# MANAGING THE EVOLUTION OF CORPORATE PORTALS

## – A USER-CENTRIC APPROACH –

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Wi.-Ing. Helmuth Elsner

Karlsruhe, 2011

Referent: Prof. Dr. C. Weinhardt
Korreferent: Prof. Dr. G. Satzger
Tag der mündlichen Prüfung: 1. Februar 2011

# Acknowledgement

I am indebted to many people for their support and collaboration in the course of my Ph.D. research. Foremost, I thank my supervisor, Prof. Dr. Christof Weinhardt for his strong and continuous support. I highly value that I had the opportunity to be part of his team for almost ten years, starting as student research assistant during my undergraduate studies. I would also like to thank Prof. Dr. Gerhard Satzger for co-advising this thesis. Prof. Dr. Oberweis and Prof. Dr. Waldmann have been so kind to serve on the board of examiners. I would like to specially thank Dr. Jan Krámer for his guidance and the valuable contributions he gave to this thesis.

Special thanks go to Bayer AG, Corporate Finance for the funding of this work and the possibility to carry out the case study.

I would like to express my gratitude to my colleagues from the Information & Market Engineering group for providing a stimulating and fun environment. Special thanks go to Arun Anandasivam, Carsten Block, Chris Flath, Florian Teschner and Wibke Michalk for proofreading and challenging discussions. Likewise, I want to thank my Brother, Hans Ulrich Elsner and my friends Tatiana Basic and Stephan Eyl.

This work is dedicated to my family, who supported and encouraged me throughout the eventful time and my friends who cheered me up when needed.

Finally, I want to thank Filipina and the rest of the team for providing the perfect environment for the continuous engineering of this work.

# Contents

# List of Figures

# List of Tables

# Motivation & Introduction

Many of the most successful Internet-based enterprises are based on the concept of web portals. Facebook[1] and MySpace[2], Yahoo[3] and MSN[4], Expedia[5] and many other portals offer a variety of services through one single gateway; thus, they allow users to execute different actions without having to switch platforms. This concept has been successfully transfered to corporate intranets; the potential has been described for the first time by Shilakes and Tylman (1998) and Murray (1999). They outline the concept of a corporate portal that supports its users in their work with all kinds of services, ranging from document management tools to integrated applications. Many authors regard the corporate portal as provider for such a broad range of services; still, most scientific discussions and case studies focus on only one subtype: knowledge management portals, which mainly offer services to search for documents, contact data or to share information. The capability of a corporate portal to serve as a main gateway not only to the company's knowledge but also to its applications is seldom described in detail, although often stated as a core feature (Murray, 1999; Dias, 2001; White, 2003). Accordingly, corporate portal definitions and classification models are narrowed on details of knowledge management but insufficiently cover other types of portals. In this work a broader approach to defining and classifying corporate portals is introduced.

Case studies of corporate portals usually describe the development process and the results from the initial start-up phase (Raol et al., 2003; Remus, 2007; Vo, 2007).

[1] http://www.facebook.com
[2] http://www.myspace.com
[3] http://www.yahoo.com
[4] http://www.msn.com
[5] http://www.expedia.com

However, long-term studies of corporate portals beyond knowledge management portals do not exist. Filling this gap is even more required because corporate portals are subject to a constant evolution (Eckerson, 1999a; Dias, 2001). Advances in web technology allow the development of complex services (Daniel 2005). Besides these technological advancements, a portal is subject to a constant evolution process due to growth, changing business requirements and changes in a companies IT infrastructure. Vo (2007) described the related requirements with regard to the software engineering aspects. The user-perspective of the evolution process is, however, not studied, especially for portals that serve to integrate applications rather than providing simple access to knowledge. The process of constant change demands attention from users as well as management. Users need to keep track of changes made to a portal's service portfolio. Management can help users in this mission by implementing support concepts such as trainings or assistance services. A further challenge lies in matching each user, new and existing, with the set of potentially relevant services. While it can be assumed that each user knows the essential portal services he requires to execute his job, it is likely that more services will exist that have the potential to contribute to his job tasks. These requirements demand for an overall discussion of the consequences of corporate portal evolution.

The first part of this work (Chapters 1 to 3) is focused on these two aspects, the lack of an appropriate classification model for corporate portals and the challenges caused by the continuous evolution of portal implementations. In Chapter 1, corporate portal definitions and classification concepts from scientific literature are discussed, based on which a new classification model is presented that incorporates the potential that the advancements in web technologies offer. The second chapter describes and analyses the challenges related to the continuous evolution of running portals. This analysis is conducted concentrating on the user perspective. Among the key challenges lies the users' knowledge of the portal's service portfolio. To meet this and the other identified challenges, several support concepts are introduced to serve as a toolset for the management of corporate portal evolution. These concepts cover a wide area ranging from the organization of trainings to the implementation of recommender system and other support services.

In Chapter 3, the findings from the first two chapters are applied for the analysis of the Corporate Financial Portal (CoFiPot), a portal being used by the finance department of Bayer AG with several thousand international users. It was implemented in 2001 and has constantly grown ever since, both, in number of users and services. The effects from portal evolution, as discussed in Chapter 2, are examined more closely with the example of CoFiPot and underlined with the findings of an expert interview with managers of Bayer Finance. Moreover, a usage analysis shows that one of the key challenges for CoFiPot is to increasing the users' knowledge about the portal.

The second part of this work (Chapters 4 and 5) describes the concept of recommender systems for corporate portals and its implementation as one of the introduced support concepts for CoFiPot.

Chapter 4 discusses various recommender system concepts based on the requirements for a corporate portal. The chapter can serve as a handbook for the development of such a system. It discusses the advantages of certain algorithms and concepts and introduces new concepts specialized for corporate portals. Differences to existing recommender system concepts are elaborated.

In Chapter 5 the chosen concept is applied to CoFiPot. A recommender system is implemented using the proposed model. By only showing recommendations to half of the users over a four month period, the effects of the system are evaluated by comparing the behavior of the groups. Moreover, a survey is presented, which points at determining the perceived usefulness of the recommender system.

The structure of this work is outlined in Figure 0.1.

## Research Questions and Contributions

This work delivers three main contributions to the further exploration of Corporate Portals. At the very base of this field of study, a comprehensive classification concept for corporate portals is introduced (Chapter 1). It takes the potential of today's portal technology into account. The concept also incorporates different evolutionary stages of corporate portals, thus allowing differentiating between portals of the same kind.

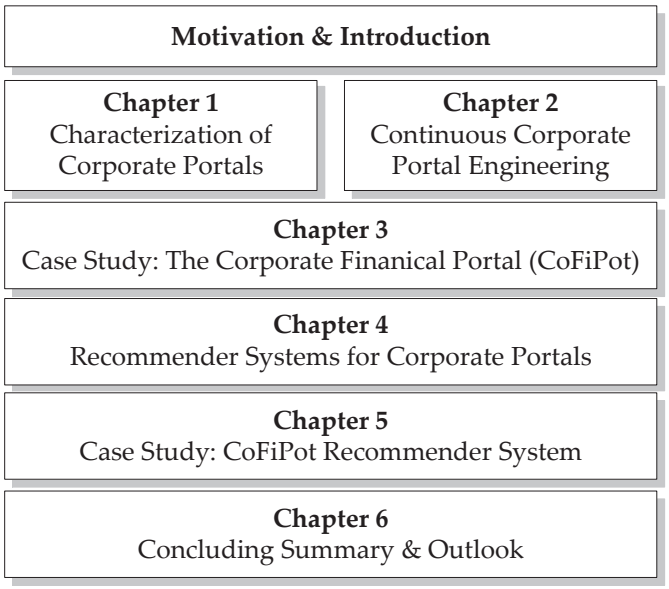


**Figure 0.1.:** Structure of this work.

Secondly, a basic concept for the management of continuous portal engineering beyond software engineering is defined for the first time (Chapter 2). To achieve this, the effects of portal evolution are analyzed taking the user perspective into account. This is supplemented with the discussion of derived tasks and support concepts.

Eventually, the model for the implementation of a corporate portal recommender system is presented (Chapter 4). The model discusses special properties of such a system compared to common recommender system and introduces a new concept to implicitly derive ratings from usage data.

Based on this work, a scientific approach to continuously improving corporate portals is possible. Furthermore, the approach of using recommender systems as one possible support system is presented and discussed in detail. Further works will need to focus on other support systems which should not be neglected when taking a holistic approach to corporate portal improvement.

# Chapter 1.

# Characterization of Corporate Portals

Portal technology became popular in the context of web portals in the 1990s. Examples are the YAHOO[1] and MSN[2] portals that offered their users a wide range of services ranging from information distribution via news-feeds to advanced services such as e-mail functionalities and online games. The portal technology allowed the provision of all these services via one central point of entry. Requiring users to register with the portal further allowed granting access to services based on the user's authorization level but also enabled the portal provider to collect more information about its users, their preferences and usage patterns.

The advantages of the portal concept were eventually picked up by companies to streamline their intranet functionalities (Aneja et al., 2000; Summer, 2006) and later to use the technology to provide all kinds of business services to their employees and customers, thus also avoiding or at least reducing the need to maintain multiple applications on each user's local computer. Starting with the initial corporate portal definitions by Shilakes and Tylman (1998) the concept was widely discussed over the following years. However, the focus of most authors is on the use of portals for knowledge management within a company (Aneja et al., 2000; Benbya et al., 2004; Tatnall, 2005). Some authors emphasize the potential for other uses such as the integration of internal and external applications and data sources (Murray, 1999; White,

[1] http://www.yahoo.com
[2] http://www.msn.com

2000b; Phifer, 2003; White, 2003) but these analyses are mainly limited to descriptions of potential benefits.

This chapter gives an overview of the corporate portal concept. The most relevant definitions are presented and categorized. Furthermore, core characteristics are identified that are common for corporate portal implementations. Following a brief introduction to portal system architecture, corporate portals are compared to public portals to identify commonalities and differences.

## 1.1. Defining Corporate Portals

Scientific literature introduces a multitude of terminologies for corporate portals that are exchangeable to a certain degree. In his overview article, Dias (2001) lists the terms *corporate portal*, *corporate information portal*, *business portal* and *enterprise information portal*, all of them used to describe the same concept. This list can be extended with Eckerson's *enterprise portal* (1999a) and White's *enterprise application suite* (2003). Firestone (2000) comments on this matter that "*the process of definition of corporate portals, as any other business-oriented strategy, is a political process, that is, an attempt to persuade the user community and the information technology (IT) investors that one definition is more adequate than another one, favoring the interests of one consultant or vendor over his competitors*". Although it is commonly agreed that Shilakes and Tylman (1998) were the first to define the concept using the term *enterprise information portals*, the other terminologies are based on different definitions of the concept (in addition to Firestone's observations on the subject). Shilakes and Tylman published their definition in a Merrill Lynch report in which they describe the potential of the portal technology as follows:

> *"Enterprise information portals are applications that enable companies to unlock internally and externally stored information, and provide users a single gateway to personalized information needed to make informed business decisions."*
> (Shilakes and Tylman, 1998)

This definition is focused on the potential of the portal technology to provide a single point of access to information from various sources within a company. However, shortly after Shilakes and Tylman, Murray provided a broader perspective:

> *"Corporate portals must connect us not only with everything we need, but with everyone we need, and provide all the tools we need to work together. This means that groupware, e-mail, workflow, and desktop applications - even critical business applications - must all be accessible through the portal. Thus, the portal is the desktop, and your commute is just a phone call away."*
> (Murray, 1999)

Murray extends the definition and scope of corporate portals to the integration of all types of applications instead of restricting it to information management aspects only. Many more definitions were published over the following years, each of them emphasizing different aspects of the potential that corporate portals offer.

Detlor (2000) strongly focuses on the information aspect. He identifies the potential of the portal technology to integrate three different *spaces* with the portal taking over the role as a *shared information work space* as shown in Figure 1.1. According to Detlor, the portal combines content with communication and coordination. Thus, its role goes beyond simple information publishing by providing functionalities for collaboration. In order to achieve this goal, he suggests focusing on the need of the people who will use the portal rather than on the technical requirements for retrieving all available information within a company. He envisions the corporate portal as an "information seeking system" rather than as "systems that merely support the retrieval of information". Portal developers should understand the requirements that people have when using the portal. Furthermore, they should not assume that people already know what exact piece of information they are looking for but rather "help them make sense of their environment, learn new ideas, or resolve their problems". He draws parallels to Taylor's (1991) *value-added model for information systems* that proposes a stronger focus on the users' information needs. Detlor also references Davenport's (1997) *information ecology* concept. Comparable to Taylor this concept criticizes the focus on technology and solutions to handle information like any other

good within a company. He proposes a holistic approach "which places emphasis on how people create, distribute, understand, and use information". Based on this analysis Detlor presents a framework to facilitate the design of corporate portals based on the users' needs. His definition and framework are limited to the aspects of information management within a company and do not elaborate on a broader perspective.

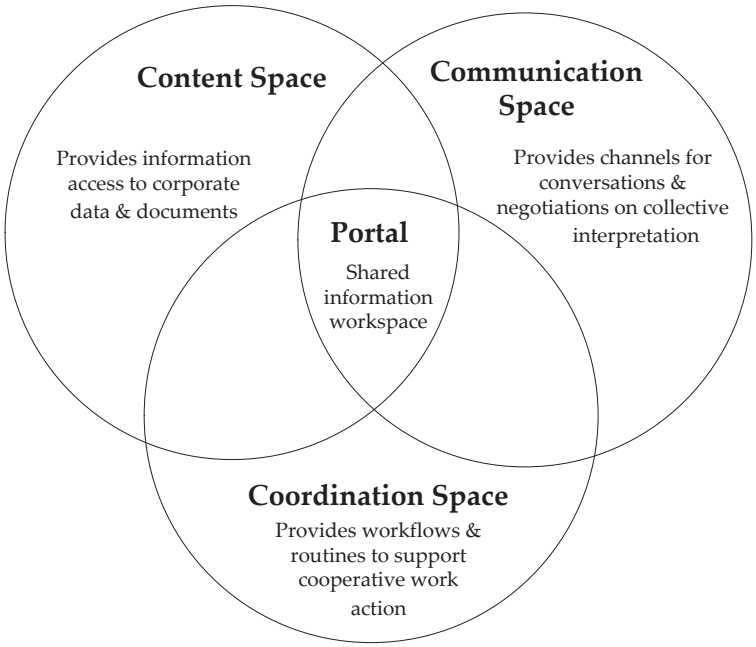


**Figure 1.1.:** The corporate portal as shared information workspace (Detlor, 2000).

A wider perspective is described by Davydov (2001, chap. 5). Following Murray's definition (1999), he picks up the information management context and combines it with the integration of application services. Yet, he remains focused on the role of information provisioning throughout his book. Still, he foresees the role of corporate portals to incorporate *intelligent agents* that "*master individual customers' or customer groups' demand priorities by learning from experience with them, and can quantitatively and qualitatively analyze those priorities*". With the preview he extends the portal's capabilities with business intelligence tools.

Daniel and Ward (2005) discuss the overall potentials of corporate portals and provide the following definition that describes the portal as a platform that integrates and connects all kind of applications and information sources:

> *"Enterprise portals are secure web locations, that can be customized or personalized, that allow staff and business partners access to, and interaction with, a range of internal and external applications and information sources. Uses of the portal may include: improved access to information, increased collaboration, greater use of existing applications and effective integration between applications."*
> (Daniel and Ward, 2005)

The authors use an empirical survey to analyze several portal implementations. Although the major part of their paper is focused on information management - due to the fact that the participating companies were using corporate portals mainly for this purpose - they conclude that companies have identified the potential of integrating further types of applications into the portal.

White provides the broadest corporate portal definition by introducing the *application server suite* (2003). This concept extends the scope for portal technology usage to also comprise the entire business infrastructure required for business processes.

Appendix A provides a comprehensive list of definitions for corporate portals (including the aforementioned ones), which in many cases describe similar concepts or have overlapping characteristics. Some are tailored to serve specific case studies or research areas (Detlor, 2000; Ryley, 2001); others encompass all potential uses into one definition (Murray, 1999; Daniel and Ward, 2005; Tatnall, 2005). In summary, most of these definitions can be categorized into two types:

- defining the corporate portal as a broad tool to cover all kinds of business purposes, following Murray's vision, and
- focusing on the distribution of information as described by Shilakes and Tylman.

While the appearance of ever new definitions is certainly explicable following Firestone's (2000) line of argumentation, another aspect might explain the ongoing appearance of new definitions: Daniel and Ward (2005) identify the continuous evolution of the portal technology as a reason for the lack of a widely accepted definition. In concurrence with this argument it is obvious that the first corporate portals

were limited by the restriction of programming languages, available frameworks and computing capacities. The fast advancements in web technologies have broadly extended the possible range of applications suitable for inclusion into portal systems (Luce, 2002). The progress of existing programming languages (such as Java or PHP) and the development of new languages, frameworks or packages (e.g. .NET, Ruby or JQuery) have opened the door for new rich web applications, which allow the implementation of more complex processes that were not feasible at the beginning of the new century. Labeled with the term *web2.0* (O'Reilly, 2005) and closely followed by *enterprise2.0* (McAfee, 2006), a new generation of web applications was introduced, which drastically transformed the way content and functionality was delivered over the World Wide Web. Standardization of technologies and the facilitation of common tasks allowed developers to create new kinds of web-based applications being able to provide highly complex functionalities. E.g., Google Docs[3] provides online document editing functionalities that were previously only available for desktop applications. This might also explain why many of the early corporate portal definitions were limited in the range of services that they deemed suitable for a portal to deliver. Based on the existing literature, it appears that most corporate portal implementations (that researchers had access to) were limited to information management functions. Creating information repositories with connections to several data sources was easier to achieve than integrating features of enterprise applications. Considering this, more specific definitions such as White's *enterprise application suite*, which is one out of three corporate portal types she lists (2003), are the logical extensions to Murray's (1999) original definition.

The multitude of corporate portal definitions and the many different views on the corporate portal technology and its role within a company repeatedly lead authors to suggest grouping patterns or to create taxonomies for corporate portals. Building on these ideas, the following subsections pinpoint corporate portals in the general portal software landscape (Section 1.1.1), present a corporate portal classification scheme based on portal type (Section 1.1.2) and evolutionary stage (Section 1.1.3).

[3] http:\docs.google.com

In the following, these approaches are taken to allocate the concept of corporate portals within the portal landscape. Starting with this allocation, a classification concept is presented that allows distinguishing corporate portals on two levels:

- the type of corporate portals with regard to its role, and
- the evolutionary stages that each type can reach.

This classification concept will help in further identifying the opportunities that the portal technology offers to support companies on many different levels.

### 1.1.1. The Position of Corporate Portals in the Portal Landscape

Several authors try to position corporate portals relative to other types of web portals in order to better understand the concept and distinctive features of corporate portals. Many papers simply distinguish based on the accessibility into *public* and *corporate* portals (Davydov, 2001, chap. 5; Dias, 2001; Gurugé, 2003, chap. 2); where only public portals are accessible for all Internet users (Gurugé, 2003, chap. 2). Portals that fall into this criterion are e-marketplaces[4], community portals[5], information portals[6] and specialized/niche portals (Tatnall, 2005). These authors have an undifferentiated view on corporate portals: While public portals are further separated into different types, corporate portals are not.

An alternative approach classification scheme distinguishes between horizontal and vertical portals ignoring public accessibility as differentiator.

*Horizontal portals* - also *general portals* (Collins, 2000) - serve as platforms for various applications. Usually they do not target a specific audience but provide a rich portfolio of all kind of services (Großmann and Koschek, 2005, chap. 1). Examples for such portals are Yahoo, MSN and also Google[7] with their broad spectrum of services.[8]

[4] E.g. Amazon (http://www.amazon.com) offering a wide variety of goods or http://www.eta.ch offerings only goods from one company.

[5] E.g. Facebook (http://www.facebook.com) or the community site for asp.net developers (http://www.asp.net).

[6] Information portals are mainly news-sites such as http://www.nytimes.com but also portals like Wikipedia (http://www.wikipedia.org).

[7] http://www.google.com

[8] Google offers services such as search, mailing or analytics via separate sites but also allows users to group all through the iGoogle portal.

*Vertical portals*, in contrast, are focused on specific processes. They offer a range of services that support these processes (Collins, 2000, chap. 2; Davydov, 2001, chap. 5; Gurugé, 2003, chap. 2; Großmann and Koschek, 2005, chap. 1; Tatnall, 2005), i.e., their focus is narrow and confined on specific goals.

Other authors further differentiate between business-to-business (B2B), business-to-employee (B2E) and business-to-consumer (B2C) portals as different types of vertical portals (Davydov, 2001, chap. 5; Gurugé, 2003, chap. 2; Knorr, 2004). Davydov calls these kinds of portals *role portals*. B2C portals either serve as selling platform for a company or provide information to attract customers. B2B portals support inter-company business operations ranging from simple information exchanges up to complex business-to-business online marketplaces (Arbin and Essler, 2002). B2E portals support employees in accomplishing their daily tasks and can also serve as information and training platforms.

