche sich die Lösung der vorgeannten Aufgabe zum Ziel gesteckt hat, mit einer hohen Erwartungshaltung und Interesse versehen. Die wirtschaftlichen Rahmenbedingungen wie Kostensenkung und Globalisierung geben zusätzlich Druck auf Lösungsansätze zur effizienten Nutzung vorhandenen Wissens.

Trotz dieser von vielen unbestrittenen Nachfrage nach Knowledge Management Lösungen ist heute sehr wenig in diesem Gebiet umgesetzt. Das hat verschiedene Ursachen. Eine der Ursachen ist, dass speziell Unternehmen die den Bedarf nach Wissensverwendung haben, sehr an der leichten Wiederverwendung von Geschäftsprozessen interessiert sind, weil es eben wichtig ist zum Beispiel bei Ausfall eines Mitarbeiters welcher sein Know How in einem Prozess repräsentiert, jenes Prozess Wissen rasch verfügbar zu haben. Das Wissen aus abgelegten Informationen zu bekommen, ist in diesem Falle primär nicht hilfreich.

Heute sind jedoch sehr viele Methoden zum Knowledge Management eben auf den Bereich, Wissen aus abgelegten Informationen/Daten zu bekommen, spezialisiert wobei diese Methoden für den wissenschaftlichen Zugang zu Dokumenten äußerst sinnvoll sind, für die Wiederverwendung von Prozess Wissen aber nicht speziell geeignet erscheinen.

Ein zwischen einem Benutzer und seinem jeweiligen IT Bearbeitungssystem zwischengeschalteter elektronischer „Butler“ zeigt hier einen neuen Weg zur Erfassung und späteren Wiederverwertung von Prozess-Wissen auf.

Der Vortrag beleuchtet zunächst grundsätzliche Anforderungen an eine hypothetische Wissensmaschine, welche in der Lage ist Wissen in jeder Form - Prozess und Informationen - verfügbar zu machen. Die Plattform elektronischer „Butler“ FRED welche intelligente Ontologie basierende Software-Agenten bei ihrer Kollaboration managed, wird dann hinsichtlich ihrer Fähigkeit Elemente der hypothetischen Wissensmaschine umzusetzen, bewertet.

Ein wesentliches Element - Prozesse welche durch FRED abgearbeitet werden - wiederverwertbar zu machen, lässt durch die Hinterlegung von Prozess Aktivitäten mit ihren semantischen Attributen die Fähigkeiten von Ontologien sehr deutlich werden. Für die Definition solcher Ontologien in FRED werden dazu die Werkzeuge und Konzepte von ONTOPRISE (Spin-off der Universität Karlsruhe) mit hohem Nutzengrad eingesetzt.

Diese ontologischen Fähigkeiten von FRED, eben bei Prozess-Aktivitäten das Prozess Wissen zu erfassen, bildet die Grundlage von Prozess Mining und kann als ein entscheidenden Schritt in Richtung Knowledge Management für Prozesse gesehen werden.

Ein Beispiel dieser Methodik wird an Hand von „Wissen aus dem Projekt-Management-Prozess“ für den Bereich „Resource Deployment“ vorgestellt und anschließend diskutiert.

Der Prozess der Erfassung von Aktivitäten durch FRED und ihre ontologisch repräsentierte Abspeicherung sowie die Wiederverwendung mit der Methode „Case Based Reasoning“ wird abschließend kurz vorgestellt.

FRED ist ein Produkt von Net Dynamics Internet Technologies Wien / Österreich

Motivationale Aspekte beim Einsatz von konventionellen und ontologiebasierten Kompetenzmanagementsystemen¹

Lars Dittmann, Malte L. Peters, Stephan Zelewski²

Institut für Produktion und Industrielles Informationsmanagement
Universität Duisburg-Essen
Universitätsstraße 9
45141 Essen

E-Mail: { lars.dittmann | malte.peters | stephan.zelewski }@pim.uni-essen.de

Abstract: Der Beitrag behandelt die denkmöglichen Auswirkungen des Einsatzes von Kompetenzmanagementsystemen (KMS) auf die Motivation von Akteuren in Unternehmen. Der Erfolg eines solchen Systems hängt dabei wesentlich von der Akzeptanz der Akteure und der Integration in deren Arbeitsabläufe ab. Insbesondere wird untersucht, inwieweit Kompetenzmanagementsysteme, die auf Ontologien basieren, vorteilhafter erscheinen als konventionelle Kompetenzmanagementsysteme. Ontologiebasierte KMS zeichnen sich dadurch aus, implizites Wissen explizieren zu können und damit gehaltvollere Ergebnisse als konventionelle KMS zu erzielen.

¹ Eine Kurzversion dieses Beitrags erscheint im Proceedings-Band zur 2. Konferenz Professionelles Wissensmanagement - Erfahrungen und Visionen (WM2003) in Luzern.

² Dieser Beitrag entstand im Rahmen der Projekte MOTIWIDI und KOWIEN. Das Projekt KOWIEN wird vom BMBF im Rahmenkonzept „Forschung für die Produktion von morgen“ gefördert und vom Projektträger PFT, Forschungszentrum Karlsruhe betreut. Das Projekt MOTIWIDI wird vom BMBF im Rahmen des Forschungsvorhabens "Wissensintensive Dienstleistungen" gefördert und vom Projektträger DLR e.V. betreut.

1. Überblick

Die denkmöglichen Auswirkungen des Einsatzes von Kompetenzmanagementsystemen (KMS) auf die Motivation von Akteuren in Unternehmen stellt eine zentrale Herausforderung an das betriebliche Wissensmanagement dar. Der Erfolg solcher Systeme hängt wesentlich von der Akzeptanz der Akteure und der Systemintegration in deren Arbeitsabläufe ab. In Abschnitt 2 werden zunächst die Begriffe Kompetenzen, konventionelles Kompetenzmanagementsystem und ontologiebasiertes Kompetenzmanagementsystem erläutert. Das zentrale Thema des 3. Abschnitts bilden motivationale Probleme, die den erfolgreichen Einsatz von KMS im betrieblichen Alltag be- oder sogar verhindern können. Hierbei wird insbesondere darauf eingegangen, über welches Potenzial zur Problemlösung konventionelle und ontologiebasierte KMS verfügen. Zusätzlich wird kurz skizziert, welche weiteren Probleme sich beim Einsatz ontologiebasierter KMS ergeben können.

2. Kompetenzen und Kompetenz-Managementsysteme

Das Wissen, über das ein Unternehmen zur Erfüllung wissensintensiver Aufgaben verfügt, spiegelt sich in den Kompetenzen seiner Akteure wider. *Kompetenzen* stellen Kenntnisse und Fertigkeiten eines Akteurs dar, die ihn zur Erreichung vorgegebener Handlungszwecke befähigen (zum hier verwendeten Kompetenzbegriff vgl. [Fa98], [Ka82]). Kompetenzen setzen also Wissen voraus, erschöpfen sich aber nicht im bloßen Besitzen von Wissen. Vielmehr wird Wissen erst dann zur Kompetenz, wenn es mit einer zusätzlichen Disposition kombiniert wird: Es ist die Fähigkeit des Akteurs erforderlich, sein Wissen in realen Handlungssituationen zielgerichtet zur Erfüllung intendierter Handlungszwecke einzusetzen. Im Gegensatz zu [No98] wird Kompetenz also weder auf den tatsächlichen Handlungserfolg („richtig handeln“) noch auf das tatsächliche Wollen in einer konkreten Handlungssituation bezogen. Vielmehr wird Kompetenz als ein reiner Dispositionsbegriff konzeptualisiert, der situationenübergreifend ein grundsätzliches Handeln-Können ausdrückt.

Ein Kompetenzmanagementsystem (KMS, vgl. [Su00], [El01]) erfüllt im Wesentlichen drei Funktionen. Erstens soll es die möglichst umfassende und systematische Akquisition von Wissen über die Kompetenzen von Akteuren eines Unternehmens durch entsprechende Akquisitionstechniken unterstützen. Zweitens hält es das akquirierte Wissen über die Kompetenzen von Akteuren zur Weiterverarbeitung vor. Da ein Akteur in der Regel über eine Vielzahl an Kompetenzen verfügt, wird das (Meta-)Wissen über diese Kompetenzen in der Gestalt von akteursspezifischen Kompetenzprofilen verwaltet. Drittens dient das KMS dazu, das vorhandene – sowohl explizite als unter Umständen auch implizite (oftmals als „tazit“ bezeichnete) – Wissen über Kompetenzen zweckbezogen anzuwenden. Die Besetzung von Projektteams für die Leistungserstellung, die Auswahl von Subkontraktoren während der Leistungserstellung und die Nachverfolgung von Gewährleistungsansprüchen nach der Leistungserstellung sind beispielhafte Anwendungsfälle, in denen das vorhandene Wissen über Kompetenzen zweckbezogen eingesetzt werden kann.

Derzeit im Einsatz befindliche *konventionelle Kompetenzmanagementsysteme* beschränken sich vornehmlich auf [Ma02]:

- ❑ manuelle Definitionen der Kompetenzen durch Mitarbeiter,
- ❑ einfache baumartige, hierarchische Darstellung der Beziehungen zwischen Kompetenzen sowie
- ❑ Bereitstellung von Informationen vornehmlich nur für den Bereich der Personalverwaltung.

Wesentliche Anforderungen an KMS aus betriebswirtschaftlicher Sicht werden hierdurch jedoch noch nicht erfüllt ([Ma02], zu einem ersten Ansatz der Evaluierung existierender KMS vgl. [DLW99]):

- ❑ Es bestehen kaum Möglichkeiten, auch die strategische Personalentwicklung zu unterstützen, da sich die KMS bisher auf das reine Verwalten von Kompetenzprofilen konzentrieren.
- ❑ Die Integration von Kompetenzmanagementsystemen in die existierende IT-Infrastruktur eines Unternehmens wird nicht unterstützt.

Um die darüber hinausgehende Lösungsleistung eines *ontologiebasierten KMS* zu verdeutlichen, wird sein schematischer Aufbau nachfolgend kurz in Anlehnung an die Onto-Broker-Architektur von STUDER ET AL. [St99] skizziert (Abbildung 1).

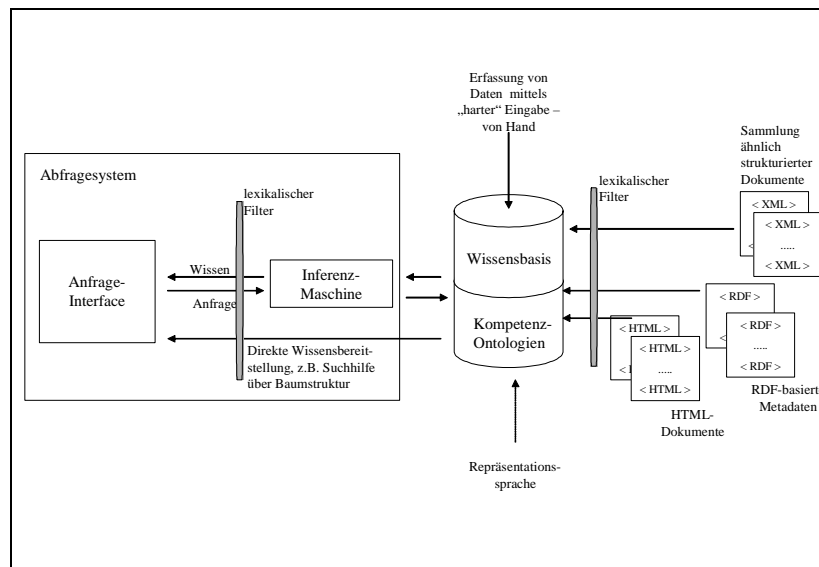


Abbildung 1: Skizzierter Aufbau eines ontologiebasierten Kompetenz-Managementsystems [St99]

Über ein *Anfrage-Interface* begibt sich der Benutzer auf die Suche nach Akteuren, deren Kompetenzprofile mit denjenigen Kompetenzen bestmöglich übereinstimmen, die zur Erfüllung einer Aufgabe für erforderlich erachtet werden. Der Benutzer kann hierbei durch Ontologien unterstützt werden, die ihm z.B. als Suchhilfe mittels einer Baumstruktur oder einer – im Sinne der Softwareergonomie vorteilhafteren – „hyperbolischen“ Benutzerschnittstelle [Fe01] zur Verfügung gestellt werden. Darüber hinaus ist ein betriebswirtschaftliches Problem aus dem Bereich der multi-attributiven Nutzentheorie zu lösen, um den „best fit“ zwischen dem aufgabenspezifischen Soll-Kompetenzprofil einerseits und den Ist-Kompetenzprofilen der Akteure andererseits festzustellen. Für die betriebliche Praxis wird oftmals eine simple Scoring-Technik oder Nutzwertanalyse als Problemlösung ausreichen.

Eine besondere Bedeutung erlangt die *Inferenz-Maschine* innerhalb ontologiebasierter KMS im Fall von *Explizierungslücken*. Solche Explizierungslücken bereiten konventionellen Kompetenzmanagementsystemen erhebliche Schwierigkeiten, weil das kompetenzrelevante Wissen nicht – wie es für konventionelle Datenbanksysteme typisch ist – in expliziten Datensätzen vorliegt, sondern in den Dokumenten lediglich implizit enthalten ist. Inferenz-Maschinen gestatten es dennoch, das angefragte Wissen über Akteure und deren Kompetenzen aus den gespeicherten, mittels eines automatischen Informationsverarbeitungssystems verarbeitbaren Dokumenten auch dann noch zu erschließen, wenn es zunächst nur in impliziter Form vorhanden ist. Der Benutzer erhält so eine höherwertige Antwort, als es bei einer herkömmlichen Datenbankabfrage möglich wäre, die nur explizit abgelegtes Wissen auszuwerten vermag. Damit leisten Inferenz-Maschinen einen wesentlichen Beitrag zur Schließung der oben erwähnten Explizierungslücke.

Durch die Anwendung deduktiver Inferenzregeln wird die Wissensbasis inhaltlich nicht erweitert. Stattdessen ermöglichen sie „nur“ die wissenserhaltende Transformation von implizitem in explizites Wissen. Eine deduktive Schlussfolgerung ist zwar allgemeingültig, weil sie unabhängig davon gilt, auf welche konkreten Inhalte sie bezogen wird. Aufgrund dieser Inhaltsunabhängigkeit besitzen die abstrakten Inferenzregeln der deduktiven Logik aber auch nur ein eng begrenztes Anwendungsfeld für KMS. Insbesondere können deduktive Inferenzregeln kein bereichsspezifisches, heuristisches „Alltagswissen“ nutzen. Dieses Alltagswissen besitzt nicht dieselbe epistemische Qualität wie allgemeingültige Schlussfolgerungen, sondern kann allenfalls Plausibilität für sich beanspruchen. Darauf aufbauende Plausibilitätsschlüsse werden den non-deduktiven Schlussfolgerungen zugeordnet.

Non-deduktive Schlussfolgerungen kommen vor allem dort zum Einsatz, wo es möglich erscheint, Einsichten zu definieren, die in speziellen Anwendungskontexten mit hoher Plausibilität zutreffen³. Solche non-deduktiven Plausibilitätsschlüsse spielen in der betrieblichen Praxis eine herausragende Rolle. Sie berücksichtigen nicht nur die syntaktische Struktur des verarbeiteten Wissens, sondern auch dessen semantischen Gehalt. Allerdings liegt es in der Eigenart solcher non-deduktiven Inferenzregeln, dass sie nur intuitiv einsichtige Schlüsse erlauben, aber nicht über die wahrheitserhaltende Stringenz der deduktiven Logik verfügen. Deswegen können sie in Einzelfällen auch zu fehlerhaften Schlussfolgerungen führen. Daher besteht ein „epistemischer Trade-off“ zwischen Folgerungsstrenge einerseits und Anwendungsbreite andererseits. Er stellt für das Design und die Anwendung von KMS eine große Herausforderung dar. Dies gilt nicht nur für ontologiebasierte KMS im Speziellen, sondern für inferenzfähige KMS generell.

Aufgrund der voranstehenden Erläuterungen sollte deutlich geworden sein, dass der Begriff „ontologiebasiert“ in Bezug auf die hier untersuchten KMS nur im Sinne eines *pars pro toto* verwendet wird. Das Attribut steht stellvertretend für KMS-Architekturen, die sich – je nach inhaltlicher Schwerpunktsetzung und verbalen Vorlieben der Autoren – als „ontologiebasierte“, „inferenzfähige“, „deduktive“ oder „intelligente“ KMS o.ä. bezeichnen lassen.

3. Motivationale Aspekte beim Einsatz von KMS

Der Einsatz von KMS stößt in der betrieblichen Praxis häufig auf den Widerstand der Mitarbeiter. Die Mitarbeiter erkennen oftmals nicht die Vorteile des Einsatzes eines KMS für das gesamte Unternehmen oder zumindest nicht für sich persönlich, sondern sie nehmen lediglich Nachteile für sich selbst wahr. Insbesondere die Motivation der Mitarbeiter zur Akzeptanz und Anwendung eines KMS spielt eine zentrale Rolle für dessen erfolgreichen Einsatz. Unter motivationalen Aspekten lassen sich vor allem die Problemfelder differenzieren, die in Abbildung 2 aufgeführt werden.

³ Zum Beispiel wäre es eine non-deduktive Schlussfolgerung, wenn man aus dem Sachverhalt, dass ein in einer bestimmten Sprache verfasster Quellcode einem bestimmten Autor zugeordnet ist, schließt, dass dieser Autor die bestimmte Sprache beherrscht. Diese Schlussfolgerung ist keineswegs allgemeingültig, weil der Autor vielleicht nicht über die erforderlichen Kenntnisse in der betroffenen Sprache verfügt und den Quellcode auch nicht verfasst hat, sondern aus anderen Gründen benannt wurde (z.B. weil er die Erstellung des Quellcodes – durch Dritte – verantwortet hat oder weil er schlicht ein Plagiat beging). Aktuelle Beispiele einiger wissenschaftlicher Arbeiten in der Forschung belegen die fehlende Allgemeingültigkeit solcher „plausibler“ Schlüsse. Dennoch führt die Schlussfolgerung im Allgemeinen zu einem verlässlichen und verwertbaren Ergebnis.

| Nr. | Problemfelder |
|-----|--|
| 1 | Einführung IT-System |
| 2 | Konsultation kompetenter Mitarbeiter |
| 3 | Evaluation der Arbeitsleistung |
| 4 | Versetzung kompetenter Mitarbeiter |
| 5 | Transparenz geschäftskritischer Kompetenzen |
| 6 | Einsatz entgegen den Präferenzen der Mitarbeiter |
| 7 | Unternehmensübergreifende Kooperationen |
| 8 | Funktionsweise und Ergebnisbegründung |

Abbildung 2: Motivationale Problemfelder beim Einsatz von KMS

Nachfolgend werden die Problemstellung eines jeden Feldes und Aspekte ihrer Lösung mittels KMS diskutiert. Neben der allgemeinen Problembeschreibung werden typische Probleme und Problemlösungspotenziale zunächst für konventionelle KMS skizziert. Anschließend wird – darauf aufbauend – erläutert, inwiefern ontologiebasierte KMS zu neuartigen Problemstellungen oder neuartigen Problemlösungsansätzen führen können.

Des Öfteren stößt man in diesem Zusammenhang auch auf das weitere Problemfeld der *Explikation von (tazitem) Wissen, das sich in den Köpfen der Mitarbeiter befindet*. Es wird befürchtet, dass ein Mitarbeiter für ein Unternehmen entbehrlich wird, sobald ein wissensbasiertes System über das Wissen dieses Mitarbeiters verfügt. Ein KMS verfügt jedoch vom Ansatz her lediglich über das (Meta-)Wissen über das (Objekt-)Wissen der Mitarbeiter, so dass ein KMS niemals zur befürchteten „kognitiven Enteignung“ der Mitarbeiter eines Unternehmens führen kann. Deshalb wird dieses Problemfeld in diesem Beitrag nicht näher untersucht.

3.1 Einführung eines IT-Systems

3.1.1 Problemstellung

Häufig stößt die Veränderung von Arbeitsabläufen auf den Widerstand der Mitarbeiter eines Unternehmens, weil sie Veränderungen grundsätzlich als Gefährdungen des vertrauten Status quo empfinden. Die Mitarbeiter befürchten in der Regel eine Arbeitszunahme – sei es durch Erlernen neuer Arbeitspraktiken oder durch eine Intensivierung der gewohnten Arbeitsabläufe – und daraus folgend eine Arbeitsüberlastung. Insbesondere bei der Einführung von IT-Systemen zur Restrukturierung von Arbeitsabläufen befürchten Mitarbeiter zusätzlich den Verlust von Kontrolle und Übersicht, langfristig sogar auch den Verlust ihrer Arbeitsplätze. Sind die Verantwortlichen für die Einführung eines IT-Systems in solchen Fällen nicht in der Lage, von Beginn an spürbare Effizienzgewinne für die betroffenen Mitarbeiter glaubhaft zu kommunizieren und die geschilderten Ängste zu zerstreuen, so ist die Systemeinführung in der Regel zum Scheitern verurteilt.

3.1.2 Konventionelle KMS

Um den erfolgreichen Einsatz eines KMS von Anfang an sicherzustellen, ist es notwendig, zu Beginn der Konzeptualisierungsphase ein Vorgehensmodell aufzusetzen. Hierbei sollte darauf geachtet werden, das Management eines Unternehmens von vornherein mit einzubinden, um ein dauerhaftes „Vorleben“ der Verwendung des IT-Systems durch das Management sicherzustellen. Des Weiteren müssen im Vorgehensmodell von Beginn an die späteren Anwender des Systems mit einbezogen werden, um frühzeitig Aversionen entgegen wirken zu können. Am Tag der Inbetriebnahme des Systems muss die Funktionalität des Systems so umfangreich wie möglich zur Verfügung stehen, damit das Wissen über den Nutzen des Systems sofort in den Köpfen der Mitarbeiter verankert wird.

3.1.3 Ontologiebasiertes KMS

Zur Konstruktion eines ontologiebasierten KMS müssen dessen spätere Anwender schon frühzeitig in die Systemgestaltung einbezogen werden, um die natürlichsprachlichen Begrifflichkeiten ihres Arbeitsumfelds in einer Ontologie zu repräsentieren und zusätzlich Inferenzregeln für Plausibilitätsschlüsse zu erfassen, die nur von den Fachleuten eines Arbeitsfelds inhaltlich beherrscht werden. Aufgrund dieser frühzeitigen Einbindung der Anwender in die Systemgestaltung besteht die begründete Aussicht, ihnen schon während der Entwicklungsphase zu verdeutlichen, welchen Funktionsumfang das System haben wird. Hinzu kommt, dass ein ontologiebasiertes System von Anfang an gehaltvollere Antworten auf Anfragen liefern kann, weil es aufgrund seiner Inferenzfähigkeiten zusätzlich implizites Wissen zu explizieren vermag. Das bedeutet, dass ein solches System tendenziell eher als ein konventionelles KMS in der Lage sein wird, von Anfang an Effizienzgewinne für seine Mitarbeiter zu realisieren. Außerdem können Ängste dadurch abgebaut werden, dass den Mitarbeitern ein Gefühl für die Wichtigkeit ihres Wissens für Aufbau und Nutzung des KMS vermittelt wird, weil nur aufgrund ihrer aktiven Mitwirkung die arbeitsplatzbezogenen Ontologien und die zugehörigen non-deduktiven Plausibilitätsregeln erhoben sowie fortentwickelt werden können. Die kontraproduktive Furcht vor „kognitiver Enteignung“ lässt sich – wie bereits kurz zuvor dargelegt wurde – dadurch vermeiden, dass die Einschränkung ontologiebasierter KMS auf Meta-Wissen hervorgehoben wird.

3.2 Konsultation kompetenter Mitarbeiter

3.2.1 Problemstellung

Aus Mitarbeiterperspektive besteht eine Gefahr beim Einsatz eines KMS darin, dass sie zu häufig von anderen Akteuren konsultiert werden. Dazu kann es kommen, weil diese anderen Akteure aufgrund der im KMS dokumentierten Kompetenzen hoffen, Antworten auf ihre Anfragen zu erhalten, die sie benötigen, um ihre eigenen Probleme zu lösen. Derartige Anfragen werden aus der Perspektive des konsultierten Mitarbeiters unter Umständen in zweifacher Hinsicht als nicht willkommen empfunden. Zum einen muss Arbeitszeit zur Beantwortung der Anfragen aufgewendet werden, die zur Erledigung der eigenen Arbeit fehlt. Dieses Problem wird verstärkt durch ungenaue Anfragen sowie un-

vollständige und gegebenenfalls falsche Erfassung der Kompetenzen im KMS. In solchen Fällen können die Anfragen vom konsultierten Mitarbeiter nicht beantwortet werden, so dass er sinnlos Arbeitszeit aufwendet. Zum anderen fühlen sich Mitarbeiter genötigt, subjektgebundenes Wissen zu explizieren. Daran besteht jedoch kein Interesse, weil subjektgebundenes Wissen nach vorherrschender Meinung die eigene Position im Unternehmen sichert.

3.2.2 Konventionelle KMS

Um den zuvor skizzierten Problemen entgegenzuwirken, können die Zugriffsrechte auf das Wissen über Kompetenzen im KMS beschränkt werden. Beispielsweise kann es nur Führungskräften erlaubt sein, sich die Kompetenzen anzeigen zu lassen. Dadurch wird die Zahl der Anfragen erheblich reduziert. Eine andere Möglichkeit besteht darin, ein Anreizsystem zu schaffen, das Mitarbeiter für beantwortete Anfragen belohnt. So können für die Beantwortung von Anfragen Punkte vergeben werden, die auf einem Konto gesammelt werden und sich zur Beantwortung eigener Fragen einlösen lassen (zur Idee eines Marktplatzes für Wissen vgl. [Be02]).

3.2.3 Ontologiebasierte KMS

Ein ontologiebasiertes KMS kann darüber hinaus gewährleisten, dass die Güte der Dokumentation von Wissen über Kompetenzen im KMS im Hinblick auf Vollständigkeit und Validität erhöht wird. Eine in diesem Sinn erhöhte Güte lässt sich erreichen, indem nicht nur auf explizites Wissen zurückgegriffen wird, das in einer Kompetenzdatenbank per Hand eingepflegt wurde, wie z.B. die Dokumentation der Aus- und Fortbildung oder die Mitgliedschaft in Gremien, sondern auch Dokumente – wie z.B. Projektberichte oder technische Dokumentationen – (nach Möglichkeit) automatisch ausgewertet werden. Diese Auswertung erfolgt mit Hilfe einer Inferenz-Maschine, die beispielsweise aus einer technischen Dokumentation eines elektronischen Bauteils schließt, dass sein Konstrukteur über die Kompetenz verfügt, die Funktionsweise dieses Bauteils zu erläutern ([AI02]). Durch die Automatisierbarkeit des Durchsuchens von Dokumenten (Data-Mining, Wrapper u.ä.) und des Erschließens von Wissen über Kompetenzen aus den Inhalten der durchsuchten Dokumente lässt sich zudem tendenziell eine höhere Aktualität der Wissensbasis im Vergleich zu einem konventionellen KMS gewährleisten, in dem die Wissensbasis händisch gepflegt werden muss. Dadurch kann im ontologiebasierten KMS sichergestellt werden, dass Mitarbeiter mit den gesuchten Kompetenzen schneller gefunden werden; und vermieden werden, dass Anfragen unbeantwortet bleiben oder auf Leistungen Unternehmensexterner unnötig zurückgegriffen wird.