Most definitions of corporate portals fall into the category of vertical portals because these focus on providing support for specific processes rather than offering a wide range of different services as horizontal portals do.

These two distinct views on corporate portals show that many authors see the corporate portal technology separated from public portals, while others distinguish by range or purpose of services offered. The second view appears more suitable because it allows identifying shared features independent of the portal's target group.

### 1.1.2. Specification of Corporate Portal Types

In an early literature overview, Dias (2001) groups corporate portals by the type of functions they support within a company. Several of the identified types are quoted and refined by other authors (Raol et al., 2003; Benbya et al., 2004; Daniel and Ward, 2005). However, due to the limited capabilities of the portal technology at that time most of these types are grouped by the possibility of handling information. For example, Dias (2001) and Firestone (2000) differentiate between portals used for decision making, decision processing, collaborative processing and hybrids of these types.

Recent definitions by White (2003) and Phifer (2003) distinguish corporate portals on a higher level and cover the broad spectrum of functionalities that the technology can offer. Based on these definitions, the following three types of portals a specified, which will serve as basis for all further analyses in the remainder of this work:

- business intelligence portals,
- application integration portals (and as sub-type application package portals),
- employee portals.

*Business intelligence portals* grant access to corporate information and provide analysis and reporting as well as collaboration and processing tools. This type includes the *collaboration/collaborative portal* (Firestone, 2000; Dias, 2001; Davydov, 2001), the *information portal* (Shilakes and Tylman, 1998; White, 2000a; Dias, 2001; Benbya et al., 2004; Detlor, 2004; Tatnall, 2005; Daniel and Ward, 2005) and other variations such as the *expertise portal* (Murray, 1999). The business intelligence portal is similar to the *Smart Enterprise Suite* as defined in a Gartner report (Phifer, 2003). Such a portal allows a company to distribute its data among decision makers, employees and business partners. In its functionality is comparable but not limited to the functionalities labeled with *groupware* in the 90s (Orlikowski, 1992; Grudin, 1994), which support users with all the information and tools required for collaboration and decision making. Business intelligence portals can enable employees to share and document their knowledge thus leveraging the sometimes hidden knowledge sources within companies (Davenport, 2005).

The *application integration portal* or *application server suite* (White, 2003) is focused on providing unified access to various applications. By integrating all types of applications and services - external and internal - the portal enables the company to connect all of them in one single platform. The main features of these types of portals are user interfaces for integrated applications, business process support and data access. It gives users the opportunity to access applications and services through one single portal rather than having to switch between systems (Schelp and Winter, 2002). It is comparable with the concept of a *business portal* as described by Eckerson (1999a) and Davydov (2001) but also includes collaboration and processing features. Examples

for portal solutions of this type are IBM's WebSphere[9] and the Plumtree portal (today part of Oracle's product portfolio[10]) (Knorr, 2004).

The *application package portal* (White, 2003) offers an out-of-the-box suite of applications by a single vendor. Examples are the SAP NetWeaver Portal[11] or the Oracle Portal 11g[12]. While such solutions can also support the integration of third-party applications they emphasize prepackaged solutions by the vendor. When implementing such a portal, the customer can select from the range of services the portal-package provides to support the underlying business needs. Different to the previously introduced *application integration portal*, such portals are specifically tailored to grant access to one single underlying system rather than integrating heterogeneous systems into a single front end (Edelman and Jussila, 2000). Thus, they often do not match the specific company's business needs (Hazra, 2002) or they become part of an application integration portal.

Still, within this classification, the application package portal is regarded as a subtype of application integration portals because it basically can be seen as a limited version of the latter one.

*Employee portals* define a separate type of corporate portal. They provide employees with various types of information that are not necessarily part of business processes. The offered services cover human resources related tools like holiday-planning, contract management, work-time tracker, and also general services such as information about the daily lunch menu or phone directories. The role of an employee portal can also be interpreted as that of an internal social network within a company. This type can partly be seen as a B2E portal (Davydov, 2001, chap. 5) with focus on non-business processes.

The presented distinction between three main types of corporate portals enables decision makers to select the right type of portal based on their business needs and current system infrastructure. Companies that seek to provide their employees with

[9] http://www-01.ibm.com/software/websphere
[10] http://www.oracle.com/bea
[11] http://www.sap.com/germany/solutions/netweaver/components/netweaverportal
[12] http://www.oracle.com/portal

a platform to collaborate, share and retrieve information will implement a business intelligence or an employee portal (the latter if the aim of the portal is to provide value-added services rather than supporting business processes). Companies that want to provide access to applications will select the application integration portal. Here they might select the application package portal in case they have a homogeneous system structure with one main business application (such as SAP), if such a solution exists for the specific system. The distinction between the three portal types is not exclusive. It is possible that a portal combines the features of multiple types.

### 1.1.3. Evolutionary Stages of Corporate Portals

Besides a type based classification, a corporate portal can further be categorized based on its evolutionary stage. Identifying the evolutionary stage of a portal is a useful addition to the planning process for a new portal implementation or the reengineering process of existing solutions. In the following, for each stage common features are grouped. Knowing the stage thus enables developers to better plan and analyze the development process knowing current and future stages of the referenced features.

Analyzing the evolution of public portals in the 90s, Eckerson (1999b) identifies three stages of portal evolution: *referential, personalized and interactive*. Dias (2001) later extended the list with a fourth level, which emphasizes that "portals have a potential to extend beyond the capabilities offered by public portals". She calls these levels *generations of portals*. The resulting four generations are shown in (Table 1.1).

The concept presented in Table 1.2 matches three stages of portal evolution with the afore defined portal types. At the referential level, all portal types merely offer references to other sites in a company's intranet and provide access to documents such as manuals or forms. The second stage, interactive portals, summarizes portals that provide various services that allow users to execute simple actions such as editing data, forwarding information or customizing reports. The specialized stage does not necessarily stand for further technological advances in the offered features but merely puts emphasis on special processes or target groups. It also stands for holistic

| *Generation* | *Category* | *Corporate portals* |
|---|---|---|
| *First* | Referential | Search engine, with hierarchical index of web content. Each index entry contains a description of the content object and a link to it. This generation emphasizes content management, mass dissemination of corporate information and decision support. |
| *Second* | Personalized | Through identification and a password, users create a personalized view of portal contents, known as "Mypage". This view shows just the categories each user is interested in viewing. The portal can notify users when new content is added to categories they have previously selected. Users can publish documents to the corporate repository so that other users may view them. This generation privileges content customized distribution. |
| *Third* | Interactive | The portal embeds applications that improve employees' productivity, such as e-mail, workflow, project management, expense reports, calendars, schedules, etc. This generation adds the collaborative character to corporate portals, providing multiple types of interactive services. |
| *Fourth* | Specialized | Portals based on professional roles, for managing specific corporate functions, such as sales, human resources, finances, etc. This generation connects corporate applications with the portal, allowing users to execute transactions, read, write and update corporate data. |

**Table 1.1.:** Generations of corporate portals (Eckerson, 1999b; adapted by Dias, 2001).

solutions, i.e., a portal that fully integrates all services available to employees of a company.

With this categorization scheme at hand, it is possible to classify existing and planned corporate portal implementations not only by type but also by their evolutionary stage. The scheme can support developers and management by providing information about features that are common in each category.

| | *Referential* | *Interactive* | *Specialized* |
|---|---|---|---|
| *Business Intelligence Portal* | Portals that provide links to documents and other sites in the intranet. Search engine and hierarchy. | Implementation of collaboration features allowing to edit and share documents. Workflow management support. | Portals focused on collaboration for project and process management. |
| *Application Integration Portal* | Portals that provide links to applications. Simple integration of application with by providing data from integrated applications. | Full integration of applications that allow the execution of all kind of processes through the portal. | Portals that support special processes within the company. E.g. portals supporting all processes of a department or role in the company. |
| *Employee Portal* | Simple sites that provide references to various sites within the intranet. Download of forms and documents. | Integration of services for HR and other employee-related tasks. E.g. management of holiday applications. | Extension to employee community portal with discussion forums, walls, etc. |

**Table 1.2.:** Corporate Portal types and evolution stages.

## 1.2. Characteristics of Corporate Portals

To describe the capabilities of corporate portals its features were summarized. Eckerson (1999a) presented a list of "15 rules for enterprise portals" that a corporate portal should comply with (cf. Appendix B). This list was picked up and extended by several authors (Raol et al., 2003; Smith, 2004; Benbya et al., 2004; Tatnall, 2005; Daniel and Ward, 2005). Raol et al. (2003) and Daniel and Ward (2005) use those characteristics to group the corporate portal solutions that were available on the market at that time.

Benbya et al. (2004) classify portal features into three categories: core capabilities, supportive capabilities and web services. As basis for the feature list they reference the corporate portal framework from Aneja et al. (2000), which is introduced to define the essential requirements to create a corporate portal with focus on information categorization and personalization (see Figure 1.2). Their paper analyzes Intel's inter-

nal knowledge management requirements based on the sheer volume of information that is available within the Intel intranet. They emphasize the need for a corporate portal to organize the available information by creating taxonomies and providing extensive search features.

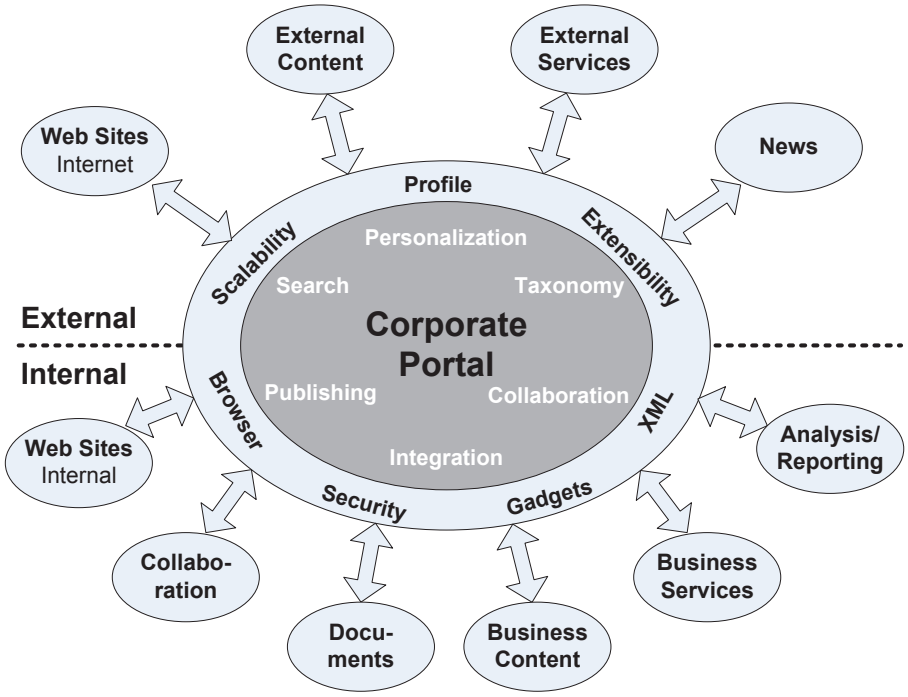


**Figure 1.2.:** Corporate Portal Framework by Aneja et al. (2000).

Most of the characteristics concern basic features while advanced concepts such as application integration are often neglected. Grouping concepts, like those introduced by Benbya et al. (2004), are - again - specific for information focused portals and do not cover the wide range of portal applications as proposed in the previous section. The following feature grouping attempts to encompass the capabilities of all corporate portal types (Krämer et al., 2010). The three groups are:

- technical functionalities,
- usability aspects,
- and business process integration.

The list picks up the characteristics that were presented in previous papers and extends the list with a strong focus on business process support capabilities.

### 1.2.1. Technical Functionalities

Most of the prominent portal functionalities to be discussed in this section rely on the available technical capabilities of the portal.

Especially in the corporate context, portals have heightened requirements for security. The security features of a portal include connection features such as secured https-connections and firewalls, authentication methods and cryptographic services to prevent unauthorized access (Eckerson, 1999a; Luce, 2002; Gurugé, 2003). With the portal serving as single point of entry to all kind of underlying systems, security issues like access management have to be handled on this level rather than separately in each application. Several portal vendors also offer the ability to integrate the portal's user administration with directory services such as Microsoft's Active Directory or Novell's Directory Service (Edelman and Jussila, 2000). This functionality saves the effort of maintaining double user lists. It also helps in keeping the user database up-to-date. New users can be added and old users are removed automatically. In this vein, the technical capabilities are closely linked to usability aspects like single sign-on, which is discussed later. Another important technical aspect is the portal's scalability, i.e., the ability to cope with a growing number of users and services (Benbya et al., 2004). Moreover, the portal can provide standardized interfaces (API) to communicate with other applications (Eckerson, 1999a; White, 2003). Through these interfaces other applications can be fed with latest information assembled by the portal and its underlying business intelligence tools, or the portal itself acquires information from other services or data sources. APIs also allow connections to other portals, for example, by using web services as interface. Finally, one of the key advantages of portals over desktop applications is the web browser interface. In particular, this means that there is no need to install any further software on the users' machines (Eckerson 2000). This greatly simplifies software deployment, because updates only need to be uploaded to the server and are effective immediately without any further need to distribute software packages to users. This is not only cost effective but also

ensures that only one version of the software is running[13] (Elsner and Vo, 2007). The software update cycle of web portals can therefore be highly dynamic and even might allow multiple updates per day.

### 1.2.2. Usability

Improving usability is a primary motivation for introducing portals in the first place and therefore highly important for their acceptance. In particular, six different dimensions can be identified which contribute to portal usability:

Foremost, *ease of use* encompasses all aspects which facilitate the process of using the portal itself. For example, this includes a common interface, such that users do not need to switch between applications or web sites to follow a course of action but can rather access all required information and services within the portal. Thereby, a single sign-on ability can allow a faster access to portal content, because users are authenticated without having to enter credentials (Edelman and Jussila, 2000; Gantner et al., 2006).

Also common design elements, *look&feel* and standardized user interfaces help users to navigate and understand the system (Nielsen, 1993). The portal may also already provide a taxonomy, which allows users to group categorize content (Aneja et al., 2000). This further enhances portal navigation and thus usability.

Additionally, users can be given the option to *customize and personalize* the portal according to user- or group-dependent requirements (White, 2000a). Moreover, portals can also provide each user with the option to adapt the appearance and to automatically create personalized output (Benbya et al., 2004).

The possibility to *search* the portal's content and services is a prominent functionality and a key feature of knowledge portals (Fagin et al., 2003; Detlor, 2004; Benbya et al., 2004). Search should cover content itself as well as META-information that describes further the content.

[13] If the software is installed on each client's desktop, several different versions of the software might run on different desktops if upgrades are not installed immediately. It has to be ensured that previous versions will work at least for a short period until all versions have been upgraded.

Complementary to the search, which requires users to *pull* information from the portal, the portal can also offer subscription features, which *push* updates and new information to users. This is achieved by allowing users to subscribe to portal services, e.g., through mail or content syndication (RSS) or customized reports.

Finally, the usability of portals is also constituted by a range of collaboration features, which facilitate the interaction between portal users. Collaboration features may range from the ability to upload and publish documents for selected users or user groups up to collaborative document writing (Benbya et al., 2004; Daniel and Ward, 2005). The latter could include wiki-features as well as more advanced document management features such as versioning. Collaboration also includes discussion forums or feedback mechanisms like commenting tools or ranking functions.

### 1.2.3. Business Process Integration/Business Intelligence

While the aforementioned usability and technical functionalities also apply to public portals, corporate portals are additionally capable of providing business process functionalities. Of course, such functionalities are highly contingent to the nature of the firm that utilizes the corporate portal. However, in more general terms, one can distinguish three broad types of business process functionalities: Integrated business intelligence, business process execution and automated reporting.

*Integrated business intelligence* describes all functionalities which are associated with the portals' ability to extract combine and analyze information from various available resources in order to generate new information (Edelman and Jussila, 2000; White, 2000a; Davydov, 2001, chap. 5; Top, 2005). For example, business intelligence tools can support business processes by assembling process status information and initializing follow-up processes. Thereby, the ability to trigger follow-up processes outlines the core concept of *business process execution* functionalities. Likewise, a simple knowledge portal can be significantly enhanced by integrated business intelligence applications which combine data and display the relevant information directly in the application. In addition, in combination with historic data, such applications can provide benchmarking and forecasting functionalities: Benchmarking enables to detect

deviations from historic developments which may indicate errors or other recognizable events. Based on this information other automatic or semi-automatic reactions can be triggered directly. Beyond event detection, interpolation of historic and current data may have predictive value and can thus help in planning operations.

Business intelligence tools are often closely coupled to *automated or scheduled reporting* features. These can either be triggered through events (Edelman and Jussila, 2000) or be based on a schedule. For example, an event-based report can result from status changes or external input such as data input or a service requests. Scheduled reports, on the contrary, are triggered independent of specific events at fixed time intervals. The role of business intelligence in the creation of reports is to provide extensive analysis of data (Benbya et al., 2004). Automatic reporting should be implemented in combination with the above described publishing capabilities (cf. Section 1.2.1). Especially event-triggered reports can be published, using, for example, RSS-feeds. In this way subscribers can be informed over multiple channels when an event has occurred. Automatic reporting also includes the functionality of automatic alerts. For instance, if the portal automatically provides prediction data, it can also issue an alert in case the current data significantly differs from predicted numbers.

| *Technical* | *Usability* | *Business Process* |
|---|---|---|
| - Security | - Ease of use | - Integrated business intelligence |
| - Global directory integration | - Taxonomy | - Business process execution |
| - Scalability | - Customization & personalization | - Automated & scheduled reporting |
| - Standardized accessibility | - Search | |
| - Browser accessibility | - Subscription | |
| - Ease of deployment | - Collaboration | |

**Table 1.3.:** Overview and hierarchy of corporate portal functionalities.

Table 1.3 summarizes the results of this section and shows the dependency of the portal functionalities. The table also shows the various features and feature combinations that corporate portals can provide. With the ongoing improvements to web-technologies the list might further evolve. As of today, it is already possible to provide

all functionalities of an integrated application through web interfaces.[14] Thus, in theory, all application interfaces can today be provided through a corporate portal using currently existing (web) technologies.

## 1.3. Corporate versus Public Portals

Some of the presented corporate portal classifications distinguished between public and corporate portals (cf. Section 1.1.1). However, there are no detailed comparisons between public and corporate portals available that go beyond superficial differentiation of the main objective of each type. However, as described before, the corporate portal development has closely followed the advances of public portals and will continue to do so. New challenges will arise due to this constant evolution that portal developers and providers have to cope with. Careful management of the portal and the services it provides becomes increasingly important. These conditions apply to corporate as well as public portals. It is therefore important to identify in detail the similarities and differences between the two types.

The topmost level for this comparison is the objective behind the implementation of a portal. For public portals, however, the attractiveness to users is the key driver. In the corporate portal context, where users are employees and the sole purpose of the portal is to increase their productivity, the challenge is to enable portal users to exploit its potential and advantages over separate applications to the fullest. In this aspect, corporate portals are very different to public portals. Identifying further differences will help in specifying useful techniques to handle challenges in the engineering process of corporate portals. In the following these differences are defined with regard to the each portal type's key objectives and portal management challenges.

[14] For example, Microsoft's Exchange Server was one of the first applications that provided a web interface with almost all functionalities that were provided by the client software Microsoft Office Outlook

### 1.3.1. Environment and Objective

One of the fundamental differences between public and corporate portals is the environment they are designed for. Public portals are usually accessible for all Internet users while corporate portals are limited to a companies' intranet or - if reachable through the Internet - only accessible for employees and partners of the company. Thus, public portals have a potential user base of millions of users while their corporate counterparts are limited to usually less than a few thousand (Zanker and Gordea, 2006). From a technical perspective this has many implications. First of all, public portals have to be highly scalable in order be able to handle up to several million user requests. Moreover, the number of users and requests is hard to predict and, for instance, in case of a rising popularity of a portal, special precautions have to be taken to be able to cope with quickly increasing user requests. Within the closed circle of a companies' network the number of users is highly predictable and usually comparably small. Scalability is therefore a minor issue. Another issue for public portals is to take much more rigid measures against malicious users. While corporate portals just need to ensure that only authorized persons have access, public portals also need to control the behavior of their users - even if they are required to register before they use the portal. Misuse of public websites is a well-known problem. For instance, in case of Wikipedia, malicious users often try to manipulate information about persons or events.[15] It is however very unlikely that employees will manipulate content within a corporate portal as the penalty for such behavior could, in the worst case, be the risk to lose their job. Therefore, it is not as important to implement processes for content validation and user monitoring as in a public portal.

With respect to objectives, public and corporate portals are very different. A public portal is usually part of the company's business model if it is not even the business model itself. For instances, portals such as Facebook, YAHOO or MySpace are the foundation for companies with hundreds of employees. Without these portals, the companies would not exist. Other companies use public portals to promote their

[15] For example, in 2009 the death of two American senators was announced on their Wikipedia-pages although they were still alive (http://voices.washingtonpost.com/capitol-briefing/2009/01/kennedy_the_latest_victim_of_w.html).

products or to provide tools related to those. Such one example is Procter & Gamble's baby-portal[16] for young parents, which provides community functions for baby-related questions. This portal follows the business goal of promoting Procter & Gamble's diapers and other baby-related products and not to earn money with user fees or advertisements as it is the case with Facebook and similar portals.

In contrast, corporate portals *support* the business model of a company without being the business model itself. Their value-add is to support employees and the management with reports and all kind of other services. This has consequences for the availability and stability of the portal. If an outage occurs in a public portal like Facebook, it immediately endangers the company in behind. Furthermore, even minor errors or flawed functionalities are usually critically commented in public. On the contrary, failures within corporate portals will not at once lead to bad publicity or endanger the company's success. Of course, it must be ensured that errors are fixed within a short period of time but even an interruption of a few hours will usually not cause severe problems. Furthermore, manual "emergency" process instructions are usually available that ensure continuation in operations.[17]

As the portal is the business model or at least part of it, it requires providers of public portals to ensure attractiveness of their product. Otherwise they will not gain new customers and might lose existing ones. Providing an attractive environment includes using the latest (web) technologies, providing popular features and content and to ensure a high reliability. Providers need to stay on par with technology advancements and also continuously need to monitor competitors in order to identify new "killer features". If competitors offer comparable services for a better price, with more features or at a better quality, users might switch portals. Within companies the attractiveness does definitely not have the same priority as in public portals. For once, the users of a corporate portal are the employees who are required to use the portal in order to fulfill certain tasks. Furthermore, there is no competition with other

[16] http://www.pampers.com

[17] In the case study, presented in Chapter 3, a failure that causes the portal to stop will lead to complaints but as long as it is fixed within 1-2 hours, no significant problems will arise. In the worst case, if the portal crashes shortly before the start of automatic processes (e.g. the end-of-day process that consolidates all trades of the day), the processes need to be run manually.

portals. There is no reason for companies to implement competing portals solutions. In case that some minor features are overlapping between different solutions, it does not matter as long as it does not hinder any processes.

Due to these considerations corporate portals have no need to provide an attractive environment with the latest features as long as the users can fulfill their tasks with the existing solution. Therefore, the priority for upgrades is usually very low. This situation, although seeming to be logical, might still hinder the full exploitation of a corporate portal, a matter, that is further discussed in the remainder of this chapter.

### 1.3.2. Differences in User Motivation

Corporate portal users are usually employees of the respective company and thus different to (anonymous) customers in public portals. This difference plays an important role when looking (i) at each user group's overall motivation and (ii) the specific motivation on how to use it.

Public portals are often competing with other portals for customers. Therefore, they will strive to provide a highly attractive interface with interesting products and/or features. Based on the available offers users choose the most attractive portal that fulfills their purposes. If another portal starts providing better services, users will simply switch to the better offer. In contrast, corporate portal users are not *attracted* to the portal but forced to use the available solution.

These differences can be seen as the foundation for differences in usage patterns of the different user groups. Public portal users are self-interested in the portal's functionality and its offers. They use it whenever they want to, e.g., just to spend some time searching for products, features, etc. They are likely to remain longer on the portal web site if there are items of interest such as interesting articles in a news portal or special deals in a shopping portal.

In contrast, an employee will access a corporate portal as soon as he needs to retrieve some information or use one of its services. In general, the visit is part of a business process. When the specific task in the portal is accomplished, the user will leave it. The likelihood that an employee will spend time in the portal to discover

new features is rather low since the visit does not occur out of personal interest but is driven by specific business need.

### 1.3.3. Complexity and Support

The complexity of the services offered by a corporate portal is often much higher than those of public portals. Corporate portals (as shown later on in the case study in Chapter 3) can integrate several applications and allow the execution of business transactions and sometimes even might completely automate processes. The complexity of such services is usually not comparable to, for example, buying a product online or editing an article at Wikipedia.

Another difference is that employees are usually trained for their job and thus understand how each portal service they need to use functions. They will also use the services on a regular basis so that they can get used to even highly complex interfaces with multiple options. Furthermore, it is possible to offer help to corporate portal users on many levels, which can range from online- or phone support up to trainings. Also it is very likely that a user will know other users, e.g. within the same department, that also have experience with the portal and can help on an informal level.

Due to costs, training or phone support (to explain services) for public portals are usually not available. Services must therefore be intuitive and are consequently limited in their usage complexity. Support is usually handled through a frequently asked questions (FAQ) page, forums or contact forms only.

Another aspect is the heterogeneity of the offered services and its users in corporate portals[18]. Such portals usually support various business processes via a variety of different services though each employee will only use a small fraction of them. For instance, in a portal that supports all aspects of order management ranging from order assembly to billing and complaint management, each involved employee will only require access to a few of the offered service. Additionally, each group of services

[18]The topic of heterogeneity is further discussed in Section 2.1.

and users might require different features or design elements thus further contributing to the complexity of the portal.

A public portal usually provides a consistent set of services to the whole group of users. Even if the variety of services is high (e.g. with YAHOO's or MSN's various services), neither the heterogeneity of the user base nor that of the services is likely to reach the level of corporate portals. Again, the most important aspect is that corporate portal users can be systematically trained to use complex service offerings.

### 1.3.4. Differences in Security Settings and Access Control

Another difference between public and corporate portals stems from the access management functionality. Public portals often require the users to authenticate themselves. In some portals users can choose between a free and a premium membership based on which some additional features of the portal become available. For example, at XING[19], a portal that allows users to create a business network by creating links to other users, standard members have access to all basic functionalities while premium members have access to several additional features such as an overview of all users who visited their profile as well as more advanced options to contact other members. Another example is Amazon's portal that differentiates between private, corporate and seller accounts. Each account offers different functionalities to the user.

Still, the number of user roles in public portals will in most cases be limited to anonymous users, those that have signed in and premium members.

In contrast to such identity-based concepts, the access management in corporate portals is handled with role- or rule-based models (cf. Section 1.4.4). The number of potential access-right combinations is high, especially as compared to public portals with two or three potential sets of access rights. Accordingly, the requirements to administrate and monitor these access rights are higher and require special care. If the portal offers access to critical data, it must be ensured that only those users have access who have the necessary job clearance. Ensuring that (i) each service implements

[19] http://www.xing.com

the access right functionality, that (ii) the rights are correctly set and that (iii) each user has the appropriate set of rights adds to a corporate portal's complexity.

Table 1.4 provides a summary of the key differences between public and corporate portals. The analysis showed that both portal types serve different purposes and target different types of users. Accordingly, they offer services with different attributes. Thus, they face challenges that differ in priority and type. Concepts for the engineering and administration of public portals cannot simply be transferred to corporate portals.

| | *Corporate Portal* | *Public Portal* |
|---|---|---|
| - *Environment* | Corporation / intranet | Internet |
| - *Objective* | Business model support | Business model |
| - *Scalability* | Predictable | Unpredictable |
| - *Users* | Employees | Users/customers |
| - *Complexity* | High | Low-medium |
| - *Support* | Training, personal, hotline, informal, mail-support | FAQ, mail-support, hotline |
| - *Access Management* | Role-based access | Simple authentication |

**Table 1.4.:** Functional differences between corporate and public portals (Krämer et al., 2010).

## 1.4. Technical Architecture of Corporate Portals

Each corporate portal implementation has its own set of main characteristics and provides different types of information using different approaches. However, the basic architecture of every portal follows the same concept. The corporate portal serves as middleware between applications, various data sources and the employees of a company. It provides a unified interface accessible through the user's web browser (cf. Section 1.2.1). Shilakes and Tylman (1998) distinguish between structured and unstructured data that corporate portals integrate. Structured data sources are applications and services that are integrated in the portal; unstructured data comprises all types of documents, e-mails and data from other websites and the corpo-

rate intranet. Access to these sources is realized through application programming interfaces (APIs) of various types. Figure 1.3 shows the basic architecture emphasizing the portal's role as middleware between internal and external systems, on the one hand, and the users, on the other. It also depicts the access control layer that ensures that services and content delivered by the portal are only accessible to authorized users.

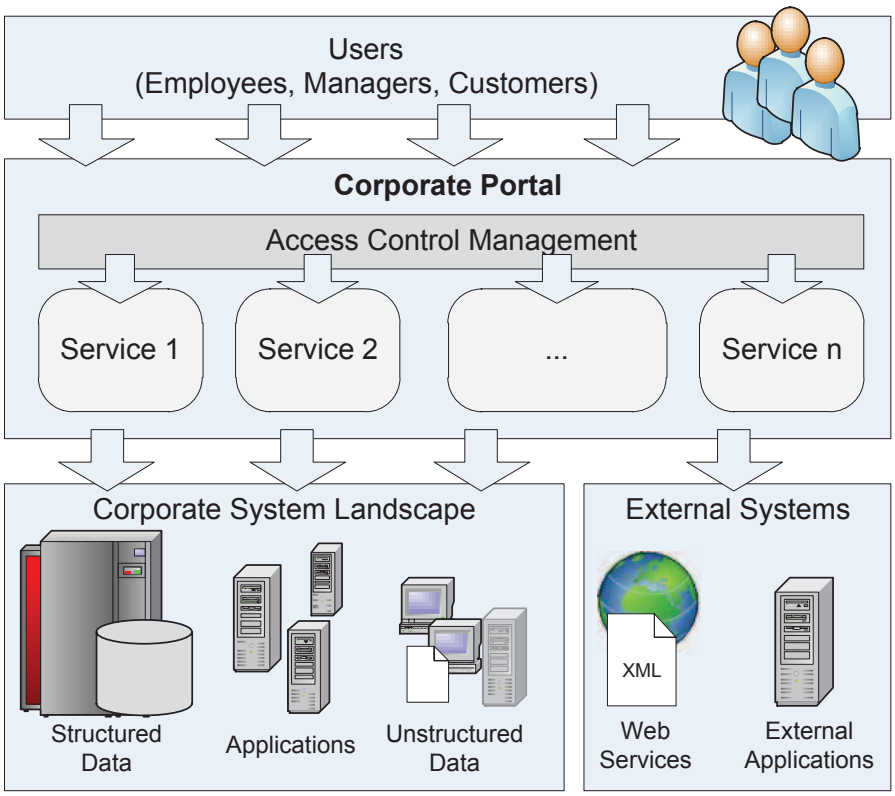


**Figure 1.3.:** Corporate portal architecture with interfaces to various sources and access control component.

### 1.4.1. Accessing Web Services

Web service interfaces allow the distribution of data and services using Extensible Markup Language (XML) messages that are transported via the Simple Object Access Protocol (SOAP) protocol. By applying standardized protocols and patterns, the technology enables service distributors on the one side to provide their services using

a widely accepted standard and service requesters on the other side to easily implement services from various providers using a common interface technology rather than having to individually adapt to each providers API.

Corporate portals can use web services as an interface to exchange data with external and internal services. External services include information sources such as stock prices, news streams or data from business partners[20], but also advanced services that allow the execution of processes or the management of data in external systems. Internal services are all those services that are offered by applications or other portals within a company.

The corporate portal itself can also provide web service interfaces to enable other applications to access the portal's data and services.

### 1.4.2. Integrating Existing Applications

To integrate an application into a portal, the system either uses interfaces that are provided by the application as described above or directly accesses the specific application's underlying databases. The decision for one or the other method depends on the availability of matching interfaces but also on performance requirements of the functionalities that need to be integrated. Even if an application provides a suitable interface (e.g. for the provision of data), the performance of the interface can be unsatisfyingly slow, especially when dealing with large amounts of data. The reason is that data provided through a web service has to be initially parsed within the application and then send to the portal that has to parse the data again. In such cases the overall process can be significantly faster if the portal gets direct access to the applications underlying database.

### 1.4.3. Accessing other Data Sources

Various other data sources are available within a company. The list includes but is not limited to network file directories, custom applications or databases, communication

[20] For example, Corporate Financial Portal (CoFiPot), the corporate portal which is portrayed in the next section, uses web services to retrieve bank account statements from financial institutes.

data (e.g. e-mails) and data from other sites in the corporate intranet. For each of those data sources a custom interface has to be developed.

### 1.4.4. Access Control

It is common practice that corporate portals provide confidential data such as trade-info, performance indicators or planning data. To ensure that only authorized users have access to such data, access control systems are part of each portal solution. Access to portal services is usually managed with role-based or rule-based models while identity-based concepts are rather not suitable. Figure 1.4 shows the basic concept of these three models.

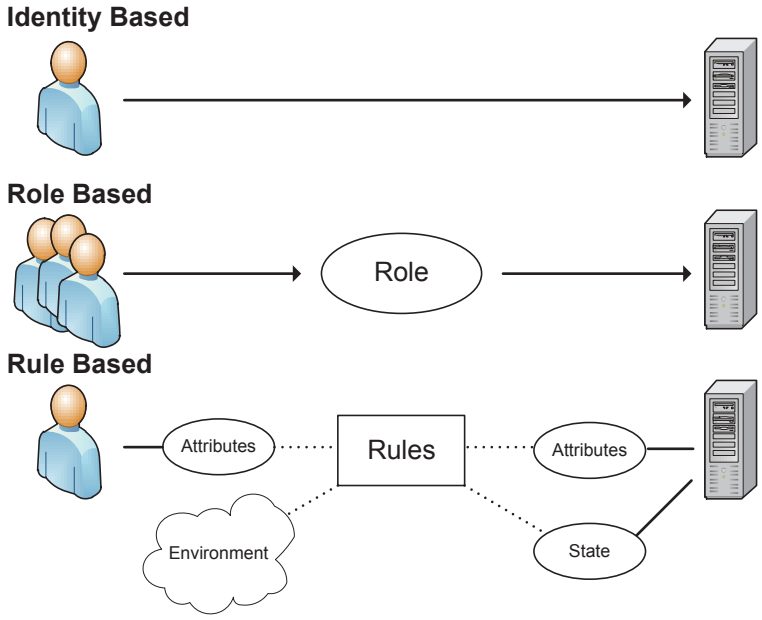


**Figure 1.4.:** Access control model types.

In *role-based models* each user is assigned several roles based on his work tasks, department and hierarchy level within the company (Benantar, 2006, p. 190 et seq.). E.g. based on their roles, managers have access to sensitive reports that should not be accessible to employees, accountants have access to bank information and dealers are able to execute trades. Using role-based authorization, the access settings can be controlled on a coarse-grain level. Certain columns of reports can be dependent on access rights. Also the visibility of fields in interfaces or the possible options to choose

in such forms can be controlled. Several concepts for the development of role-based access control are available which can be grouped into top-down and bottom-up approaches (Molloy et al., 2008). In the top-down approach, the real-world business is analyzed and roles are derived from it (Fernandez and Hawkins, 1997; Roeckle et al., 2000; Neumann and Strembeck, 2002). Bottom-up models, on the other hand, target existing systems that already have an identity-based access control and need to be migrated to role-based access control. Their goal is to analyze the existing system and propose a set of roles (Molloy et al., 2008).

In *rule-based access control* models the decision on whether or not to grant a user access to a resource is based on the evaluation of rules which are based on the user's attributes such as the position within the company (Denning, 1976; Department of Defense, 1985; Clark and Wilson, 1987; Moffett and Sloman, 1991; Bell, 2005). Based on those rules, similar to role-based models, access is granted or denied to resources.

In *identity-based models*, direct connections between users and permissions are established. Interdependencies are shown with the access-matrix model (Lampson, 1974; Harrison et al., 1976) that contain a row for every subject (here: user) and a column for every object (here: service). Each field of the matrix contains a list of access rights that the respective subject possesses for the respective object. Access matrices can become very large and sparse. For performance reasons, they are therefore often implemented in the form of either access control lists or capability lists. For portals with several hundred or more users and multiple access rights, the administration of identity-based systems quickly becomes very complex. Therefore, they are not suitable for portals that require multiple security levels. On the other hand, the model is suitable for public portals where access rights seldom will extend beyond authorized and premium users (cf. Section 1.3).

The corporate portal literature provides further specifications for corporate portal architecture such as the specification of common elements that portals offer (Shilakes and Tylman, 1998; Dias, 2001; Davydov, 2001) or capability lists Aneja et al. (2000). Some of those aspects have already been introduced in Section 1.2.1. This work focuses on management and usage aspects of corporate portals. Portal architecture

hereby plays a subordinate role and is therefore not further described in detail. More information on portal frameworks and architecture can be found at Davydov (2001), Zykov (2004) and Großmann and Koschek (2005).

## 1.5. Chapter Summary

This chapter gave an overview of corporate portal definitions, categorizations for portals and capabilities that corporate portal solutions can provide. Most literature sources are focused on portals that provide service for corporate knowledge management only. The possibilities that the portal technology offers due to the continuous developments in web-technology have not been reflected in scientific literature so far. Although some authors such as Murray (1999) or White (2003) have emphasized the potential range of applications for corporate portals, no detailed review of the changing corporate portal environment is available yet. Thus, the existing approaches to classify portals become obsolete and do not reflect today's possibilities.

To close this gap, three main types of corporate portals were presented to reflect the range of applications for the corporate portal technology:

- business intelligence portals,
- application integration portals, and
- employee portals.

Furthermore, three evolution stages for each of these types were identified to provide another dimension for the classification of new or existing portals. This classification concept also takes into account that similar types of portals may follow different objectives. While one company only requires a business intelligence portal that offers its employees the possibility to find relevant documents, another company might require an advanced solution enabling users to use the portal as main collaboration tool for project management. To cope with these differences, the distinction between different portal evolutionary stages is introduced as a third classification criterion.

In the remainder of the chapter, the focus is shifted to the comparison of public and corporate portals. Both portal types are subject to the same technological advances and thus face comparable challenges to development and portal management. However, it is unclear whether the same solutions are applicable in both environments. Thus, different approaches to handle challenges related to continuous portal evolution are discussed.

Finally, as a foundation for technical considerations in the following chapters, the basic technical concepts of corporate portal architecture are introduced.

# Chapter 2.

# Continuous Corporate Portal Engineering

Once a corporate portal has been initially implemented, a continuous (re-)engineering process has to be established. Its content and service portfolio likely has to be extended or changed to cope with changing business requirements. This development is influenced by multiple factors, which are described and discussed in this chapter. The process of constant portal (re-)engineering has to fulfill several criteria. Besides regular software development tasks (e.g. bug fixing), the quality and usability of the portal has to be continuously ensured and reviewed. Especially for portals whose service base is constantly growing a variety of additional challenges has to be addressed. Most importantly, users have to become aware of newly offered services. This awareness issue and several related aspects are the core focus of this work. The basic software engineering process for corporate portals is not further discussed. As several authors provide good descriptions for both, the initial development phase (Davydov, 2001; Hazra, 2002; Sullivan, 2003; Remus, 2007) as well as the continuous software engineering process (Vo, 2007). Other aspects, however, are not discussed in detail in scientific literature. Moreover, case studies are rather focused on the implementation phase (Raol et al., 2003; Counsell, 2004; Daniel and Ward, 2006; Remus, 2007) or the first phase of a portal's operation (Sullivan, 2003; Cloete and Snyman, 2003; Top, 2005; Remus, 2007).

The research agenda covered in this chapter are threefold: The first part focuses on the factors influencing the continuous portal engineering process and analyses them with regard to the different portal types, which were presented in the previous chapter. The second part (Section 2.2) looks at the resulting challenges and groups them into two sets. Finally, in Section 2.4, different portal engineering support concepts are introduced that have the potential to facilitate the handling of a corporate portal for both, users and responsible persons for the portal. Each of these support concepts addresses one (or more) of the previously identified challenges in portal engineering. Their applicability further depends on the type of portal and its evolution stage.

The analysis in this work is restricted to services that are offered by corporate portals. Considerations regarding knowledge management aspects such as the quality of content or technical aspects like performance are not discussed. For simplicity reasons, it is assumed that a service in the portal is represented by a page. Thus, each service can be accessed through the portal navigation. A more granular definition of a service in the portal context is possible but would not contribute to the context of this section. Moreover, the concepts being presented in the following can also be applied for a more granular approach to define a portal service.

## 2.1. Challenges in daily Portal Operation

Corporate portals posses certain properties that have to be addressed in daily portal operation. For example, the complexity of offered services, which was described in Section 1.3.3, usually requires additional training effort for target users. This and other challenges such as continuously changing business requirements are described in more detail in the following subsections. Three groups of challenges can be distinguished:

- challenges related to the services offered by a portal,
- challenges related to user requirements and changes in the user base,
- challenges related to the overall portal structure and objectives.