Probleme können sich gegebenenfalls dadurch ergeben, dass die Schlussfolgerungen der Inferenz-Maschine nicht stringent sein müssen, wenn sie auf non-deduktiven Plausibilitätsregeln beruhen (s.o.). Die Belastbarkeit von Aussagen lässt sich jedoch erhöhen, indem man nicht einer einzelnen Schlussfolgerung vertraut, sondern einem Bündel von Schlussfolgerungen, die zum gleichen Ergebnis führen. Eine analoge Struktur findet sich beim Vertrauen in zwischenmenschlichen Beziehungen. Dort tendieren die Akteure zur Vertrauensbildung, wenn nicht ein einzelner, sondern ein Bündel von Vertrauensindikatoren⁴ positiv ausgeprägt ist.

3.3 Evaluation der Kompetenzen

3.3.1 Problemstellung

Aus der Mitarbeiterperspektive gibt es oftmals Befürchtungen, dass ein KMS eingesetzt wird, um Kompetenzen zu evaluieren. Hieraus werden negative Auswirkungen abgeleitet, wie beispielsweise die Nichtberücksichtigung bei anstehenden Beförderungen. Oftmals wirken sich die Evaluationen der Kompetenzen von Mitarbeitern auch demotivierend aus, wenn intersubjektiv nachvollziehbare Begründungen für die Kompetenzurteile vermisst werden.

Aus den vorgenannten Gründen kann der Einsatz von KMS auf Widerstand bei Gewerkschaften stoßen. Ebenso werden oftmals datenschutzrechtliche Bedenken gegen den Einsatz von KMS angeführt (zur Beurteilung von KMS aus Sicht von Gewerkschaften und Datenschutz vgl. [HZ00]).

3.3.2 Konventionelle KMS

Einerseits kann eine Evaluation der Kompetenzen von Mitarbeitern durch die Unternehmensleitung von vornherein ausgeschlossen werden, um bei den Mitarbeitern die Akzeptanz für den Einsatz eines KMS zu erhöhen. Dies kann so weit führen, dass das KMS nicht die Möglichkeit bietet, den Datensatz eines bestimmten Mitarbeiters anzusehen, sondern nur die Option besteht, nach Kompetenzen zu suchen. Als Ergebnis der Suche erhält der Anwender eine Liste, welche die (im strengsten Falle: anonymisierten) Mitarbeiter mit der gesuchten Kompetenz auflistet.

Andererseits lässt sich auch versuchen, den Mitarbeitern einen Anreiz zu bieten, damit sie die Verwendung eines KMS zur Evaluation ihrer Kompetenzen akzeptieren. Ein derartiger Anreiz kann beispielsweise geschaffen werden, indem sich zusätzliche, im KMS hinterlegte Kompetenzbewertungen für den jeweiligen Mitarbeiter in Form höherer Bezüge oder in Form erweiterter Karrieremöglichkeiten lohnen.

⁴ Ein Beispiel für einen Vertrauensindikator stellt die Übereinstimmung der Ankündigung einer Handlung mit der tatsächlichen Durchführung der angekündigten Handlung dar. Liegt eine derartige Übereinstimmung vor, so wird dem Akteur eher vertraut, als wenn dieser entgegen seiner Ankündigung handelt.

Um die Urteilsfindung bei den Evaluationen nicht demotivierend wirken zu lassen, ist es notwendig, das Vorgehen so transparent wie möglich zu gestalten. Werden Evaluationen von Dritten durchgeführt, so bleibt stets ein „Restrisiko“, dass diese Fremdevaluationen als nicht einwandfrei intersubjektiv nachvollziehbar empfunden werden. Daher wird des Öfteren empfohlen, in erster Linie auf das Instrument der Selbsteinschätzung zurückzugreifen, bei dem die einzelnen Mitarbeiter selbst für die jeweiligen Ausprägungen der Kompetenzurteile verantwortlich zeichnen.

3.3.3 Ontologiebasierte KMS

Insbesondere bei einem ontologiebasierten KMS müssen die vom KMS erkannten neuen Kompetenzen hinsichtlich ihrer Relevanz durch Dritte (beispielsweise Wissensmanager oder die Personalabteilung) evaluiert werden. Diese Evaluation ist erforderlich, um bei den „automatisch“ vom KMS ermittelten Kompetenzen zum einen den (Datenschutz-) Richtlinien zu entsprechen und zum anderen Vertrauen innerhalb des Unternehmens in die Güte der Dokumentation der Kompetenzen aufzubauen. Generell sollte das KMS nur neue Kompetenzen ermitteln, die ausschließlich als positive Erweiterung der Einsatzfähigkeit eines Mitarbeiters angesehen werden können. Wie bereits erwähnt, sollte das System auf die Selbsteinschätzung des betroffenen Mitarbeiters hinsichtlich der Ausprägung der ermittelten Kompetenz zurückgreifen.

3.4 Versetzung kompetenter Mitarbeiter

3.4.1 Problemstellung

Aus der Perspektive von Projekt- und Abteilungsleitern kann der Einsatz eines KMS als nachteilig empfunden werden. Es wird befürchtet, der Einsatz des KMS mache Kompetenzen von Mitarbeitern in ihren Projekten bzw. Abteilungen für andere Führungskräfte transparent. Dadurch erwachse die Gefahr, dass sie eher kompetente Mitarbeiter aus ihren Projekten bzw. Abteilungen verlieren als gewinnen und ihnen somit Nachteile bei ihrer zukünftigen Arbeit entstehen.

3.4.2 Konventionelle und ontologiebasierte KMS

Konventionelle und ontologiebasierte KMS unterscheiden sich hinsichtlich ihrer Problemlösungspotenziale für das hier angesprochene Problem nicht. Grundsätzlich sollte nach der Maßgabe gehandelt werden, dass der einzelne Mitarbeiter dort eingesetzt wird, wo er die größtmögliche Produktivität erreichen kann. So ist es nur natürlich, dass kompetente Mitarbeiter fluktuieren. Deshalb muss in solchen Fällen als erstes versucht werden, die Unternehmenskultur dahin zu entwickeln, dass der Wechsel in ein neues Projekt oder eine neue Abteilung nicht als schädlich, sondern als „besser“ für das Gesamtunternehmen empfunden wird. Sollte dieses nicht praktikabel sein, so können derartige Befürchtungen durch organisatorische Regelungen entkräftet werden. Beispielsweise kommt eine Regel in Betracht, dass ein Mitarbeiter immer für die gesamte Projektlaufzeit einem Projekt zugeordnet bleibt. Ein anderer denkmöglicher Ansatz besteht darin, den

Abzug eines kompetenten Mitarbeiters von einem Projekt durch eine Gegenleistung – wie z.B. eine Erhöhung des Projektbudgets oder eine Verlängerung der Projektlaufzeit – zu kompensieren.

3.5 Transparenz geschäftskritischer Kompetenzen

3.5.1 Problemstellung

Die Unternehmensleitung hat ein Interesse daran, dass Kompetenzen bestimmter Mitarbeiter für andere Mitarbeiter nicht evident werden. So können die im KMS dokumentierten Kompetenzen Aufschluss darüber geben, welche Mitarbeiter in geschäftskritische Prozesse des Unternehmens involviert sind. Aus Angst vor Missbrauch des KMS – wie z.B. durch Spionage und/oder Wissensweitergabe an Headhunter – kann die Unternehmensleitung dazu tendieren, den Einsatz eines KMS abzulehnen.

3.5.2 Konventionelle KMS

Um einem Missbrauch des KMS entgegenzuwirken, lässt sich durch entsprechende Zugriffsrechte sicherstellen, dass nur eine geringe Anzahl an Mitarbeitern – z.B. die Unternehmensleitung – Zugriff auf das KMS besitzt. Des Weiteren können die Zugriffsrechte so ausgestaltet werden, dass nicht alle Mitarbeiter alle Kompetenzen angezeigt bekommen.

3.5.3 Ontologiebasierte KMS

Zu diesem Punkt lässt sich kein spezifisches Problemlösungspotenzial ontologiebasierter KMS ermitteln, das über das Potenzial eines konventionellen KMS hinausgeht. Allerdings kommen zur Repräsentation von Ontologien in der Regel Skriptsprachen zum Einsatz (wie z.B. F Logic oder XML), die sich im Allgemeinen leichter nachvollziehen lassen, als es bei einem kompilierten Programmcode der Fall ist. Deshalb müssen insbesondere die Server, auf denen eine Ontologie schließlich hinterlegt wird, vor missbräuchlichem Zugang gesichert werden.

3.6 Einsatz entgegen den Präferenzen der Mitarbeiter

3.6.1 Problemstellung

Aus der Mitarbeiterperspektive kann die erhöhte Transparenz der Kompetenzen beim Einsatz eines KMS abgelehnt werden, wenn die Mitarbeiter befürchten, dass sie von Führungskräften entgegen ihren eigenen Präferenzen in bestimmten Projekten und/oder Abteilungen aufgrund der im KMS vorhandenen Kompetenzprofilaten eingesetzt werden.

Beispielsweise kann eine Führungskraft anhand des KMS erfahren, dass ein bestimmter Ingenieur über Kompetenzen im internen Rechnungswesen verfügt. Aufgrund einer Kapazitätsüberauslastung im internen Rechnungswesen des Unternehmens wird der Ingenieur – entgegen seinen Präferenzen – aus der Forschungs- & Entwicklungsabteilung des Unternehmens mit demotivierender Wirkung in die Abteilung für internes Rechnungswesen versetzt. Die Gefahr des Einsatzes entgegen den eigenen Präferenzen ist beim Verzicht auf die Nutzung eines KMS möglicherweise noch größer. Ein solcher Nichteinsatz von KMS gehört jedoch nicht zum Untersuchungsgegenstand dieses Beitrags.

3.6.2 Konventionelle KMS

Um die Akzeptanz des KMS unter den Mitarbeitern zu erhöhen, wird die Möglichkeit gegeben, Präferenzen zu den im KMS dokumentierten Kompetenzen zu hinterlegen. Für das o.a. Beispiel des Ingenieurs heißt dies, es müsste ihm möglich sein, im KMS zu hinterlegen, dass er einen Einsatz in der Forschungs- & Entwicklungsabteilung gegenüber einem Einsatz in anderen Abteilungen präferiert. Bei der Planung, in welchen Projekten und/oder Abteilungen der jeweilige Mitarbeiter eingesetzt werden soll, können seine Präferenzen dann berücksichtigt werden. Durch den Einsatz des Mitarbeiters gemäß seinen Präferenzen soll erreicht werden, dass der Mitarbeiter seine Aufgaben mit einer höheren Motivation – und somit tendenziell effizienter – erledigt, als wenn er entgegen seinen Präferenzen eingesetzt wird.

Neben der bereits skizzierten Möglichkeit der Hinterlegung von Präferenzen der Mitarbeiter im Hinblick auf ihre Kompetenzen könnte die Akzeptanz der Mitarbeiter erhöht werden, indem ihr Einsatz in von ihnen nicht präferierten Abteilungen und/oder Projekten belohnt wird. Dazu kommen beispielsweise materielle Anreize wie höhere Bezüge oder auch immaterielle Anreize wie die Teilnahme an Fortbildungen in Betracht.

3.6.3 Ontologiebasierte KMS

Der Einsatz ontologiebasierter KMS kann aus der Perspektive der Mitarbeiter insofern als positiv erachtet werden, als dass durch das ontologiebasierte KMS zusätzliche Kompetenzen expliziert und somit für Führungskräfte wahrnehmbar werden, die ohne das ontologiebasierte KMS den Führungskräften verborgen geblieben wären. Dies trifft jedoch nur so weit zu, wie es sich um Kompetenzen handelt, aufgrund derer ein Mitarbeiter gern eingesetzt werden würde. Demgegenüber stehen Befürchtungen der Mitarbeiter vor negativen Auswirkungen. Denn das ontologiebasierte KMS könnte Kompetenzen ermitteln, von denen sich der jeweilige Mitarbeiter nicht wünscht, dass diese transparent gemacht werden, weil ein Einsatz gemäß diesen Kompetenzen nicht seinen Präferenzen entspricht. Beispielsweise könnte der o.a. Ingenieur sich wünschen, dass seine kaufmännischen Kompetenzen nicht durch das ontologiebasierte KMS expliziert werden, um einem Einsatz im kaufmännischen Bereich zu entgehen. Um eine einwandfreie Feststellung der Präferenzen zu gewährleisten, bietet es sich an, nach der automatischen Erschließung einer „neuen“ Kompetenz durch das System eine Nachricht an den betroffenen Mitarbeiter mit der Aufforderung zu senden, sich hinsichtlich dieser Kompetenz – insbesondere der Selbsteinschätzung ihrer Ausprägung – und seiner dazugehörigen Präferenz zu äußern. Anschließend werden die Kompetenz, ihre Ausprägung und die dazu-

gehörige Präferenz des Kompetenzträgers dem KMS zur Verfügung gestellt.

3.7 Unternehmensübergreifende Kooperationen

3.7.1 Problemstellung

In der betrieblichen Praxis kann es im Rahmen unternehmensübergreifender Kooperationen mit anderen Unternehmen dazu kommen, dass Mitarbeiter zwischen einzelnen Unternehmen ausgetauscht werden. Dies kann beispielsweise der Fall sein, wenn ein Zulieferer nicht über Mitarbeiter mit einer Kompetenz verfügt, die benötigt wird, um sein Bauteil in ein Produkt seines Abnehmers zu integrieren. Dann kann der Zulieferer durch Recherche im KMS, in dem die Kompetenzprofile der an der Kooperation beteiligten Unternehmen abgelegt sind, ermitteln, dass der Abnehmer Mitarbeiter beschäftigt, die über die benötigte Kompetenz verfügen. Beim Einsatz eines derartigen unternehmensübergreifender KMS deutet sich an, dass die in den Abschnitten 3.2 bis 3.6 dargestellten Probleme verschärft werden können, weil sich beispielsweise die Grundgesamtheit an Mitarbeitern erhöht. Durch eine größere Grundgesamtheit intensiviert sich der Wettbewerb bei der Evaluation der Kompetenzen. Ferner können Mitarbeiter – gegen ihren Willen – in einem anderen Unternehmen eingesetzt werden, sofern das entsendende Unternehmen über den externen Einsatz seiner Mitarbeiter entsprechend disponiert.

3.7.2 Konventionelle KMS

Zum einen können hier wieder die Möglichkeiten zur Problemlösung angeführt werden, die bereits in den vorangegangenen Abschnitten dargestellt wurden. Darüber hinaus lässt sich vorstellen, Mitarbeiter durch Anreize – wie z.B. höhere Bezüge oder die Teilnahme an Fortbildungen – zu ihrem Einverständnis zu motivieren, dass kooperierende Unternehmen auf ihre Kompetenzprofile zugreifen können. Die gleichen Anreize können auch für den späteren Einsatz in einem anderen Unternehmen gewährt werden. Ein Mitarbeiter kann auch von sich aus den Einsatz in einem anderen Unternehmen als positiv werten, weil z.B. hiermit prestigehaltige Geschäftsreisen oder Gelegenheiten verbunden sind, für die eigene Karriere förderliche Kontakte in anderen Arbeitsumgebungen zu knüpfen. Analog zur Hinterlegung von Präferenzen im Hinblick auf Kompetenzen sollte auch hier den Mitarbeitern die Möglichkeit gegeben werden, im KMS zu hinterlegen, ob sie den Einsatz in anderen Unternehmen begrüßen würden.

Darüber hinaus ist zu berücksichtigen, dass aufgrund „neuerer“, im KMS abrufbarer Kompetenzen eines Unternehmens von dessen Kooperationspartnern auf aktuelle oder zukünftige Konkurrenzvorteile des Unternehmens geschlossen werden kann. Dies kann insbesondere im Fall von Kernkompetenzen die Wettbewerbsfähigkeit eines Unternehmens ernsthaft gefährden, so dass auf der Ebene inter-organisatorischer Beziehungen ein großer Anreiz besteht, den Zugriff auf das eigene KMS durch Kooperationspartner grundsätzlich zu verhindern. Andernfalls müssen erhebliche Sicherheitsvorkehrungen ergriffen werden, wie z.B. spezielle Freigaberegeln und Zugriffskontrollen.

3.7.3 Ontologiebasierte KMS

Die Anwendungsmöglichkeiten eines ontologiebasierten KMS sind bei unternehmensübergreifenden Kooperationen eingeschränkt, weil ein Unternehmen seinem Kooperationspartner allenfalls nur Zugriff auf das Wissen über Kompetenzen geben wird, das im KMS vorliegt. Dagegen wird im Regelfall kein erweiterter Zugriff auf sämtliche Dokumente des Unternehmens gewährt werden. Daher können ontologiebasierte KMS ihre spezifische Stärke, implizit vorhandenes Wissen über Kompetenzen mittels Inferenzen aus dem gesamten zugreifbaren Dokumentenbestand eines Unternehmens zu erschließen, in Unternehmenskooperationen gewöhnlich nicht ausspielen. Sie bieten also in dieser Hinsicht keine nennenswerten Vorzüge gegenüber konventionellen KMS.

3.8 Funktionsweise und Ergebnisbegründung von KMS

3.8.1 Problemstellung

KMS wirken auf Akteure in Unternehmen teilweise befremdlich, weil ihre Funktionsweisen oftmals nur schwer nachvollziehbar erscheinen. Es besteht somit die Gefahr, dass die von einem KMS angezeigten Ergebnisse (Kompetenzen) bei den Akteuren auf wenig Akzeptanz stoßen. Insbesondere dann, wenn die Begründung von ermittelten Ergebnissen im Unklaren bleibt und daher durch die Anwender nicht nachvollzogen werden kann, kommt es zu Akzeptanzproblemen. Mangelnde Akzeptanz eines KMS motiviert nicht zur Benutzung. In diesem Fall besteht die Gefahr, dass das KMS zu einer Investitionsruine degeneriert.

3.8.2 Konventionelle KMS

Um die Funktionsweise von KMS für die Benutzer transparent zu gestalten, können Schulungen durchgeführt werden. Ein KMS kann durch Dokumentationen, Hilfefunktionen und interaktive Lernkomponenten mit entsprechenden Erläuterungen ergänzt werden. Darüber hinaus lässt sich die Akzeptanz fördern, indem die Darstellung der Ergebnisse an der Benutzerschnittstelle anschaulich gestaltet wird. Beispielsweise lassen sich Kompetenzen graphisch in der Form von intuitiv übersichtlichen Kompetenzdiagrammen visualisieren.

3.8.3 Ontologiebasierte KMS

Bei einem ontologiebasierten KMS verstärken sich die Probleme hinsichtlich der Ergebnisbegründung im Vergleich zu einem konventionellen KMS. Denn zum Verständnis der Funktionsweise eines ontologiebasierten KMS reicht es – im Gegensatz zu einem konventionellen KMS – nicht aus zu wissen, wie eine Datenbank funktioniert. Stattdessen müsste der Anwender eines ontologiebasierten KMS auch über Inferenzmechanismen im Allgemeinen sowie über die speziellen, oftmals non-deduktiven Inferenzregeln des KMS Bescheid wissen. Dieses Wissen über Aspekte des Knowledge Engineerings kann jedoch von betrieblichen Mitarbeitern in der Regel nicht erwartet werden – selbst dann nicht, wenn sie mit der Funktionsweise gewöhnlicher Datenbanken vertraut sind. Weil implizi-

tes Wissen expliziert wird, fällt das Nachvollziehen von Ergebnissen eines ontologiebasierten KMS besonders schwer. Deshalb wird es notwendig, im System eine Erklärungskomponente [Ha00] zu verankern, welche die Inferenz-Maschine mit besonderem erklärungsbefähigenden Wissen und dem Abfrage-Interface des KMS verknüpft. Diese Komponente verdeutlicht auf Wunsch des Anwenders – unter Umständen sogar grafisch unterstützt – zumindest den Weg, auf dem eine Antwort für seine Anfrage generiert wurde, und gewährt ihm somit die Möglichkeit, das Ergebnis hinsichtlich seiner Plausibilität zu prüfen. Gerade hierin liegt eine Stärke von ontologiebasierten KMS, weil es für den Anwender nachvollziehbar ist, unter Anwendung welcher Regeln das System zu seinem Ergebnis gekommen ist. Das KMS ermöglicht hierdurch eine anschauliche Darstellung seiner Schlussfolgerungen.

4. Fazit

Der vorliegende Beitrag zeigt, dass sich motivationale Probleme beim Einsatz von KMS zum Teil innerhalb konventioneller KMS lösen lassen. Durch den Einsatz ontologiebasierter KMS ist es möglich, weitere Probleme zu überwinden. Dies beruht insbesondere auf der höheren Güte des dokumentierten und/oder erschließbaren Wissens über Kompetenzen, wie anhand von Präferenzen der Kompetenzträger bzw. mittels plausibilitätsgeleiteter Inferenzen aus implizitem Wissen exemplarisch belegt wurde. Es wird jedoch auch deutlich, dass ontologiebasierte KMS für einige Problemstellungen kein eigenständiges Problemlösungspotenzial bieten und dass ihr Einsatz teilweise sogar weitere motivationale Probleme hervorbringt, die über die Probleme konventioneller KMS hinausgehen. Für diese neuartigen Motivationsprobleme müssen zum einen organisationale und arbeitsrechtliche sowie zum anderen system-konzeptionelle Lösungen erst noch gefunden werden.

Literaturverzeichnis

- [Al02] Alparslan, A.; Dittmann, L.; Ilgen, A.; Zelewski, S.: Wissensmanagement im Anlagenbau. In: *Industrie Management*, 6 (18), 2002, S. 45-48.
- [Be02] Benjamins, V.R.; Cobo, J.M.L.; Contreras, J. et al.: *Skills Management in Knowledge Intensive Organizations*. ISOCO Whitepaper, Madrid 2002.
- [DLW99] Deiters, W.; Lucas, R.; Weber, T.: *ISST-Bericht 50 – Skill-Management: Ein Baustein für das Management flexibler Teams*. Dortmund, 1999.
- [El01] Elbert, S.: Einführung eines Management-Support-Systems zum effektiven Skill-Management bei Bertelsmann mediaSystems. In (Schnurr, H.-P.; Staab, S.; Studer, R.; Stumme, G.; Sure, Y. Hrsg.): *Professionelles Wissensmanagement – Erfahrungen und Visionen*. Shaker Verlag, Aachen, 2001, S. 129-144.
- [Fa98] Faulstich, P.: *Strategien der betrieblichen Weiterbildung – Kompetenz und Organisation*. Verlag Franz Vahlen, München, 1998.
- [Fe01] Fensel, D.: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer Verlag, Berlin, 2001.
- [Ha00] Haun, M.: *Wissensbasierte Systeme – eine praxisorientierte Einführung*. Expert-Verlag, Renningen, 2000.
- [HZ00] Hüneke, K.; Zimmermann, B.: Skill-Datenbanken. In: *Computer Fachwissen*, 8/9 (9), 2000, S. 51-55.
- [Ka82] Kaiser, P.: *Kompetenz als erlernbare Fähigkeit zur Analyse und Bewältigung von Lebenssituationen auf mehreren Ebenen*. BIS-Verlag, Oldenburg, 1982.
- [Ma02] Maedche, A.: *Skillmanagement der nächsten Generation*. Vortrag: AK Wissensmanagement 2002/06/01, Karlsruhe, 2002.
- [No98] North, K.: *Wissensorientierte Unternehmensführung – Wertschöpfung durch Wissen*, Gabler Verlag, Wiesbaden, 1998.
- [St99] Studer, R.; Abecker, A.; Decker, S.: *Informatik-Methoden für das Wissensmanagement*. In (Lausen, G.; Oberweis, A.; Schlageter, G. Hrsg.): *Angewandte Informatik und Formale Beschreibungsverfahren*. Teubner Verlag, Stuttgart, 1999, S. 263-274.
- [Su00] Sure, Y.; Maedche, A.; Staab, S.: *Leveraging Corporate Skill Knowledge – From ProPer to OntoProPer*. In (Mahling, D.; Reimer, U. Hrsg.): *Proceedings of the Third International Conference on Practical Aspects of Knowledge Management*. Basel, 2000, S. 221-229.

Ontologiebasierte Wissensräume *

Dipl.-Kfm. Yilmaz Alan
Univ.-Prof. Dr. Stephan Zelewski

Institut für Produktion und Industrielles Informationsmanagement
Universität Duisburg-Essen
Universitätsstraße 9
45141 Essen
Yilmaz.Alan@pim.uni-essen.de
Stephan.Zelewski@pim.uni-essen.de

Abstract: Die Theorie der Wissensräume von DOIGNON/FALMAGNE wird auf ihre Verwertbarkeit als formale Grundlage für ontologiebasierte Kompetenzprofile untersucht. Der Fokus liegt hier zum einen in der Spezifikation von Begriffen, mittels derer die Kompetenzprofile von Akteuren beschrieben werden können. Zum anderen wird von der Konstruktion von Surmise-Systemen eine Erhöhung der Effizienz von Wissensakquisition, -strukturierung und -repräsentation durch Inferenzregeln erwartet.

1. Problemstellung

Um innerbetriebliche Erfolgspotenziale ausschöpfen zu können, versuchen Unternehmen zunehmend, Tätigkeiten in ihre Geschäftsprozesse zu integrieren, von denen sie sich einen Beitrag zum *Wissensmanagement* erhoffen. Es wird hierbei versucht, die teilweise personengebundenen Kenntnisse und Fähigkeiten von Teams und Abteilungen dem gesamten Unternehmen zur Verfügung zu stellen. Untersuchungen reichen hierbei von soziokulturellen Arbeiten über organisationstheoretische Analysen bis hin zu IT-Lösungen für das Wissensmanagement. Im Rahmen der Erforschung von IT-Lösungen für das Wissensmanagement haben *ontologiebasierte* Ansätze an Bedeutung zugenommen [St02]. In den letzten Jahren haben sich Unternehmen zunehmend zum Ziel gesetzt, Ontologien zur Strukturierung der Begrifflichkeiten zu verwenden, mit denen das Wissen über die Kompetenzen ihrer Mitarbeiter („personal knowledge“) und ihrer Organisation insgesamt („organizational knowledge“) beschrieben werden kann. Ein besonderer Schwerpunkt wurde hierbei auf ontologiebasierte Lösungen für *Kompetenzmanagement-Systeme* gelegt [LaSu02]. Von der formalen Spezifikation der begrifflichen Strukturierung betrieblicher Kompetenzen wird dabei erhofft, eine Grundlage für die effektive Kommunikation auf der Basis eines gemeinsamen Begriffsverständnisses zu schaffen.

Ein großer Vorteil, der mit dem Einsatz von Ontologien einhergeht, ist die Möglichkeit der Definition von *Inferenzregeln*, mittels derer „neue“ Fakten zu einer ontologiebasierten Wissensbasis hinzugefügt werden können. Die *inhaltliche* Spezifizierung von Inferenzregeln, die in einem Kompetenzmanagement-System eingesetzt werden können, ist allerdings bislang größtenteils unerforscht geblieben. Mit dem vorliegenden Beitrag wird ein Ansatz aus der mathematischen Psychologie vorgestellt, mit dem diese Spezifizierungslücke – zumindest teilweise – geschlossen werden kann.

Dieser Beitrag gliedert sich wie folgt: In Kapitel 2 werden die konzeptionellen Grundlagen erläutert. Als erstes wird dabei auf den Einsatz von Kompetenzprofilen im Rahmen des betrieblichen Wissensmanagements eingegangen. Als zweites wird das Konzept der Wissensräume erläutert. Anschließend werden Ontologien als formale Grundlage für Begriffssysteme vorgestellt, die sich in computergestützten Kompetenzmanagement-Systemen anwenden lassen. In Kapitel 3 werden Auszüge aus einer prototypischen Ontologie für Kompetenzmanagement-Systeme vorgestellt, in der das Konzept der Wissensräume implementiert wurde. Die Arbeit wird mit einem Ausblick in Kapitel 4 abgeschlossen.