### 2.1.1. Heterogeneity and Evolution of Services

A corporate portal provides a variety of services to its users. E.g. in a business intelligence portal, these services oftentimes include download and edit functionalities for shared documents or real time collaboration tools. Application integration portals offer services that provide functionalities and data from underlying applications. Employee portals provide services for HR-related tasks or offer information for employees such as company news.

#### Heterogeneity of Services

The service portfolio of a corporate portal can support multiple processes within a company, targeted at users with different business tasks and within different departments. Thus, not all services are relevant for all users (cf. Section 1.3) but also very different types of services or views on the underlying applications might coexist within one portal.

Two types of service groups can be specified: the first consists of services that are *stand-alone*. They are focused on one process or provide access to a single functionality and do not depend on other services. On the contrary, the second group contains services *supporting process chains*. Here, multiple services have to be accessed in certain sequences in order to fulfill a task. A strong interrelationship exists between services of this type. Business intelligence and employee portals usually consist of stand-alone service whiles application integration portals normally offer both types.

A document library offering electronic documents for a specific topic is a typical example for a stand-alone service in business intelligence or employee portals. A service that offers functionality to edit the content of a database or to execute an action in an underlying system is another service example, which is commonly found in all types of portals. Stand-alone, in this context, does not imply simplicity. The key characteristic is the fact that a user can fulfill a task by solely using the particular service.

Process-oriented services, in contrast, provide functionalities for one process step in a bigger process chain. In order to finish a process, multiple of these services have

to be accessed. An example is the service process chain that supports the order process of a company. One service provides the functionality to administrate incoming orders, another is used to oversee the packing and shipment process and a third provides access to the billing process. Each of these services depends on the others, they are all part of the order management process.

The distinction between the two groups strongly depends on each case and the definition of a service. It is possible to divide the features of one service into two services and also to summarize the required functionality for an entire process chain in one service as it is described in this work. Therefore, it is not possible to strictly assign certain services to each group.

#### Extension of Services

Once a service has been implemented into a portal, it is likely to be subject to further changes that extend the simple bug fixing. Usually, the first version of a new service supports key features only. E.g. a document management service allows uploading and downloading documents and an order administration service offers to simply add and edit order entries. As soon as the core features are supported, the service will be added to the portal's portfolio and made accessible. Subsequently, features will be added because new requirements arise from user experiences with the service or to extend the scope of the offered features. In the example of a document management service, new features could be versioning of documents or the possibility to comment or categorize entries. The continuous extensions to an existing service can even lead to a service split-up, so that the functionalities of one service, which was initially developed as stand-alone service, are separated into two interdependent services.

The change process can also be triggered through changes in the underlying applications. If an application is upgraded to a new version or replaced by a similar system from a different software provider, the service's access interface to this system needs to be updated.[1] In the best case, only minor changes to the overall appearance of affected services are required, in the worst case, some services need to be entirely

[1] The case study in Chapter 3 presents such a scenario in Section 3.1.2.

redeveloped. Assuming the example of an order administration service and also assuming that the underlying order management software is replaced by a different solution, which provides more features than the previous one: the new features then need to be integrated into the portal service's interface and business logic. E.g. if the new software provides more data fields that have to be filled for each order, these need to be integrated into the service interface.

#### Changes to the Service Portfolio

Changes in underlying systems might also cause the addition of new services and the removal of obsolete ones. The new ones have to be announced to portal users and integrated into the portal's navigation structure; the old ones have to be removed.

Similar such changes occur based on management decision to extend the range of a portal or to follow different priorities. For example, it could be decided that a portal that supports processes of one department should be extended to support processes involving other departments too. This decision would not only affect the number of offered services but also the user base.

### 2.1.2. Portal User Base

Changes in a portal's user base as well as the characteristics of the different types of users also strongly influence the continuous portal engineering process.

#### Changes to the User Base

Similar to the evolution of a portal's service portfolio, its user base evolves over time as (i) new users are added, (ii) new users groups(with corresponding business requirements and permission rights evolve and (iii) existing users quit.

Each change in the user base can influence the requirements the portal needs to address and might also shift priorities. Also the perception of the portal within a company might be affected. Important factors to handle those cases are the general design of services, the structure of the portal and portal documentation.

**Addition of Users:** The portal will gain new individual users in several situations. *New employees* are a common scenario in which users are added to the portal who are entirely unfamiliar not only with the portal but also with the processes within the company. Being introduced to their new job-responsibilities, they will enter the portal for the first time and need to identify the relevant service set for their work. This familiarization process can be challenging since these users are not only required to learn and acquire experiences with their new business tasks but also have to familiarize themselves with the portal's functionality. First of all, they have to understand how their services work in context of the underlying business processes. Secondly, they need to get used to nomenclature and usage patterns within the portal. Often within a company a large set of abbreviations and special terms exist that are consequently used in the portal too. Also the interface design of the services (e.g. grouping of buttons or fields) will take some time of getting used to.

Within larger companies the group of *random visitors* is another source for new users. By following links from other (intranet) websites or because they have seen a service "in action" at someone else's computer, these users enter the portal and, in case they find services that can support them in their work, become regular users. Quite similar are new users who visit the portal *by active recommendation* from colleagues. These two types of new users can be grouped as *unplanned users*, who are not in the group of people the portal's services where originally implemented for. However, in evolving portals features quite likely exist that can also facilitate the work of users who execute different business processes than those the service was originally designed to support.[2] It is disputable whether to consider this group of users when designing the portal. One the one hand, the focus of the development should target the users the particular service is originally designed for. On the other hand, such services could also add value to other functionally not considered processes of other users and would thus contribute to the company's overall performance at no additionally effort. In larger companies with hundreds if not thousands of portal users it is often not possible to exactly determine the group of target users a service is devel-

[2] E.g., a business process might require users to look up information from a list. This feature will be implemented as a service. The look-up service might also facilitate other user's work in case they require the same information - just for another process.

oped for. Therefore, it should be carefully explored whether the design of the portal should incorporate the possibility of unexpected user groups.

**New User Groups:** Entire groups of new users are added to the portal when new processes are mapped through portal services or, in case of the integration of new subsidiaries, into a corporation. The effort to familiarize the new users with the portal is similar to the addition of single users. If these new users are in contact with others who already use the portal (e.g. from the same department) the transition will be easier as they can get informal support with general questions concerning the handling of the portal. In the other case, for example, if new services add an entire department to the portal's user base, some effort might be required to get these new users trained to work with the portal. Basically, dependent on the processes that are moved to the portal, they will have to adapt their daily working patterns and include the portal into their set of tools. This transition process might take some time and training.

The addition of users from departments who did not use the portal before can pose further challenges. Often departments differ in their working patterns or the nomenclature (terms and abbreviations) they use. New users will need to understand these abbreviations or adapt to the other working patterns or the services need to be upgraded to support the different patterns.

**Leaving Users:** Some users will leave the company or change positions so that they will not use the portal any more or at least have a very different focus on it. Portal management has to ensure that their user accounts are removed or their access rights updated respectively (in case they change their position within the company but still use the portal). As soon as business processes change the portal might as well lose some users. This can occur when the new process does not require some of the previously used portal services or even the portal at all.

### User Characteristics

Similar to services, users can be quite different with regard to their behavior and requirements. The differences are most obvious in the general acceptance of web technology, the familiarity with it and the confidence in handling portal services.

On a rather different level users are distinguishable by their background. Some users feel more comfortable with using web-based systems than others. Those who frequently visit websites are more familiar with the general concept and less hesitant in using it than others. User attitude towards portal also depends on previous experience, e.g. with the processes they use the portal for. Long-time employees might be required to change their working patterns when the portal system is introduced. They have to relearn processes in the portal that they were used to execute for years with a certain working patterns. Whether this learning process is easier or harder if compared to new employees, who have to understand the process as well as the portal's functionalities in general, strongly depends on the aforementioned affinity to web technologies. Already knowing the underlying business process can facility the learning process because the basic intention is clear but it might also lead to confusion if a user is not used to using web portal technologies. Another important differentiation is whether a user trusts the portal in a sense that provided services are showing reliable information and are really supporting the business.

Although such attitude based differentiation can further result in very different usage patterns and acceptance levels of a portal, some user differences have its roots in the background of user groups. Users with differing educational background can act quite differently and thus have differing demands on the design of processes within a portal. Such differences often occur between different departments within a company. The legal department, for example, usually consists of lawyers whose education is focused on the exact description and formulation of facts. Those users typically demand services that are extremely accurate and well explained. On the other hand, engineers, whose studies are rather focused on "doing" will be annoyed by overly documented interfaces and oftentimes prefer a "click and try" approach for exploring the portal functionality and for developing trust in a service.

Such rather drastic examples show the heterogeneity that exists especially between users from different departments. Although each group is forced to accept the portal the way it is (there is usually no alternative), considering each user's requirements will lead to a better individual acceptance of the portal and thus a higher motivation to use the portal.

A further distinction can be made based on the tasks each user requires the portal for. This distinction is closely related to the origin of the user within the company, e.g. whether a user is manager, accountant or secretary and from which different department or subsidiary he comes from. For example, managers are likely to use the portal services to retrieve data such as reports or meeting minutes and will rarely execute tasks that involve entering data. The latter is a more likely task for accountants and similar users. Thus, these groups are accustomed to use the portal for very different types of actions. Accordingly they will perceive the portal very differently and have different demands. A manager might ask for more complex reports while accountants prioritize features that facilitate data entry.

### 2.1.3. Overall Portal Evolution

Changes to services and to the user base as well as the heterogeneity of both are strong influence factors in the evolution process of a corporate portal. In addition, more factors, which are discussed in the following, have to be considered. They are related to several of the previously discussed aspects.

#### Portal Structure

The heterogeneity of services and users puts a high demand on the structure of the portal. Each group of users should be able to access its required services in one (logical) place.[3] The more different groups and services a portal has, the more complex this task becomes especially if some services are shared across several groups.

[3]Services can be accessed through the portal menu. Similar services are usually grouped in one branch of the menu. Switching between services in the same branch is usually more intuitive and faster than between branches. Especially in large menu structures with multiple levels of subbranches switching requires knowledge of the structure.

The task becomes even more complex with the extension of a portal's service portfolio (cf. Section 2.1.1). The initial navigation structure of a portal will reflect the set of services that was available at that time. Similar services will be grouped together in logical groups of acceptable size. Additions to the portal will not occur similarly distributed for each group of services. It is rather likely that some service groups will be strongly affected while others are not at all. Thus, the initial service grouping and structure will become imbalanced and needs to be updated to reflect the new service portfolio.

### 2.1.4. Primary and Secondary Services

Related to the task of managing navigational structure and portal growth is the challenge of matching each user with all services relevant to them. The purpose of a corporate portal is to support each user in its daily work (cf. Section 1.3). Each employee who uses the portal will have one or more services that he is required to use. These services are his main motivation for accessing the portal. However, especially in a large portal that offers a wide range of services, the chance is that there are several services that have the potential to support tasks of an employee other than these he would typically require when asked for the first time. These services might not be essential for his work, but using them would facility or improve it. Back in the example of an order administration process, an example would be a service that allows users the automatic retrieval of reports from the portal based on a schedule: Users not knowing this service have to retrieve the reports manually and thus experience some overhead. Knowing the scheduling service would facilitate their work. Another example is common to services providing simple sets of data such as deadlines, customer lists, exchange rates, etc. Often this information is also distributed via email or available on shared network drives. Users who use this information might often not know that they could get the same data through a portal service and instead use the (potentially outdated) local version. They will only learn about changes to the data in case of errors.

The above problem can be classified by distinguishing between *primary services* that users have to use and *secondary services* that have the potential to further support a user. A common reason for a user's unawareness of secondary services goes back to the evolution of portals. Each service is developed with the goal of supporting a certain process or group of users. These users will be informed about the service. However, often the new service could also benefit other users. The above example of services that provide simple data sets demonstrate this situation: several departments use a simple list, e.g., of customers which is accessible through an arbitrary and stand alone customer management software. One department decides to develop a service for the portal that provides the list but further enhances it with other information such as the customer's order history. If the other departments are not informed about this new service, they will continue to use the customer management software and thus will not have the additional information that is provided by the new service.

A corporate portal fulfills its purpose if every user knows and uses his set of *primary services*. However, the full potential is only unleashed, if users learn about the *secondary services* which will facilitate their work and which might even improve the resulting quality or their work performance.

## 2.2. Overcoming the Main User Challenges

The presented factors that influence the continuous engineering process stand for a wide range of challenges. The resulting goals for the engineering process can be grouped by two main *missions*:

- find relevant services,
- provide a high usability.

The first mission comprises all tasks that are related to enabling users to learn about all relevant services of the portal, primary and secondary (cf. Section 2.1.4). The second mission is to design and organize these services in way that allows for efficient usage. In the following, several key tasks are identified for each mission. Each task is

briefly introduced in the following. A discussion of concepts to accomplish the tasks is provided in Section 2.4.

### 2.2.1. Help Users find the Relevant Services

The easiest case to match a user with all services that have the potential to support his work, is a situation where he actively *searches* for them. This is likely to be the case in business information portals. Users access such portals to retrieve information. Even if they know where to find the required data, they will be aware that there is other information available in the portal and will actively look for it. The focus of this work is not on the semantic aspect of finding information but on finding other relevant services for users. If users actively search for information they might also search for other services. And even if they do not explicitly intend to search for a service, the results of a search for information can be extended with results pointing to useful services too.

An approach to actively push information to users is to *recommend* services similar to recommender systems in online shops (e.g. Amazon's book recommender system) or movie databases (Basu et al., 1998) that offer related products. For this purpose recommendations need to be automatically computed or manually entered. By actively recommending services, the user is guided through the portal to all relevant services. However, from a portal operator point of view, it requires knowledge about possibly relevant services for different users and groups, which can be difficult to obtain.

A more direct concept to familiarize users with the portal's service is to actively *train* users. Organizing workshops or personal introductions to the portal is a direct way of informing users about the entire service portfolio. However, it requires significant effort from both, users and operators to join the training.

An alternative to active training is to *document* the portal and its services and provide the documentation to the users. Such documentation can provide a brief overview of the overall functionalities that the portal supports with descriptions for the key functionalities of each service. Although the documentation can be pushed to each user, it requires them to read it.

Finally, each user has to be able to easily *navigate* through the portal. Regardless whether he is trying to access a service he learned about before, or because he wants to discover the portal on its own or only switch between services he already knows: Providing the possibility to easily find and access services with a good menu structure or quick links will help users to better use a portal.

### 2.2.2. Ensure Good Usability

The second mission is to ensure a good usability of the portal. This requirement is based on potential problems due to heterogeneous requirements by users, the possibly wide range of different types of services and the potentially high complexity of some of the services (cf. Section 2.1).

The key task for a high usability of services and overall portal functions is the *design* process. Using coherent design elements and extensive usability testing of interfaces is in the core of this process.

A closely related task is to coherently *structure* services themselves and within the portal. This task is important to help users quickly understand what functionalities are offered and also how a process works.

In addition, by *documenting* each service in detail, users can be enabled to learn more about its functionalities. An important aspect here is the way the documentation is made available. The best documentation is worthless if users do not know that it exists.

By *training* users in the usage of certain services the best results can be achieved. Training will enable them to even understand highly complex services. It should, however, not be seen as a substitute to the effort of designing services to be as easy to use as possible. Trainings only reach a small number of users. Considering the discussion about secondary services shows that users might exist who benefit from a service although it was not implemented for them. Thus, relying on training for identified users as a replacement for easy-to-use service design potentially alienates a large group of unidentified users.

Table 2.1 summarizes the missions and the related tasks. It shows that the first mission consists of tasks users can by accomplish by active participation (pull) and those that push information to users. The second mission, to provide a high usability, only consists of push tasks.

| *Mission* | *Task* | *Information* |
|---|---|---|
| *Find the right service* | - search | pull |
| | - recommend | push |
| | - document | pull |
| | - train | push |
| | - navigate | push & pull |
| *Provide a high usability* | - design | push |
| | - structure | push |
| | - document | push |
| | - train | push |

**Table 2.1.:** Missions and tasks for the portal engineering process.

### 2.2.3. Relevant Tasks for each Portal Type

The relevance of each of the previously introduced tasks depends on the type of portal it is being applied for. Each portal is used differently and thus some tasks may be less effective for one type than for another.

An important difference between application integration portals, on the one side, and business intelligence and employee portals, on the other side, is the expected usage frequency. Application integration portals support users in their regular tasks. Therefore, they are likely to be used regularly and users exhibit coherent usage patterns. They will access these portal knowing what they want to do and most likely which service they need to use in order to accomplish this goal.

Although business intelligence portals can also provide support for regular tasks they are more likely to be used less regularly or at least not with constant usage patterns in such a portal. Information and collaboration features are merely accessed on demand and not due to repeating job requirements as in application integration portals.

Employee portals provide services that are rather used infrequently and on demand, depending on the range of services they offer. A service that publishes internal company news might be accessed daily, but others such as HR-services more seldom.

Based in these assumptions, the relevance of each task for the different portal types can be deduced. Table 2.2 presents a model that distinguishes the relevance of a task for each portal on three relevance levels: high, medium and low. The relevance values have to be seen in comparison with the other tasks for each portal type. A task assigned with a medium relevance does not necessarily mean that this task would not benefit the missions, the other tasks, however, are more likely to better contribute in comparison.

| | *Business Intelligence Portal* | *Application Integration Portal* | *Employee Portal* |
|---|---|---|---|
| *Search* | ● | ○ | ● |
| *Recommend* | ● | ● | ● |
| *Document* | ◐ | ● | ○ |
| *Train* | ◐ | ● | ○ |
| *Structure* | ● | ◐ | ● |
| *Navigate* | ● | ◐ | ● |
| *Design* | ◐ | ● | ◐ |

●: high relevance, ◐: medium relevance, ○: low relevance

**Table 2.2.:** Relevance of tasks per portal type.

Supporting user search is more suitable for business intelligence and employee portals since users are more likely to use such features as previously described. Recommending services can have a good effect for all types of portals. It does not require any effort from users and also does not require any training in order to use recommendations. Although documentation has the potential to provide very detailed information, it requires users to read it. The more complex the offered features of the portal are, the more likely users might read the documentation. Therefore, this task is better suited for business applications portals that offer such complex services. Similar considerations apply for trainings, which are only required to explain complex

services or provide an overview of larger portals. Also, a good design is more important to encourage usage of complex services than it is for usage of comparably simple ones because the latter can still be used even if the interface design is not of a high quality. A good portal structure and good navigation options will have more effect for portals in which users probably search for services but are less relevant for application integration portals.

## 2.3. Methods for Analyzing Corporate Portals

For the decision on which tasks to prioritize in the portal engineering process, a portal needs to be analyzed to gain knowledge about the users of the portal and their usage patterns. The core process is the analysis of the usage logs with web usage mining concepts. Furthermore, users can actively be integrated in the process by explicitly requesting their feedback.

### 2.3.1. Web Usage Mining

The concept of *web usage mining* summarizes the methodology for analyzing visits to web sites. Clusters of pages that are jointly accessed during a user session as well as clusters of users who have similar visiting patterns can be identified with these methods. Especially within enterprises the question of user privacy plays an important role in this context.

Web usage mining is a sub-category of general web mining methods (Kosala and Blockeel, 2000). The concept is based on the analysis of the log data which is collected for each action that a user performs on a web site. Whenever he enters a web site or clicks on a button on this site, the web server stores this page request (*hit*) in a raw data log. This log usually contains information about the requested site, the user's Internet Protocol (IP) address, the date and time of the request and further Meta data. The analysis process of this log data is divided into three phases (Srivastava et al., 2000): preprocessing, pattern discovery and pattern analysis as shown in Figure 2.1.