*) Kurzfassung eines Beitrags, der in längerer Fassung als Projektbericht 3/2003, Projekt KOWIEN (Kooperatives Wissensmanagement in Engineering-Netzwerken) vorliegt. Nähere Informationen zu diesem Projekt, das vom BMBF finanziell gefördert und vom Projektträger Kernforschungszentrum Karlsruhe GmbH betreut wird (Förderkennzeichen Hauptband 02 PD1060), finden sich im Internet unter der URL: <http://www.kowien.uni-essen.de>.

2. Konzeptionelle Grundlagen

2.1. Kompetenzprofile im Rahmen des betrieblichen Wissensmanagements

Im Rahmen des *Resource Based View* [Gr91] werden unternehmensinterne Ressourcen als Vorlaufgrößen für zukünftige Erfolgspotenziale anerkannt. In der Ausprägung des *Competence Based View* [CoMc02, Fr00, HePr00] erfolgt dabei eine Konzentration auf eine spezielle Kategorie von Ressourcen, die als „intangible assets“ zum Human- und Wissenskapital eines Unternehmens gehören und zumeist als *Kompetenzen* (oder *Kernkompetenzen*) thematisiert werden. Unter Kompetenzen werden hier sämtliche *handlungsbefähigenden* Kenntnisse (deklarative Wissenskomponente) und Fähigkeiten (prozedurale Wissenskomponente) verstanden. *Kompetenzprofile* stellen eine Agglomeration von Kompetenzen in Bezug auf einen Akteur dar. Als Akteure kommen im Bereich des Wissensmanagements sowohl Individuen als auch Kollektive (Teams, Abteilungen, Unternehmen als Gesamtheiten) in Betracht.

Die Motivation für den Einsatz von Kompetenzprofilen im Rahmen des betrieblichen Wissensmanagements ist vielschichtig [El01]. Es verschafft Transparenz über die unternehmensintern verfügbaren Kompetenzen, ohne die Explikation des zugrunde liegenden (Objekt-)Wissens einzufordern. Den Nutzern eines Kompetenzprofils wird lediglich ein Hinweis auf Kompetenzträger gegeben. Kompetenzprofile sind somit *Wissensbasen für Metawissen*, die Wissen *über* handlungsbefähigendes Objektwissen und die zugehörigen Wissensträger repräsentieren, ohne selbst das jeweils betroffene Objektwissen der Wissensträger zu enthalten.

Kompetenzprofile geben Auskunft darüber, welche Kompetenzen von welchen Akteuren im Unternehmen in welchen Ausprägungen vorliegen. Dieses Wissen kann beispielsweise genutzt werden, um Kompetenzlücken zu identifizieren. So können mit der Hilfe von Kompetenzprofilen Weiterbildungsmaßnahmen des Unternehmens zielgerichteter geplant werden. Darüber hinaus bieten Kompetenzprofile die Möglichkeit, ein informelles Netzwerk aufzubauen, welches nicht auf zufälligen Begegnungen beruht, sondern die Kompetenznachfrager mit den Kompetenzanbietern auf direktem Weg zusammenführt. Durch Kompetenzprofile wird Akteuren die Möglichkeit geboten, andere Akteure auszumachen, deren Kompetenzen für eine aktuelle oder zukünftige Aufgabe erforderlich sind. Auf personaler Ebene kann dies zum Beispiel in einer Situation erfolgen, in der ein Mitarbeiter für eine konkrete Fragestellung die Hilfe von Kollegen braucht, die über die dafür nötigen Kompetenzen verfügen.

Insbesondere im *Projektgeschäft* erlangen Kompetenzprofile eine besondere Bedeutung. Durch sie wird der Projektleitung ermöglicht, für wissensintensive Teilprozesse hinsichtlich mehrerer Punkte einen „strategic fit“ zwischen prozessinduzierten Wissensanforderungen einerseits und verfügbarem handlungsbefähigenden Wissen andererseits zu gewährleisten. Kompetenzprofile unterstützen z.B. den Teilprozess der Teambildung, indem sie Auskünfte darüber ermöglichen, welche potenziellen Teammitglieder welche Kompetenzen in ein Team einbringen könnten. Somit kann die teambildende Instanz die Konfiguration eines Teams auf der Grundlage einer breiten Wissensbasis vornehmen. Kompetenzprofile sind ebenso in der Durchführungsphase eines Projekts von wesentlicher Bedeutung. Beispielsweise kann bei einem Ausfall eines Teammitglieds schnell nach Ersatz gesucht werden, der zumindest über die im Projekt benötigten Kompetenzen des Teammitglieds verfügt. Des Weiteren bieten sich Kompetenzprofile für die Suche nach ergänzenden Teammitgliedern an, wenn in einer Phase der Projektdurchführung erkannt wird, dass die derzeitige (Kompetenz-) Konfiguration des Projektteams für die Problembewältigung nicht mehr ausreicht.

2.2. Konzept der Wissensräume

Das Konzept der Wissensräume lässt sich auf Arbeiten von DOIGNON & FALMAGNE zurückführen [DoFa99]. Im Folgenden wird ein Interpretationsvorschlag für diejenigen Aspekte wiedergegeben, die für die hier diskutierte Thematik von Kompetenzmanagement-Systemen Relevanz besitzen.

Die *Probleme*, die innerhalb einer Domäne von einer Gruppe von Akteuren als relevant erachtet werden, konstituieren die Problememenge P . Das handlungsbefähigende Wissen – also die Kompetenz – eines Akteurs wird mit der Teilmenge aller Probleme identifiziert, die er zu lösen vermag [Ko99]. Handlungsbefähigendes Wissen wird also im Folgenden als Wissen verstanden, das zur Lösung von Problemen befähigt. Diese Spezialisierung des Kompetenzbegriffs auf Problemlösungskompetenz erweist sich aus betriebswirtschaftlicher Perspektive als sehr leistungsfähig, weil sich nahezu alle Aufgaben, die in Unternehmen zu erfüllen sind, auf das Lösen von (Entscheidungs-) Problemen zurückführen lassen. Es wird in der Betriebswirtschaftslehre dabei auch von einem (entscheidungs- oder) problemorientierten Paradigma gesprochen. Zur formalen Repräsentation des Wissens ü-

ber Fähigkeiten, Probleme zu lösen, wird die (binäre) *Surmise-Relation* R aus dem kartesischen Produkt der Problememenge P mit sich selbst eingeführt:

$$R \subseteq P \times P \quad (1)$$

Die Surmise-Relation (oder *Vorhersagerelation* [Sc93]) bildet das Grundgerüst der Wissensstrukturtheorie, indem sie ein partielles Ordnungsverhältnis – also eine Halbordnung – zwischen den Problemen in der Menge P konstituiert. Diese Relation kann auf unterschiedliche Weise inhaltlich interpretiert werden.

Im hier betrachteten Anwendungskontext von Kompetenzmanagement-Systemen wird die Surmise-Relation R auf die Problemlösungskompetenz von Akteuren zurückgeführt: Die Aussage $(p_1, p_2) \in R$ gilt genau dann, wenn von jedem Akteur, der ein Problem $p_1 \in P$ richtig zu lösen vermag, mit triftigen Gründen auch eine richtige Lösung für das Problem $p_2 \in P$ erwartet werden kann. Sie beruht auf der heuristischen, von Erfahrungswissen gestützten *Vermutung*, dass die Fähigkeit, das Problem p_1 zu lösen, auch die Fähigkeit zur Lösung des anderen Problems p_2 einschließt. Somit ist die Lösungsfähigkeit für das „anspruchsvollere“ Problem p_1 eine *hinreichende* Bedingung für die Lösungsfähigkeit für das „einfachere“ Problem p_2 . Dies bedeutet allerdings, dass die Probleme p_1 und p_2 nur dann in der Beziehung $(p_1, p_2) \in R$ stehen können, wenn es sich zwar um unterschiedlich schwere, aber inhaltlich so eng verwandte Probleme handelt, dass die Lösungsfähigkeit für das eine Problem (p_1) die Lösungsfähigkeit für das andere Problem (p_2) begründet erwarten lässt. Mittels der Surmise-Relation R wird also detailliertes Erfahrungswissen über die inhaltliche Verwandtschaft und die Lösungsschwierigkeit der Probleme des Problemraums P repräsentiert; sie drückt Wissen über den Zusammenhang von Problemlösungsfähigkeiten aus. Darüber hinaus stellt die Lösungsfähigkeit für das „einfachere“ Problem p_2 eine *notwendige* Bedingung für die Lösungsfähigkeit für das „anspruchsvollere“ Problem p_1 dar. Die Lösungsfähigkeit für p_2 ist allerdings *nicht* hinreichend für die Lösungsfähigkeit für p_1 , da der Umkehrschluss nicht zutrifft.

Die (partielle) Ordnungsrelation R erfüllt folgende Eigenschaften:

- $\forall p \in P: (p, p) \in R$ (Reflexivität)
- $\forall p_1, p_2, p_3 \in P: (p_1, p_2) \in R \wedge (p_2, p_3) \in R \rightarrow (p_1, p_3) \in R$ (Transitivität)
- $\forall p_1, p_2 \in P: (p_1, p_2) \in R \wedge (p_2, p_1) \in R \rightarrow p_1 = p_2$ (Antisymmetrie)

Die *Reflexivität* beinhaltet die Regel, dass von der Fähigkeit zur Lösung eines Problems p genau auf diese Fähigkeit geschlossen werden kann. Diese intuitiv überflüssig erscheinende Regel ist notwendig, um die Vollständigkeit des Konzepts zu gewährleisten. Von größerer Bedeutung ist die *Transitivität*. Sie erlaubt es, implizites Wissen über den mittelbaren Zusammenhang von Problemlösungsfähigkeiten zu erschließen, auch wenn zwei Problemlösungsfähigkeiten in keiner unmittelbaren Beziehung der Surmise-Relation zueinander stehen. Die *Antisymmetrie* ist dagegen von geringerer Bedeutung, da sie nur dazu beiträgt, die Integrität der Ordnung des Wissensraums zu bewahren. Allerdings gestattet sie nicht, durch Schlussfolgerungen „neue“ explizite Fakten über Beziehungen zwischen Problemlösungsfähigkeiten der Wissensbasis hinzuzufügen.

Die Menge von Problemen, die ein Akteur zu lösen vermag, wird als *Wissenszustand* WZ des Akteurs bezeichnet. Dieser Zustand kennzeichnet dasjenige handlungsbefähigende Wissen, das hinreichend ist, um alle Probleme aus der Problememenge (richtig) lösen zu können. Ein Wissenszustand kann daher aus der Perspektive von Kompetenzmanagement-Systemen auch als ein *Kompetenzzustand* bezeichnet werden.

Durch die Vorgabe der Surmise-Relation R werden einige Wissenszustände – oder synonym: Kompetenzzustände – für alle Akteure ausgeschlossen. Für das oben angeführte Beispiel mit $(p_1, p_2) \in R$ bedeutet dies beispielsweise, dass die einelementige Menge $\{p_1\}$ keinen zulässigen Wissenszustand bilden kann. Denn aus der Lösungsfähigkeit für das Problem p_1 folgt wegen $(p_1, p_2) \in R$ stets auch die Lösungsfähigkeit für das Problem p_2 . Also ist es sachlogisch unmöglich, angesichts dieser Surmise-Relation R die Menge $\{p_1\}$ als einen Wissenszustand irgendeines Akteurs auszuzeichnen. Lediglich $\{p_1, p_2\}$ kommt als ein R -zulässiger Wissenszustand in Betracht. Formal lautet die Definition von Wissenszuständen, die hinsichtlich der Surmise-Relation R zulässig oder „wohlgeformt“ sind:

$$WZ \subseteq P : \Leftrightarrow \forall p_1, p_2 \in P: (p_1 \in WZ \wedge (p_1, p_2) \in R \rightarrow p_2 \in WZ) \quad (2)$$

Eine Teilmenge WZ der Problememenge P ist also genau dann ein R -zulässiger oder *wohlgeformter Wissenszustand*, wenn für alle Lösungsfähigkeiten für Probleme p_1, p_2 aus der Problememenge P gilt: Falls ein Akteur das

Problem p_1 aus dem Wissenszustand WZ zu lösen vermag und die Lösungsfähigkeit für dieses Problem p_1 gemäß der Surmise-Relation R die Lösungsfähigkeit für das Problem p_2 einschließt, dann muss auch jenes Problem p_2 in dem Wissenszustand WZ enthalten sein.

BEISPIEL 1: Die Menge aller Probleme ist durch $P=\{p_1,p_2,p_3,p_4\}$ gegeben. Die Surmise-Relation über dieser Problememenge ist durch $R=\{(p_1,p_2),(p_1,p_3),(p_4,p_3)\}$ definiert. Sie kann durch ein HASSE-Diagramm visualisiert werden, das in Abbildung 1 wiedergegeben wird. In einem solchen Diagramm repräsentiert jeder Knoten eine Teilmenge der Problememenge P mit der Problememenge P selbst als „oberstem“ und der leeren Menge \emptyset als „unterstem“ Knoten. Jede Kante, die im HASSE-Diagramm von einem Knoten mit dem Problem p_a zu einem Knoten mit dem Problem p_b gerichtet ist, drückt die Beziehung aus der Surmise-Relation R aus, dass die Fähigkeit zur Lösung des Problems p_a die Fähigkeit zur Lösung des Problems p_b einschließt:

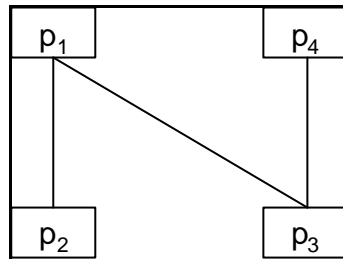


Abbildung 1: HASSE-Diagramm zur Surmise-Relation

Die Potenzmenge 2^P der Problememenge P gibt die Menge aller potenziellen – aber nicht notwendig wohlgeformten – Wissenszustände wieder. Im o.a. Beispiel entspricht dies $2^4=16$ Zuständen. Bei Einführung der Surmise-Relation R wird die Menge W aller wohlgeformten Wissenszustände WZ allerdings eingeschränkt auf:

$$W=\{\emptyset,\{p_2\},\{p_3\},\{p_2,p_3\},\{p_3,p_4\},\{p_1,p_2,p_3\},\{p_2,p_3,p_4\},\{p_1,p_2,p_3,p_4\}\}$$

In Bezug auf die Surmise-Relation $R=\{(p_1,p_2),(p_1,p_3),(p_4,p_3)\}$ erweisen sich z.B. die Zustände $\{p_1\}$, $\{p_4\}$, $\{p_1,p_2\}$, $\{p_1,p_3\}$, $\{p_1,p_4\}$, $\{p_1,p_2\}$ und $\{p_1,p_3,p_4\}$ als „unzulässig“.

Die Menge W aller wohlgeformten Wissenszustände zeichnet sich durch folgende Eigenschaften aus [DoFa99, Fa90]: Zum einen ergibt die Vereinigung von zwei wohlgeformten Wissenszuständen stets einen neuen wohlgeformten Wissenszustand. Es gilt somit:

$$\forall WZ_1,WZ_2: (WZ_1,WZ_2 \in W) \rightarrow WZ_1 \cup WZ_2 \in W \quad (3)$$

Zum anderen ist jede Schnittmenge von zwei wohlgeformten Wissenszuständen stets ein neuer wohlgeformter Wissenszustand:

$$\forall WZ_1,WZ_2: (WZ_1,WZ_2 \in W) \rightarrow WZ_1 \cap WZ_2 \in W \quad (4)$$

Durch die beiden Eigenschaften wird belegt, dass die Menge W aller wohlgeformten Wissenszustände bezüglich Disjunktion (3) und Konjunktion (4) der involvierten Problemlösungsfähigkeiten abgeschlossen ist.

Wird die Forderung nach der konjunktiven Abgeschlossenheit (4) aufgegeben¹⁾, so gelangt man zu dem Konzept der Wissensräume. Die Visualisierung eines Wissensraumes erfolgt durch einen UND/ODER-Graph, wie er in der Forschung zur Künstlichen Intelligenz eingesetzt wird [Ri98]. Die Formalisierung des UND/ODER-Graphen

1) Die Vernachlässigung der konjunktiven Abgeschlossenheit erweist sich bei dem hiesigen Vorhaben keineswegs als eine „Schwächung“ des Konzepts. Vielmehr erweist sich diese Vernachlässigung im Kontext von Kompetenzmanagement-Systemen sogar als vorteilhaft, weil sie die Einbeziehung von Lernprozessen gestattet. Ein Beispiel aus [Fa90] verdeutlicht dies: Mann nehme an, dass n Akteure mit den jeweiligen Wissenszuständen WZ_1, WZ_2, \dots, WZ_n an einer gemeinschaftlichen Wertschöpfungskette oder einem Virtuellen Unternehmen beteiligt sind. Wenn ein einzelner Akteur die Problemlösungsfähigkeiten aller anderen Akteure während der gemeinsamen Wertschöpfung erlernt, so lässt sich sein neuer Wissenszustand als die Vereinigung $WZ_1 \cup WZ_2 \cup \dots \cup WZ_n$ darstellen. Der Durchschnitt aller Wissenszustände würde hingegen einer Reduktion der persönlichen Problemlösungsfähigkeiten gleichkommen und an der betrieblichen Realität vorbeigehen.

erfolgt durch die Einführung einer Abbildung σ , die jedem Problem $p \in P$ eine Familie von Teilmengen aus der Problememenge P zuordnet:

$$\sigma(p) := \{K_1, K_2, \dots, K_n\} \text{ mit } K_i \subseteq P \text{ (} i=1, \dots, n; n \in \mathbb{N} \text{)}. \quad (5)$$

Die Abbildung σ wird als *Surmise-Funktion* (oder *Vorhersagefunktion* [Sc93]) bezeichnet. Die Surmise-Funktion ordnet jedem Problem p aus der Problememenge P eine Menge von Teilmengen $K_i \subseteq P$ zu, die als *Klauseln* bezeichnet werden. Es wird dabei angenommen, dass die Fähigkeit, das Problem p zu lösen, auch die Fähigkeit zur Lösung aller Probleme in mindestens einer Klausel $K_i \in \sigma(p)$ einschließt. Die Interpretation der Begriffe „mindestens“ und „aller“ entspricht dabei den ODER- und UND-Knoten des Graphen [Ko93]. Während sich jeder UND-Knoten unmittelbar in so viele Elemente der Surmise-Relation R übersetzen lässt, wie der UND-Knoten im UND/ODER-Graphen Folgeknoten besitzt, stellen die ODER-Knoten eine echte Erweiterung der Surmise-Relation R dar.

BEISPIEL 2: Die nachfolgende Abbildung 2 veranschaulicht das Konzept der Wissensräume. In der Abbildung kann z.B. aus der Fähigkeit zur Lösung von Problem p_2 auf die Fähigkeit zur Lösung von Problem p_1 oder von Problem p_3 geschlossen werden. Dagegen schließt die Fähigkeit, das Problem p_5 zu lösen, auch die Fähigkeit ein, alle Probleme p_1, p_2 und p_3 zu lösen.

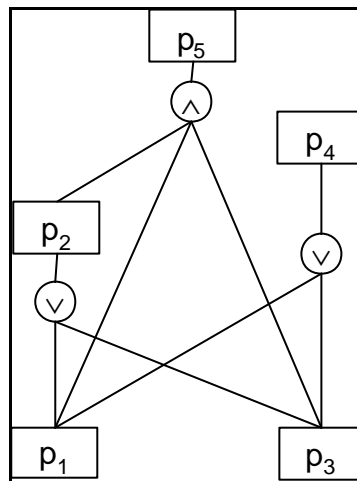


Abbildung 2: Exemplarischer UND/ODER-Graph für einen Wissensraum

In diesem Beispiel lässt sich der Wert $\sigma(p_2) = \{\{p_1\}, \{p_3\}\}$ der Abbildung σ für das Problem p_2 auf folgende Weise interpretieren: Es ist plausibel zu erwarten, dass jeder Akteur, der das Problem p_2 zu lösen vermag, auch das Problem p_1 zu lösen vermag (Klausel $K_1 = \{p_1\}$) oder das Problem p_3 zu lösen vermag (Klausel $K_2 = \{p_3\}$). Für das Beispiel lauten alle Werte der Abbildung σ :

$$\begin{aligned} \sigma(p_1) &= \{\{\emptyset\}\} \\ \sigma(p_2) &= \{\{p_1\}, \{p_3\}\} \\ \sigma(p_3) &= \{\{\emptyset\}\} \\ \sigma(p_4) &= \{\{p_1\}, \{p_3\}\} \\ \sigma(p_5) &= \{\{p_1, p_2, p_3\}\} \end{aligned}$$

Für dieses Beispiel der Abbildung 2 nimmt die zugehörige Menge W aller wohlgeformten Wissenszustände folgende Gestalt an:

$$W = \{\emptyset, \{p_1\}, \{p_3\}, \{p_1, p_2\}, \{p_1, p_3\}, \{p_1, p_4\}, \{p_2, p_3\}, \{p_3, p_4\}, \{p_1, p_2, p_3\}, \{p_1, p_2, p_4\}, \{p_1, p_3, p_4\}, \{p_2, p_3, p_4\}, \{p_1, p_2, p_3, p_4\}, \{p_1, p_2, p_3, p_5\}, \{p_1, p_2, p_3, p_4, p_5\}\}$$

Eine Besonderheit des hier vorgestellten Konzepts der Wissensräume ist es, einen Wissenszustand nicht „intensional“ durch die Inhalte des zugehörigen Wissens zu spezifizieren, sondern auf „extensionale“ Weise durch die Auflistung aller Probleme, die mit Hilfe des im betroffenen Zustand verfügbaren Wissens (richtig) gelöst werden können. Es handelt sich also um keine inhaltsbezogene, sondern um eine leistungsbezogene Definition von Wissenszuständen. Dies erweist sich aus der hier verfolgten Perspektive von Kompetenzmanagement-Systemen als bemerkenswert. Denn diese leistungsbezogene Definition von Wissenszuständen entspricht der Eigenart von Wissen über Kompetenzen, nicht die Inhalte von handlungsbefähigendem Objektwissen zu umfassen, sondern

nur das Metawissen, welche handlungsbefähigenden Kenntnisse und Fähigkeiten – gegebenenfalls bei welchen Akteuren – existieren. Dabei entspricht das Metawissen über Kompetenzen dem Metawissen über diejenigen Probleme, die mit einem – inhaltlich nicht näher spezifizierten – Objektwissen gelöst werden können. Dies gilt allerdings nur dann, wenn die handlungsbefähigende Eigenschaft von Wissen (im Kompetenzbegriff) mit der Eigenschaft von Wissen gleichgesetzt wird, mit Hilfe dieses Wissens Probleme (richtig) lösen zu können (im Begriff des Wissenszustands). Diese abstrakte Gleichsetzung zwischen der Handlungsbefähigung durch Wissen und der Problemlösungsfähigkeit durch Wissen ist zulässig, solange Handeln als problemlösendes Handeln konzeptualisiert wird. Dies entspricht – wie bereits an früherer Stelle kurz angemerkt wurde – dem weit verbreiteten problem- und entscheidungsorientierten Paradigma der Betriebswirtschaftslehre. Aufgrund dieses betriebswirtschaftlich vertrauten Denkmusters, das die Lösung betrieblicher Probleme in den Fokus von Aufgaben des Wissensmanagements rückt, erweist sich das hier ausgewählte Konzept der Wissensräume auch aus ökonomischer Sicht als reizvoll und fruchtbar.

2.3. Ontologien zur Strukturierung von Domänenwissen

Durch die formale Spezifikation von Begriffen kann mit Ontologien das Wissen einer Domäne durch die Einbettung der Begriffe in ihren semantischen Kontext *maschinenlesbar* strukturiert werden. Somit wird die Kommunikation von Menschen und Softwareagenten sowohl miteinander als auch untereinander – aufbauend auf einem gemeinsamen Verständnis für die verwendeten Begriffe – ermöglicht [Gu98]. Werden die in einer Ontologie definierten Begriffe (Klassen) anhand konkreter Elemente aus einem Anwendungsfall instanziiert, entsteht eine Wissensbasis, die auf einem – auch zur Laufzeit des ontologiebasierten Kompetenzmanagement-Systems – explizit vorliegenden Schema aufbaut. Dieses Schema erweist sich im Hinblick auf Anforderungen an leicht wartbare und wiederverwendbare Software, die seit längerem und mit wachsendem Nachdruck geäußert werden, als vorteilhaft.

Seitens der Forschung zu Ontologien wird erhofft, eine *semantische Anreicherung* der Fakten in einer ontologiebasierten Wissensbasis zu erreichen. Dies soll durch die Explikation des Umfangs eines Begriffs einerseits und seines Inhalts andererseits erreicht werden. Der *Umfang* eines Begriffs entspricht dabei seinem *extensionalen* Aspekt. Hierin sind sämtliche Sachverhalte (Elemente) enthalten, die unter dem Begriff aufgezählt werden können. Der *Inhalt* gibt wiederum die *Intension* eines Begriffs wieder. Dieser intensionale Aspekt eines Begriffs wird in einer Ontologie durch die Aufzählung aller „konstitutiven“ oder „definitiven“ Eigenschaften gewährleistet, die den Elementen, die unter dem Begriff subsumiert werden können, zukommen müssen. Durch die Formalisierung dieses intensionalen und/oder extensionalen *Kontextes* von Begriffen wird ein interindividuelles, gemeinsames Verständnis über diese Begriffe geschaffen, das die Akteure – Menschen oder Softwareagenten – teilen, die an der arbeitsteiligen Erfüllung einer gemeinsamen Aufgabe zusammenwirken.

Die formale Spezifikation von Ontologien erweist sich unter anderem in zweierlei Formen als vorteilhaft: Zum einen können teilweise – in Abhängigkeit von der verwendeten Sprache – *Inferenzregeln* definiert werden, die über das bereits explizit vorliegende Wissen hinaus *Schlussfolgerungen* erlauben. Diese Möglichkeit der Ableitung „neuer“ Fakten, die bereits aus der logischen Programmierung (z.B. mittels PROLOG) bekannt ist, erlaubt die Gestaltung „intelligenter“ Software. Die folgende exemplarische Inferenzregel setzt die Plausibilitätsannahme, dass jede Person, die einen Projektbericht zu einem Sachgebiet verfasst hat, auch als kompetent in diesem Gebiet gilt, in der Notation der weit verbreiteten F-Logic [KLW95] um:

FORALL X,Y,Z
 $X[\text{hat_Kompetenz} \rightarrow Z] \rightarrow$
 $(X:\text{Person}[\text{Autor_von} \rightarrow Y] \text{ AND}$
 $Y:\text{Projektbericht}[\text{betrifft_Sachgebiet} \rightarrow Z])$

Zum anderen kann durch *Integritätsregeln* vermieden werden, dass Fakten in der Wissensbasis aufgenommen werden, die im Widerspruch zu bereits enthaltenen Fakten stehen würden. So wäre beispielsweise eine Schlussfolgerung auf Basis der obigen Inferenzregel unzulässig, wenn zuvor schon das explizite Wissen vorgelegen hätte, dass sich die betroffene Person in dem angesprochenen Fachgebiet als „inkompetent“ erwiesen hat. Solche Integritätsregeln sind insbesondere dann erforderlich, wenn in einem Kompetenzmanagement-System Inferenzregeln für plausibles Schließen verwendet werden. Denn solche Plausibilitätsregeln drücken nur Erfahrungswissen aus, das sich in der Vergangenheit zwar als tendenziell zutreffend bewährt hat, das jedoch keine Allgemeingültigkeit für sich in Anspruch nehmen kann. Bei der Verwendung solcher Plausibilitätsregeln sollte für jede Möglichkeit, einen Plausibilitätsschluss zu ziehen, eine korrespondierende Integritätsregel dafür Sorge leisten, dass das Inferenzresultat niemals in Widerspruch zu bereits vorhandenem Wissen treten kann.

3. Ontologiebasierte Spezifikation von Wissensräumen

Im Folgenden werden Auszüge¹⁾ aus einer Ontologie vorgestellt, die für die betriebswirtschaftliche Domäne der Kompetenzmanagement-Systeme an der Universität Essen im Rahmen des BMBF-Projekts KOWIEN entwickelt wurde. Sie baut auf dem Konzept der Wissensräume auf. Die für das Projekt erforderlichen (Teil-)Ontologien wurden in der Ontologie-Entwicklungsumgebung OntoEdit [Su02, Ma00] konstruiert (vgl. Abbildung 3). Die Spezifikation der Ontologien erfolgte in F-Logic [KLW95], RDF(S) [LaSw99, BrGu02] und DAML+Oil [Ho02]²⁾.

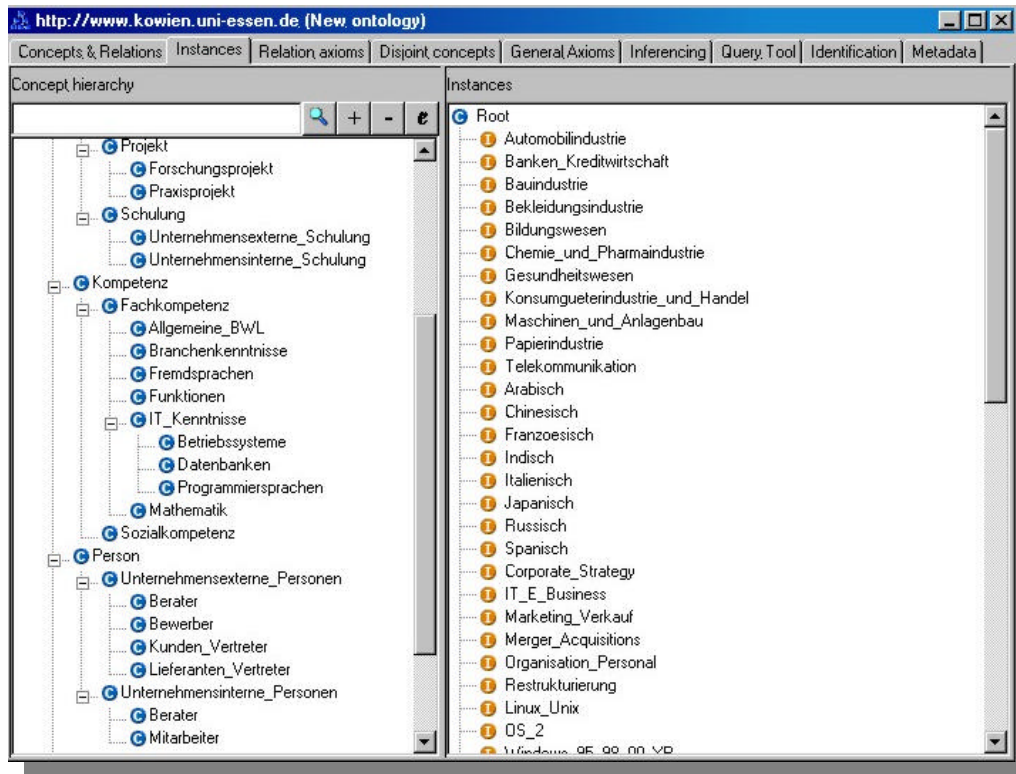


Abbildung 3: Screenshot OntoEdit

Zu Beginn der Erstellung der Ontologien wurde eine graphische Visualisierung der Konzepte angestrebt, um – darauf aufbauend – eine formale Spezifikation der verwendeten Begrifflichkeiten und semantischen Regeln zu erleichtern.

1) Die vollständige Ontologie kann unter <http://www.pim.uni-essen.de/mitarbeiter/pimyal/Kompetenzontologie.flo> oder <http://www.pim.uni-essen.de/mitarbeiter/pimyal/Kompetenzontologie.rdf> bezogen werden.

2) Es wird darauf hingewiesen, dass mit dem Einsatz der unterschiedlichen Sprachen Vor- und Nachteile verbunden sind. Die Repräsentation der Ontologien erfolgt hier in F-Logic, da die Möglichkeit zur Definition non-deduktiver Inferenzregeln konstitutiv für die Implementierung der Surmise-Relation ist. Die übrigen Sprachen stellen hierfür zur Zeit keine Mechanismen zur Verfügung.

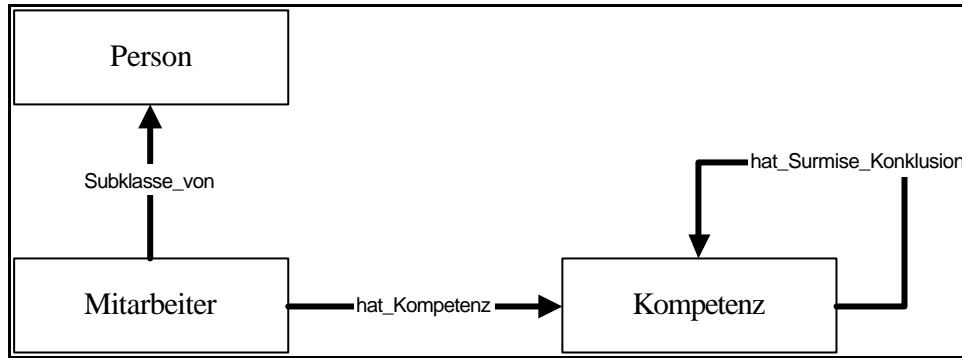


Abbildung 4: Struktur der Surmise-Relation-Ontologie

Die voranstehende Abbildung 4 skizziert eine erste (Teil-)Ontologie, mit deren Hilfe das Konzept der Surmise-Relation R implementiert wird. Stehen zwei Probleme p_1 und p_2 in einer Beziehung der Surmise-Relation – d.h. gilt $(p_1, p_2) \in R$ –, so lautet die F-Logic-Spezifikation dieses Sachverhalts:

$$p_1: \text{Kompetenz}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow p_2]$$

Die Eigenschaften der Surmise-Relation R werden spezifiziert als:

$$\text{FORALL Kompetenz1 Kompetenz1:Kompetenz}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz1}] \quad (\text{Reflexivität})$$

und:

$$\begin{aligned} &\text{FORALL Kompetenz1, Kompetenz2, Kompetenz3} \\ &\text{Kompetenz1}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz3}] \neg \\ &(\text{Kompetenz1:Kompetenz}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz2}] \text{ AND} \\ &\text{Kompetenz2:Kompetenz}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz3}]) \quad (\text{Transitivität}). \end{aligned}$$

Die dritte Eigenschaft der Antisymmetrie lässt sich hingegen nicht mittels einfacher Subjugaten in der voranstehenden Form in F-Logic implementieren, da sie ihrem „Wesen“ nach nicht die Folgerbarkeit neuer Fakten, sondern den Ausschluss unzulässiger Fakten bedeutet: Denn die o.a. „Definition“ der Antisymmetrie mittels der Formel „ $\forall p_1, p_2 \in P: ((p_1, p_2) \in R \wedge (p_2, p_1) \in R) \rightarrow p_1 = p_2$ “ stellt nur die allgemein übliche, als Subjugat „geformte“ Umschreibung für folgenden Sachverhalt dar: $\forall p_1, p_2 \in P: ((p_1, p_2) \in R \wedge p_1 \neq p_2) \rightarrow (p_2, p_1) \notin R$. Die Nicht-Zugehörigkeit zur Surmise-Relation R lässt sich aber in einer gewöhnlichen Wissensbasis nicht als Faktum repräsentieren. Um die Eigenschaft der Antisymmetrie korrekt darstellen zu können, muss daher eine Relation „ $\text{hat_nicht_Surmise_Konklusion}$ “ komplementär zur bereits eingeführten Relation „ $\text{hat_Surmise_Konklusion}$ “ definiert werden. Zusätzlich ist es erforderlich, mittels einer Integritätsregel den o.a. Sachverhalt auszudrücken: In der Wissensbasis eines Kompetenzmanagement-Systems darf niemals das faktische Wissen repräsentiert werden, dass hinsichtlich der Surmise-Relation R durch die Fähigkeit, ein Problem (p_1) zu lösen, eine andere Problemlösungsfähigkeit (p_2) sowohl eingeschlossen ($(p_1, p_2) \in R$) als auch ausgeschlossen ($(p_1, p_2) \notin R$) wird. Diese Anforderungen werden durch folgende Festlegungen erfüllt:

$$\begin{aligned} &\text{FORALL Kompetenz1, Kompetenz2} \\ &\text{Kompetenz2}[\text{hat_nicht_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz1}] \neg \\ &(\text{Kompetenz1:Kompetenz}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz2}] \text{ AND} \\ &\text{NOT equal}(\text{Kompetenz1}, \text{Kompetenz2})) \end{aligned}$$

$$\begin{aligned} &\text{FORALL Kompetenz1, Kompetenz2} \\ &\text{NOT}(\text{Kompetenz2}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz1}] \text{ AND} \\ &\text{Kompetenz2}[\text{hat_nicht_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz1}]) \neg \\ &\text{NOT equal}(\text{Kompetenz1}, \text{Kompetenz2}) \quad (\text{Antisymmetrie}). \end{aligned}$$

Die erste subjugatförmige Regel zur Implementierung der Antisymmetrie-Eigenschaft in F-Logic stellt eine konventionelle Inferenzregel dar, weil sie das Schlussfolgern neuer Fakten – hier in Bezug auf die Nicht-Zugehörigkeit eines Paares von Problemlösungsfähigkeiten zur Surmise-Relation R – gestattet. Die zweite subjugatförmige Regel zur Implementierung der Antisymmetrie-Eigenschaft besitzt jedoch eine andere epistemische Qualität, weil sie in ihrer Konklusion kein ableitbares neues Faktum, sondern eine *unzulässige* Koexistenz zwei-

er Fakten – d.h. hier $(p_2, p_1) \in R$ und $(p_2, p_1) \notin R$ – in der Wissensbasis ausdrückt. Um die *Unzulässigkeit* einer Formel – hier: $(p_2, p_1) \in R \wedge (p_2, p_1) \notin R$ – in einer Wissensbasis operativ handhaben zu können, bedarf es zusätzlicher Mechanismen, wie z.B. eines „Trigger-Mechanismus“: Er zeigt im Falle des Vorliegens einer solchen unzulässigen Formel den Eintritt der Integritätsverletzung an und setzt alle vorangegangenen Schlussfolgerungen zurück, die zu dieser Integritätsverletzung geführt haben. Beispielsweise bietet die *Petrinetz*-Technik mittels ihrer Fakt-Transitionen und den darauf aufbauenden Fakt-Netzen ein Kalkül, mit dessen Hilfe solche Unzulässigkeiten von Formeln in einer Wissensbasis ausgedrückt und im Falle ihres Eintritts angezeigt werden können, um entsprechende Zurücksetzungsmechanismen anzustoßen (zu „triggern“).

Bei der Konstruktion der KOWIEN-Ontologien hat es sich als vorteilhafter erwiesen, die verwendeten Relationen zu klassifizieren und charakteristische Eigenschaften aller Mitglieder einer solchen Relationenklasse mittels „Reifizierung“ durch Inferenzregeln für die gesamte Klasse zu definieren¹⁾. Auf diese Weise wird es möglich, häufig wiederkehrende Relationeigenschaften nur einmal für alle klassenzugehörigen Relationen zu definieren, anstatt sie für jede einzelne Relation explizit angeben zu müssen. Dies trägt erheblich zur Kompaktheit und Übersichtlichkeit der Ontologien bei. Als pars pro toto wird die Relationseigenschaft der Transitivität betrachtet. Die Surmise-Relation R – hier als zweistellige Relation „*hat_Surmise_Konklusion*“ in Infix-Notation wiedergegeben – wird in diesem Fall mittels des Meta-Prädikats *transitiv* als transitive Relation klassifiziert: *transitiv(hat_Surmise_Konklusion)*. Die charakteristische Regel für die „reifizierte“ Klasse *aller* transitiven Relationen REL lautet dann mit X, Y und Z als beliebigen „normalen“ Klassen mit Individuen als Elementen:

$$\begin{aligned} & \text{FORALL } X, Y, Z, \text{REL} \\ & \quad X[\text{REL} \rightarrow \rightarrow Z] \leftarrow \\ & \quad \quad (X[\text{REL} \rightarrow \rightarrow Y] \text{ AND } Y[\text{REL} \rightarrow \rightarrow Z] \text{ AND } \text{transitiv}(\text{REL})) \end{aligned}$$

Zur weiteren Illustration der Ontologie für Kompetenzmanagement-Systeme dient eine Inferenzregel, die sicherstellt, dass die Kompetenzprofile der Mitarbeiter mit der Strukturierung der Problemlösungsfähigkeiten durch die Surmise-Relation R übereinstimmen. Sie drückt aus, dass alle Mitarbeiter, die eine bestimmte Kompetenz (Problemlösungsfähigkeit) besitzen, auch über alle Kompetenzen (Problemlösungsfähigkeiten) verfügen müssen, die in Surmise-Beziehungen zu der erstgenannten Kompetenz stehen. Diese Inferenzregel, die auch als eine Integritätsregel „höherer Ordnung“ zwischen Kompetenzprofilen einerseits und Surmise-Relation R andererseits aufgefasst werden kann, lautet:

$$\begin{aligned} & \text{FORALL } \text{Mitarbeiter1}, \text{Kompetenz1}, \text{Kompetenz2} \\ & \quad \text{Mitarbeiter1}[\text{hat_Kompetenz} \rightarrow \rightarrow \text{Kompetenz2}] \leftarrow \\ & \quad \quad (\text{Mitarbeiter1}:\text{Mitarbeiter}[\text{hat_Kompetenz} \rightarrow \rightarrow \text{Kompetenz1}] \text{ AND} \\ & \quad \quad \text{Kompetenz1}:\text{Kompetenz}[\text{hat_Surmise_Konklusion} \rightarrow \rightarrow \text{Kompetenz2}]) \end{aligned}$$

Wenn beispielsweise das Kompetenzprofil für einen Mitarbeiter *Meier* die Lösungsfähigkeit für ein Problem p umfasst, so wird eine Anfrage der Form:

$$\begin{aligned} & \text{FORALL } X \\ & \quad \text{Meier}[\text{hat_Kompetenz} \rightarrow \rightarrow X] \end{aligned}$$

entsprechend der Surmise-Relation in Abbildung 1 zu der Antwort eines Kompetenzmanagement-Systems führen, dass der Mitarbeiter *Meier* neben der Fähigkeit, das Problem p_1 lösen zu können, auch über die Kompetenzen verfügt, die Probleme p_2 und p_3 zu lösen.

Wenn von der „einfachen“ Surmise-Relation R zur Darstellung von Kompetenzen-Zusammenhängen in UND/ODER-Graphen übergegangen wird, die auch neuartige ODER-Verknüpfungen zwischen Problemlösungsfähigkeiten zulassen, wird von einem Surmise-System S gesprochen. Die Visualisierung einer Ontologie, die auf dem Gerüst eines solchen Surmise-Systems konstruiert wurde, erfolgt in der nachfolgenden Abbildung 5.

1) Es wird darauf hingewiesen, dass bei solch einem Vorgehen der Bereich der Prädikatenlogik *erster* Stufe verlassen wird, in der nur Individuen-Konstanten und -Variablen als Subjekte von Prädikaten und Relationen bzw. als Objekte von Quantifizierungen zulässig sind. Prädikats- und Relations-Konstanten als Subjekte von (höherstufigen) sowie Prädikats- und Relations-Variablen als Objekte von Quantifizierungen sind dagegen erst innerhalb der Prädikatenlogik *zweiter* Stufe zulässig. Die Prädikatenlogik zweiter Stufe wird hier erforderlich, um quantifizierte Formeln über Relations-Variablen ausdrücken zu können.

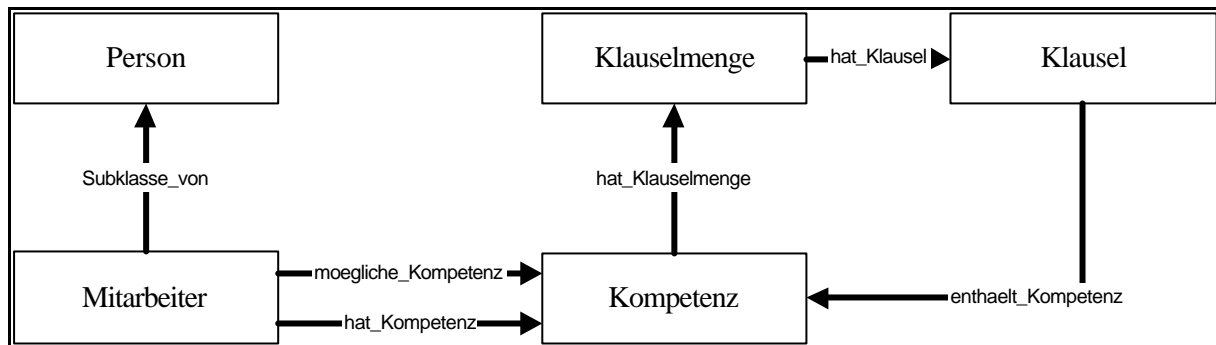


Abbildung 5: Struktur der Surmise-System-Ontologie

Da in einem Surmise-System die Klauselmenge aus Kompetenzen (Problemlösungsfähigkeiten), die von einer Kompetenz eingeschlossen werden, in disjunktiver Normalform dargestellt wird, ist es für Kompetenzen, die zu *unterschiedlichen* Klauseln gehören, nicht mehr möglich, stringent auf eine bestimmte Kompetenzenmenge zu schließen. Stattdessen können nur noch Mengen *möglicher* Kompetenzen angegeben werden, über die ein Akteur aufgrund seiner explizit angegebenen Kompetenz verfügen könnte. Die Inferenzregel, die zu einer solchen Surmise-System-Ontologie gehört, lautet entsprechend:

FORALL Mitarbeiter1, Kompetenz1, Kompetenz2, Klauselmenge1, Klausel1
 Mitarbeiter1[hat_moegliche_Kompetenz->>Kompetenz2] ←
 (Mitarbeiter1:Mitarbeiter[hat_Kompetenz->>Kompetenz1] AND
 Kompetenz1:Kompetenz[ist_enthalten_in->>Klausel1] AND
 Klausel1:Klausel[enthaelt_Kompetenz->>Kompetenz2])

Die epistemische Qualität dieser Inferenzregel für das Surmise-System S ist geringer als die epistemische Qualität der entsprechenden Inferenzregel für die Surmise-Relation R , da die erstgenannte Inferenzregel lediglich Wissen über *mögliche* Kompetenzen eines Akteurs zu erschließen vermag¹⁾, während die zweitgenannte Inferenzregel gestattet, Wissen über eine *tatsächlich vorhandene* Kompetenz zu folgern. Die Erweiterung von Surmise-Relationen R auf Surmise-Systeme S ermöglicht also eine *größere Ausdrucksmächtigkeit* hinsichtlich Zusammenhängen zwischen Problemlösungsfähigkeiten (Kompetenzen), bedeutet aber zugleich eine *geringere Schlussfolgerungsqualität*. Dieser *Trade-off* zwischen Ausdrucksmächtigkeit und Schlussfolgerungsqualität der Konstruktionsweise einer Ontologie stellt einen Ontologie-Designer vor das Problem, zwischen diesen beiden gegenläufigen Tendenzen abzuwägen. Dieses Evaluationsproblem konstituiert ein ökonomisches Optimierungsproblem, für das bislang noch keine Lösungsansätze bekannt sind. Zu seiner Lösung bedarf es einer Bewertungsfunktion sowohl für unterschiedliche Ausdrucksmächtigkeiten als auch für unterschiedliche Schlussfolgerungsqualitäten von ontologiebasierten Kompetenzmanagement-Systemen. Bewertungsfunktionen beider Art müssen erst noch durch weiter führende, interdisziplinäre Zusammenarbeit zwischen Informatik und Betriebswirtschaftslehre aufgestellt und hinsichtlich ihrer charakteristischen Eigenschaften analysiert werden.

4. Ausblick

In dem Beitrag wurde aufgezeigt, wie das Konzept der Wissensräume als formale Grundlage für die Konstruktion von ontologiebasierten Kompetenzmanagement-Systemen verwendet werden kann. Als vorteilhaft erweist sich dieses Vorgehen, da innerhalb von Kompetenzmanagement-Systemen Wissen über Kompetenzen abgelegt werden soll, das möglicherweise lediglich *implizit* vorliegt. Durch die formale Vorstrukturierung von Begriffen, die mittels Ontologien erzielt wird, können Inferenzregeln konstruiert werden. Diese Inferenzregeln ermöglichen die Explikation zuvor impliziten Wissens über Kompetenzen. Darüber hinaus lassen sich auch Integritätsregeln spezifizieren, die für die Konsistenz des Wissens über Kompetenzen sorgen und im Falle einer drohenden Inkonsistenz die erforderlichen „Reparaturmechanismen“ auslösen.

1) Diese Einschränkung lässt sich so lange nicht vermeiden, wie gängige Inferenz-Maschinen auf die Verarbeitung von HORN-Klauseln beschränkt bleiben. Sobald diese Einschränkung auf HORN-Klauseln überwunden ist und die Konklusion von Subjunkt-Formeln auch komplexe, u.a. disjunktiv zusammengesetzte Formeln umfassen darf, ist die Schlussfolgerungsfähigkeit nicht mehr auf lediglich mögliche Kompetenzen restringiert. Die nachfolgenden Anmerkungen zum epistemischen Trade-off und dem daraus folgenden ökonomischen Optimierungsproblem würde entfallen.

Im Weiteren kann das hier vorgestellte Vorgehen auch in anderen Domänen angewendet werden. Als Beispiele kommen hierfür in Frage:

- ❑ *Fehlerdiagnose-Systeme*: Wenn potenzielle Fehlerquellen zuvor in einer Surmise-Relation oder einem Surmise-System eingebettet werden, kann eine effiziente Fehlerdiagnose durchgeführt werden. Systemanalytikern würde hiermit die Möglichkeit geboten, explizit vorliegende Fakten in die Wissensbasis einzugeben um im Anschluss den Raum der möglichen Fehlerquellen für eine erste Untersuchung zu minimieren.
- ❑ *Medizinische Diagnose*: Ontologiegestützte Wissensbasen, in denen Krankheits-Symptome mittels einer Surmise-Relation oder eines Surmise-Systems in einem Ordnungsraster strukturiert wurden, können in der medizinischen Praxis als Diagnosehilfen dienen. Ähnlich der Analyse technischer Systeme können auch hierbei Krankheits-Symptome als Fakten abgelegt werden, aus denen sich mittels computergestützter Inferenzen mutmaßliche Krankheits-Ursachen ableiten lassen.
- ❑ *E-Learning*: In einem E-Learning-System wird die Modellierung des Wissens über den Benutzer wichtig, wenn das System mit dem Benutzer keinen vordefinierten Ablauf von Lernaufgaben durcharbeiten kann, sondern die Lernsequenz auf die spezifische Wissensstruktur eines jeden Benutzers individuell zugeschnitten werden soll [PeTh00]. Beispielsweise können einzelne Aufgabenstellungen „übergangen“ werden, für die der Benutzer aufgrund bekannter Surmise-Beziehungen bereits bewiesen hat, dass er sie erfolgreich zu erfüllen vermag.

Als Erweiterung des Konzepts der Wissensräume kommt zudem ihre Verknüpfung mit der *formalen Begriffsanalyse* in Betracht [GaWi96]. Dadurch würde neben der extensionalen Auflistung aller Objekte, die sich unter einen Begriff subsumieren lassen, auch die intensionale Beschreibung des Begriffsinhalts durch die Auflistung aller Merkmale, die den Designata eines Begriffs zukommen, erreicht werden [RuWi96]. In dem hier zugrunde gelegten problemorientierten Paradigma der Betriebswirtschaftslehre würde dies bedeuten, dass zur Kompetenzspezifikation neben der extensionalen Auflistung aller Probleme, die von einem Akteur gelöst werden können, es ebenso möglich wird, auf intensionale Weise die gemeinsamen Merkmale aller Probleme anzugeben, die ein Akteur mit einer bestimmten Kompetenz zu lösen vermag.

Literaturverzeichnis

- [BrGu02] Brickley, D.; Guha, R.V.: RDF Vocabulary Description Language 1.0: RDF Schema. W3C Working Draft 30 April 2002. URL: <http://www.w3.org/TR/rdf-schema/>
- [CoMc02] Coates, T.T.; McDermott, C.M.: An exploratory analysis of new competencies: a resource based view perspective. In: Journal of Operations Management, 20. Jg. (2002), S. 435-450.
- [DoFa99] Doignon J.P.; Falmagne, J.C.: Knowledge Spaces. Berlin et al. 1999.
- [El01] Elbert, S.: Einführung eines Management-Support-Systems zum effektiven Skill-Management bei Bertelsmann mediaSystems. In: Schnurr, H.-P.; Staab, S.; Studer, R.; Stumme, G.; Sure, Y. (Hrsg.): Professionelles Wissensmanagement. Erfahrungen und Visionen. Aachen 2001, S. 129-144.
- [Fa90] Falmagne, J.C.; Doignon, J.P.; Koppen, M.; Villano, M.; Johannesen, L.: Introduction to Knowledge Spaces: How to Build, Test, and Search Them. In: Psychological Review, Vol. 97 (1990), No. 2, S. 201-224.
- [Fr00] Freiling, J.: Entwicklungslinien und Perspektiven des Strategischen Kompetenz-Managements. In: Hammann, P.; Freiling, J. (Hrsg.): Die Ressourcen- und Kompetenzperspektive des Strategischen Managements. Wiesbaden 2000, S. 13-45.
- [GaWi96] Ganter, B.; Wille, R.: Formale Begriffsanalyse. Berlin et al. 1996.
- [Gr91] Grant, R.M.: The Resource-Based Theory of Competitive Advantage: Implications for Strategy Formulation. In: California Management Review, 33. Jg. (1991), Nr. 3, S. 114-135.
- [Gu98] Guarino, N.: Formal Ontology and Information Systems. In: Guarino, N. (Hrsg.): Formal Ontology in Information Systems. Proceedings of the First International Conference (FOIS '98), June 6-8, Trento, Italy. Amsterdam et al. 1998, S. 3-15.