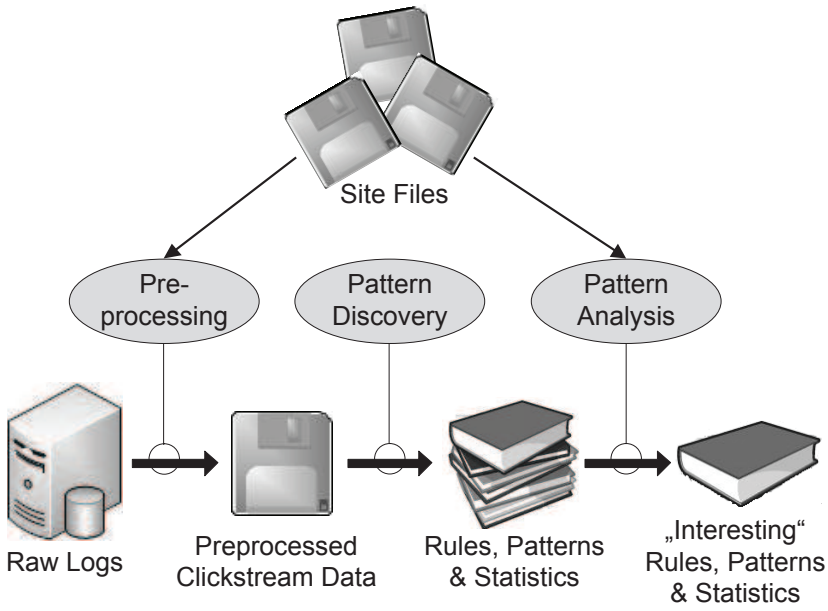


**Figure 2.1.:** High level web usage mining process (Srivastava et al., 2000).

During the preprocessing phase the hits of each user are aggregated to *visits* that show how long a user stayed on a page. For example, if a user enters a page and clicks three times to execute an action on the same page, the resulting four hits (initial hit plus three clicks) are aggregated to one page visit. Based on the collected visit data *usage paths*, clusters of similar visits or users can be determined during the pattern discovery phase. Usage paths show how a user navigated between pages during a visit in the portal. By grouping such paths, popular navigation patterns can be detected (Spiliopoulou and Faulstich, 1999). Comparably, users with similar visit patterns can be grouped together (cf. Section 2.3.1). Further simple statistics such as most popular pages, average duration of page or portal visits and many other are determined during this phase (Cooley et al., 1999). Pattern analysis is the final step of the entire process. The objective is to filter out uninteresting data and patterns from the set found in the previous phase. Here, analyses can be run that focus on key indicators. For example, usage paths can be highlighted that contain certain pages or sequences.

More advanced methods not only identify visits to pages but also track each user's actions on a page (Shahabi et al., 1997) even to a detail level that shows the user's mouse movements (Card et al., 2001; Arroyo et al., 2006).

In the corporate portal context such analyses allow the identification of service usage in general. With the help of advanced analysis methods even field selections in interfaces (e.g. report types, currencies, etc.) could be analyzed to reveal further information about feature usage.

For the interpretation of service usage in the corporate portal context some specific issues have to be taken into consideration that relate to access rights and target user groups. When comparing visits between different services, the number of users who have access to each service must be taken into account. Also, it is questionable whether the number of pages is a relevant measuring parameter. Some services might just be tailored for a small group of users or only for special events or time periods.[4] Because of this, the simple comparison of visit numbers is not the key result that web usage mining delivers. It is rather a basis for the identification of user and service clusters.

#### User Clustering

Based on the web usage mining results similar users can be identified and grouped together in several profile types. These profiles can be solely based on visit patterns (Mobasher, 1999; Shahabi et al., 1997) but can also include advanced user data (Mobasher et al., 2002; Srour et al., 2007). Advanced user data might include information about the user's company (e.g. in CoFiPot the subsidiary a user is working in), his career level within the company, preferences for searches or reports and all kinds of additional personal information. Determining user profiles opens the door to several ways of providing a better portal experience to the user but also raises questions concerning the user's privacy (cf. Section 2.3.1).

[4]The portal presented as case study in Chapter 3, for example, offers some services that are only relevant for the generation of the annual report and thus are only in use for a short period of time towards the end of each year.

Even simple data analysis can help to better structure, for example, the portal navigation to better serve the most common user groups. More complex analysis scenarios include the generation of customized portals for each profile group or personalized navigation and interfaces as described later in Section 2.4.5.

Adomavicius and Tuzhilin (2001) divide user data in *factual* and *transactional*. Factual data describes who the user is and transactional data describes what he does. They propose to combine both data sets to improve the quality of user information in order to ultimately offer completely personalized portals. This approach is very useful to because it combines static data (profile information) with dynamically generated data which can enhance the knowledge about the user.

**Privacy Considerations**

Privacy of user profiles and log data is a very important issue that has recently attracted a lot of attention due to abuse of such data sets within companies.[5] Therefore, clear privacy rules have to be defined that comply with national laws (Aneja et al., 2000). Users in general are extremely averse against any data collection (Srivastava et al., 2000; Eirinaki and Vazirgiannis, 2003). The best way to cope with this issue is to inform them about what data is collected and how it is used (Kobsa, 2007). Also, each portal should maintain a written privacy policy that explains what kind of data is collected and for which purposes.

Kobsa recommends to restrict the collection of data to the absolute necessary attributes and to store the data only as long as needed. Also, mechanisms should be implemented that evaluate and remove personal data if not needed or not in compliance with the privacy rules (Cingil et al., 2000; Kobsa, 2007).

The trade-off with privacy restrictions is that advanced usage analyzes and personalization tools depend on personal data. Portal personalization is barely possible without having individual information about a user. Also the quality of recommendations increases with the detail level of the available usage data. However, before

[5] Recently, at Deutsche Telekom and Deutsche Bahn the abuse of employee data has lead to public attention and started a discussion about stricter privacy rules for employees.

the implementation of new services that might need data, which impacts personal privacy, alternatives with less impact should be considered (Kobsa, 2007).

An important aspect within companies is the pro-active involvement of the staff council before introducing new services that use personal data or ahead of changing the privacy policy of a portal.

### 2.3.2. User Feedback

Interviews or surveys with portal users help collecting additional information about their experiences and satisfaction with the portal. While interviews allows reacting to each user's answers and enquiring follow-up information, a survey reaches more users with less effort but also with less information about individual user experiences

A useful model to design a survey is the *Technology Acceptance Model (TAM)* originally developed by Davis (1989). It measures the perceived usefulness and perceived ease of use of IT systems. The *perceived usefulness* is defined by Davis as *"the degree to which a person believes that using a particular system would enhance his or her job performance."*. *Perceived ease of use*, in contrast, is taken as indicator for the degree *"to which a person believes that using a particular system would be free of effort"*. TAM has been widely accepted (David and Venkatesh, 1996; Doll et al., 1998; Sharp, 2007) and has been used in many studies to measure the usability of websites especially in the area of knowledge management tools (Lederer et al., 2000; Mathieson et al., 2001; Money and Turner, 2004; Abdinnour-Helm et al., 2005). Some extensions exist that have been adapted to changed requirements or single aspects of the model. For example, TAM2 by Venkatesh and Davis (2000) emphasizes the role of the perceived usefulness with regard to mandatory versus voluntary usage.

TAM can serve as basis for designing the questionnaire for a survey and also for structuring an interview. An approach to apply the model to measure the usability of a corporate portal or its features is e.g. described by Vo (2007).

Interviews with portal users deliver a very high data quality. The possibility to immediately react to the interviewed user's demands and concerns allows further discovering previously unknown problems and questions. As a positive side-effect,

conducting interviews furthermore integrates users into the portal engineering process. By reacting to their wishes, they get the feeling that they can contribute to the portal development process and might even serve as advocates for new features.

Compared to surveys, the effort of conducting interviews is higher with regard to the number of users who can be questioned. On the other hand, short interviews with a few users can be organized faster since interviews do not require a fully structured questionnaire. Also, the number of surveys should be limited in order not to annoy users with too many requests. Interviews, in contrast can be conducted more often by selecting different users.

## 2.4. Concepts to Support Portal Engineering

This section describes several different concept that have the potential to support the ongoing engineering process of corporate portals and that can help fulfilling the missions and related tasks described in Section 2.2. The main emphasis is put in concepts that support system operation. Concepts that are required or useful for the initial development of a corporate portal or the implementation are not discussed here. Information on these topics can be found in Vo (2007), Hazra (2002) or Gurugé (2003).

The presented concepts are:

- Service and Portal Documentation
- User Trainings
- Internal marketing
- Recommender Systems
- Personalized Portals
- Design and usability reviews

Some of these concepts require the integration of new features into a portal, e.g. a recommender system, while others (such as trainings) are concepts that have to be implemented by portal responsibles or management. Each concept is introduced and analyzed according to the following criteria:

- Which tasks does it support?
- How much initial effort is required to realize the concept?
- How much maintenance effort does it require?
- Which are the key prerequisites for the application of the concept?

### 2.4.1. Service and Portal Documentation

Documentation of services and portal features allows users to learn about the offered functionalities. It has to be distinguished between the documentation of single services and a global documentation. The latter can give users an overview of all offered services, which can be especially helpful for users who visit the portal for the first time. Such documentation could be designed as a *demo-tour*. Similarly a short introduction for each service can help users to identify its core features, thus enabling them to decide whether it is of relevance to them. A detailed description, on the other hand, is helpful for advanced users who want to learn about particular features.

Providing documentation is very helpful and can lower the number of inquiries that are sent to portal support staff. However, the concept is also prone to a high level of inefficiency. The challenge it to maintain an up-to-date set of documents that addresses the requirements of the target group in the right granularity (Polanyi, 1997; Davenport et al., 1998; Hahn and Subramani, 2000; Orlikowski, 2002; Stenmark, 2004).

#### Documentation Technologies

Systems used to organize knowledge have significantly changed over the past years. The most basic form, static documents that are stored on network drives or in knowledge management systems, are still widely in use but prone to inefficiency (Davenport et al., 1998; Hahn and Subramani, 2000; Gold et al., 2001). A general problem with static documents is the process of writing and keeping them up-to-date. On the one hand, access rules have to be established which not only control who is permitted to change documents but also ensure that only one person can edit a document at a time. On the other hand, the process of updating an entire document might appear too much of an effort for performing minor corrections. Such obstacles can be low-

ered with advanced document management systems[6]; however, other technologies have the potential to better serve the purpose of documenting certain processes and systems.

The incentives for users to share their knowledge are not further discussed. In depth descriptions and analysis of this topic are provided by various authors (Davenport and Prusak, 1997; Holtshouse, 1998; Leonard and Sensiper, 1998; Baumard and Baumard, 1999; Lam, 2000; Stenmark, 2004; Davenport, 2005).

Wiki-systems are a very efficient way to document services and processes because they allow each user to easily add and update information. They also enable authors to add cross-references to other documentations. Another advantage with a high relevance for the portal technology is that users can directly access a wiki-system through their browser, i.e. using the same application they use to access the portal itself. The users do not have to download a document in order to access it. Wiki systems allow documentations to mature over time and to adapt to changes (Braun and Schmidt, 2007). The implementation of wiki-systems in company intranets has been widely discussed in general (Leuf and Cunningham, 2001; Cunningham, 2003; Stenmark, 2005; Buffa, 2006; Hasan and Pfaff, 2006) and with special focus on why this type of documentation succeed or fails(Désilets et al., 2005; Chui et al., 2009; Bughin et al., 2009).

A very different kind of documentation is achieved through the provisioning of a forum for portal users. In such a forum they can post questions and discuss them with other users and administrators. Over time, a quasi-documentation of the system is created that is focused on answering common questions about it (Stenmark, 2005). The concept is very close to FAQ-sections and also to wiki systems although less structured as compared to the latter.[7]

[6] An example for such a document management system is Microsoft's SharePoint Portal that allows user to directly download and edit documents within their browser.

[7] In a (maintained) wiki-system each topic is represented by one article whereas in forums one topic might be discussed in several threads.

**Documentation Creation and Maintenance**

The creation and maintenance of portal documentation requires regular attention and effort. It has to be ensured that the content of each description is up to date in case that changes to the system affect the interface or process of a service. Although no constant effort is required, a regular check should be scheduled to fulfill these tasks. Otherwise the usage of the documentation will decline and users will rather contact the portal responsibles in case of questions.

Whatever system is used for the provisioning of documentation, a further challenge is to inform users about its existence (Davenport et al., 1998; Benbya et al., 2004). If the documentation is stored in a system that users do not know they will rather ask all their questions to the portal administrators or follow the informal path of asking colleagues. Creating a system that integrates knowledge in working processes (Stenmark and Lindgren, 2004) will help to prevent such scenarios. In corporate portals it is easily possible to display links to the documentation system and also to show links to relevant documentation directly next to the service they document.

The process of informing users about the existence of documentation is more critical than having the most sophisticated documentation system implemented and running. Therefore, portal administrators should first aim at providing direct links to existing documentation rather than changing the documentation system.

In summary, the key demands for the documentation concept are:

- implement procedures to keep the documentation up-to-date,
- allow users to easily change and add content on their own,
- provide both, short introductions and in-detail descriptions,
- ensure a high visibility and accessibility of the documentation.

### 2.4.2. User Trainings

A very effective but also expensive concept is the training of users in form of workshops or even as part of the introduction process of new employees. In such training

sessions, all parts of a portal can be introduced. The trainings can focus on two aspects: the introduction of users to specific services and an overall introduction of the general services available in the portal.

The realization of trainings requires a lot of effort. On the one hand, a person is required who is able to explain the portal's services, on the other hand, portal users have to take the time to attend such a training session. In larger international companies these preconditions can become very expensive. However, this concept has the highest potential of properly introducing users to the primary and secondary services they need to accomplish their work.

The key demands to trainings are:

- identify all users who can benefit from training,
- determine the requirements of participants to keep trainings to the point,
- include a brief portal overview in each training session.

### 2.4.3. Internal Marketing

A very different concept is to actively promote portal services within the company. Such marketing activities can take place inside and outside the portal.

Marketing within the portal can be rather subtle, for example, by highlighting links to certain services or by moving them high up in the portal navigation structure. More direct forms include displaying messages to promote a service. Such messages can be shown on the start page, constantly in a banner, footer or even as pop-ups similar to advertisements on Internet pages. The difficulty is to decide between the levels of attention that each form can gain and to trade it off against its annoyance factor.

Outside the portal specific services can be advertised in newsletters, personal emails or with links in other (intranet) sites of a company. Also, the mentioning of services in the documentation of business processes (cf. Section 2.4.1) can be seen as marketing measure. By providing a link to a service where a user can get the data as described in the documentation, the reader might be motivated to visit the portal. Similarly each report or other document that is downloaded from the portal could in-

clude short information about the service that was used to produce this data as well as link to its origin.

The difficulty with the marketing concept lies (i) in the decision, which services to promote and (ii) in the possibly annoying effect of too many messages. The first problem especially occurs in portals that serve multiple users from different areas of work. Services should only be promoted if they are possibly relevant to the majority of these users or the marketing measures need to focus in specific user groups. The second issue is similar to problems with SPAM mails. If too many messages appear, users will simply ignore them. Therefore, such measures have to be clearly limited.

The effort for internal marketing is relatively low. It has to be decided which services to promote and how to transmit the message to users.

When implementing the concept of internal marketing, the following suggestions should be considered:

- find a good balance between visible messages and annoying messages,
- ensure the relevance of each message for the majority the addressed users,
- limit the target group to possibly relevant users.

### 2.4.4. Recommender Systems

In contrast to the previously described concepts, a recommender system requires less input from users or portal operators but much higher realization efforts. Based on usage data or by comparing user profiles, a recommender system is able to automatically compute personalized recommendations for services. By analyzing usage patterns not only each user's preferred services can be identified but also users who have overlapping or identical usage patterns can be grouped together. By comparing users within a group and by focusing on the differences in their preferences, possible usage improvements can be determined. Figure 2.2 shows an example. Users A, B and C all use services 1 and 2. Users A and B additionally use service 3. Since A, B and C have a similar usage pattern, it can be assumed that service 3 might be of interest to user C. A recommendation could be made to this user suggesting to him

that he might want to test the service. Service 4 might also be of relevance to user C but since only user B is using it, the relevance is most likely lower (it can be also of relevance for user A, even more likely than to user C, because A and B have more similar usage patterns).

A well-known example for such recommendations is the recommender system in Amazon's shopping portal. Whenever a user looks at a product, the system will recommend him a set of products that other users who also bought this product have previously bought. These recommendations are shown with an image of each product and a short description.

Within a corporate portal a recommendation system could be used to recommend services, which might be of interest for a user, based on similarities in portal usage patterns with other users. Such recommendations can be communicated in different ways. They can be shown within the portal as additional links in the menu, in the banner, footer or anywhere else on a page. More obtrusive would be to randomly show pop-up windows. Other forms of communication might include personal emails containing all recommendations with a short explanation.

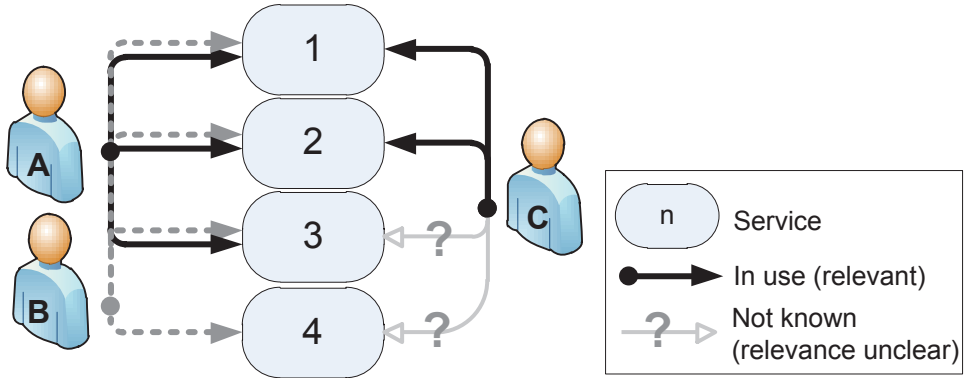


**Figure 2.2.:** Usage-based recommendations. Based on similar usage patterns of users A and B, services 3 and 4 could be recommended to user C who does not know the two services yet.

Implementing a recommender system requires a higher initial effort for the implementation but little maintenance effort. The system can automatically refresh its recommendations and thus also automatically react to changes to the service portfo-

lio. A big advantage is that it does not require any knowledge about the portal and its structure since it learns from user behavior.

Almost no literature exists on the application of a recommender system to recommend services in a corporate portal. Zanker and Gordea (2006) discuss the benefits of implementing a recommender system in a knowledge portal. Raghavan (2002) vaguely proposes the application of recommender systems to improve knowledge discovery in enterprise portals. However, no case studies and in detail discussions are available. This gap is addressed in this thesis.

The key demands to a corporate portal recommender system are:

- to inform users about the portal's service portfolio,
- to identify and suggest the most relevant services for each user,
- to require little to no input from users,
- to run automatically without requiring constant attention by portal management,
- to cope with heterogeneous users and services.

### 2.4.5. Portal Personalization

The provision of personalization features is a concept to better support users in navigating the portal and to structure it so that it best fits each user's requirement. *Personalization* features can be defined manually but can also be based on explicit user profile data (Mobasher et al., 1999; Adomavicius and Tuzhilin, 2001) or on implicit information such as usage statistics (Mobasher et al., 2000).

Several degrees of personalization are possible which range from feature based personalization to global settings. A relevant aspect in the context of corporate portals is the personalization of navigation. A very basic approach is showing a list of quick-links containing each user's most actively used services. Through quick-links, users are able to switch between their favorite services without having to navigate through the menu. A more advanced approach is the personalization of the entire portal structure (Mobasher et al., 2000). This personalization does not necessarily

have to be provided for each user but rather for user groups with similar interests. E.g. if a portal is used by several departments, a personalized portal could be offered for each of them. Each customized portal view would be structured such that it best supports the individual department's needs.

Disadvantages of personalization are the high effort which is required to implement such features but also possible confusion among the users. Moreover, using explicit user data as basis for personalization would require this information to be collected in the first place. When offering multiple personalized portals with the same set of services possible exchange between users from these portals has to be considered. If a user recommends another to have a look at a certain service, confusion will arise when the other user sees a different navigation structure due to personalized portals. However, these disadvantages are mainly based on communication matters. The key question is the evaluation of the benefits that personalization can bring compared to the required effort.

Personalization features should meet the following demands:

- require little to no input from users (e.g. profile information),
- keep personalization at a normal level (no over-specialization),
- inform users about personalization features.

### 2.4.6. Design and Usability Reviews

Design and usability engineering are usually considered as tasks of the initial development phase rather than being seen as part of the continuous engineering process. Authors such as Nielsen (1999), Abdinnour-Helm et al. (2005) or Shneiderman (1998) provide very resourceful descriptions of this process. In the context of this thesis, design and usability are listed with regard to the nature of corporate portal services and their user base. It has been described that corporate portals can provide different types of services to very heterogeneous users with different requirements and backgrounds. Taking into account that both, the service portfolio and the user base, are

subject to changes, e.g. the implementation of a new set of services, the usage patterns and the requirements to the portal's design might change.

If, for example, the initial version of a portal was designed for a specialized group with few users, it is likely that no major effort was taken to ensure a good interface design and high usability. The reason is that users are specialized in this process and also can easily be trained due to the small size of the group. With the addition of new services and new users, the services might need to be redesigned to be usable for other users who are less specialized than the initial user base. Whenever the design is changed to provide a better usability, all users of the service should be informed so that they can adapt to the changes. Although a higher usability is expected with the new design, they will have to change their usage patterns.