- [HePr00] Herling, R.W.; Provo, J.: Knowledge, Competence, and Expertise in Organizations. In: Herling, R.; Provo, J. (Hrsg.): Strategic Perspectives on Knowledge, Competence, and Expertise. Baton Rouge 2000. S. 1- 7.
- [Ho02] Horrocks, I.: DAML+OIL: A Reason-able Web Ontology Language. In: Jensen, C.S.; Jeffery, K.G.; Pokorný, J.; Saltenis, S.; Bertino, E.; Böhm, K.; Jarke, M. (Hrsg.): Advances in Database Technology - EDBT 2002, 8th International Conference on Extending Database Technology, Prague, Czech Republic, March 25-27, Proceedings. Berlin et al. 2002, S. 2-13.
- [KLW95] Kifer, M; Lausen, G.; Wu, J.: Logical Foundations of Object-Oriented and Frame-Based Languages. In: Journal of the ACM, Vol. 42 (1995), No. 4, S. 741-843.
- [Ko93] Korossy, K.: Modellierung von Wissen als Kompetenz und Performanz. Heidelberg 1993.
- [Ko99] Korossy, K.: Qualitativ-strukturelle Wissensmodellierung in der elementaren Teilbarkeitslehre. In: Zeitschrift für experimentelle Psychologie, 46. Jg. (1999), Heft 1, S. 28-52.
- [LaSu02] Lau, T.; Sure, Y.: Introducing Ontology-based Skills Management at a large Insurance Company. In: Glinz, M.; Müller-Luschnat, G. (Hrsg.): Modellierung 2002. Arbeitstagung der Gesellschaft für Informatik e.V. (GI). Bonn 2002, S. 123-134.
- [LaSw99] Lasilla, O.; Swick, R.R.: Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999. URL: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- [Ma00] Maedche, A.; Schnurr, H.P.; Staab, S.; Studer, R.: Representation Language-Neutral Modeling of Ontologies. In: Ebert, J.; Frank, U. (Hrsg.): Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik. Beiträge des Workshops „Modellierung 2000“, St. Goar, 5.-7. April 2000, Koblenz 2000. S. 129-142.
- [PeTh00] Peylo, C; Thelen, T.: Skills und Konzepte als Grundlage für Benutzermodellierung in einem Prolog ITS. Technical Report of the Institute for Semantic Information Processing, Osnabrück 2000.
- [Ri98] Rittgen, P.: Prozeßtheorie der Ablaufplanung - Algebraische Modellierung von Prozessen, Ressourcenrestriktionen und Zeit. Stuttgart - Leipzig 1998.
- [RuWi96] Rusche, A.; Wille, R.: Knowledge Spaces and Formal Concept Analysis. In: Bock, H.H.; Polasek, E. (Hrsg.): Data Analysis and Information Systems. Berlin et al. 1996, S. 427-436.
- [Sc93] Schrepp, M.: Über die Beziehung zwischen kognitiven Prozessen und Wissensräumen beim Problemlösen. Heidelberg 1993.
- [St02] Staab, S.: Wissensmanagement mit Ontologien und Metadaten. In: Informatik Spektrum, Band 25 (2002), Heft 3, S. 194-209.
- [Su02] Sure, Y.; Erdmann, M.; Angele, J.; Staab, S.; Studer R.; Wenke, D.: OntoEdit: Collaborative Ontology Engineering for the Semantic Web. In: Horrocks, I.; Hendler, J.A. (Hrsg.): The Semantic Web - ISWC 2002. First International Semantic Web Conference, Sardinia, Italy, June 9-12, 2002. Berlin et al. 2002, S. 221-235.

On the role of a Librarian Agent in Ontology-based Knowledge Management Systems

Nenad Stojanovic
Institute AIFB,
University of Karlsruhe,
76128 Karlsruhe, Germany
nst@aifb.uni-karlsruhe.de

Abstract: In this paper, we present an agent-based approach for the improvement of searching in an ontology-based knowledge management system. The system implements a library scenario in which users query the repository for knowledge resources. Consequently, the so-called Librarian Agent plays the role of the human librarian in the traditional library – it uses all possible information, about the domain vocabulary, the behaviour of previous users and the capacity of the knowledge repository, in order to help users find the resources they are interested in. We partially implemented the approach in the Web Portal of our Institute and some initial evaluation results are shown.

1. Introduction

The basic problem is that the searching for knowledge provided by traditional IR systems only partially reflects the process which humans use in searching for goods in the bricks-and-mortar environment [Jan98]. Briefly, in the non-virtual search, there exist a shop assistant, who helps the user to express his need more clearly and guides the user through the searching space.

Although some recent work is done in the query expansions [Bru97] and recommendation systems [Bal97], the whole process of simulating bricks-and-mortar environment is not modelled so far. The most crucial differences between searching for information in a knowledge management system and the non-virtual searching are:

- 1) the absence of the support for more precise explicitisation of the user's information needs [Sar75]
- 2) not targeting the capacity and organisation of the information repository at all.

In this paper, we present an approach for coping with these two issues, in order to enable more efficient searching for knowledge in a knowledge management system. As in the bricks-and-mortar environment, the main module is a (software) agent which analyses the information about the user's activities in the portal and the capacity of the information repository. Based on these analyses, the agent, through an interactive dialogue, guides the users in more efficient searching for information. Particularly, for a

query given by a user, the agent measures its ambiguity, regarding the underlying vocabulary (i.e. ontology), as well as the content (capacity) of the information repository. In case of high ambiguity, the agent suggests the user the most effective reformulation of the query, considering the underlying vocabulary, the information repository and the agents' experience (past behaviour of users). Last, the agent analyses the users' requests off-line and compares the users' interests with the capacity of the information repository, in order to find which "new titles" should be obtained or which topics are no more interesting for users. As in the real scenario, the agent does not perform interviewing of users, i.e. it does not require explicit feedback of users. In order to avoid disturbing users, the agent uses the information stored in the log file of the KM system. The logging system is based on the logging web-server transactions and it is independent of the concrete KM system. The traditional web-server logging approach is extended by capturing the semantic information about pages which the user has visited so far, so that the agent can perform various semantic analyses about the users' behaviour. Since the agent learns from the behaviour of all users, he can make the personalised recommendations, but also the recommendations to the anonymous users. Moreover, the agent can evaluate his own recommendations and consequently improve (learn) his service continually.

The paper is organised as follows: In Section 2, we elaborate on the requirements for more efficient searching in an ontology-based knowledge management system, whereas in Section 3, we discuss the role of a Librarian Agent in fulfilling these requirements. Section 4 contains concluding remarks.

2. The efficient querying in a KM system – the requirements

The problem of satisfying a user's requirement posted to the Information Portal [Bae99] is the question of whether a relevant information resource exists in the information repository, and if the answer is positive, whether the resource can be found by a user. Therefore, the efficient searching depends on:

1. the "quality" of the information repository in the portal,
 - if information resources reflect the needs of users, e.g. if the information repository contains information resources which users are interested in and
2. the "quality" of the searching process, i.e. when a relevant information resource exists in the repository, how easily (if any) the resource can be found. This problem can be divided into two sub-problems:
 - a) if a resource which is relevant for the user's information need can be found by the querying mechanism and
 - b) if the resource which is highly relevant for the user's information need can be found easily by the user in the list of retrieved results.

The first criterion (1) is the matter of the so-called "collection management policy"¹, which manages the deletion of old information resources and enter of new ones, corresponding to the changes in the user's interests.

¹ For an example, see [SSI02]

The retrieval of resources which are relevant for the user's need (2a) depends on:

- 1) the clarity of the expression of the need in the query which is posted to the system [Bae99], [Wen01]
- 2) the quality of the annotation (indexing) information resources in the repository

The part of this problem, a so-called prediction game between providers and users of information, can be resolved by using a commonly-agreed vocabulary, i.e. an ontology [Gua95] as the semantic backbone of the portal. We assume that such an ontology exists in the given domain, and that the system, consequently, benefits from using such a conceptual structure in searching for information [Gua99].

Finding (easily) an information resource which satisfies the users' information need (2b) depends on the capability of the system to interpret the term "relevance" in the right manner.

In the rest of the paper, we are focused on the two issues elaborated in the introduction: the quality of the knowledge repository and the clarity of the users' queries.

3. The Librarian Agent

3.1 The role of the Librarian Agent

As more elaborated in the introduction, the role of the Librarian Agent is:

- a) to support the disambiguation of the queries posted by users (query management) and
- b) to enable the changes in the knowledge repository regarding the users' information needs (collection management)

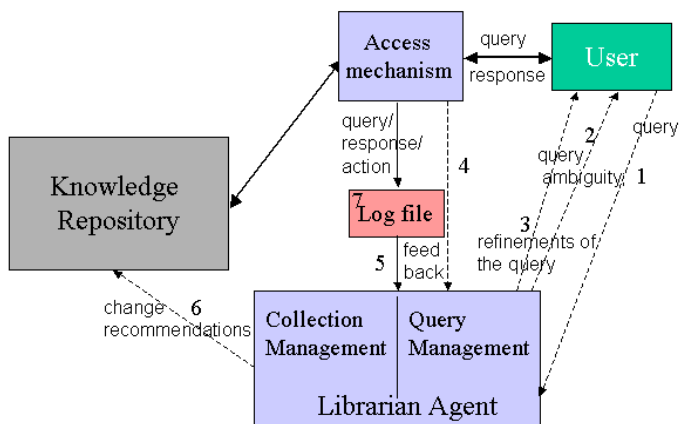


Figure 1. The roles of the Librarian Agent in the process of searching for knowledge

Figure 1 sketched the application scenario for the Librarian Agent. A user posts the query (cf. 1 in the Figure 1) which is first processed by the Librarian Agent. The Agent measures the ambiguity of the query (cf. 2 in the Figure 1) by considering the capacity of the knowledge repository and the domain vocabulary - ontology (for more details, see the next section). In case that the ambiguity of the query exceeds the given threshold, the agent recommends the user some changes (refinements) in the query (cf. 3 in the Figure

1). The Agent receives the feedback information about which suggestions the user has accepted (cf. 4 in the Figure 1), and it uses this information to refine its own strategies for creating recommendations

As the result of the querying, the access mechanism retrieves a list of resources, which is analysed (cf. 5 in the Figure 1) by the Librarian Agent in order to make recommendations for the changes in the collection (cf. 6 in the Figure 1). This recommendation takes into account the analysis of the queries posted by users and the used vocabulary, as well.

In order to avoid disturbing the users by additional questioning, all feedback information is collected implicitly by analysing the activities of the users captured in the log file of the system (cf. 6 in the Figure 1).

In the next section, we give more details about our approaches for the query management.

3.2 The Query Management

3.2.1 Measuring Query Ambiguity

Recent analyses [Cro02] have shown that the precision in searching for information depends strongly on the clarity of the query which a user posts to the system. When the query is formulated in an ambiguous manner, one can expect that a high percent of irrelevant results can be retrieved, independently of the mechanism which is used for searching. Therefore, we see the query disambiguation as the initial step in searching for information in a KM system.

The Librarian Agent observes the query ambiguity in two dimensions:

- 1) the structure of the query
- 2) the content of the knowledge repository

1) The structure of the query

Regarding ambiguities in the structure of the query, two other issues are defined:

a) structural ambiguity

the structure of a user's query is analysed regarding the underlying ontology; in [Sto02a] we define three such criterions: compactness, completeness and aggregation.

b) semantic ambiguity

the terms form a user's query is analysed regarding the relation which exist in the underlying ontology. In [Sto02b] we have defined two measures for estimating the semantic ambiguity of an ontology-based query: i.e. Clarity and ContextClarity. Very briefly, the Clarity is inversely proportional to the number of subconcepts of a concept and ContextClarity is inversely proportional to the number of relations between two concepts.

2) The content of the knowledge repository

The ambiguity of a query posted in a knowledge repository is obviously repository-dependent. For example, when a user interested in the competitors in 2002 Soccer World Cup gives the query “World Cup” against the collection of the news articles in which the articles about Chess World Cup Tournament are predominant, it is simply impossible for the system to return soccer articles consistently ranked higher than related to chess.

We introduce the Response factor for taking into account the specificities of the knowledge repository content in determining the ambiguity of a query.

The Response factor of a query Q is the measure how the terms from that query cluster the resources in the underlying knowledge repository (KR)

$$\begin{aligned} \text{Response}(Q, KR) &= \min_{\text{all } Q'} P(\text{not Relevant}(Q, KR) / (\text{Relevant}(Q', KR))) = \\ &= \max_{\text{all } Q'} \frac{\text{NumberOf Relevant}(Q', KR) - \text{NumberOf Relevant}(Q, KR)}{\text{NumberOf Relevant}(Q', KR)} \end{aligned}$$

whereas $\text{NumberOf Relevant}(X, KR)$ denotes the set of knowledge resources stored in the KR which are annotated with the X – in other words, it is the number of results by querying for X the repository KR and $Q' = \{x, x \subset Q, x' \neq \{\}\}$ is the set of all non-empty subsets of the Q. The special case is when the Q contains just one term

$$\text{Response}(Q, KR) = P(Q) = \frac{\text{NumberOf Relevant}(Q, KR)}{\text{TotalNumber}(KR)}$$

whereas $\text{TotalNumber}(KR)$ is the total number of resources in the repository.

The Response factor describes the probability that a knowledge resource relevant for the query Q will not be relevant for the one of the non-empty subsets of the Q –

$$P(\text{not Relevant}(Q, KR) / (\text{Relevant}(Q', KR)))$$

When this probability is very low, it means that the query Q is “covered” by a Q', i.e. that query for Q results in almost all results for the query Q' and that query Q should be extended (refined), in order to return more precise results. The difference between Q' and Q is the set of terms which effect the querying process very low, and probably they should be refined (see the next section). However, this is only a recommendation how to get results which are closer to the information need of the user – it is possible that the user is satisfied with the original query.

3.2.2 Query Refinement

For each query posted to the system, Librarian Agent checks the structural ambiguities, and if they are present, it suggests the improvements of the structure of the query. Next, the Agent calculates ambiguity of the query and gives the users the recommendations how to change the query in order to refine their information needs. Nevertheless, the users can initiate the query refinement on their own.

The Librarian Agent uses three strategies for the query refinement:

- A. according to the structure of the underlying ontology
- B. according to the content of the knowledge repository
- C. according to the users' behaviour (usage - query refinement done by users)

A. The structure of the ontology

In case that the query contains a structural ambiguity, it has to be resolved by considering the structure of the ontology. For example, when the query is incomplete, the Librarian Agent recommends the expansion of the query in such a manner that completeness is achieved. For more details see [Sto02b].

B. The capacity of the knowledge repository

The extension of the query terms should correspond to the characteristics of the document term space. The most popular method in the information retrieval is the so-called local context analysis [Sal90] in which the top-ranked documents are used for the query extension. The extension is usually done by using a variation of the Rocchio coefficients [Roc71]. The query is extended by increasing the influence of the most frequently appearing terms from the top-ranked (i.e. relevant) documents, and by decreasing the influence of the terms from non-relevant documents.

In Section 4.2.1, we have defined the Response factor as the measure of the ambiguity regarding the knowledge repository. Obviously, it is very useful for refining the query.

In case that a query Q is ambiguous because the Response factor is too low, i.e. there is a sub-query Q' for which the list of retrieved results is almost identical to the list of Q , then such a query can be refined by refining the terms for the set containing the difference between Q and Q' .

However, since our approach is ontology-based, the extension we provide is ontology-based, as well. We are not interested in the frequently appearing terms per se, but in the semantic extensions of those terms. For example, when the query is about “ontology+researcher” and the most frequently appearing terms in the top-ranked documents (beside “ontology” and “research”) are “professor”, “assistant” and “student”, which belong to the hierarchy of the concept “researcher”, probably the relevant strategy is to expand the query with the information about the “researcher”.

C. The users' behaviour

By searching the portal, a user makes a query, observes the list of retrieved results, probably refines the query in some manner, then observes the new list and “clicks” on the information resource when she notes a relevant one; when not, she refines the query again... This is the ordinary user's behaviour, and can be very useful for predicting what can be relevant for a user in a situation. By analysing such information, the system can learn how to rewrite a query in case the user is not satisfied with the retrieved results.

We define three types of query-rewriting patterns based on the users' behaviour described in the rest of the section: expansion-, reduction- and generalization/specialization- pattern.

Considering that a query represents an interest of the user, we can assume that two users who make the same query have the similar interests, regarding the query (situation). They also have the same goal in the searching – to find an information resource about the topic of interest. This assumption allows us to make another one, about the behaviour of users during searching: users with similar interests (goals) should behave in a similar manner. Consequently, for a given query, the system can suggest a user to repeat the

behaviour of the users who have already posted the same query. For example, when a lot of users **expand** the initial query “aspirin and headache” with the term “young” or “old”, in order to get more precise results, we can conclude that the treatment of headache by aspirin heavily depends on the age of a patient. Every time the user makes such ambiguity in a query, the system should suggest the user to expand the query with the information about age. It is worth noting that this analysis is performed on the ontological level – “young” and “old” are only two values for the property “age” of the concept “patient”. The user should be asked to expand the query not only with the terms “young” or “old”, but to select any valid value for “age”, e.g. “middle-aged”.

The same principle can be used for **reducing** too specific queries, in case no results were retrieved for such a query. The case that users often reduce the query “aspirin, headache, female, young” to the query “aspirin, headache, young” can be interpreted as the irrelevance of the patient’s gender for curing the headache by using aspirin. The system should recognise this reduction pattern and recommend such a change every time a user makes this ambiguity in a query. By generalising this pattern on the ontological level, the system can process/treat the previously unseen examples. For example, in case the discovered recommendation is that “queries about side-effects of using aspirin in the patient who suffers from rheumatoid arthritis should not contain information about “age”, the query “side-effect, rheumatoid arthritis, aspirin, young, male” can be reduced to “side-effect, rheumatoid arthritis, aspirin, male”, although the initial query has not been seen previously.

4. Conclusion

In this paper, we presented an agent-based approach for the improvement of the searching in an ontology-based knowledge management system. The system treats a library scenario in which users query the repository for knowledge resources. Consequently, the so-called Librarian Agent plays the role of the human librarian in the traditional library – it uses all possible information, about the domain vocabulary, the behaviour of previous users and the capacity of the knowledge repository, in order to help users to find the resources they are interested in. Based on various analyses, the agent, through an interactive dialogue, guides the users in more efficient searching for information. Particularly, for a query given by a user, the agent measures its ambiguity, regarding the underlying vocabulary (i.e. ontology), as well as the content (capacity) of the information repository. In case of high ambiguity, the agent suggests the user the most effective reformulation of the query. Last, the agent analyses the users’ requests off-line and compares the users’ interests with the capacity of the information repository, in order to find which “new titles” should be obtained or which topics are no more interesting for users.

We find that this approach represents a very important step in simulating the brick-and-mortar environment and benefiting from applying the practical results obtained in that area in the searching for information in the virtual world. Moreover, this approach leads to the self-adaptive knowledge portals, which can discover some changes from the user’s interactions with the system automatically and evolve their structure correspondingly.

5. References

- [Bae99] Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Addison-Wesley-Longman Publishing co., 1999.
- [Bal97] Marko Balabanovic, Yoav Shoham: Content-Based, Collaborative Recommendation. *CACM* 40 (3): 66-72 (1997)
- [Bru97] Bruza PD and Dennis S. Query Reformulation on the Internet: Empirical Data and the Hyperindex Search Engine. In: Proceedings of RIAO97, Computer-Assisted Information Searching on Internet, Montreal, June 1997
- [Cro02] Cronen-Townsend, S. and Croft, W.B., Quantifying Query Ambiguity, in the Proceedings of HLT 2002, pp. 94-98.
- [Jan98] Janiszewski, C. "The Influence of Display Characteristics on Visual Exploratory Search Behavior," *Journal of Consumer Research*, 25 (3), 290-301, 1998
- [Gua95] Guarino, N. and Giaretta, P. 1995. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars (ed.) *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995*. IOS Press, Amsterdam: 25-32.
- [Gua99] N. Guarino, C. Masolo, and G. Vetere, "OntoSeek: Content-Based Access to the Web", *IEEE Intelligent Systems*, 14(3), pp. 70-80, (May 1999).
- [Roc71] Rocchio. J. 1971. *Relevance feedback in information retrieval. The Smart Retrieval system---Experiments in Automatic Document Processing*. G. Salton. Ed. Prentice-Hall Englewood Cliffs. NJ. 313-323.
- [Sal90] Salton, G. and Buckley, C. *Improving retrieval performance by relevance feedback*. *Journal of the American Society for Information Science*. 41(4): 288-297, 1990.
- [Sar75] Saracevic, T. (1975). Relevance: A Review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, (6), 321-343
- [SSI02] SOSIG Internet Catalogue Social Science Information Gateway Collection Management Policy, <http://www.sosig.ac.uk/desire/collect.html>
- [Sto02a] Stojanovic, N., Stojanovic, L. Usage-oriented Evolution of Ontology-based Knowledge Management Systems, ODBASE 2002,
- [Sto02b] Stojanovic, L., Stojanovic, N., Maedche, A. A Framework for Change Discovery in Ontology-based Systems. Second International Workshop on Evolution and Change in Data Management (ECDM 2002), held in conjunction with the 21st International Conference on Conceptual Modelling, ER 2002, Tampere, Finland
- [Wen01] Wen, J.-R., Nie, J.-Y. and Zhang, H.-J. Clustering User Queries of a Search Engine. WWW10, May 1-5, 2001, Hong Kong.

SWAP: Ontology-based Knowledge Management with Peer-to-Peer Technology

¹Marc Ehrig ¹Christoph Tempich ²Jeen Broekstra
²Frank van Harmelen ²Marta Sabou ²Ronny Siebes ¹Steffen Staab
 ²Heiner Stuckenschmidt

¹{meh,cte,sst}@aifb.uni-karlsruhe.de

Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany

²{jbroeks, frankh, marta, ronny, heiner}@cs.vu.nl

Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

1 Introduction

In today's knowledge-based economy, the competitiveness of enterprises and the quality of work life are directly tied to the ability to effectively create and share knowledge both within and across organizations. Many enterprises have therefore spent huge amounts of money to implement centralized knowledge management systems (KMSs).

Ontologies have shown to be the right answer to knowledge structuring and modelling by providing a formal conceptualization of a particular domain that is shared by a group of people in an organization [O'L98]. However, KMSs based on centralized ontologies need a long development phase and are difficult to maintain.

[BBT02] suggest a distributed approach to Knowledge Management which better fits the true situation in organizations and the processes people are used to. From a technological point of view **peer-to-peer** (P2P) solutions are particularly well suited, because they make it possible for different participants (organizations, individuals, or departments) to maintain their own knowledge structure while exchanging information. However, today's P2P solutions are extremely limited (they mostly rely on keyword search) and not appropriate for the high requirements of a KMSs.

The creation of ontologies is still the bottleneck for a fast development. **Emergent Semantics**[Mae02] builds on lightweight (e.g. a file structure with files as instances) and/or heavyweight ontologies that different participants have created. It considers the overlap between simple ontology definitions in order to build shared ontologies. As new semantical structures emerge from known structures, knowledge management can occur in a distributed fashion without overhead through central administration.

The Semantic Web and Peer-to-Peer (SWAP) project¹ demonstrates that taking the successful technologies of the above research areas, will allow support for decentralized envi-

¹funded by EU under contract No. IST-2001-34103; <http://swap.semanticweb.org>

ronments. Participants can maintain individual knowledge structures on their peers (PCs), while sharing knowledge in ways such that administration efforts are low, but knowledge sharing and finding is easy. This paper gives a general presentation of the project: the baseline technologies, an overview of the system, the case studies, and related work.

2 The SWAP System

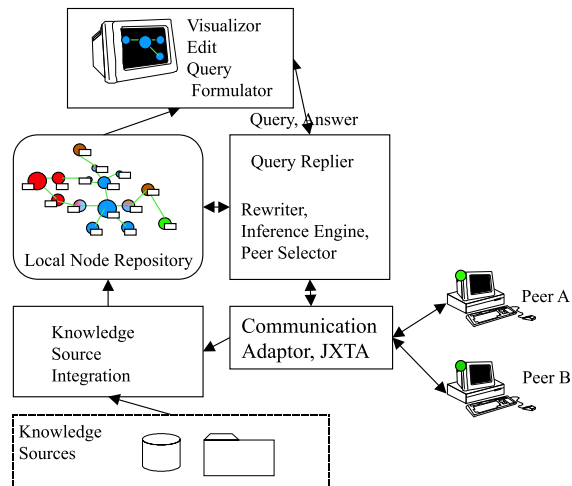


Figure 1: Architecture

The system consists of a set of peers called “SWAP Nodes”. The knowledge of a particular peer is extracted from several Knowledge Sources, then integrated and stored in the Local Node Repository LR. A user interface ensures that the user can edit/browse/query the knowledge. Queries that cannot be answered by the available knowledge are sent to the whole system. A specialized component deals with rewriting these queries and selecting the peers which are likely to know the answer. Further, we describe several major issues.

Providing Knowledge: Every participant in the SWAP system has to be enabled to provide his knowledge. In this module the user can select among existing knowledge sources from his personal computer for addition (emails, files, folders, databases, bookmarks, or whole ontologies). A special component will extract ontology-like structures from the selected sources. These sources will then be integrated automatically into the LR, which is represented in RDF(S). Merging routines will be required at this point. Content is therefore stored as an ontology, which allows high-quality processing. Besides the content we also include technical information about each entity i.e. an ID and the location (e.g. peer123//outlook//abcdef1234). The trust in a specific information is another important measure which is saved and maintained. The process guarantees that even a novice can start quickly.

Updates and Changes: The LR is an integrated view on all known information sources, both local sources and other peers. Any change in the knowledge sources is propagated to the LR. The edit functionality is thought to be useful for advanced users and knowledge engineers. Those who are familiar with ontologies and their graphical structures can

intervene and append their own structures by adding, deleting, or renaming concepts and relations between the concepts. Changes to the LR due to editing are currently not propagated to the underlying structures, since they can provoke undesirable contradictions. Changes will also occur when answers from other peers provide information return. This information is incorporated to get an overview of the knowledge in the P2P network.

Views: A view is a pre-defined perspective of the knowledge. It can be generated out of selected structures (e.g. only email folders, peer123 perspective) or the entire LR. They are implemented using different visualization techniques (topic hierarchies, thematic maps) [FSvH02]. The user can easily browse through this graphical knowledge representation.

Query generation, distribution, and answering: Finally, the most important aspect for the user is to get answers to specific queries. Queries can be entered by clicking in the views graph or manually as text. The query itself can have various degrees of complexity from simple conjunction to recursion formulated in an RQL-related query language. Eventually, it is sent to the internal inference engine which tries to solve the request. If the inference engine can not get an answer from the local repository, it splits the query and distributes the sub-queries in the P2P network. In this case the query first has to be rewritten in order to fit the underlying knowledge structures on other peers. The routing is based on meta-information about their knowledge and trust figures. The other peers will answer the queries in the same fashion and finally return answers, which are put together and presented back to the user. He can then decide if he wants to add the answer to his own knowledge representation. Answers consist of statements which can also link to e.g. files. While the peer communicates with others, information about the network is stored which can then be used for finding better paths when having an own query.

File sharing: If the answer received from the SWAP system is a file title or an email subject the user might like to get the entire file from the answering peer. The request is sent back to the responding peer with all meta information from the answer and the peer solves this file request with a separate module.

3 Case studies

First, in the Knowledge Management Case Study we are going to investigate the work of investment analysts at **Dresdner Bank**, one of the largest German banks. The environment for this case study is fairly closed in terms of topics and technology, and comprises only a defined and limited number of participants. The case study surveys their use of current central knowledge management technology. The analysts store big amounts of knowledge in various formats on their personal computers. The SWAP P2P based solution will be installed and test-driven in order to come up with a realistic estimation of the work that might be saved (or wasted) by the P2P system.

Second, in the virtual enterprize case study, contractor **IBIT** (Fundació de les Illes Balears per a la Innovació Tecnològica) will use P2P technology to enable knowledge sharing between different small and medium-sized enterprizes that all have a stake in sustainable tourism. This topic is of high importance to the local economy of the Balearic Islands, and it involves a large number of very different peers (local government, tour operators, hotel operators, university researches etc.), and involves very heterogenous knowledge sources. Also, the peers are geographically dispersed, they are very different in nature, and they

will be intermittently connected to the network.

4 Related Work and Summary

Edutella[ANS02] also provides a P2P infrastructure for exchanging metadata but focuses on the education community. Lecturers can publish their notes without losing ownership and students can easily access them. The InfoQuilt System[ASA01] provides a framework for formulating complex information requests, involving multiple ontologies, and supporting a form of knowledge discovery. From the local ontologies less quoted ones eventually disappear through evolution. A set of InfoSleuth[FPNB99] agents collaborate at a semantic level to execute information gathering and analysis tasks, where the underlying information sources can be diverse both in their structure and content, as in SWAP. LARKS[SWKL02] is an agent capability description language that enables advertising, requesting and matching over the Internet. The aim of EDAMOK[BBT02] is to develop research in information technology and software tools that support the Distributed and Autonomous Management of Knowledge, not using an ontology premise though.

SWAP however has a different approach. The combination of Semantic Web and P2P is highly innovative with prospective benefits to the individualization of work views as well as to the facilitation of knowledge sharing between peers. SWAP takes on the challenges brought up by this novel combination of ontologies and P2P computing such that knowledge finding and sharing is effectively possible. It also considers how Emergent Semantics knowledge is constructed from the distributed environment. The technology is implemented into a system of which the architecture has been shown. The case studies from different companies for this specific problem define the application.

References

- [ANS02] B. Ahlborn, W. Nejdl, and W. Siberski. OAI-P2P: A Peer-to-Peer Network for Open Archives. In *Workshop on Distributed Computing Architectures for Digital Libraries - ICPP2002*, 2002.
- [ASA01] M. Arumugam, A. Sheth, and I. B. Arpinar. Towards Peer-to-Peer Semantic Web: A Distributed Environment for Sharing Semantic Knowledge on the Web. Technical report, Large Scale Distributed Information Systems Lab, University of Georgia, 2001.
- [BBT02] M. Bonifacio, P. Bouquet, and P. Traverso. Enabling Distributed Knowledge Management: Managerial and Technological Implications. *Novatica and Informatik/Informatique*, III(1), 2002.
- [FPNB99] J. Fowler, B. Perry, M. H. Nodine, and B. Bargmeyer. Agent-Based Semantic Interoperability in InfoSleuth. *SIGMOD Record*, 28(1):60–67, 1999.
- [FSvH02] C. Fluit, M. Sabou, and F. van Harmelen. Ontology-based Information Visualisation. In Vladimir Geroimenko, editor, *Visualising the Semantic Web*. Springer Verlag, 2002.
- [Mae02] A. Maedche. Emergent Semantics for Ontologies. *IEEE Intelligent Systems*, 2002.
- [O’L98] D. O’Leary. Using AI in knowledge management: Knowledge bases and ontologies. *IEEE Intelligent Systems*, May 1998.
- [SWKL02] K. Sycara, S. Widoff, M. Klusch, and J. Lu. *LARKS: Dynamic Matchmaking Among Heterogeneous Software Agents in Cyberspace*, chapter 5. Kluwer, 2002.

Integration with Ontologies

Conference Paper WM2003, April 2003, Luzern

author: Andreas Maier¹ (maier@ontoprise.de)
co-authors: J. Aguado² (jessica@miramon.net)
A. Bernaras² (amaia@miramon.es)
I. Laresgoiti³ (lares@labein.es)
C. Pedinaci² (carlos@miramon.net)
N. Peña³ (npena@labein.es)
T. Smithers² (tim@miramon.net)

Abstract: One of today's hottest IT topics is integration, as bringing together information from different sources and structures is not completely solved. The approach outlined here wants to illustrate how *ontologies* [Gr93] could help to support the integration process.

The main benefits for an ontology-based approach are

- the ability to picture all occurring data structures, for ontologies can be seen as nowadays most advanced knowledge representation model
- the combination of deduction and relational database systems, which extends the mapping and business logic capabilities
- a higher degree of abstraction, as the model is separated from the data storage
- its extendibility and reusability

First we will give a motivation for our approach and build the requirement specification (1.). After describing the foundations (2.) we will introduce our ideas about new software tools supporting the ontology-based integration (3.) and present a case study (4.), where the ontology-based integration is going to be realized - both from the view of the software producer¹ and of the teams^{2,3}. Closing remarks to limitations of our approach and to related publications will complete this work.

1. Motivation for an Ontology-based Approach

Today's users and IT professionals have high expectations towards software applications:

- they want to access the content they need
- this content must be accurate and free of redundancy
- the application must be intuitive and easy to use
- the application must be reusable and extendable
- the application must be implemented in a short and inexpensive way and within the current IT legacy environment

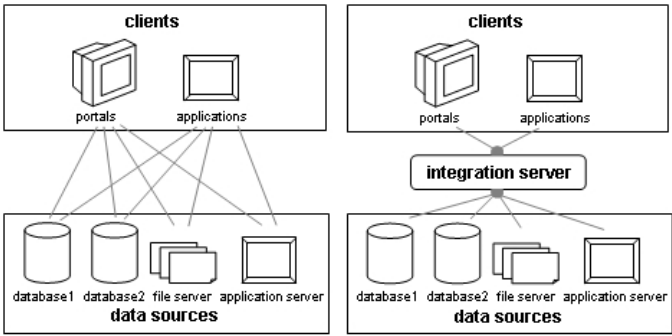
To meet these expectations, the content has to be identified from the different sources (i.e. databases, applications, XML-Files, unstructured text files ...), and then to be integrated. But this means not building only connectors [Kr99] between applications, because

¹ ontoprise GmbH

² Parque Tecnológico de San Sebastian, Spain

³ LABEIN, Zamudio, Spain

syntactical incompatibilities could be reduced by approaches like SQL [KK01] or XML [An03]; nor it's only tying up diverse data sources and displaying them on a common interface (Picture 1 left).

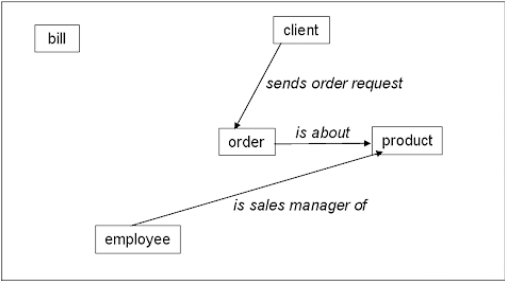


Picture 1: point-to-point-connections vs. a “real” integration solution

The goal of integration is to consolidate distributed information intelligently, free of redundancy, processed and operated by the right business logic to deliver the appropriate and condensed answer and offer the end user a simple access to it, without him needing knowledge about the underlying data structures (Picture 1 right). We believe that with ontologies there's now a model at hand to fit for this goal.

1.1. Introducing Example

A typical integration scenario found in large organisations is the management of product data and contacts (clients, suppliers, employees), which proves to be very difficult. These information lie widespread in different departures and there in different sources like ERP/PPS⁴-systems, CRM⁵-applications [Sc00] [UB03], databases, email programs, documents, organizers, etc., - often redundantly.



Picture 2: an “integration” ontology

⁴ ERP=Enterprise Resource Planning, PPS= Production Planning and Scheduling

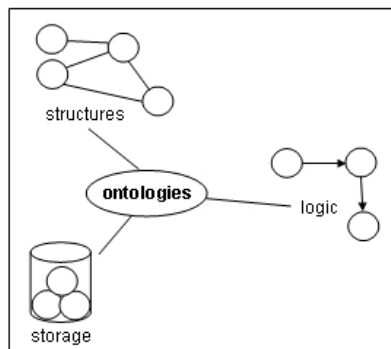
⁵ CRM=Customer Relationship Management

An ontology (Picture 2) could now catch up the different concepts that we want to integrate: **clients** (i.e. from a CRM-application), **orders**, **products** (i.e. from a PPS-System) and **employees** (i.e. from an intern “telephone and birthdays” list). A fifth concept, **bill**, has not been taken into consideration yet by the existing systems; we will generate its instances automatically by a rule (2.4).

1.2. Defining the Requirements

In our integration process we then have

- to picture all existing data **structures** (*requirement 1*), which can be simple table structures up to complex hierarchical structured data with deep inheritance,
- to **map and merge** these schemas (*requirement 2*),
- to define the **logic** for the whole new application (*requirement 3*) (hereby we will be supported by deductive inference mechanisms) and
- to provide a performant data **storage** for the information (*requirement 4*).



Picture 3: **ontologies meet the requirements**

In our view ontologies are the best representation model to meet these requirements (Picture 3). In the next chapter (2.) we want to prove this statement.

2. Foundations: Enabling the Ontology-based Integration

The foundations talked about next – knowledge representation, mapping, deductive logic and databases – are not new for themselves. It’s their interaction what makes it necessary to shortly describe them. Thereby we will especially go into their impact on the ontology-based integration.

2.1. Requirement 1: Picture all Data Structures

In [Ma01] we compared several knowledge representation models and discussed the advantages and weaknesses of them. As a conclusion we found that ontologies are the most advanced model of all of them, summing up most of the qualities of the others:

- Like **Taxonomies** [Pe89], ontologies are able to picture *hierarchies*.

- Like **Thesauri** [Me95], **Semantic Nets** [Ho86] and **Topic Maps**⁶ [PH02], ontologies contain *relations*. With them, complex contexts can be modelled and visualized in nets. *Linguistic contexts* (i.e. multilingualism or synonym relations), terminologies and classifications can be described, through which the semantic of the integration solution is increased [An01].
- Like the **EntityRelationship-Model** (ER) [BCB91] and unlike the others mentioned above, ontologies have a *data model* distinguishing schema information from facts. This is essential for storing the facts (Requirement 4: Provide a Data Storage). As relational databases do not provide an object model, they have some difficulties in picturing taxonomies and must define primary keys themselves. Every ER-model can be transformed in an ontology and, with some expense and limitations, vice versa.
- As an **object based model**, ontologies support *inheritance* and *multiple inheritance* of attributes.

2.2. Logic Models

With the ability to define mapping and deductive rules respectively axioms⁷, logic brings in an important impact for the integration. Logic models (i.e. **Prolog** [Sp96], **Datalog** [DL91], **F-Logic** [KL90] [An02], **Description Logic** [Ba02]) have to be seen complementary to ontologies.

Especially F-Logic⁸ [AL03] acts as a bridge between the model and logic, because it covers the ontological information as well as the rules. It can also be used to query the system similarly to SQL.

Logic models help us

- to connect the different data sources by **mapping** or **merging** rules
- to easily build a deductive “**business logic**” upon our integrated information base⁹
- to check the consistence of the knowledge base [AS92a].

2.3. Requirement 2: Mapping and Merging

If only integrating databases, an ontology-based mapping would not seem to be necessary at first. But in comparison to pure database mappings, which could be used instead, the ontological approach leads to a higher degree of abstraction, connecting not only primary keys and table rows, but working on a more conceptual level.

Before starting the mapping procedure the structures respectively the schemas have to be imported into the ontology. For thus, an ontology modelling tool must provide various

⁶ As the dates of the references (6.) show, Taxonomies, Thesauri, Semantic Nets and the ER-model are comparatively old models. Topic Maps are the newest of them and may merge with ontologies.

⁷ As a special type of axioms, constraints help us to define restrictions within our system.

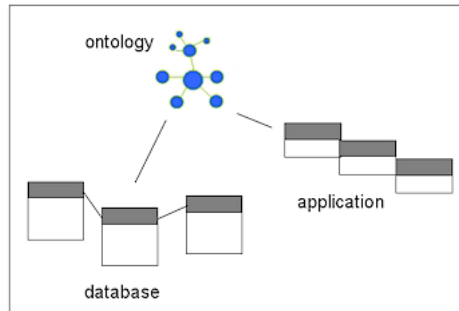
⁸ Comparing the mentioned logic models, we see F-Logic as the most capable of them; Prolog is rather a programming language and contains artefacts, Description Logic misses instance reasoning and Datalog isn't expressive enough.

⁹ Deduction can also be very effective in a stand-alone application, but much more in an integration scenario. By consolidating applications and their underlying processes often new contexts are created and require a new business logic on top.

schema import filters for different formats (i.e. for all relevant commercial databases). Such a tool must also support the fundamental mapping types

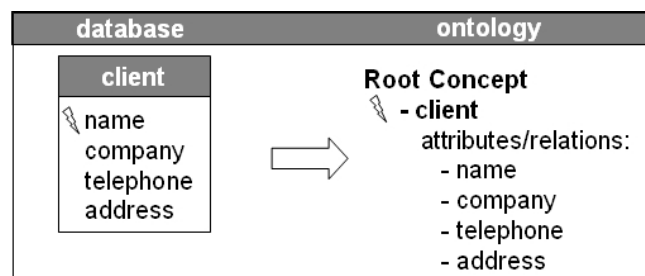
- concept-to-concept mapping
- attribute-to-attribute mapping
- attribute-to-concept mapping as well as
- conditions and constraints¹⁰ on the mapping rules (which is not explained further here)

2.3.1. Schema Import



Picture 4: Importing diverse schemas

The equivalent to the schema import described next is the schema export (2.5), that we will need for *requirement 3* below. Beneath other schema imports (i.e. for formats like RDF¹¹ or DAML¹²), the SQL import plays a very important role. As the commercial databases¹³ usually a different syntax of SQL, different imports for each database are needed¹⁴.



¹⁰ i.e. unit conversions

¹¹ <http://www.w3.org/RDF/>

¹² <http://www.daml.org/>

¹³ i.e.: MSSQL Server (Microsoft), DB2 (IBM), Oracle, MySQL

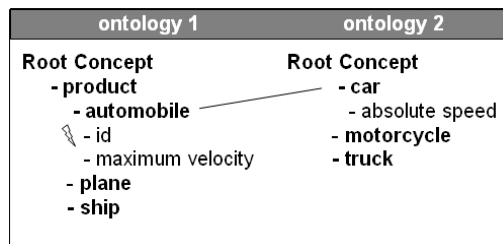
¹⁴ An SQL import has already been realized in *OntoEdit*, ontoprise's ontology modelling tool. For detailed information and screenshots please read the product tutorial at http://www.ontoprise.de/documents/tutorial_ontoedit.pdf.

Picture 5: Importing a database table

After an import of a database table, this one is embedded as a concept into the concept taxonomy. The former primary key from the attribute “name” has moved to the object id of the concept “client”. Tables are interpreted as objects or concepts, as they usually contain information about a distinct entity. Rows typically describe attributes of that entity and are coherently interpreted as attributes of concepts.

2.3.2. Concept-to-Concept Mapping

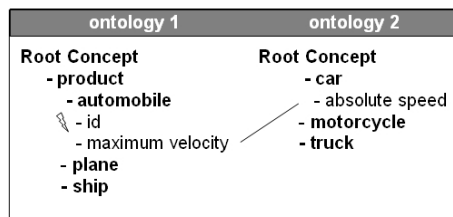
An ontology mapping process¹⁵ is very similar to pure database respectively XML-mapping [Bo01]. In each case the two schemas, which are going to be mapped, are displayed in vertical rows parallel to each other (picture 6: concept-to-concept-mapping). If two concepts of two different sources contain the same type of information, a concept-to-concept mapping can be drawn (i.e. “automobile” and “car”).



picture 6: concept-to-concept-mapping

2.3.3. Attribute-to-Attribute Mapping

An attribute-to-attribute mapping connects two attributes, stating that these contain the same information (i.e. “maximum velocity” and “absolute speed”) (picture 7: attribute-to-attribute-mapping). A previous concept-to-concept mapping is prerequisite for that.

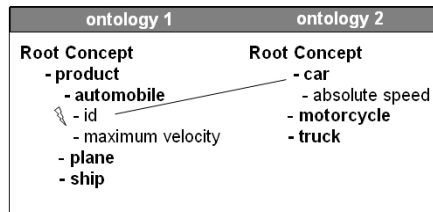


picture 7: attribute-to-attribute-mapping

¹⁵ *OntoEdit* contains *OntoMap*, a mapping tool for ontologies. More information can be found at http://www.ontoprise.de/documents/tutorial_ontoedit.pdf.

2.3.4. Attribute-to-Concept Mapping

Another important aspect is the mapping of table rows (attributes) that are represented as concepts in other formats. The attribute-to-concept mapping shown below states that the primary key “id” of “automobile” in connected to the object id of “car” (picture 8: attribute-to-attribute-mapping).



picture 8: attribute-to-attribute-mapping

2.4. Requirement 3: Deductive Logic

An often asked question is: “Why using logic? Didn’t databases solve all problems decades ago?”

On the one hand applications with lots of logical dependencies (i.e. configuration or variant management systems, solutions representing extensive knowledge domains, expert systems) can be realized much better with rule-based systems.

On the other hand deductive logic reduces complexity. It’s a difference, if you ask

- “Who is the contact person of client ‘Smith’?” or
- “Who is the employee that handles the orders of the product, that client ‘Smith’ has ordered?” (
- Picture 9: **using deductive rules**).

As this is just a small example, in really complex contexts with many relations between the concepts of the ontology the effort and complexity to realize in SQL quickly gets too high.

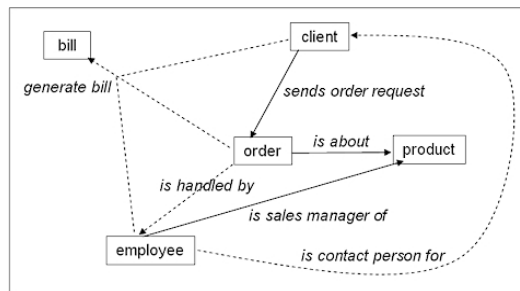
As a third reason, the user doesn’t need to know the underlying data structures. For example he only knows, that he can ask for “contact persons of clients”, and not the whole conceptual structure that lies behind this question.

2.4.1. Rule Modelling

It’s an often underestimated fact for today’s rule-based-systems in commercial environments that users do not want to encode these rules by hand, just as they dislike to write SQL queries. Therefore we will propose our idea of a visual rule editor in (3.1).

After having tied together the different data sources (2.3.1-2.3.3), we want to illustrate the extendibility that deductive logic rules offer to applications. Our basic ontology (Picture 2: an “integration” ontology) has been enriched by three simple rules to expand the knowledge base (

Picture 9: **using deductive rules**):

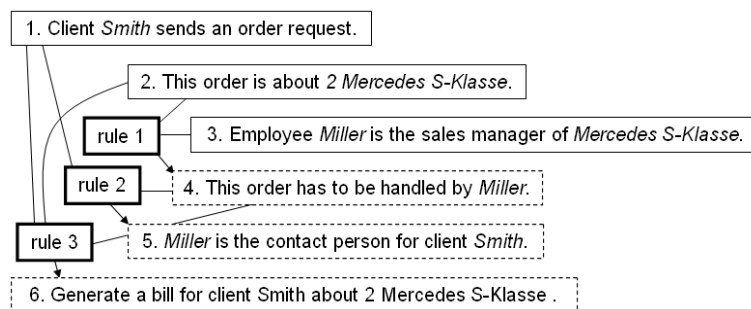


Picture 9: using deductive rules

- If an order is about a product and an employee is the sales manager of this product, this order **is handled by** the employee. (*rule 1*)
- If a client sends an order request, which is handled by an employee, this employee is the **contact person for** the client. (*rule 2*) As we see, *rule 2* is based on *rule 1*.
- If a client sends an order request, which is handled by an employee, then **generate a bill** with all available information: client, name and quantity of the product, the responsible employee ... (*rule 3*)

2.4.2. Inferencing

By an inferencing process¹⁶ the rules are applied to the given facts (1.,2.,3.) and extend the knowledge base by the newly created facts (4.,5.,6.). picture 10 visualizes this process.



picture 10: inferencing graph

In analogy to the graphical rule editor (3.1), there has to be a visual rule debugger for the modeller to show the outcome of his rules. In (3.2), we propose our idea of a visual rule debugger.

¹⁶ Inferencing has been realized in *OntoBroker™*, ontoprise's core ontology server. Please read the product tutorial at http://www.ontoprise.de/documents/tutorial_ontobroker.pdf to get detailed information.

After the inferencing you can decide whether you want to materialize the attained new facts into the data storage (2.5.), or to keep them only virtually in the inference server.

2.5. Requirement 4: Provide a Data Storage

As a name for a mainly unchanging pool of information the terms *Knowledge Base* or *Repository* are often used. We prefer to say *data storage* instead, emphasizing that in integration solutions the content is permanently changing.

For maintenance reasons, the data itself should be kept only once, preferably in the origin application. If this application isn't able to query, we propose migrating it to a database. Although the ER-Model has its weaknesses (2.1), we suggest using relational databases as storage because of its widely spread and mature solutions. In comparison to other repositories there's no alternative concerning performance and compatibility. Therefore we need an SQL-Export for creating the database schema out of the ontology.

3. Introducing a Toolbox for the Ontology-based Integration

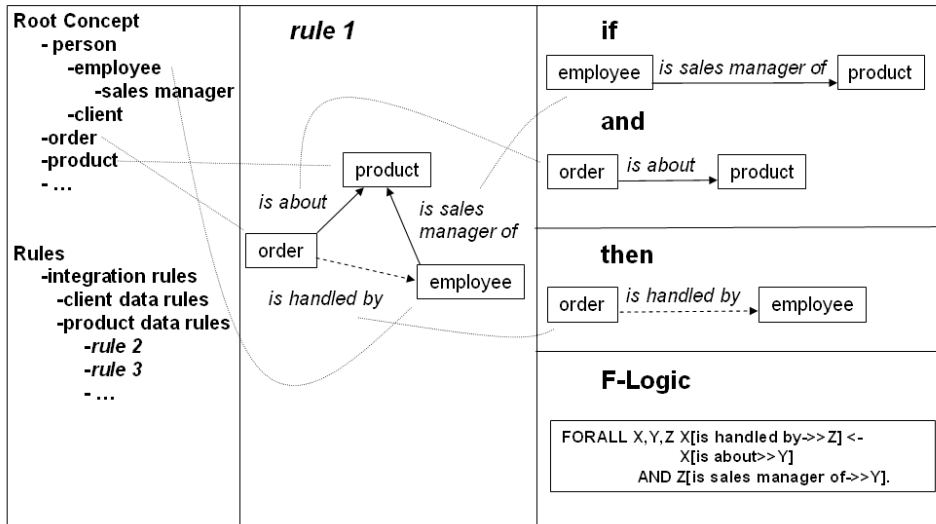
As we found in (2.) and (3.), the following components are needed within an ontology modelling environment [AS2002], for meeting the requirements for an integration solution:

- a core modelling component for concepts, attributes, relations, instances, multilingual representations and domain entries
- a schema import and export (2.3.1, 2.5) supporting various formats, particularly SQL
- a mapping tool (2.3.2-2.3.4)
- a rule editor (2.4.1)
- a rule debugger (2.4.2)

The first three points have been realized in recent modelling tools more or less. For the last two points we will introduce a rough idea of visualization, not covering all functionalities coming up.

3.1. Visual Rule Editor

In our example, *rule 1* (2.4.1) is a composition relation rule, connecting the three concepts *product*, *order* and *employee*. In our proposal (picture 11: a visual rule **editor**), a user would select them by drag&drop from a left window, where all concepts are listed in a "is-a"-hierarchy, and move them to the center window.

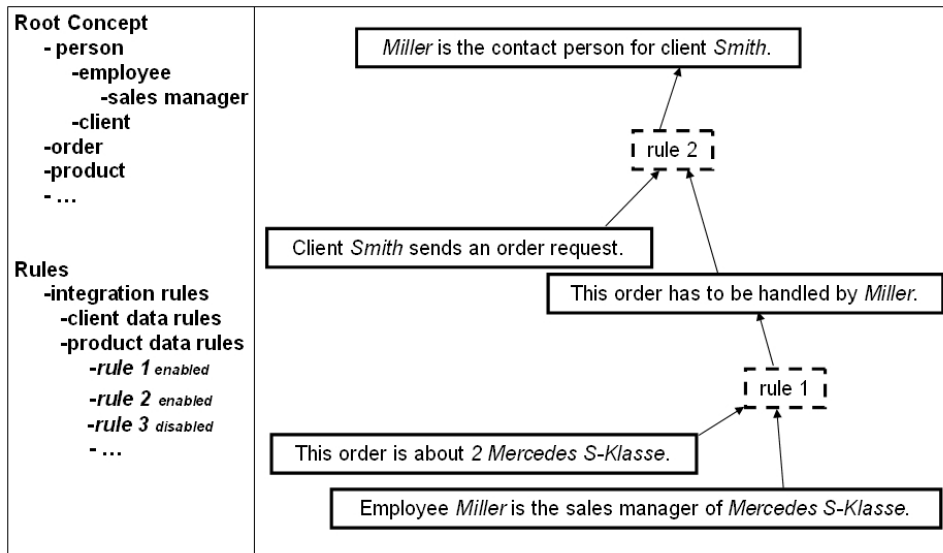


picture 11: a visual rule editor

There the modelled relations (*is about*, *is sales manager of*, *is handled by*) will appear. By moving them by drag&drop into the fields on the right side (*if*, *then*), you would create the rule shown in F-Logic code below. There you could change the rule also by hand. In picture 11 the ability to define attribute conditions (i.e. `employee.name="Miller"`) or operators (+, -, *, /, NOT, EXISTS, ...; i.e. `price=quantity*[price per unit]*discount`) is missing and has to be added to the draft yet.

3.2. Visual Rule Debugger

The visual rule debugger is an important tool for the IT professional. It's supposed to support him during the rule modelling phase, showing him the outcome of the rules. Thereby it visualizes the inference process for one selected new fact (picture 12: **a visual rule debugger**). If the concerning rules (*rule 1*, *rule 2*) for this new fact (i.e.: "*Miller is the contact person for client Smith*") have been enabled in the window on the right side, a graph would appear showing the course of conclusion.

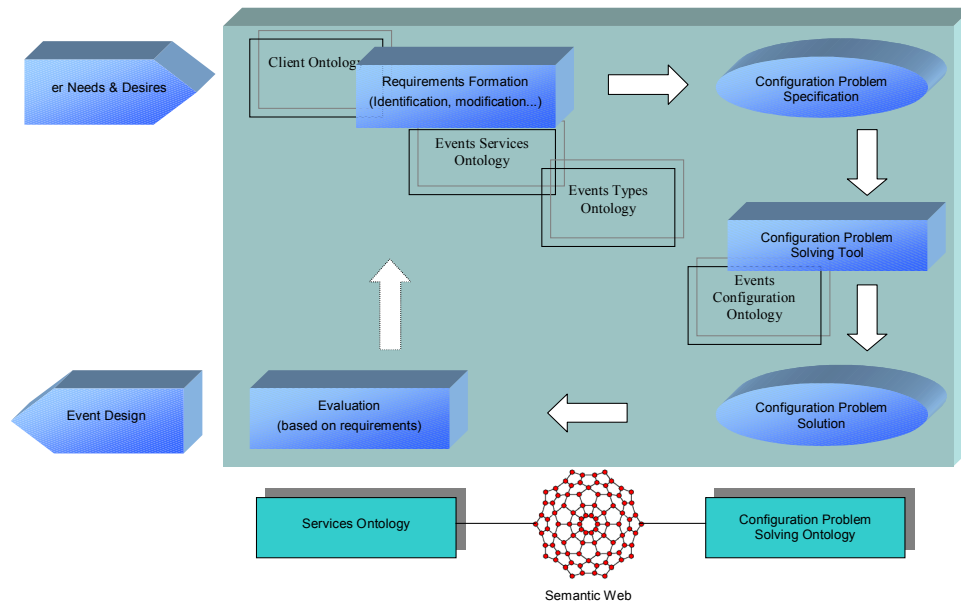


picture 12: a visual rule debugger

4. Case Study: Online Design of Events

In this section we introduce an event design support application in which complex knowledge integration is required. This application, developed within the **OBELIX project (IST-2001-33144)**, aims to support event organisers and suppliers in the process of designing events such as symposiums, conferences, exhibitions, workshops or meetings, and where the service provided, i.e., a running event, should be configured and customised to the needs of the type of event and client.