In summary, this concept is more relevant for a corporate portal that offers complex services such as the application integration type. The key demands are:

- review design and usability with focus on unspecialized untrained users,
- communicate design changes to users.

### 2.4.7. Matching Concepts and Tasks by Portal Types

Each of the presented concepts has the potential to support certain tasks. E.g. documentation is, naturally, assigned to the task of documenting the portal but also helps users to search for services or features and further can be used to train users in the usage of services. By adding links to related services into a service description, it can also help in the task of recommending. Table 2.3 shows the relationships between tasks and concepts.

Except for documentation, each concept is closely related to one task. E.g. a recommender system is good to recommend services but has no major benefit for the other tasks. Offering trainings can also be seen as a recommendation because training sessions can be used to teach users about other services.

The decision of which concept to follow also depends on the type of portal and its evolutionary stage (cf. Section 1.1.3). Table 2.4 shows a matrix for each portal type and concept with a short estimate with the relevance for each combination.

| | *Search* | *Recommend* | *Document* | *Train* | *Navigate* | *Design* |
|---|---|---|---|---|---|---|
| *Documentation* | ◐ | ◐ | ● | ● | ○ | ○ |
| *Trainings* | ○ | ◐ | ○ | ● | ○ | ○ |
| *Internal Marketing* | ○ | ● | ○ | ◐ | ○ | ○ |
| *Recommender System* | ○ | ● | ○ | ○ | ○ | ○ |
| *Personalization* | ◐ | ○ | ○ | ○ | ● | ◐ |
| *Design & Usability* | ○ | ○ | ○ | ○ | ◐ | ● |

●: high relevance, ◐: medium relevance, ○: low relevance

**Table 2.3.:** Relationship between tasks in portal in engineering and related concepts.

Although documentation appears as good concept for all types of portals regardless of the evolutionary stage, it is more useful in case of application integration portals where complex services might lead to more questions as compared to an employee portal. Still, the more a portal grows, the more helpful documentation concepts such as quick demo tours through the portal become.

Similarly, trainings are helpful to explain complex services in an application integration portal but are not required for an employee portal. For business integration portals that offer collaboration features, they can be helpful to enable users in better using their individual service portfolio but are not necessarily required to explain general tasks such as uploading a document into the portal through a corresponding web form.

Internal marketing and recommender systems have comparable effects and can be applied for all types of portals. Their relevance is higher if users do not search the portal for other services.[8] Both concepts, however, are only useful when the portal is providing a large amount of services so that users can no longer oversee the entire portfolio by looking at the navigation bar.

[8]Internal marketing is also useful even if users regularly search in the portal. However, it will have comparably more effects in portals where users do not search.

| | *Business Intelligence Portal* | *Application integration portal* | *Employee portal* |
|---|---|---|---|
| *Documentation* | Useful for specialized portals with advanced services. | Useful and required for all types. | Only useful if complex tools are offered. |
| *Trainings* | Good to explain collaboration tools and complex services. | Especially helpful to teach usage of complex services. | Not required. |
| *Internal Marketing* | Useful for advanced portals. | | |
| *Recommender System* | Useful for advanced portals. | | |
| *Personalization* | Good to provide specialized portals for user groups in large portals. | | Not required. |
| *Design & Usability Reviews* | Useful in case of larger changes. | Important to cope with portal growth. | Useful but not important. |

**Table 2.4.:** Support concepts analyzed for the applicability for different portal types.

Design and usability engineering in terms of re-evaluating the implemented services is relevant for portals at a higher evolutionary stage (cf. Section 2.4.6). Similar, personalization concepts should only be applied when the portal has reached a certain complexity level.

## 2.5. Chapter Summary

Constantly evolving corporate portals pose several challenges for the accompanying engineering process. These challenges are identified and discussed in this chapter. The focus in this chapter were tasks and challenges that occur during portal operation. Challenges related to the development process are well discussed in related software engineering literature. The identified challenges have not yet been discussed in scientific literature about corporate portals although the complexity of such portals has already been identified by many authors.

The challenges were derived by looking at different factors that result from portal evolution. An import aspect is the complexity of corporate portals with heterogeneous user groups and service offerings as well as the user's knowledge about the offered services. Two missions are specified that help determining tasks that portal responsibles should execute to meet the described challenges: (i) to support users in finding relevant services and (ii) to ensure a good usability of the offered services.

The subsequently presented concepts to fulfill these missions cover a wide area of procedures and tools. Concepts such as documentation and user trainings require significant organizational effort and strongly depend on the participation of users while others concepts such as recommender systems push information to users without requiring any additionally effort from their side.

The decision to utilize a certain concept is case specific and depends not only on portal type and evolutionary stage but also on the environment the portal is embedded in and on other concepts that are already in place. This chapter was structured in a way that it can serve as a handbook for analyzing an existing portal for identifying current challenges and for selecting appropriate concepts that address the identified challenges.

# Chapter 3.

# Case Study: The Corporate Financial Portal (CoFiPot)

This chapter presents the Corporate Financial Portal (CoFiPot) which is used by the Finance Department of the Bayer AG for almost ten years. The portal is classified as an application integration portal. The case study allows the analysis of the constant engineering process, which was discussed in the previous chapter. CoFiPot was subject to constant change, leading to today's role as the major tool within the Finance Department.

The evolution of the portal is summarized and followed by an analysis of usage patterns that allows the identification of several growth-related challenges. This evaluation is further strengthened with the results from an expert interview among managers of the Finance Department.

Based on the findings the concept of a recommender system is selected as the most suitable tool for the current stage of the portal's development. The core requirements to such a system are therefore determined to serve as starting point for implementation.

## 3.1. A Portal to Support Bayer Finance Department

The Bayer Group is a German chemical and pharmaceutical company with its headquarter in Leverkusen. It is organized as a holding company with three subgroups

representing the core businesses - HealthCare, CropScience and MaterialScience - and three service companies. These subgroups and service companies operate independently, led by the corporate center in Leverkusen. Together they consist of more than 350 subsidiaries globally. Among the large product portfolio are well-known products such as Aspirin, Alca-Seltzer, many types of paints and agricultural pesticides.

The company employs a workforce of 108,400 employees.[1] In 2009, the Bayer AG generated revenue of €31.168 billion with an operating income (EBIT) of €3.006 billion.[1]

The Bayer Group Holding is globally responsible for the Group's financial and tax management. The principal objective of the Finance Department is to ensure optimal management of financial risk in the Bayer Group. This should ensure that the Bayer Group can implement its strategy and meet its payment obligations. It is responsible for all funding operations and thus steers all procurement and allocation activities centrally. The performance of financial and tax management is continuously monitored against benchmarks. The core activities of the Finance Department are focused on the following value drivers:[2]

- optimizing the capital structure/minimizing the cost of capital,
- minimizing tax liability,
- improving financial cash flow,
- portfolio optimization,
- optimal asset management for the pension fund and
- minimizing the organization costs of the Finance Department.

The Finance Department is structured in six departments which are shown in Figure 3.1. A risk committee, which meets on a monthly basis, has been established to monitor central risk management and define strategies. Therefore, monthly status, risk, and performance reports have to be created. Country and regional F&A (Finance and Administration) managers coordinate local activities with the holding's corporate center.

[1] Annual Report, http://www.annualreport2009.bayer.com/, retrieved March 2010.
[2] Based on Bayer Group Regulation No. 1053, effective July 1, 2007.

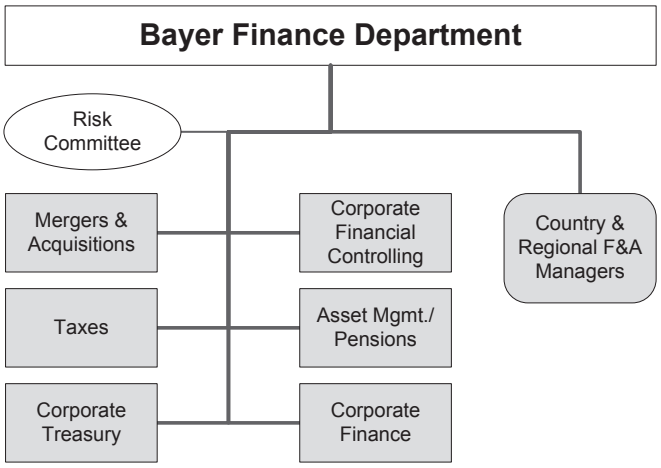


**Figure 3.1.:** Bayer Corporate Finance organizational structure.

Being responsible for the non-operating result and risk management, an important part of the Finance department's work consists of aggregating and processing data from and delivering information to the subsidiaries and the F&A managers. Some of the key data requirements are financial planning data, interest rates, bank account balances, predicted cash flows and commodity information.

### 3.1.1. Implementation and Evolution of CoFiPot

Several applications are running at the corporate center to collect and process all financial information. Further resources that are often linked to the central systems are used at many other locations. In addition to these applications external and internal services as well as internal databases serve as data sources such as internal organizational data, market data, bank account information and many more. To distribute the data and services from the different sources to all subsidiaries worldwide the Corporate Financial Portal was developed and made accessible for all employees in 2001. The primary objective of this portal is to provide a central point of entry for all kinds of financial services for the global financial community of Bayer (Vo, 2007). Figure 3.2 shows the portal's start page. Is position within the system landscape of Bayer as of 2010 is displayed in Figure 3.3 which demonstrates its role as gateway to the various systems and databases as described in Section 1.4.

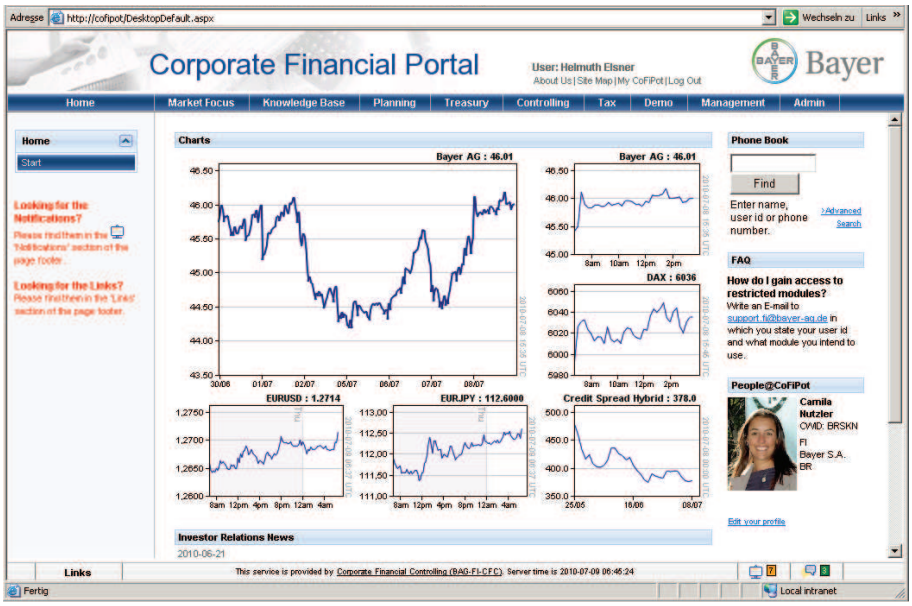


**Figure 3.2.:** CoFiPot start page (June 2010) with market data, random user picture and links to other services.

The operation and development of CoFiPot is one of the responsibilities of the Corporate Financial Controlling (CFC) department which is also responsible for all other finance systems within the corporate center.

Since its start in 2001 CoFiPot has been continuously extended with new services as well as existing services have been updated to meet new requirements. This evolution can be separated into three phases which took partially place in parallel (Elsner and Krämer, 2010b):

- application integration and data collection,
- development of intelligent services and
- extension with assistance services.

### Phase 1: Application Integration and Data Collection

Initially the development of CoFiPot was focused on the task of implementing a central interface to all stakeholders. This system should allow access to the most important financial applications and data sources. Previous to the implementation employees from subsidiaries (especially those from non-European countries) did not

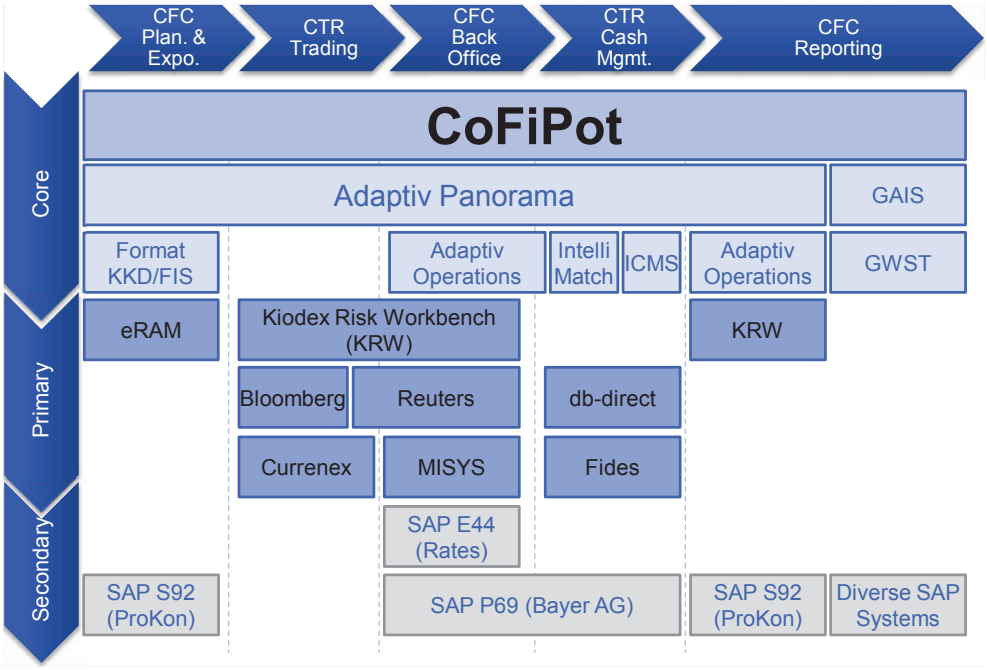


**Figure 3.3.:** Bayer Corporate Finance system landscape (as of 2009).

necessarily have direct access to these and had to rely on the finance department to forward information. Many applications and data sources were simply not known to them. A further goal was to streamline the collection process of financial data from subsidiaries by implementing a central database with interfaces accessible for all subsidiaries. The idea was to replace the process of manual data collection via email exchange. Instead of creating a new application for this task, the corporate portal concept was chosen to implement the required interfaces for those databases - at this time a very new technology.

The key factors for the implementation of a corporate portal solution were fourfold. The technology allowed the development of *optimized interfaces* for the management of financial data. Instead of having to train all employees with the handling of the rather complex interfaces of the underlying financial systems, which allow the execution even of highly complex financial transactions (in the following: trades), simplified interfaces were generated covering the majority of the trade types that were entered by users outside the corporate center (cf. Shneiderman, 1998; Aneja et al., 2000). Thus only few employees remained that had to be trained in the handling

of the underlying applications, most of them situated in Leverkusen. Furthermore the portal prevented the need to *install and maintain software* on the workplaces of all employees who require access to the system (cf. Section 1.2.1).

With the *integration of external service providers* the portal could distribute external data throughout all finance areas. These services included up-to-date market information such as exchange rates and stock prices and also account balances. The data from these services is also included in the generation of simple reports that, for example, combine data from the finance systems with current exchange rates. Several services were created to generate such simple reports and charts. For these, several interfaces between the various systems and CoFiPot were implemented. Thus the portal could access and aggregate all available information and process it to reports.

The fourth factor was the afore mentioned creation of *databases for the aggregation of data* from all subsidiaries. In contrast to the interfaces to existing systems, new processes were implemented that replaced the manual collection of this data. The new process did not only streamline the procedure by abolishing the need for an intermediate in the corporate center, who had to communicate with all subsidiaries. Responsible persons for local data could now not only enter the data by themselves, the process also enabled them to check and adapt their data rather without having to request the latest status from headquarters. The portal technology facilitated these processes to such an extent that new demands for more detailed data emerged.

**Phase 2: Intelligent Services**

The integration of several applications and services leveraged the potential of the portal to combine all sources for the development of intelligent services. These new set of services followed three basic concepts: the *compilation of complex reports* which themselves could serve as new input to other services and processes, the *creation of intelligent services* that are able to execute processes semi- or fully automatically and the *interaction of services* following a business process chain.

*Complex reporting services* leverage the availability of connections to multiple data sources within the company. With Bayer's heterogeneous system landscape within a

constantly changing environment[3] creating reports was not possible without having to manually collect the data first from various systems. A variety of reports, especially for risk management purposes, were implemented in CoFiPot. Before, these reports had to be manually compiled, which required several days. With the new services that provide these reports it is a matter of minutes not requiring any manual input. In addition to the consolidation of similar data sets from various sources[4] the quality of the reports is enhanced by joining further information such as accounting rates for variable report dates. The new simplicity of generating such reports allowed the implementation of subscription services. Users can choose to subscribe to any report and have it delivered to their mailbox at scheduled times (cf. Section 1.2.3).

Beside those reports *intelligent services* were introduced in the second phase. They analyze data and provide recommendations for transaction or even execute such transactions automatically. Such services derive proposals from monitoring the status of transactions or other key indicators. In case of events such as the maturity of a transaction or the input of a new deal they analyze the new state and issue proposals based on predefined rules. Such proposals are either sent to responsible employees or even automatically executed (cf. Section 1.2.3).

An example for such a service is the *auto-hedging service* (Vo et al., 2005). This service analyzes the data of all upcoming foreign exchange (FX) deals of all participating companies within the Bayer Group. Usually all deals in a foreign currency have to be insured against currency fluctuations by closing hedging deals. When such hedging deals are closed with financial institutions or other trading partners, fees have to be paid. These fees can be saved by internally clearing deals in the same currency. Assuming that company A (A is a subsidiary of the Bayer Group) will receive a payment of 200 million USD and company B (B is also a subsidiary of the Bayer Group) has to pay an amount of 150 million USD, these two deals can be internally allocated so that the external risk only consists of the remaining 50 million USD. The Bayer Group

[3] For example, with the takeover of the Schering Group in 2006 several new systems had to be linked to the infrastructure.

[4] An example for such similar data sets that is stored at various sources are bank account balances. They are either entered (through a CoFiPot service) into a central database or directly imported through external services from the banks.

thus saves the fees if both deals had been hedged separately. All subsidiaries are required to report FX deals over a specific amount to Finance (using another service in CoFiPot). The auto-hedging services continuously analyses all available data and suggests internal hedges to the financial accountants. They can review all relevant trade information through the portal and decide whether to settle the internal hedge. In case the decision is taken, they only have to click once to start the trade execution process in the underlying financial application.

The benefit of this service is that it (i) analyzes data from different sources and (ii) directly allows execute the hedges rather than having to enter the trade in a separate application. Vo et al. (2008) show the effects of this service compared to the previous process and conclude that, based on the auto-hedging service, the number of internal hedges has increased to a level which would not have been achievable with the old processes.

*Supporting business processes with multiple services* is the third type of intelligent services that was introduced in the second phase. The concept allows that several services can exchange data to support complex process chains. An example is the integration of services for taxes. Initially stand alone services were implemented that allow to record tax assessments, to perform tax computations and to manage tax declarations. During the second phase those services were interlinked in such a way that tax assessment data could be compared to the original computation and declaration. In contrast to other services that also shared same data sets, the explicit connection between services introduced a higher level of complexity. Two services that share the same data do not necessarily need to know that the other service uses the same data set. However, explicitly linked services are dependent on each other in a manner that one cannot be used without the other. In other words: those services do not only share data sets but also sets of rules.

**Phase 3: Assistance Services**

All the services of the first and second phase are primarily focused on finance processes. Third phase services aim at providing value added functionalities. These services can be split into *knowledge management* and *general support* services.

**Knowledge Management Services:** Parallel to the implementation of CoFiPot a knowledge management system was implemented for Finance. This system was supposed to replace shared network drives and collect documentations, meeting minutes and all other types of documents. Often these documents included manuals for CoFiPot services or underlying processes but also reports that were created in CoFiPot. In contrast to CoFiPot services the usage of this system was not necessarily required for employees to fulfill their tasks. In order to enter a deal, the stakeholder had to use the corresponding service in the portal; however, often there was no knowledge about the documents that explained the service and the underlying process. On the one hand, users contacted support if they had any troubles with the handling of CoFiPot services even though these questions would have been answered by reading the documents in the knowledge management system. On the other hand, the documentations were not updated by the persons responsible for a service since they were contacted with general questions anyway. In the end, this vicious cycle lead to a high number of outdated documents in the knowledge management system as well as many support requests.