The aim of this application is to develop a Web-based system that will support the collaborative process of designing customised events, thus greatly improving this type of service.



Picture 13: **System Architecture**

(Picture 13: **System Architecture**) presents a view of the system architecture. It shows how the different ontologies needed are integrated so as to provide the desired support of event designers and clients. This system architecture is derived from a Knowledge Level theory of designing, known as KLDE (see [Sm96, Sm98, Sm02a, Sm02b]).

As a Knowledge Level theory, KLDE defines the necessary and sufficient kinds of knowledge needed and generated in any kind of designing, together with the roles each kind of knowledge plays, and the relationships between them. Applying this theory to the design of the event design support system described here, made it possible to quickly and easily identify the kinds of knowledge-based support needed, and the way the different kinds of knowledge needs to be integrated.

Starting with the needs and desires of a client, the event design support system first supports the identification of a set of requirements that, when satisfied, would result in an event design acceptable to the client. It further supports the development of well formed problem specifications that are essentially attempts to operationalise some or all of the requirements. In this event design application, the problems are formed as configuration problems, which are then passed to a configuration problem solver being developed by LABEIN.

The solutions that are returned are then evaluated with respect to the current requirements. And the outcome of this evaluation then leads either to the identification of further or modified requirements, a different operationalisation, or a final design.

To support this process, the event design support system thus integrates a "Client Ontology", to support the identification and construction of the requirements set, an "Event Services Ontology" and "Event Type Ontology" to support the formulation of well formed problems that operationalise effectively some or all of the current requirements, an

"Event Configuration Problem Solving Toolbox", which in turn uses the "Events Configuration Ontology."

Each of these ontologies model different kinds of knowledge. The "Events Types Ontology", contains the different types of events that can be configured by the application. These are defined in terms of resources used, such as networks, computers, rooms, layouts, etc. The "Client Ontology" contains information about type of clients and their needs, and the "Events Services Ontology" refers to considering the design of events as an event service. In the "Configuration Problem Solving Ontology" different types of configuration problems are reflected. In the "Service Ontology" the different kinds of services are reflected. The "Events Configuration Ontology" represents different types of event configurations problems, such as the configuration of layouts in a room assigned to an event or the different resources requested by the client.

These kinds of knowledge are integrated within the system. At the same time, (and again informed by KLDE), relationships can be identified between them. For example, the "Event Service Ontology" is related to both the "Events Types Ontology" and the "Service Ontology". In the same way, the "Events Configuration Ontology" is related to both the "Events Types Ontology" and the "Configuration Problem Solving Ontology".

Thus, in this event design support application we can see clearly the need for effective integration of different ontologies, and how what ontologies are needed and how they should be integrated has been effectively established by applying the KLDE theory of designing.

5. Closing Remarks

Looking at today's available software for ontologies, [NM02] compare different tools distinguishing editors from mappers. As shown above (2.), this must be no opposite¹⁷.

Another mapping mechanism between distributed ontologies¹⁸ is introduced in MAFRA [MMSV02], an interactive, incremental and dynamic framework for ontology mapping.

Of course, integration across different sources depends on the compatibility of their data structures. Just as you can't map Newton's 3-dimensional world by 100% to Einstein's 4-dimensional one, the result will always be an approximation. In the worst case, when ontologies are totally orthogonal to each other, a mapping can be impossible. But should a mapping be feasible, the manual approach described here (2.3) might be supplemented by semi- or full automatic approaches. [MDH02] deliver new findings on this subject.

6. References

[AL03] J. Angele, G. Lausen: Handbook on Ontologies in Information Systems, Springer (editors: S. Staab, R. Studer), 2003/2004

[An01] J. Angele: Semantik und Wissensmodelle, Proceedings Deutscher Internet Kongress, Karlsruhe, 2001

¹⁷ In OntoEdit, there's a modelling component as well as a mapping plug-in.

¹⁸ i.e. in the vision of the Semantic Web

- [An02] J. Angele: F-Logic Tutorial, http://www.ontoprise.de/documents/tutorial_flogic.pdf, Karlsruhe, 2002
- [An03] J. Angele: XML reicht nicht aus, Springer, 2003
- [AS92a] J. Angele, R. Studer: Konsistenzprüfung in wissensbasierten Systemen, Universität Karlsruhe, 1992
- [AS92b] J. Angele, R. Studer: Inferenzmechanismen in wissensbasierten Systemen, Universität Karlsruhe, 1992
- [AS2002] J. Angele, Y. Sure: Evaluation of Ontology based Tools, Siguenza, 2002
- [Ba02] F. Baader: The description logic handbook: theory, implementation and applications Cambridge University Press, 2002
- [BCB91] C. Batini, S. Ceri, C. Batini: Conceptual Database Design: An Entity-Relationship Approach, Addison Wesley Publishing Company, 1991
- [Br92] K. Brinker: Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden, E. Schmidt, Berlin
- [Bo01] R. Bourret: Mapping DTDs to Databases, 2001, at <http://www.xml.com/pub/a/2001/05/09/dtdtodbs.html>
- [DL91] M. Dahr, K. Lautenbach: Towards a formal theory of datalog nets, Koblenz, 1991
- [Gr93] T. R. Gruber: A translation approach to portable ontology specifications; Knowledge Acquisition; 1993;
- [Ho86] H. Hoffmann: On the visualization of design notions, of notion instantiations, and of structural relationships in a design data base realized as a semantic net, Darmstadt, 1986
- [KL90] M. Kifer, G. Lausen: F-Logic - A Higher-Order Language for Reasoning about Objects, Inheritance, and Scheme, Freiburg, 1990
- [KK01] K. Kline, D. Kline: SQL in a Nutshell, O'Reilly, 2001
- [Kr99] D. Kreuz: Formale Semantik von Konnektoren, Techn. Univ. Hamburg, Dissertation, 1999
- [Ma01] A. Maier, Vergleich von Wissensmodellen, December 2001
- [MDH02] A. Doan, J. Madhavan, P. Domingos, A. Halevy: Learning to Map between Ontologies on the Semantic Web, Honolulu, 2002
- [Me95] N. Meder, Konstruktion und Retrieval von Wissen: "Thesauri als Terminologische Lexika", Weilburg, 1995

- [MMSV02] A. Mädche, B. Motik, N. Silva, R. Volz: MAFRA — A Mapping FRAMework for Distributed Ontologies in the Semantic Web, FZI - University of Karlsruhe, 2002
- [MS00] A. Maier, H.-P. Schnurr: Wissensmanagement bei einer Investmentbank, 2000
- [NM02] N. Noy, M. Musen: Evaluating Ontology Mapping Tools - Requirements and Experience, Stanford University, 2002
- [Pe89] Christof Peltason: Wissensrepräsentation für Entwurfssysteme : d. Behandlung von Klassifikation und Taxonomie, Berlin, Techn. Univ., Diss., 1989.
- [PH02] J. Park, S. Hunting: XML Topic Maps, Addison-Wesley Professional, 2002
- [Sc00] W. Schwetz: Customer Relationship Management, Gabler, 2000
- [SM02] Y. Sure, E. Moench: Semantic Miner: Smarter Knowledge Retrieval. Poster Session at: First International Semantic Web Conference 2002 (ISWC 2002), June 9-12 2002, Sardinia, Italia.
- [Sm96] T. Smithers, 1986: On Knowledge level theories of design process, in J S Gero an F Sudweeks (Eds.) Artificial Intelligence in Design '96, Dordrecht: Kluwer Academic Publishers.
- [Sm98] T. Smithers, 1988: Towards a knowledge level theory of design process, in J S Gero an F Sudweeks (Eds.) Artificial Intelligence in Design '98, Dordrecht: Kluwer Academic Publishers.
- [Sm02a] T. Smithers, 2002: On Knowledge Level Theories and the Knowledge Management of Designing. International design conference - Design 2002, Dubrovnik, May 2002.
- [Sm02b] T Smithers 2002: Synthesis in Designing, in J S Gero (ed), Artificial Intelligence in Design '02, Kluwer Academic Publishers, Dordrecht, pp 3--26.
- [Sp96] J. M. Spivey: An introduction to logic programming through Prolog, Prentice Hall London, 1996
- [SSN01] R. Studer, H.-P. Schnurr, A. Nierlich: Semantisches Knowledge Retrieval, ontoprise Whitepaper Series, 2001
- [SSS02] S. Staab, R. Studer, Y. Sure: Knowledge Processes and Meta Processes in Ontology-based Knowledge Management. In: Handbook on Knowledge Management. C. W. Holsapple (ed.), Springer, 2002.
- [UB03] M. Ullrich, S. v.d.Bergh: Sales Knowledge Manager - Kombination von statistischen und semantischen Ansätzen zur Verkaufsberatung, Karlsruhe, 2003

Metadaten-Abhilfe

für das „Überinformationsproblem“?

Ursula Schneider

Institut für Internationales Management
Karl Franzens Universität Graz
Universitätsstraße 15
8010 Graz
ursula.schneider@uni-graz.at

Abstract: This paper will look at the endeavor to structure and classify knowledge by establishing a system of categorical pointers, from a critical perspective. For that purpose it will integrate three lines of critical arguments that have been discussed at the margin of mainstream efforts but have hardly been taken seriously. Had those arguments really been turned into knowing, research efforts would have been redirected from technical or more of the same to socio-technical paradigms. The three lines of argument refer to the doubtful nature of knowledge in principle which is aggravated by disciplinary specialization, they refer to a revival of reduced concepts of rationalization which ignores or suppresses results of psychological and neurological research and it will finally briefly discuss the possibility of animated machines constituting the next step of evolution as confronted with emotionally and socially competent humans as another path evolution could take.

As the purpose of the paper is to enlighten by de-construction it will not contribute to schemes or representations of knowledge. It will rather develop requirements such representations should fulfill. It will plead for caution and modesty and claim an analysis from the point of view of theories of power. As to the last affront it will suggest that some public funds are redirected to the project of emotional and social learning.

Kurzfassung: Die schiere Fülle an potenziellem Wissen, das durch die Öffnung von Grenzen, durch die Diffusionskapazität neuer Technologien und durch eine höhere Anzahl an lebenden Forscher/innen produziert wird, als in historischen Epochen zusammengenommen je gelebt haben, ruft geradezu danach, dieses Wissen zu ordnen und auf einer Metaebene so zu bezeichnen, dass es leicht wieder auffindbar wird. In einem solchen Ordnungsprozess müsste es zweifellos auch bereinigt werden, wofür die selbst referenziell funktionierende Wissenschaftsgemeinde nur theoretisch sorgt.

Deshalb ist dieses Paper kritisch angelegt, obwohl es die Plausibilität und Notwendigkeit des Metadaten-Generierungsvorhabens anerkennt. Es wird drei Argumentationslinien zusammenführen, die zwar an den Rändern der Mainstreamforschung immer wieder vorgelegt, aber nie so ernst genommen wurden, dass sie die Klassifizierungs- und Bereinigungsprojekte berührt hätten.

Als erstes wird es sich der Umkehrung des Ausspruchs von F. Bacon widmen und darauf verweisen, dass das herrschende Wissen, das nicht ohne Grund so heißt, epistemologisch immer nur als höchst zweifelhafter Stand des Irrtums aufzufassen ist, in der Geschichte aber immer wieder in den Dienst von Mächtigen und Profiteuren gestellt wurde. Dies gelang eben dadurch am besten, dass Wissen um die Zweifelhaftigkeit verboten, tabuisiert oder wie heute – in der Fülle von Belanglosigkeiten schlicht ausgeblendet wurde, was der Zerfall in Teildisziplinen mit systemisch notwendigen Scheuklappen sehr begünstigte. Eine zweite Argumentationslinie widmet sich der Wiederkehr technisch verengter Zweckrationalitätskonzepte, die alte Irrwege jenseits der Erkenntnisse psychologischer und neurologischer Forschungsergebnisse mit neuen Mitteln beschreiten.

Schließlich werde ich drittens noch auf die Vorstellung eingehen, die nächste Evolutionsstufe könne ein vom Menschen geschaffenes, ihn aber übertreffendes Wesen sein, nämlich die animierte Maschine. Man könnte sich auch andere Entwicklungspfade vorstellen, z. B. die intentionsgeleitete Weiterentwicklung der noch recht steinzeitlichen emotionalen und sozialen Fähigkeiten, was allerdings zu einer Umlenkung menschlicher Forschungsanstrengungen und von Forschungsmitteln führen müsste. Der Aufsatz schließt mit Konsequenzen, die aus seiner Dekonstruktionsarbeit gezogen werden können.

1 Vorspann: Wissen und Kritik

Die folgenden Überlegungen gründen sich auf ein Wissenschafts- und Kritikverständnis, das nicht auf die Akkumulation von Verfügungswissen abzielt. Verfügungswissen geht von Istzuständen aus, deren Zustandekommen nicht interessieren kann, da seine Erforschung Fortschritt zu weiteren Erkenntnissen nur aufhalten würde. Es muss daher notgedrungen immer affirmativ von der normativen Kraft bestehender Verhältnisse ausgehen. Meine Überlegungen sind allerdings auch nicht dialektisch angelegt als Verneinung dieser Verhältnisse. Verneinung bedarf nämlich eines Bezugspunktes, der nur eingenommen werden kann, wenn man sich entweder im Besitz der Wahrheit wähnt oder eine normative Position ohne Letztbegründung als überlegen durchzusetzen sucht. Damit vollzieht man explizit, was die Anhänger/innen einer rein technischen Funktionslogik, die übrigens oft als vernünftige Akzeptanz des unaufhaltsamen Rades der Geschichte auftritt, welche das Beste unter gegebenen Bedingungen herausholt, implizit tun: Sie postulieren eine normative Position, nämlich die den gegebenen Bedingungen inhärente, als unhintergebar.

Vielmehr verstehe ich Kritik mit Foucault als eine soziale Praxis, die sich den herrschenden Wahrheiten entziehen will und dies tut, indem sie so weit als möglich, Urteile außer Kraft setzt, um Elemente der Willkürlichkeit und Kontingenz aufzuzeigen, die sich in den herrschenden Wahrheiten und ihren Gegenpositionen verbergen (vgl. Foucault, 1990). An die Stelle einer transzendentalen Kritik im Sinne Kants, setze ich also eine historische Kritik gesellschaftlichen Seins in Form der – nie abschließbaren – Problematisierung. Dabei geht es nicht um Denunziation bestehender Verhältnisse, sondern um ihre Diagnostik, in der Erwartung, dass aus einer differenzierenden Diagnostik Möglichkeiten einer anderen Praxis erkennbar werden.

Skepsis, so meinte ein Reviewer, reiche als Beitrag zur Wissensschaffung nicht aus (da keine Begründung mitgeliefert wurde, kann nur implizit über eine Grundhaltung spekuliert werden, die sich auf Funktionslogik beschränkt).

Dem setze ich die Frage nach der Verantwortlichkeit für Aussagen und Gestaltungsvorschläge im engen Korsett einer isolierend-reduktionistischen Zugangsweise gegenüber, die nachweisbar immer zu gravierenden Neben-, Rück- und Fernwirkungen mit teilweise irreversiblen Charakter führt, welche dann von der nächsten Generation affirmativer Forscher/innen als unvermeidlicher Sachzwang aufgegriffen werden kann: Wie anders als skeptisch kann man Wissenschaft betreiben?

Im Dialog um das Thema Wissensontologien z. B. geht es – aus aufgeklärt erkenntnistheoretischer Perspektive – tatsächlich um eine Umkehrung von Bacon. Wer Metakategorien als Weltstandardsoftware anstrebt, muss sie festschreiben. Begriffe allerdings nehmen die Wirklichkeit in einen bestimmten Griff – der auch anders ausfallen könnte. Wie wird das ablaufen? Zufällig, traditions- oder machtgesteuert, wobei letzteres am wahrscheinlichsten ist. Wer die Macht hat, macht Wissen. Wer dies zu verschleiern trachtet, muss sich gefallen lassen, als Diener der Macht bezeichnet zu werden.

Nun noch zur Frage, ab wann ein Kollektiv weiß oder erkannt hat. Man sollte meinen, es genüge, ein Problem einmal zu erwähnen, danach mögen noch einige stilistische Variationen das Verständnis fördern – und das Kollektiv weiß, im Sinne eines Verständnisses der ersten Aufklärung, die stillschweigend und kontrafaktisch davon ausgeht, dass Menschen ihrer Erkenntnis gemäß handeln. Dass dies deutlich nicht der Fall und durch more-of-the-same-Strategien, nämlich mehr „Wissensaufbereitung“ auch nicht zu heilen ist, ist unter anderem Thema dieser Ausführungen.

Das Kollektiv weiß eben nicht, sondern schiebt die Erkenntnis als unbequem zur Seite, um an den more-of-the-same-Strategien Einkommens-, und Prestigeziele bzw. die Hoffnung auszuagieren, dass persönliche Entwicklung durch von der eigenen Person abgehobenes Denken möglich ist, ohne sich der ungewissen Wirkung nicht gut beherrschter Emotionen auszusetzen. Dass dies nicht gelingen kann und Denken immer affektbegleitet ist, ist ebenfalls eine Aussage dieses Papers mit Konsequenzen, die aus dieser Sicht durchaus als Lösungen gelten dürfen, nämlich z. B. einer Umleitung der Mittel vom Technischen in die Erforschung des Psychischen und Sozialen.

Dass diese Lösungen nicht auf persönliche Präferenz stoßen, sollte für einen offenen Geist eigentlich kein Anlass sein, sie unbelegt zu diffamieren und dabei einen zentralen Begriff der Veranstaltung so unscharf zu verwenden. Was weiß (!) man(n) längst? Man weiß, nach den gängigen Definitionen nicht, denn die Information wird nicht handlungswirksam.

Ganz abgesehen davon, dass ein streng angelegter Maßstab der Originalität, dem ich als unter einer Fülle an rein reproduzierten und recycelten Arbeiten leidende Zeitgenossin durchaus viel abgewinnen könnte, 95 % aller Publikationen vom Markt fegen würde. Wenn also z. B. über Reviewprozesse Markteintrittsbarrieren errichtet werden und eine Review wie im vorliegenden Fall emotional argumentiert im Sinne nicht offen gelegter Präferenzen, weder begründet noch belegt, sondern nur behauptet, dann würde gerade gegenüber Nachwuchswissenschaftlern von denen Innovationen ausgehen sollen, deutlich, wie sehr die Umkehrung von Bacon gilt. Wer Macht hat (und keine Verantwortung, was hoffentlich nicht auf eine Mehrzahl der Markteintrittsschrankenhüter zutrifft) macht Wissen. Solche Fragen sind im Allgemeinen natürlich tabuisiert. Hätte sich im vorliegenden Fall eine vermutlich unter Zeitdruck zusammengestoppelte Review nicht so als Paradebeispiel schlechter Community Praxis angeboten – und hätte diese schlechte Praxis nicht geradezu einen weiteren Beleg für die Ausführungen abgegeben, hätte man es auf sich beruhen lassen können. So kann nur gefolgert werden: quod erat demonstrandum.

2 Wissensontologien: Verweise auf bewegte Ziele

Aus dem Gesagten geht hervor, dass das Zeichen nicht mit seinem Sinn, seiner Deutung identisch ist und letztere auch nicht mit dem Bezeichneten übereinstimmt: Wie Wittgenstein und nach ihm Derrida betont haben, sind die Grenzen der Sprache, die Grenzen der Welt (Wittgenstein, 1994) bzw. schwingen in jeder Unterscheidung immer auch alle jene Begriffe mit, zu denen Unterscheidungen gebildet werden (Derrida, 2001). Dies stellt eine prinzipielle Schwierigkeit für ein Projekt dar, das in eindeutiger Weise ein möglichst auch noch global gültiges Schema von Verweisen entwickeln will. Obwohl es zunächst flexibler scheint als eine Dokumentation des Wissens selbst und z. B. auf unterschiedliche Lehrmeinungen im bezeichneten Behälter verweisen kann, verschiebt das Projekt die epistemologische Problematik nur auf die nächste Ebene: Es muss entschieden werden, welche Unterscheidungen von einem Verweis ein-, und welche ausgeschlossen werden, es müssen Begrifflichkeiten und Deutungen vor anderen ausgezeichnet werden, was im multikulturellen Kontext zumindest für alle, geistes- und sozialwissenschaftlichen Phänomene eine zusätzliche imperialistische oder kolonialistische Problematik aufwirft.

Diese Problematik mag für naturwissenschaftliche Phänomene geringer sein und/oder manchmal recht pragmatisch gelöst werden können, weil der Vereinheitlichungsnutzen größer ist als das Beharren auf kulturellen Besonderheiten. Nichts anderes geschieht in internationalen Normungsausschüssen, die sich z. B. um Messsysteme oder technische Protokolle kümmern. Dennoch sind mit dem „gewinnenden“ System, so es kein künstlich konstruiertes ist, ökonomische und qualifikatorische Gewinne verbunden, denen auf Seiten der sich nicht durchsetzenden Standards gesunkene Investitions- und Umstellungskosten gegenüberstehen. Es ist weder beliebig noch durch Sachzwang vorentschieden, worauf man sich einigt. Wie am internationalen Normierungsgeschehen nachvollziehbar, wird dies entweder offen verhandelt, wobei der Ausgang einerseits durch Wertvorstellungen über faires Verhandeln, andererseits durch Machtverhältnisse entschieden wird, oder es entscheidet sich am Markt durch selbstreferenzielle Popularität.

Auch hier wiederum gilt: Wer Macht hat, macht Wissen (und Ignoranz). Es sind immer die gesellschaftlichen Machtverhältnisse, die darüber entscheiden, was selbstverständlich unterstellt, was tabuisiert und was sanktioniert wird. Eine Kultur des Gleichheitszeichens, d. h. ein identitätsfixiertes Forschungskonzept, welches die Welt als Gleichungssysteme modelliert, hat nicht nur bereits Karl Popper stark irritiert, sondern muss auch regelmäßig scheitern, wenn es darum geht zu erklären, wie das Neue in die Welt kommt. Von Festschreibungen und Standardsetzungen ist wie von Paradigmen im Allgemeinen also einerseits eine produktive Kanalisierung des Denkens, andererseits aber auch ein hohes Maß an Konservativität und Innovationsfeindlichkeit zu erwarten.

Dabei ist in der bisherigen Argumentation noch unterstellt, dass Wissenschaft ein einheitliches System, mit etwaig einheitlichen ethischen Maßstäben darstelle. Mit der Spezialisierung jedoch zerrinnt, wie bereits Cusanus befürchtete, die Erkenntnis in Einzelwahrheiten, denen jeder sinnstiftende Zusammenhang fehlt. Oder, mit Jaspers gesprochen: *„Geht das mit dieser glaubenden Grundverfassung wirksame geistige Niveau verloren, dann wandelt sich die Wissenschaft selbst. Sie ist in Gefahr, in dem organisierten Betrieb des technischen Zeitalters sich zu verlieren, an die Endlosigkeit der Feststellungen, die nicht mehr gedanklich durchdrungen werden, an die Verfahren, der nichts anderes als ihren beschränkten Gegenstand sehenden Spezialisten“* (Jaspers, 1958).

Die Aufgabe der Repräsentation der Welt (Wissensontologien im Verständnis der Informatik) steht somit vor drei Problemen.

Erstens, jenem gültige Klassen, Funktionen und Relationen zu definieren, die eine Domäne so beschreiben, dass die Interpretation der Objekte eindeutig bestimmt wird.

Zweitens, diese Repräsentation etwaig domänenübergreifend zu schaffen, um die aus der Spezialisierung resultierende Problematik negativer Externalitäten zu mildern: Im interdisziplinären Kontext geht es nicht nur um unterschiedliche Terminologien, sondern um unterschiedliche Epistemologien oder nach Wittgenstein, Sprachspiele.

Drittens, die Ontologie so zu gestalten, dass sie in der Lage ist, neue Erkenntnisse aufzunehmen, was im Fall konstanter Paradigmen kein Problem sein dürfte, im Fall des Paradigmenwandels allerdings Schwierigkeiten aufwirft.

Wie an anderer Stelle dargelegt (vgl. Schneider, 2001), beruht Wissensmanagement auf einer Vision der Wiedergewinnung der Aufklärung nach den Verwirrungen der Postmoderne: Durch Methodik, konsequentes Denken und Forschen, vor allem aber durch eine Herauslösung aus dem praktischen Kontext sollen die Dinge in Form gebracht, ihr „Wesen“ herausfiltriert werden. Das Netz, als Metapher der Ortlosigkeit, bietet hierfür beste Voraussetzungen, Digitalisierung bringt den Logos auf die aus dem natürlichen Sein herausgelöste Zahl.

Die Vorstellung von einer hinter den Dingen liegenden Ordnung, von einem eigentlichen Wesen der Phänomene ist eine Setzung, was die sogenannte natürliche Welt betrifft. Für eine künstlich geschaffene Welt lässt sich eine entsprechende Setzung allerdings durchsetzen: hier liegt der tiefere Sinn der Rationalitätszumutung, die hinter einem Projekt von Wissensmanagement steckt, das Sprache nur bezüglich ihres Informationscharakters akzeptiert. Dieses reine epistemologische Projekt (als Repräsentationslehre fällt die Ontologie der künstlichen Intelligenz notwendig mit der Lehre von der Erkennbarkeit des Existierenden zusammen) vollzieht sich nun allerdings in einer praktischen Umgebung, in der Wissen auch als System sozialer Relationen zu interpretieren ist. Für diese praktische Welt soll das Repräsentationsprojekt Nutzen stiften, weshalb es sich auch auf sie einlassen muss.

2.1 A state of confusion

Wie im Folgenden deutlich wird, herrscht über das praktische Anliegen von Wissensmanagement keineswegs Klarheit.

“Knowledge Management (KM) is one now of the most ramified topics in the business arena. Much of this amplitude can be attributed to the number of fields that lay claim to the idea, or some part of it, including computer and information science, business strategy, macro economics and interpersonal dynamics to mention only a few. Proponents generally claim that this symphony of schemes is appropriate given the important, transversal and imminently practicable nature of Knowledge Management. Its critics, on the other hand, are hearing either a re-mix of older refrains, schizoid melodies or an outright cacophony. All parties agree that we are nonetheless witnessing an explosion of interest in the term Knowledge Management and all that it may or may not imply.” (Despres, et. al., 2001, 20)

Im Rahmen verschiedener organisationstheoretischer Zugänge scheint ein Großteil der Protagonisten den Fokus auf (meist nicht genauer definiertes) implizites Wissen zu legen, während einige wenige Stimmen sich erheben und meinen, diese Art von Wissen sei ökonomisch vernachlässigbar (z. B. Hakanson, 2000) bzw. gar nicht als Wissen bezeichnbar (Schreyögg und Geiger, 2002).