The concept for knowledge management services aims at providing the knowledge where it is required. The first goal is to allow referencing items from the knowledge management system to CoFiPot services. Thus a list of relevant documents can be displayed next to a service enabling the users to access those with one click. The next step will be the implementation of smart processes that automatically identify these relationships. By integrating knowledge management services directly into the portal, the incentive for users to read documentations and to keep the content up-to-date will be higher due to the higher visibility of this data. Another benefit can be lower requirements for the administration of documentations. The higher visibility of doc-

uments can cause users to notify management about obsolete pieces of information and thus prevent the knowledge base from containing outdated documents.

**General Support Services:** Several functions were integrated into the portal to provide quick access to common applications but also to make users feel more familiar with CoFiPot. The first group of services consists of several links to other resources within the Bayer intranet, an organigram of all employees in finance and the possibility to search phone numbers in the Bayer phone directory directly from CoFiPot. A rather simple service became quite popular. It randomly displays images of other CoFiPot users on the start page (only of those users who approved (cf. Figure 3.2). This feature even caught the attention of top-level management and lead to a suggestion to other intranet application providers to implement a similar feature.

Although these services provide no direct benefit for Finance they can save users the effort of changing between applications[5] or simply make them feel more familiar with the portal because they get to know the faces of others that use the portal. The popularity of the organigram service indicates the relevance of such services (cf. Section 3.2.3).

### 3.1.2. Major Milestones

The three development phases did not run consecutively but were rather overlapping or in parallel. They can best be viewed as the progress that most of the larger service groups within CoFiPot followed. While smaller services were implemented and upgraded at all times in the ten years since CoFiPot's initial implementation, several larger projects can serve as milestones along the history of the project. Figure 3.4 shows the evolution of CoFiPot and emphasizes the key projects.

**Tax Information System (TaxIS):** In 2006 the Tax Information System (TaxIS) project was started to support tax-processes through services in CoFiPot. The final goal of

[5]Many users have CoFiPot open in their browser at all times in order to follow the development of the Bayer share or to use the direct link to the phone book or the organigram for the identification of incoming calls.

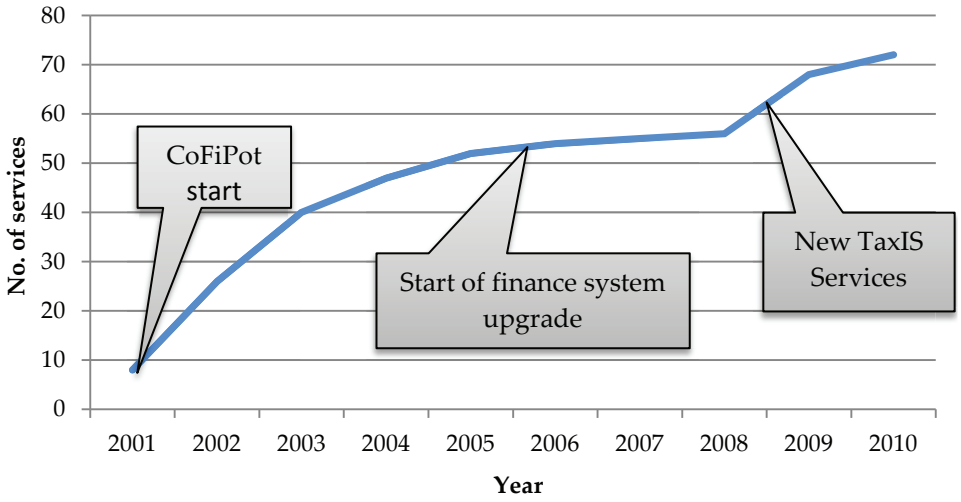


**Figure 3.4.:** Evolution of CoFiPot: Number of services and milestones.

the project was the support of the entire tax compliance process from tax computation over tax declaration management up to tax assessment validation and appeal management. During the project's initial phase two services were implemented to allow the registration of tax assessments and the management of tax audit related issues. In the second phase services for each of the remaining process steps were developed and all services connected as described in Section 3.1.1.

The TaxIS project added an entire new group of six services to the portal's portfolio. It also increased the user base with employees from the tax department who had not used the portal before. Extending the user base to a department that previously did not use portal technology posed some challenges to portal management because the new users were used to different working patterns and had not used any portal services before (cf. Section 2.1.2). Also the type of the support process differed significantly from the other services in the portal. The tax services are *strongly interdependent* while most of the other services were rather *stand-alone* (cf. Section 2.1.1).

**New Finance System:** One of the reasons for the development of CoFiPot was the wish to allow users to connect to the core finance application without local software installations and extensive trainings on the software. Most of the services are using data from this application and often allow editing or entering its data. Starting in 2006, the underlying application was changed to a different application. Since

the original application was highly interwoven with almost every service and process, this undertaking required a complete redesign of the system. The entire process lasted almost two years. Although many services had to be significantly changed, especially the interfaces for users from outside the corporate center could be upgraded with minor changes thus keeping training requirements at a minimum. During the upgrade process the development of major new services was stopped.

**Compliance and Incentive Services:** Some services were implemented that are not direct part of the finance processes. The first was a service that allows entering compliance related data such as gifts or invitations from business partners. Another lately added service shows the development of incentive programs that are based on the performance of Bayer shares. Through this service all authorized users[6] can follow the status of their bonus base on a daily basis.

Both services draw users to the portal that originally would not know about CoFiPot. This development, on the one hand, bears the chance of promoting the portal among those users but, on the other hand, requires special attention in design since those users are not necessarily used to the technology.

These three examples represent only a part of the services which were implemented in the 10 years since CoFiPot's introduction. They emphasize the challenges and effects of a growing portal environment.

### 3.1.3. CoFiPot Characteristics: An Application Integration Portal

The Corporate Financial Portal can serve as an example for an *application integration portal* as described in Section 1.1.2. It demonstrates how various applications and other sources are made accessible through one single gateway, thus saving the effort of distributing the applications to all users and of training them in the handling of different systems. Furthermore, it shows the potential that the connection of a company's various systems at one place has. The portal's reporting services combine the

[6] Authorized are all users above a certain career level within the company.

data from various sources allowing instant in-depth analysis whose creation previously required an effort of several days.

With the latest addition of knowledge management components as described in Section 3.1.1, the portal was enhanced with services that are inherent to *business intelligence portals*. It shows that the different portal types are not exclusive and that constant portal evolution can even lead to a change of its characteristic. However, in case of CoFiPot the focus of the portal remains on the integration of applications while the knowledge management services only serve as assistance services.

The evolutionary stage as specified in Section 1.1.3 of CoFiPot can be defined as *specialized*. The portal is focused on the finance department and offers services for all types of processes of that department.

## 3.2. CoFiPot Key Figures and Usage Analysis

The portal has 5971 registered users and offers 72 services. These numbers and all others in this section are taken as of April 30, 2010. The usage of the portal is constantly monitored. In the first years of operations usage analysis was performed using server log files. These log files did not include any information about the active user except his computer's IP address. Since autumn 2008 the monitoring tools have been extended to include user information such as the roles of the user and some further profile data (of all users who have logged in to the portal). All data however is automatically anonymized so that the real user cannot be derived from this set of data. This section presents key figures and statistics taken from the log files and historic figures about the portal.

### 3.2.1. User Base

In theory every person with access to the Bayer intranet can access CoFiPot. These visitors will only see services which are not protected (cf. Section 3.2.3). In order to provide access to protected services, users are assigned to roles (cf. Section 3.2.2). Users are automatically authorized based on their intranet identification; they are not

required to manually sign in (single sign-on - cf. Section 1.2.2).[7] In case the user already exists in the portal database, his credentials are automatically loaded. Thus, he has direct access to all authorized services without the need to sign in. New users are added to the database without getting any special access permissions. Administrators can later assign additional rights. They can also add new users who have not yet visited CoFiPot. When those users visit the portal for the first time they will automatically get the rights that were already assigned in advance. For every user the date of the last visit is stored.

Based on this, the 5971 registered users can be distinguished by their activity level. Table 3.1 shows the numbers accordingly grouped by the number of visits per month. The table further shows the share of users in each group who are assigned to any roles so that they have access to more services.

All users who have not accessed the portal within half a year are not included in these and all further statistics.

| *Activity Level* | *Users* | | *With roles* | |
|---|---|---|---|---|
| | *Per group* | *Total* | *Per group* | *(%)* |
| *Every month (>10 visits)* | 93 | 93 | 86 | (92.47%) |
| *Every month (<= 10 visits)* | 825 | 918 | 336 | (40.73%) |
| *Every other month* | 1592 | 2510 | 356 | (22.36%) |
| *Remaining* | 3407 | 5917 | 317 | (9.30%) |
| ***Total*** | **5971** | | **1095** | **(18.34%)** |

**Table 3.1.:** Active and inactive CoFiPot users and share of users who are assigned to roles (based on data from November to April 2010).

The data shows that 93 (1.56%) of all 5971 active users visit the portal more than 10 times per month with 4999 (83.72%) users accessing the portal less than once per month. This high rate of seldom visits is explicable with the fact that many users access the portal to retrieve information from its start page. This page provides important exchange rates, the share price development of the Bayer stock and access

[7] Before October 2009, every user who required access to secure services had to sign in manually to access protected services. Therefore, two visits from a single user with one being signed in and the other anonymously could not be linked to each other up to this date.

to the Bayer phone directory and other sites within the Bayer intranet. Each person who visits the portal is automatically registered in the user database the moment he accesses any of its pages. Visits to the start page are, however, not included in the analysis of portal visits because it is impossible to distinguish between real visits to the portal and "wrong" visits caused by users who set CoFiPot as start page in their browser.[8]

The high percentage of users with roles in the more active groups appears logical. Users who need to be authorized to access protected service will require these services for their work. It can therefore be expected that these users are rather active.

**User Profile Data:** For each user information such as his user ID, his assigned roles and the date of the first and last visit to the portal are available. In addition each user can provide further information (and upload a picture that is randomly displayed to all users on the start page). Among these additional pieces of information are some of relevance to identify users groups with similar preferences or tasks: the department the user is working for[9], the types of processes he is involved in[10] and information whether the user is working at the corporate center or at a subsidiary.

Users are encouraged to provide this additional information. However, only 13.5% of all users have chosen to do so.

The log files of the portal also include this additional information except for the user's ID. Thus for each visitor to the portal, information about his roles (always), departments and working area (if specified) is available for usage analysis.

### 3.2.2. Access Rights Management

CoFiPot uses a role based access management concept to control access to its services. Some services are visible to every user but access to the remaining is limited to mem-

[8] It is not possible to determine whether users "use" any of the services on the start page since they show all information without requiring any interaction from the user. Thus, even a planned visit to the start page only consists of the initial opening hit (cf. Section 2.3.1).

[9] The department does not indicate the location of the user. For example, if a users specifies 'treasury' as its department, it can be the treasury department of the holding or any subsidiary.

[10] In this case the selectable types cover broad areas of work such as accounting, taxes or FX-trading.

bers with specific roles (cf. Section 1.4.4). In 2010 there were 68 roles. The process of assigning users to roles can be initiated by management or by users themselves. For new employees the required set of roles is often derived from their predecessors as soon as the information about the new employee is forwarded to portal management. In addition users request access to a service by sending an email or submitting a predefined form in the portal. Services to which users have no access are not visible in the menu but there is an additional site map that shows all services.[11] Using this site map a user can visit any service. However, if he is not authorized to access it, a short description of the service is displayed together with a form allowing the user to request access for it. By submitting the form, portal management will receive a notification containing the message as well as additional information about the users which allows them to decide whether to grant access or not.

### 3.2.3. Service Portfolio and Usage

The number of offered services in CoFiPot has grown to 72 services in its ten year existence (cf. Figure 3.4). This number does not include special services such as the start page (which shows key figures such as Bayer's stock exchange value and important currency rates - cf. Figure 3.2) and portal specific services such as those to update its own profile or access stored reports and documents as well as all portal management services. Table 3.3 lists the services, grouped by the number of users who are authorized to use them, together with additional information about the active usage which is further discussed in the following section.

#### Service Usage

Based on the usage logs the number of users who actively use a service can be identified. *Actively using a service* in this context is defined as visiting the service at least twice per month on different days. Table 3.2 shows the results of this analysis from

[11] There are a few services that are not listed in the site map. These services are too special to be of benefit for anybody except the target group. They include all portal management services but also special reporting services that are tailored for high management levels.

the user perspective. It groups active users (cf. Section 3.2.1) by the number of services they have access to and shows how many services they use regularly.

| *Used services* | *Users* | | *Actual usage** | |
|---|---|---|---|---|
| | *Per group* | *Total* | *Avg.* | *StdDev.* |
| *> 15 services* | 8 | 8 | 22.00 | 10.16 |
| *> 10 services* | 11 | 19 | 12.27 | 1.42 |
| *> 5 services* | 67 | 86 | 7.24 | 1.27 |
| *> 2 services* | 236 | 322 | 3.52 | 0.72 |

**Table 3.2.:** Users grouped by actively used services (November-April 2010).
* Average number of services being per user in the group

The data does not include the majority of users who use less than two services. The remaining 322 constitute 5.39% of the active users. One reason for these numbers has already been explained with the fact, that visits to the start page are not included in these numbers. It also shows that for the majority of users only one service is of relevance and that only 86 (1.44%) use more than five services on a regular basis.

### Privacy Issues

The previous sections showed that extensive information about visits to the portal is logged. The log contains the visited service, the length of the visit and user data which includes the user's roles and, if available, additional information as described in Section 3.2.1. Although the user's ID is not saved in the log files, users could in theory be identified by comparing the available data with the user database.[12] To prevent any issues regarding user privacy, each user has to accept the portal disclaimer which was drafted in coordination with staff council. This disclaimer states that the log data is not and will never be used to track working patterns of single users, to use any of the log data for legal issues or to measure working performance of groups or single users. This approach is in compliance with European (Art. 6 ff. RL 95/46/EG) and German law (§§ 3 ff. BDSG). Privacy issues are further discussed in Section 2.3.1.

[12]The more information a user has provided, the easier the identification will be, due to the limited number of users and high number of possible role-department information.

### 3.2.4. Usage Pattern Analysis

To understand the overall usage patterns, Table 3.3 shows a list of all services grouped by the number of users who are authorized to access these. For each group the number of users who actually use the service is provided. For example, there are six services to which 250-500 users have access to. On average 444 users have access to these services but only 80 are using them. The numbers show that (i) less than a third of the authorized users use the services and (ii) that the usage of the services strongly varies within each group. Further analysis of the usage numbers compared to the number of authorized users reveal a group of popular services being used by many users while others are only accessed by a minority of the authorized users. On average 16% of each service's authorized users are actively using it.[13]

| *Group* | *No. of authorized users* | | *Services* | | *Used by* | |
|---|---|---|---|---|---|---|
| | *Avg* | *StdDev* | *Count* | *Total* | *Avg* | *StdDev* |
| All | - | - | 23 | 23 | 471.33 | 879.42 |
| 1000+ | 2440.00 | 1059.25 | 2 | 25 | 264.00 | 117.38 |
| 500-1000 | 623.33 | 34.02 | 3 | 28 | 173.00 | 53.51 |
| 250-500 | 443.67 | 55.87 | 6 | 34 | 79.67 | 60.36 |
| 100-250 | 184.67 | 50.90 | 3 | 37 | 74.67 | 30.92 |
| 50-100 | 61.20 | 8.65 | 6 | 43 | 11.30 | 10.94 |
| 20-50 | 31.00 | 7.82 | 15 | 58 | 9.94 | 11.12 |
| 10-20 | 14.27 | 2.16 | 7 | 65 | 9.77 | 37.44 |
| <10 | 9.00 | 0.00 | 7 | 72 | 1.70 | 2.36 |
| ***Total*** | 136.82 | 407.85 | 72 | | 124.79 | 449.11 |

**Table 3.3.:** CoFiPot services grouped by number of authorized users (November-April 2010). E.g. three services can be accessed by 100 to 250 authorized users with an average of 184.67 users for the group. On average 74.67 of these use the service.

Service usage does not only differ with regard to number of users but also depends on the time of year for some services. As an example, Figure 3.5 shows the usage numbers of four services for an interval of one year. One service (Cash Planning) is used intensively towards the end of each quarter while the usage of the remaining

[13]This number excludes the services that are visible to all users.

ones only shows minor fluctuations. The fluctuations are mainly caused by the reporting requirements of subsidiaries for the end-of-quarter and end-of-year closings.

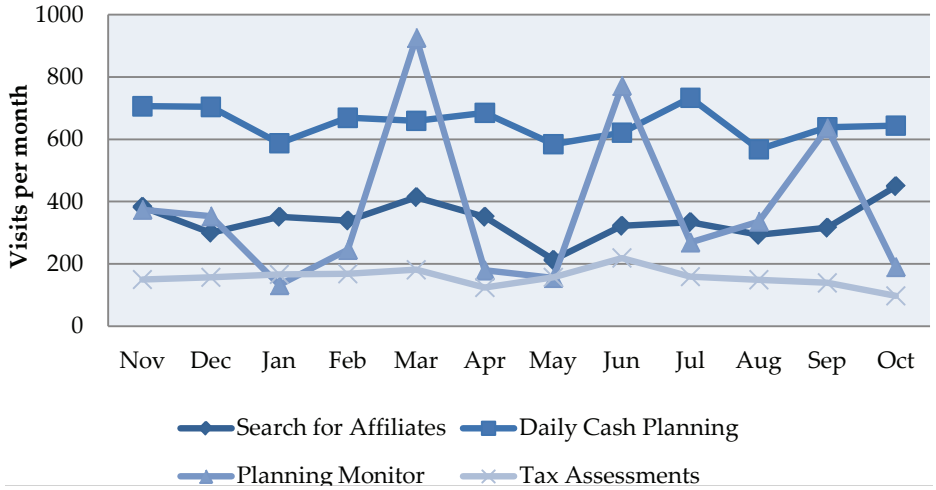


**Figure 3.5.:** Service usage fluctuations in CoFiPot for four services in the period from November 2008 to October 2009.

### New User Behaviour

To understand how users explore the portal and whether they look for other services besides their primary set (cf. Section 2.1.4), the visits of new users during their first six months after their first visit to CoFiPot were analyzed. Figure 3.6 shows the average number of services that new users visited per month. It shows the visits of all users (A) and also filtered for those among them who visited the portal at least once per month in the monitored time period (B). Each column in the graph is split into the number of services that users visited for the first time in the given month (1) and those that they revisited. For example, in the third month after their first visit to the portal, active users visited three services, 1.9 of which they had already visited in any or both of the previous two months.

The analysis shows that new users barely browse through the portal to discover its service portfolio. In the first month they just visit on average 2.4 services (Chart A). The number of newly visited services afterwards declines. After six months new users have visited 3.7 services in total although each new users initially has access to 23 services (cf. Section 3.2.3). Focusing on active users (Chart B) increases the

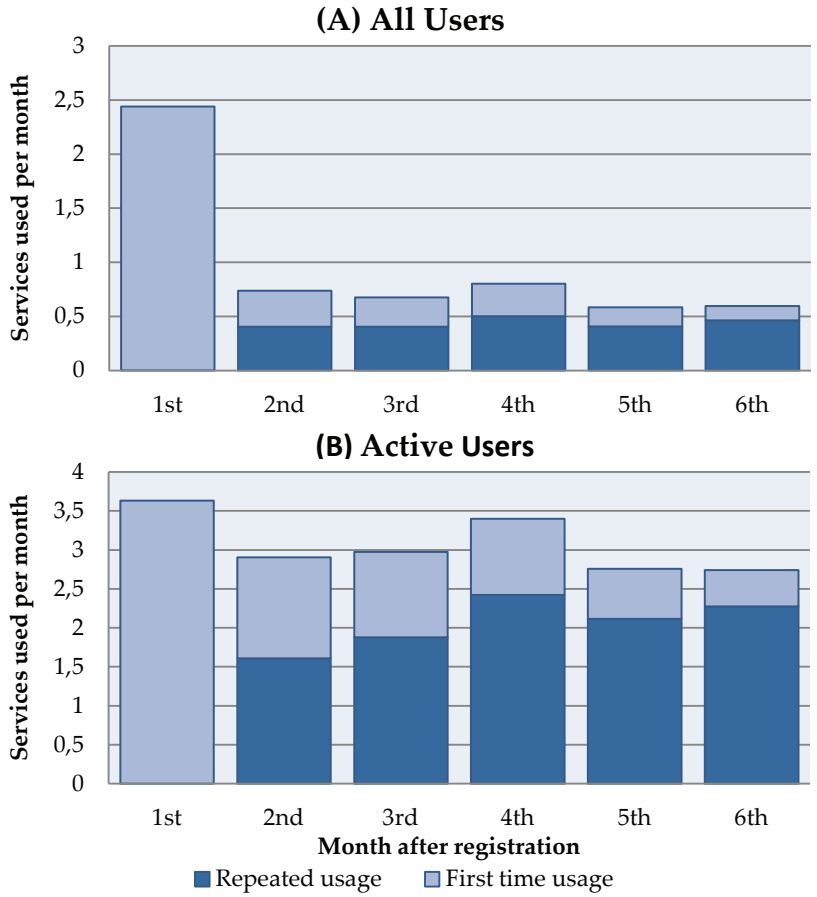


**Figure 3.6.:** Visited services per month by all (A) and active (B) users in the first half year after registration. Values are split into services that were visited by the user in any of the previous months and new ones.

numbers to 3.6 visited services in the first month and 8.1 in total. The fact that they do not even once visit any of the other services strongly indicates that there is no interest in finding more services or that they are not well informed about the potential of the portal (Elsner and Krämer, 2010a). This confirms the assumptions that were made in Chapter 2 that users of application integration portals do rather not browse the portal to find other services, which might be of interest to them.