Offenbar haben wir es bei der verteilten Produktion von Wissen mit massiven inter- und intra-disziplinären Problemen zu tun, die aus drei Quellen gespeist werden. Zum einen bestimmen bewusst selektierte Blickwinkel und Vorkenntnisse das, was erkannt werden kann. *Die Weltformel, die Metatheorie oder das Modell*, in die alle Aussagen unterschiedlicher Disziplinen integriert werden könnten, sind nach gegenwärtigem Stand des Irrtums aus prinzipiellen Gründen nicht ermittelbar. Zum anderen bringt die Feindifferenzierung im wissenschaftlichen System mit ihrer hohen Spezialisierung auf isolierte Teilbereiche Fülle- und Schnittstellenprobleme mit sich, welche die Übertragung, Übersetzung und letztlich gemeinsame Formung von Erkenntnis zwischen Mitgliedern verschiedener Teilsysteme erschwert. Vom Standpunkt der Einzelwissenschaften betrachtet, sorgen konkurrierende Theorien und eine ausgeprägte Feindifferenzierung dafür, dass die Menge an Publikationen unüberschaubar und kaum noch aufeinander bezogen ist, was einen – vom Prinzip her Innovation versprechenden – interdisziplinären Dialog erschwert.

Der gut gemeinte Ruf nach Interdisziplinarität, weil Wissensmanagement als Anliegen philosophische, psychologische, neurologische, soziologische, ökonomische und technische Fragen aufwirft, scheitert häufig daran, dass Interdisziplinarität noch nicht stattfindet, wenn, die Monologe artig hintereinander her aufgesagt werden. Interdisziplinarität ist ein Prozess der den Erkenntnisfortschritt zunächst verlangsamt, da eine gemeinsame Syntax entwickelt und eine gemeinsame Semantik verhandelt werden muss, ehe im pragmatischen Raum gemeinsam exploriert werden kann. Drittens, sind mit Wissensprojekten aktuell massive kommerzielle Interessen verbunden. Die Logik der Erkenntnis wird mit einer Marktlogik vermischt, welche sich in Geldflüssen ausdrückt. Der genannte Schwerpunkt auf implizitem Wissen könnte in diesem Sinn auch als Marktstrategie potenzieller Berater interpretiert werden: Das schwer fassbare, organisationstypische, in Routinen eingebettete Wissen erschwert die Schaffung von Standardlösungen im Umgang mit diesem Wissen und erfordert daher maßgeschneiderte Beratung im Einzelfall.

2.2 Metadaten als Abhilfe?

Für den Kontext lebensweltlichen Handelns macht es keinen Unterschied, ob Funktionen, Klassen und Relationen tatsächlich „existieren“ oder ob sie nur eine konsensuell validierte Erkenntnisweise darstellen, die es Menschen erlaubt, ihre Handlungen aufeinander zu beziehen. Es geht also nicht darum, was wahr ist, sondern was funktioniert.

Allerdings ist damit noch keine Klarheit gewonnen: Funktionieren kann nur auf einen Kontext bezogen definiert werden, wobei Kontext immer auch durch Interessen bestimmt wird: Wenn Geld, Dokumentationsaufwand und Forschungsanstrengungen sich auf nukleare Energie konzentrieren, kann weniger über Wind- und Solarenergie gewusst werden.

Wenn ein liberales volkswirtschaftliches System für sich selbst verstärkenden Reichtum ebenso sorgt wie für sich selbst verstärkende Armut, sind seine Funktionen, Klassen, Relationen nicht neutral, sondern interessenbezogen. Jedes Metadaten-Projekt hat demzufolge offenzulegen, auf welchen Unterstellungen und Vorentscheidungen es beruht. Da es aber von Kontexten losgelöst und daher scheinbar neutral auftritt, erschwert es Kritik.

Kann dies ein Argument gegen Wissensontologien sein? Ist es unvorstellbar, eine Wissensbasis für ein bestimmtes Wissensgebiet weltweit einheitlich zu gestalten und zu nutzen? Empirische Evidenz zeigt uns, dass dies nicht der Fall ist. Das medizinische Wissen¹ ist weltweit schon recht zufriedenstellend harmonisiert, weshalb es realisierbar und ökonomisch vorteilhaft ist, Auswertungen von Schichtbildern einer Magnetresonanz – oder Computertomographie nach Indien zu senden.

Allerdings ließe sich auch hier einwenden, dass medizinisches Wissen selbst innerhalb eines reduktionistisch-isolierenden Paradigmas nicht eindeutig ist, ganz zu schweigen von den unterschiedlichen Zugängen zwischen letzteren und ganzheitlichen Paradigmen. Im Bereich der täglichen medialen Berichterstattung wird die Festschreibungsproblematik noch deutlicher: Wenn etwa Bildspeicher- und Bilderkennungsverfahren anhand von CNN-Material entwickelt werden ist zwar prinzipiell nicht ausgeschlossen, dass Metakategorien wie Ort, Zeit, sowie inhaltliche Deutungen von Ereignissen (wie z. B. Terroristische (!) Angriffe) auch auf andere Quellen verweisen, doch ist dies sozial und ökonomisch unwahrscheinlich: Da digitale Aufbereitung nicht kostenfrei ist, werden wir es mit Dominanzen zu tun bekommen, die allerdings als neutrale Inhalte erscheinen.

Dies ist kein Argument gegen Metadaten. Letztere sind ein notwendiger und sinnvoller Weg, sich der aus einer Vielzahl von Elementen und Verbindungen zwischen ihnen resultierenden Komplexität zu stellen. Da ein solches Projekt höherer Ordnung im politisch sozialen Raum stattfindet und pfadabhängig ist, wäre jedoch ein Meta-Metadatenprojekt quasi als Supervision nötig, um etwaig unerwünschten Entwicklungen gegenzusteuern.

Bei dieser Argumentation wurde unterstellt, dass Metadatenprojekte geplant verlaufen in internationalen Projektteams, in denen zunächst vor allem Wissenschaftler tätig werden. Möglicherweise werden mehrere Projekte in einem Coopetition-Verhältnis tätig, die dann als offene oder proprietäre Standards zur Verfügung stehen. Aus Kompatibilitätsgründen mag ein Standard sich zuletzt durchsetzen, was erkenntnistheoretisch eine Verarmung und politisch problematisch wäre.

¹ Das medizinische Wissen? Anhänger der Homöopathie und chinesischen Gesundheitsleere würden wohl widersprechen.

3 Das paradoxe Dreieck der Rationalisierung

Eine erste Zumutung besteht in der Notwendigkeit, bislang relativ unverbundene Paralleluniversen zu verbinden, was für unterschiedliche wissenschaftliche Disziplinen ebenso gilt wie für unterschiedliche praktische Felder und unterschiedliche Kulturen. Dies impliziert psychologisch Verunsicherung und lerntheoretisch Doppelschleifigkeit: Man wird sich der Skurrilitäten der eigenen Weltbilder bewusst. Eine Verständigung ist umso schwieriger, je größer die Unterschiede in den Begriffs- (syntaktische Ebene) Denk- (semantische Ebene) und Handlungsmodellen (pragmatische Ebene) ausgeprägt sind (vgl. Schneider, 1990, Carlile, 2002, Rooney/Schneider, 2001) (vgl. Abb.1).

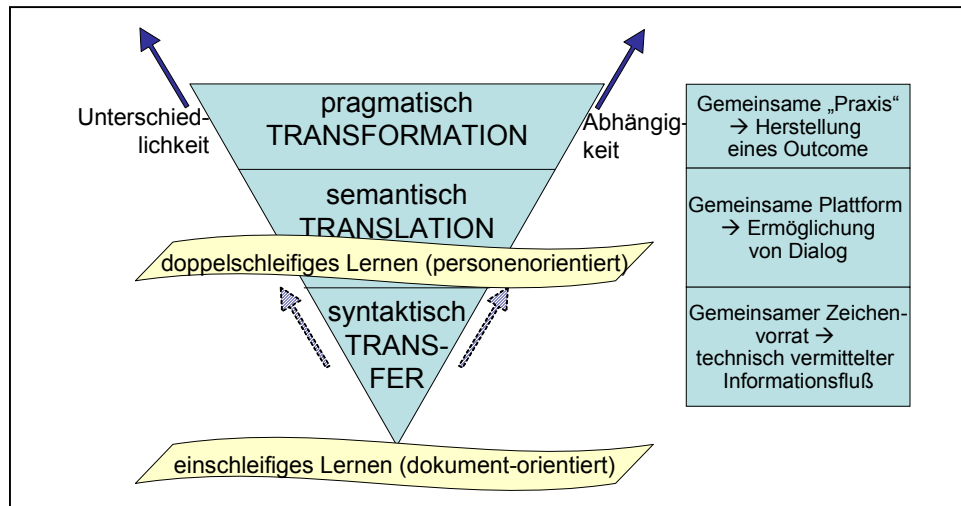


Abbildung 1: Wissensorientierte Gestaltung von Grenzen (in Erweiterung von Carlile, 1997 und 2002), vgl. auch Schneider, 1990; 152

Mehrdeutigkeit, Kontextgebundenheit und hohe implizite Anteile erweisen sich als Hindernis des persönlichen Transfers, besonders aber der über Dokumentation vermittelten Diffusion von Wissen. Seit dem Positivismusstreit ist wohl nur noch über die Nützlichkeit von Hypothesen zu befinden, nicht mehr über ihren Wahrheitsgehalt – und erstere ist deutlich zeitgebunden.

Hier ist die besondere Stellung der Sozialwissenschaften hervorzuheben, die – unter ausgeprägten innerdisziplinären Konflikten – um eine Positionierung zwischen den a-historisch, reduzierend, analytisch vorgehenden Naturwissenschaften und den historisch, ganzheitlich, interpretierend ausgerichteten Geisteswissenschaften ringen. Weil sie es mit menschlichem Verhalten zu tun haben, das sich in die Zukunft entfaltet, gilt notwendig die Ungewissheit doppelter Kontingenz (vgl. Luhmann, 1996).

Damit wird viel für Management relevantes Wissen (wie jenes über Partner, Konkurrenten, Kunden, Lieferanten, Gegebenheiten der Politik in Zielmärkten) zwangsläufig mehrdeutig und ungewiss bzw. kurzlebig, was hohe Aufwendungen für strukturierte Dokumentation und Bearbeitung ökonomisch unrentabel gestaltet.

Indizierungen, Klassifikationen, Metadaten sind eine gemeinsame Syntax, die ohne gemeinsame Semantik und Pragmatik nicht zustande kommen kann, deshalb ist die Schaffung von Wissensontologien kein triviales Problem. Je größer die Interdependenz umso drängender wird das Problem. Wissensmanagement, das die Grundsätze der Industrialisierung für Produktivitätsgewinn nutzen und in „Blockbuster“ – oder Killer-Applications umsetzen möchte, muss an Normen, Standards, Wiederholbarkeit und einem weltweit einheitlichen Stand von (noch) nicht falsifizierten Hypothesen interessiert sein (vgl. Popper, 1973).

Einer kritisch-rationalen Position eines durch wissenschaftliche Konkurrenz und Opposition dialektisch geschärften Wissens, welche den Fortschritt von einer ideosynkratischen und an Ausführungspraxis gebundenen Handwerkskunst zum automatisierbaren, bis in den kleinsten Vollzug explizierten begründeten industriellen Produktionswissen (vgl. Hakanson, 2000) in „Verwissenschaftlichung“ ortet, stehen die Beobachtungen von Kuhn (vgl. 1962) und vor allem der Konstruktivisten theoretisch gegenüber. Praktisch finden sich Gegenpositionen im Bemühen um Innovation im Schumpeterschen Sinn (vgl. 1942, Aussagen zur kreativen Destruktivität des Unternehmers).

Neben erkenntnistheoretischen und ökonomischen Fragestellung werfen Indizierungs- und Klassifikationsprojekte immer auch politische Fragen auf, die machttheoretisch zu untersuchen sind (vgl. Foucault, 1992 und 1984). Wenn das Mainstream Wissen, das im Deutschen nicht ganz zu Unrecht als das herrschende bezeichnet wird, weltweit standardisiert wird, stellen sich Macht- und Innovationsfragen. So stehen etwa, wie geschildert, bei der Bildauswertung aktuell amerikanische Forschungen an der Spitze, die Millionen von CNN Bildern so indizieren, dass unter unterschiedlichsten Suchkategorien darauf zugegriffen werden kann (vgl. Wactlar, 2002). Wenn sich ein solches System zum Standard entwickelt, werden künftige Generationen von Historikern in ihrer Forschung nur noch auf die CNN Sicht zugreifen können, was nicht ganz unproblematisch erscheint, jedenfalls problematischer als die standardisierte Auswertung von 125 Schichtbildern eines Artrrosegeschädigten Knies?

Unter der Annahme, dass Wissen immer voraussetzungsgeladen ist, ergibt sich die Erkenntnis, dass transdisziplinäre Ontologien nur für einfach bzw. hoch strukturierte Wissensgebiete mit einer normierenden Einigung auf eine gemeinsame Syntax das Auslangen finden. Für komplexer strukturierte Wissensgebiete ist die Involvierung in eine gemeinsame Praxis unverzichtbar.

3.1 Informationsfülle, Datenschrott

Als zweites Problem werde ich das mit dem ersten Problem verbundene Phänomen der Informationsdichte und –fülle behandeln, das u. a. durch die hochspezialisierte Ausdifferenzierung des wissenschaftenden Systems (zur Systemdifferenzierung vgl. Luhmann, 1996) zustande kommt. Eine Jahresproduktion von 750 Terabyte im Netz (vgl. Ebert, 2002) und weiteren 500 Terabyte in den Medien liegt schlicht jenseits der Bewältigungsgrenze durch eine menschliche „CPU“.

Waren 1932 noch 202 Dokumente für einen Typ der Marke Opel ausreichend, wuchs deren Menge bis 1999 auf fast 14.000 Dokumente an. Wie Theo Wehner launig bemerkt könnte ein Airbus die Menge an Dokumenten nicht transportieren, die bei seiner Produktion anfallen (vgl. Dick/Wehner, 2002). Wo früher ein Abtasten oder 2 Röntgenbilder genühten, fallen heute für eine Abdomenuntersuchung 700 Schichtbilder an: Wenn der diagnostizierende Arzt 10 Patienten à 700 Bilder an einem Vormittag zu therapieren hat, was eine realistische Annahme darstellt, ergibt sich ein ebenfalls nicht triviales Fülleproblem (vgl. Schneider, 2002).

Diesen Überlegungen könnte man entgegenhalten, dass auch die damals bahnbrechende Erfindung des Telegraphen schon Bedenken auslöste, die durch Telegramme bedrängten Geschäftsleute kämen nicht mehr zur Ruhe, worunter ihr Familienleben leide. (vgl. Cairncross, 2001; 46). Inzwischen ist der Telegraph durch das Telefon und in der Folge durch elektronische Medien ersetzt und die Möglichkeit von Echtzeitinformation ganz gut bewältigt worden. Der Kern der damals geübten Kritik bleibt allerdings gültig und wird durch den im Vergleich ungleich höheren Vorrat an Daten, bzw. Informationen verschärft.

Menschen können lernen, mit der Fülle umzugehen: Höhere Konzentration, stärkere Fokussierung durch klare Willensbildung und technische Denkhilfen werden für eine Teilbewältigung des Problems sorgen. Diese werden den Menschen aber nicht in gleicher Art und Weise zur Verfügung stehen, wodurch sich wieder ein Machtproblem ergibt. Ferner ist es trotz der Hilfen problematisch, in der Datendichte informiert zu bleiben und den Überblick zu behalten.

Wie psychologische und neurologische Untersuchungen zeigen, sind Menschen unter Informationsstress schlechte Entscheider: Technische Denkhilfen beziehen Affekte nicht mit ein – für menschliches Denken gilt allerdings, dass es immer affektbegleitet ist (vgl. Dörner, 1998, Damasio, 1997) und dass Menschen ohne eine Verbindung zwischen Großhirnrinde und Mandelkern über keine Heuristiken und Anhaltspunkte verfügen, daher nur noch rechnen würden, ohne je zu handeln.

Solange Menschen entscheiden, was bei jeder Handlung stattfindet, werden sie bewussten und unbewussten Einschränkungen unterliegen. Dabei ist plausibel zu vermuten, dass sozial verträgliche Beziehungen Intransparenz voraussetzen. Letztere wird durch Metadatenprojekte auch nicht gefährdet, die Grenze wird nur ein wenig in Richtung höherer Überprüfbarkeit menschlicher Entscheidungshandeln verschoben, was im Sozialen entsprechende Neuausrichtungen erfordert. Insgesamt kann festgehalten werden: Metadatenstrukturen im Sinne von Ordnungsschemata entschärfen die Retrievalproblematik, nicht jedoch die begrenzte Verarbeitungskapazität des Menschen.

Als drittes Problem möchte ich die soziale, existenzphilosophische und psychologische Spannung ansprechen, die aus der Rationalitätszumutung künstlicher Intelligenz erwächst: Wenn Maschinen über sehr große Mengen geordneter (indizierter) Information verlässlich verfügen und auch vernetzte, hoch komplexe Inferenz Mechanismen auf sie anwenden können, müssten sie streng genommen, besser entscheiden als Menschen (vgl. Kurzweil, 2001)

Damit ist ein Identitäts- und Autonomieproblem berührt, es stellt sich die Frage, ob Menschen sich nach dem Ergebnis der Maschine richten oder sie gelegentlich aus Autonomiegründen sabotieren werden. Ferner sind alle jene, die derzeit von Unschärfen und Mehrdeutigkeiten profitieren (als Deuter oder als „Minder“- oder Leerlaufleister) zutiefst bedroht und mit ihnen das soziale Gleichgewicht. In der Regel werden solche NRF (Neben-, Rück- und Fernwirkungen) bei technischem „Fortschritt“ nicht mitberücksichtigt. Man sieht zunächst nur, wie langsam und marginal die Denkprothesenentwicklung voranschreitet und kann sich daher nicht vorstellen, wie eine soziale Welt aussehen könnte, nachdem eine Mehrheit regelmäßig Denkprothesen verwendet hat und dadurch zwar an Bedienungs- und Schnittstellenwissen gewonnen, selbständige Denkfähigkeit aber zurückgebildet hat.

Zusammengefasst ergibt sich ein Dreiecksproblem der Rationalitätszumutung (vgl. Abb. 2), das mit der Relativität wissenschaftliche Erkenntnis und des Zweckrationalitätsbegriffs zu tun hat (Mehrdeutigkeits-, Spezialisierungs- und Interdisziplinaritätsproblem). Jegliche Festschreibung in Ontologien wäre daher dahingehend zu analysieren, welche Interessen damit festgeschrieben werden und welche Neben-, Rück- und Fernwirkungen ein Weiterarbeiten mit verengten Festschreibungen produzieren wird.

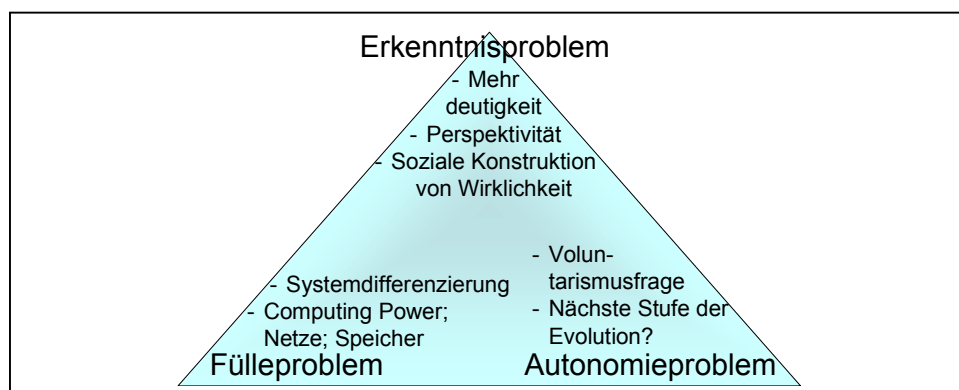


Abbildung 2: Dreiecksparadoxon der Rationalitätszumutung

Zum anderen sorgt die Fülle an Daten und Orientierungsangeboten sowohl für paradoxe Willensbildung als auch für eingeschränkte und verzerrte Informationsverarbeitung. Während H. Simon versuchte, auf Basis dieser Erkenntnis lediglich die Annahme von Zweckrationalität und individuell unabhängigem Denken, das Handeln vorausgeht, im Begriff der bounded rationality zu lockern (Simon, 1945), zeigen psychologische Forschungen, dass es sich nicht nur um eine Einschränkung von Rationalität handelt, sondern dass unter Informationsstress kontraproduktive Reaktionen erfolgen.

Ferner dürfte die kartesianische Annahme der Trennung von Körper und Geist, die bei Simon unangefochten bleibt, im Lichte neuerer Erkenntnisse neurologischer Forschung (Damasio, 1997) nicht aufrecht zu erhalten sein.

Drittens, geht es um das Auseinanderklaffen von System- und individueller Zweckrationalität: Individuen haben Identitäts-, Autonomie- und soziale Reziprozitätserfordernisse auszubalancieren, weshalb sie u. U. kein Interesse an verbesserten Informationsgrundlagen zur Steigerung der Zweckrationalität ihrer Entscheidungen haben.

4 So what? – Welche Konsequenzen haben die angestellten Überlegungen

Wie anders wo dargelegt, kann es nicht darum gehen, technische „Fortschritt“ aufhalten zu wollen. Es geht vielmehr darum das Konzept eines sozio-technischen Systemansatzes ernst zu nehmen (vgl. Schneider, 2001; 114 – 125). Das würde allerdings bedeuten, dass die Politik, die Steuergelder verteilt, mit denen alle technologischen Weichenstellungen zunächst finanziert werden, auch in stark deregulierten Staaten wie den USA, einen Teil der Gelder der Erforschung des sozialen Kontextes widmet und nicht, wie bisher, alles in scheinbar wohldefinierte rein technische Vorhaben pumpt, die später der sozialen Nachsorge bedürfen, wobei die Betreiber der technischen Projekte sich für die aus Komplexität resultierenden Neben-, Rück- und Folgewirkungen regelmäßig für unzuständig erklären bzw. ohnehin längst nicht mehr greifbar sind.

Konkret würde dies bedeuten, stärkere Zurückhaltung beim Dokumentieren und Festschreiben von Inhalten zu üben, mehr Geld für Sichtung, Prüfung, Zusammenfassung, Strukturierung, Verdichtung und Vernichtung von Inhalten bereitzustellen.

Was derzeit zu geschehen scheint, ist ein hype-gesteuertes, das Bestehende jeweils affirmativ als unverrückbar zur Kenntnis nehmendes Vorwärtsstürmen, das dem Konzept der Meme nach Dawkins zu entsprechen scheint (vgl. 1989) Zweck ist die Entfaltung einer technischen Evolution, für die Menschen nur „Wirte“ sind, welche die mit ihr verbundenen Ideen ausbrüten und weitergeben, als Einzelschicksale aber keine Bedeutung haben. Dass dabei einige wieder mal gleicher sind als andere und sich überproportional am Steuertopf bedienen, erklärt vermutlich, warum sie sich den Projekten zur Verfügung stellen. Die Sintflut findet erst nach ihnen statt.

Literaturverzeichnis

- [Ca01] Cairncross, F.: The Death of Distance 2.0. How the Communications Revolution will change our Lives. Texere, London, 2001.
- [Ca02] Carlile, P.: Transferring, Translating and transforming: An Integrative Framework for Managing Knowledge Across Boundaries. Working Paper, MIT, Sloan School of Management, Boston, 2002.
- [Da97] Damasio, Antonio, R.: Descartes Irrtum. Fühlen, Denken und das menschliche Gehirn. dtv, München, 1997 (5. Auflage, 2000)
- [Da89] Dawkins, Richard: The selfish gene. Second Edition. Oxford, Oxford University Press, 1989
- [De01] Derrida, Jacques: Die unbedingte Universität. Suhrkamp, Frankfurt, 2001
- [DC01] Despres, Charles/Chanvel, Daniele (2001): The Thinking Behind the Action in Knowledge Management. In: Proceedings of the Second European conference on Knowledge Management. Bled (Slovenia), 8 – 9 November 2001 ; pp 133 – 152.
- [DW02] Dick, M./Wehner, Th.: Wissensmanagement zur Einführung; Bedeutung, Definition, Konzepte. In: Lüthy, W./Voit, E./Wehner, T. (Hrsg.): Wissensmanagement – Praxis, vdf Hochschulverlag AG an der ETH Zürich, 2002.
- [Dö98] Dörner, Dietrich: Die Logik des Misslingens. Strategisches Denken in komplexen Situationen. Rowohlt Verlag, Reinbek bei Hamburg, 1998.
- [Eb02] Ebert, Andreas: Podiumsdiskussion – Informationsgesellschaft, Alpbacher Technologiegespräche, August 2002.
- [Fo84] Foucault, M.: Du pouvoir. Gespräch mit P. Boncenne von 1978. In: L'Express, N° 1722, 6, 12.7.1984; S. 56 – 68.
- [Fo90] Foucault, M.: Was ist Aufklärung? In: Eva Erdmann et al. (Hrsg.): Ethos der Moderne, Suhrkamp, Frankfurt, 1990.
- [Fo92] Foucault, M.: Was ist Kritik? Merve, Berlin, 1992.
- [Ha00] Hakanson, Lars: From Tacit Knowledge to Scientific Theory: The Power and Logic of Articulation. Working Paper, Johannes Kepler Universität Linz, Center for Research in International Management, 2000.
- [Ja58] Jaspers, Karl: Die Atombombe und die Zukunft des Menschen. Politisches Bewusstsein in unserer Zeit. Piper, München, 1958.
- [Ku62] Kuhn, Th. S.: The Structure of Scientific Revolutions, Chicago, 1962.
- [Ku01] Kurzweil, R.: Homo sapiens. Leben im 21. Jahrhundert. Was bleibt vom Menschen? 3. Auflage, Ullstein, München, 2001.
- [Lu96] Luhmann, Niklas: Soziale Systeme – Grundriss einer allgemeinen Theorie, 6. Auflage, Suhrkamp Verlag, Frankfurt, 1996.
- [Po73] Popper, Karl: Logik der Forschung. 5. Auflage, J.C.B. Mohr (Paul Siebeck), Tübingen, 1973; S 76.
- [RS01] Rooney, David/Schneider, Ursula: Rethinking Tacit Knowledge. In: Proceedings of the Second European conference on Knowledge Management, Bled (Slovenia), 8 – 9 November 2001; pp 583 – 594.
- [Sc90] Schneider, Ursula: Kulturbewusstes Informationsmanagement. Oldenburg Verlag, München, 1990.
- [Sc01] Schneider, Ursula: Die 7 Todsünden im Wissensmanagement. FAZ Verlag, Frankfurt, 2001.
- [Sc02] Schneider, Ursula: The Knowledge-Attention-Gap: Do we Underestimate the Problem of Information Overload in Knowledge Management? In: Tochtermann, Klaus/Maurer, Hermann (Eds.): J.UCS – Journal of Universal Computer Science, Vol. 8, Issue 5, 2002; S. 55 – 63.
- [Sc42] Schumpeter, J.: Capitalism, Socialism and Democracy; New York, 1942.
- [Si45] Simon, H. A.: Administrative behavior, New York, 1945.

- [Wa02] Wactlar, H. D.: Extracting and Visualizing Knowledge from Film and Video Archives. In: Tochtermann, Klaus/Maurer, Hermann (Eds.): J.UCS – Journal of Universal Computer Science, Vol. 8, Issue 6, 2002; 64ff.
- [Wi94] Wittgenstein, Ludwig (posthum): Philosophische Betrachtungen. Philosophische Bemerkungen. In: Nedo, M. (Hrsg.): Ludwig Wittgenstein. Wiener Ausgabe, Band 2, Springer, Wien, New York, 1994.