The following results of the expert interviews among managers of the Finance Department will further support this analysis and provide information about the relevance of CoFiPot as well as the perception of the portal in general.

## 3.3. Identifying Key Challenges: Expert Interview

In autumn 2009 an expert interview was conducted targeting managers at Bayer Finance. This group included all financial managers whose subordinates use the portal and who all are heads of finance departments at the corporate center as well as the regional responsible persons - in total this group contains 15 managers. Nine managers (60%) agreed to participate in the one hour semi-structured interview. The interview aimed at analyzing three main questions:

- How do the managers see the relevance of the portal?
- What knowledge do they have about the service portfolio?
- Do they see the benefit of enhancing the portal with a recommender system?

In the first part the interview partners were asked to estimate the relevance of the portal for their departments, to share their experience with the portal and to provide ideas for improvements or to criticize aspects, processes or developments in the portal's evolution. The second part focused on the knowledge about the portal and its services that the interview partners and their subordinates have. In the third part, the idea of a recommender system for portal services was presented to collect feedback about certain aspects of the system. In total the interview consisted of 24 main questions with several subquestions. Six of the questions required the interview partner to classify in total 25 statements on a 5-point Likert scale from 1-5. The interview-sheet is presented in Appendix D. The detailed results of these questions are summarized in Table 3.4 at the end of this section.

**Relevance of CoFiPot**

Being asked for their general impressions of the relevance of CoFiPot the acceptance of the portal within the Bayer finance community became obvious. The statements ranged from "a tool that facilitates the execution of working steps" to the definition of CoFiPot as "THE source for financial data at Bayer". Among the benefits of the portal the relevance for business processes is highest ranking while the role as a knowledge

management tool is described as "less relevant". All interviewed managers emphasized the high quality and reliability of the financial data in the portal's many reports. Two of them also see an important aspect of the portals success in its ability to "hide the complexity of the underlying systems", thus lowering the costs for entering and retrieving data and lowering the effort for training. However one statement warned that with the constant evolution of some services their complexity had become too high, preventing inexperienced users from fully exploiting their benefits.

As shown in Figure 3.7, on a 5-point Likert scale (5 being the highest) the relevance was classified as 4.2 (StdDev: 0.97) for the interview partners themselves and 4.1 (0.98) for their subordinates. This result confirms the afore listed statements emphasizing the important role of the portal for Bayer Finance. A third question asked whether the portal supports the managers and their subordinates at the optimal level. The average answer of 3.9 (0.95) indicates a high level but leaves room for improvements.

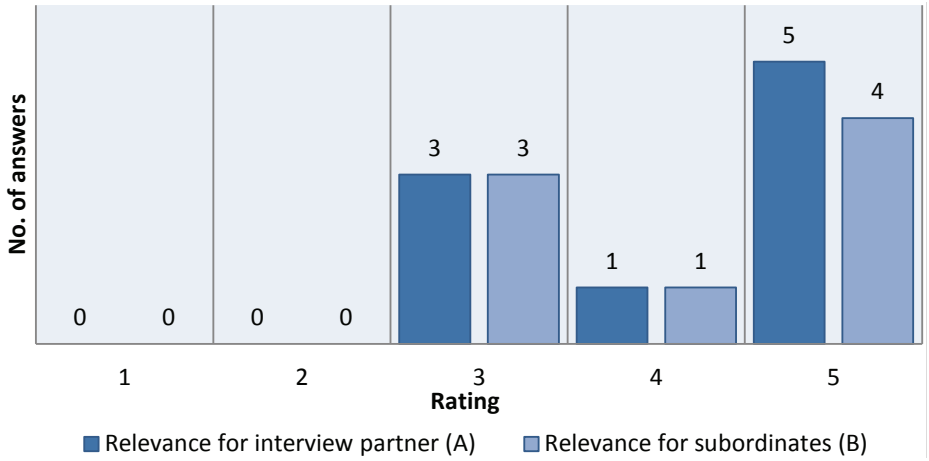


**Figure 3.7.:** Perceived relevance of CoFiPot for employees (A) and their subordinates (B). Scale from 1 (no relevance) to 5 (high relevance).

To further determine the relevance of the portal one questions asked for the alternative systems or processes that could deliver the same services as CoFiPot. According to the answers the main replacement for the portal would be a high use of Microsoft Excel (six answers) and more effort for coordination via email and phone (four answers). Three managers stated that the data-quality of the reports would be lower if

the numbers were to be taken from the primary systems. Further investigation explained these statements based on the fact that many of the reports accumulate data from various sources.[14]

The final question of the first part aimed at determining how managers value the usage of the portal by their subordinates. Therefore, they were asked to rate the two statements whether users *could* or *should* use the portal more often. The answers in Figure 3.8 show that the first question (could) received higher agreement with 3.8 (0.75) than the second (should) with 3.0 (1.29). The higher estimate for the first part is simply explicable since it would not make sense to want users to use more services while similarly not seeing the potential. The overall indication from these answers, however, is that the managers obviously do not see the usage of the portal at its optimum level. None of them disagrees with the statement that it could be used more.

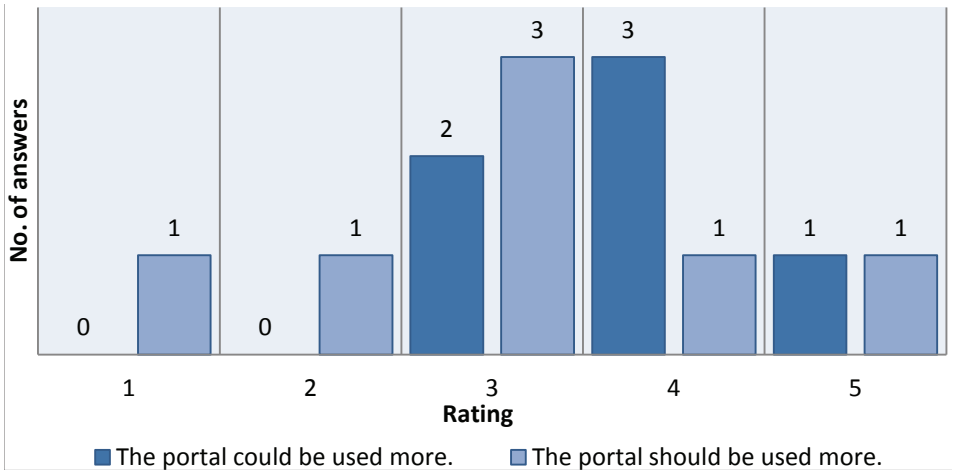


**Figure 3.8.:** Expert interview question: "The portal could/should be used more often.". Scale from 1 (totally disagree) to 5 (totally agree).

The answers to the first part of the interview show that the portal is of high relevance for Finance. However, although management sees the various benefits of the portal, the usage of the portal is estimated to be below its potential.

[14]Some data sets are stored on multiple systems with different attributes. Manually exporting each attribute from the best source would require too much effort. Thus, some values would be missing or be slightly different from the best possible result.

### Knowledge about CoFiPot

Managers estimate their overview about the portal and its services slightly better with 3.5 (StdDev: 1.20) compared to their subordinates with 3.1 (1.25). This rather neutral results show that the overall knowledge of the portal is not very good. Especially the results for the manager's knowledge could be seen as alarming. If they do not have a good overview of the portal and its services, they cannot advise their subordinates in the usage of the portal. This is further supported by the demand for more information about changes to the portal. Figure 3.9 shows the discrepancy between the perception level of information about new services in the portal that is currently available and the perceived need. The average information level is specified with 2.4 (1.36) while the wish for more information is at an average of 3.9 (1.42). Similarly the need for information about changes to existing services specified with 3.9 (1.36) differs from the current level of 2.3 (1.03).

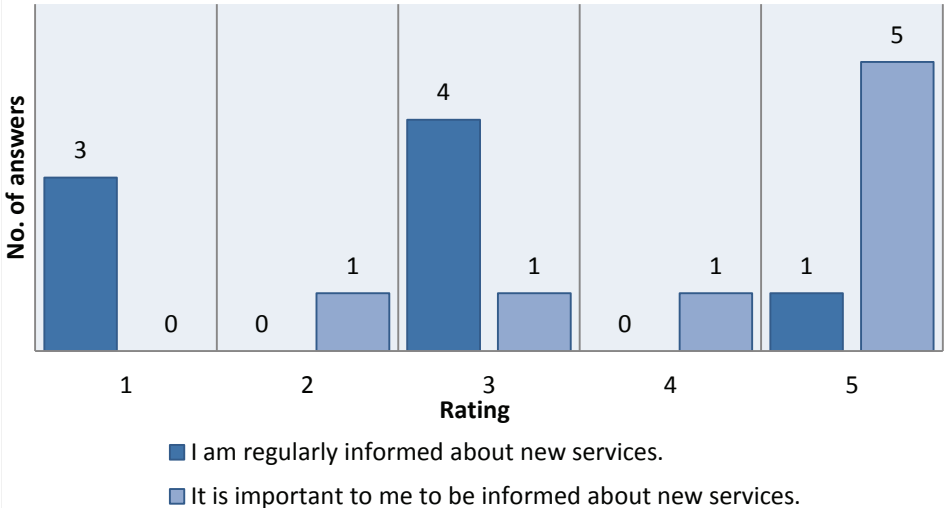


**Figure 3.9.:** Information about new services: Status and requirements. Scale from 1 (totally disagree) to 5 (totally agree).

In order to support the results from the log analysis, which showed that even new users barely look at the portal's service portfolio (cf. Section 3.2.3), the interviewed managers were asked whether they themselves browse the portal for new service. Only three replied with a value of 4 by an average of 2.1 (1.45). Comparing this result to the estimate of their knowledge about the portal, which was specified with 3.5

(1.20), can be used to speculate whether the entire service portfolio is even known to the management level. Taking together these values leaves room for speculation whether the answers about the manager's knowledge of the portal reflect the real level. Unfortunately the question of how many services they think the portal offers was only answered by two interview partners. Therefore, no relevant comparison between the perceived and real knowledge of the portal could be drawn.

To better understand how employees get to know the portal, each interview partner was asked how their subordinates get to know the services in the portal that are relevant for their work. Only two referred to process descriptions as the main source of learning about the portal services. "Learning by doing" (four answers) and recommendations by colleagues (three answers) seem to be the main factors that lead new users to the portal.

Several proposals were made to raise the awareness of the portal itself and certain services. Supporting the result from Figure 3.9 it was demanded by three managers to increase the transparency of the portal and its development process. Also three managers independently proposed to provide screencasts about the portal with one respondent stating that "screencasts are way more efficient than any documentation". Offering workshops or personal trainings is seen as the most efficient way of finding the right services for users; however, the high effort for such workshops was emphasized as a reason against this measure.

The expert interview showed the high relevance that CoFiPot has for Bayer finance. Still a lack of information about the portal is obvious.

The third part of the expert interview regarding a recommender system is summarized in Section 3.4.2.

## 3.4. Overcoming the Key Challenges: Introduction of a Recommender System

The findings of the usage analysis show that the majority of CoFiPot users regularly use one or two services. Furthermore, it was shown that even new users do barely

| No | *Question* | *Avg* | *StdDev* |
|---|---|---|---|
| *1* | I am regularly informed about new services | 2.39 | 1.36 |
| *2* | I want to receive more information about new services | 3.94 | 1.42 |
| *3* | I am regularly informed about service updates | 2.28 | 1.03 |
| *4* | I want to receive more information about service updates | 3.89 | 1.36 |
| *5* | I browse the portal for services | 2.11 | 1.45 |
| *6* | Relevance of CoFiPot for interview partner | 4.22 | 0.97 |
| *7* | Relevance of CoFiPot for subordinates | 4.13 | 0.99 |
| *8* | The portal supports my work at optimum | 3.94 | 0.95 |
| *9* | The portal could be used more by my subordinates | 3.83 | 0.75 |
| *10* | The portal should be used more by my subordinates | 3.00 | 1.29 |
| *11* | I have a good overview of the portal | 3.50 | 1.20 |
| *12* | Subordinates have a good overview of the portal | 3.13 | 1.25 |

**Table 3.4.:** Results of the expert interviews on a 5-point Likert scale.

browse the portal for other services. This underlines the characteristics of an application integration portal being used as business tool to execute certain tasks. Users access the portal to do this and immediately leave without looking at any other service. Based on these findings it can already be concluded that the user's knowledge of the portal's service portfolio is not on a very high level. This assumption has been confirmed by the answers in the expert interview. The interviewed managers estimated the knowledge level of themselves and of their subordinates to a neutral level. Moreover, they agreed that users could and should use the portal more often.

The results show that a key factor to further increase the relevance and benefit of the portal for the Finance Department, which is already on a high level as the interview showed, is to better inform users (and management) about CoFiPot's services. This demand is equal to the mission to find relevant services as defined in Section 2.2.

### 3.4.1. Selecting a Recommender System as the right Concept

Comparing the requirement with the concepts that were presented in Section 2.4 and also considering local realities narrows the list to internal marketing, recommender systems and trainings. Design and usability engineering does not apply for the mis-

sion. Personalized portals are not suitable due to the heterogeneous services that CoFiPot offers. Many of the services are of relevance for multiple processes. Thus, creating personalized portals with focus on special processes would be difficult. Also, personalization still requires users to browse for new portals - a matter that is not given in CoFiPot. Documentation was excluded for two reasons: (i) there is already a knowledge management system in place that offers documentation of several services and processes and (ii) experience has shown that users barely look at the documentation but rather contact support if there are any questions.

Of the remaining concepts, the *recommender system* was selected as most suitable concept. Internal marketing was not considered because the very different user requirements would have made the selection of the services to be promoted very difficult. Moreover, any actions including sending emails to users were refused by management. Trainings with the aim to present users all services in the portal would require a considerable effort in order to reach all users who are not located at Bayer headquarter. Table 3.5 summarizes the considerations for all concepts.

| *Concept* | *Comment* |
|---|---|
| *Documentation* | Already available but seldom used. |
| *Trainings* | Expensive to organize, especially for international users. |
| *Internal Marketing* | Difficult to select services for promotion in heterogeneous portals. Sending emails rejected by management. |
| *Recommender System* | Good match. Requires no input. Able to cope with heterogeneous content. |
| *Personalization* | Difficult to provide. Still requires users to search by themselves for services. |
| *Design & Usability* | Not applicable to increase service knowledge. |

**Table 3.5.:** Analysis of concept to increase service knowledge among CoFiPot users.

A recommender system matches the requirements of the above presented challenges. It can automatically identify possibly relevant service without the need to

provide explicit information such as user preferences. Showing personalized recommendations for services might motivate users to have a look at them compared to browsing the menu.

### 3.4.2. Expert's Evaluation of the Recommender System Concept

In the third part of the expert interview, the managers had to evaluate the concept of a recommender system for its suitability for CoFiPot. They highly valued the potential of the concept for all users with 4.1 (StdDev: 1.27) and especially for new users with 4.6 (0.53) as shown in Figure 3.10. In addition, being asked about the effects of such a system, they positively rated the effect that it can increase the usability of the portal with 3.7 (1.03).

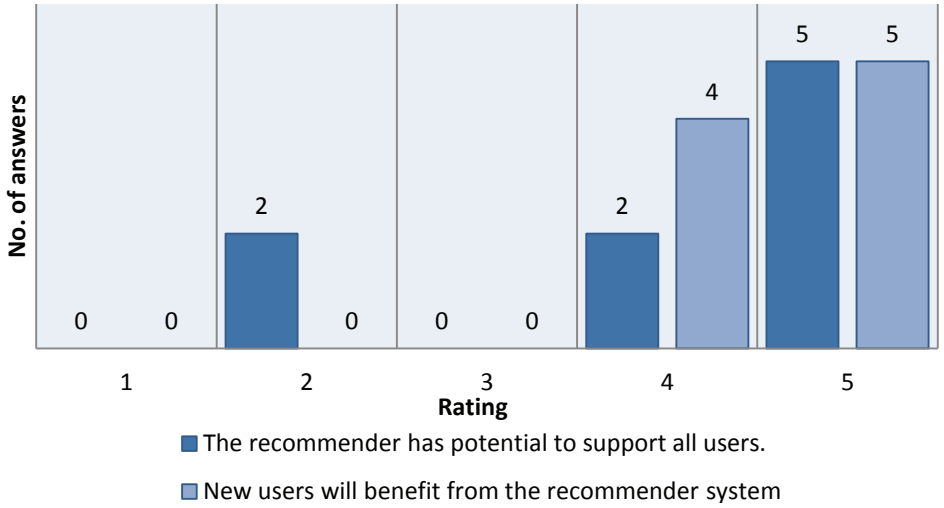


**Figure 3.10.:** Estimate of the potential relevance of a recommender systems for all and new users. Scale from 1 (totally disagree) to 5 (totally agree).

Comments made during the interviews gave further insights regarding the wishes and expectations to a recommender system. Four managers mentioned the high potential for new users and for the introduction of new services. One confirmed the before discussed challenges by stating that the complexity of the portal is very high and that a recommender system is very suitable because it pushed users to explore new services, on the one hand, but also supports them in their search for relevant

services.[15] However, three managers emphasized that despite the potential benefit of a recommender system, its effects should also not be overestimated especially with regard to frequent users of the portal. One mentioned trainings as an alternative although he also identified the difficulty and costs related to this concept.

The results of the interview strongly support the idea of implementing a recommender system into CoFiPot. Although such a system can have positive effects, they should not be overestimated. The interview answers also emphasized the difficulty of identifying relevant services.

### 3.4.3. Requirements to the Recommender System

In Section 2.4.4 the following key demands to a recommender system were defined:

- to inform users about the portal's service portfolio,
- to identify and suggest the most relevant services for each user,
- to require little to no input from users,
- to run automatically without requiring constant attention by portal management, and
- to cope with heterogeneous users and services.

Most of these can be copied to the requirements list for a CoFiPot recommender system. There should, however, be no requirements at all to users in order to derive recommendations. The high heterogeneity level of the portal is likely to prove a challenge to the recommender algorithms. Also, the role of protected services puts new demands to the system. Thereby the key question is whether protected services should be recommended to users even if they do not have access, given the fact, that the users are only authorized for protected services when the relevance has been identified combined with the analysis that the existence of secondary services which are not necessarily known (cf. Section 2.1.4).[16]

[15] Quote: "Halb zog es ihn, halb sank er hin."
[16] Whether and how to recommend protected services is discussed in Section 4.1.2.

## 3.5. Chapter Summary

In this chapter the Corporate Financial Portal was introduced and classified as a specialized application integration portal. The relevance of the portal for the Finance department is very high as it is used for most processes. Findings from the expert interview underlined the high relevance of CoFiPot for operations of Bayer Finance. The portal offers 72 services to 5917 active users from multiple departments. Supporting business processes from all areas of the Finance department, the portal offers highly heterogeneous services to users with not only heterogeneous requirements but also different backgrounds. While some services are specifically tailored to one group of users, many others have the potential to serve various groups.

As a current key challenge, which is caused by the growth of the portal since its implementation in 2001, the users' knowledge about the portal's service portfolio was identified. Analysis of usage patterns and an expert interview further revealed that despite the high relevance of the portal users do not search the portal for services, which might deliver additional benefits to their work. Accordingly, the managers that were interviewed for the expert interview stated that they see potential for a higher use of the portal. In the following sections and chapters focus will be laid on developing a recommender system that copes with the challenge of introducing service to users based on comparison to similar users.

# Chapter 4.

# Recommender Systems for Corporate Portals

The popularity of recommender systems has constantly increased since first implementations by Goldberg et al. (1992) and Resnick et al. (1994). These implementations are not limited to scientific projects[1] but are widely used in several web portals. Popular examples are Amazon's book recommendation system, music recommendations by last.fm[2], Netflix's movie recommendation system[3] and Facebook's friend recommendation[4] service.

Although different recommender systems exist for various fields of operations, the purpose of all systems is similar. Mulvenna et al. (2000) describe the basic concept of recommender systems "to provide users with what they want or need without requiring them to ask for it explicitly".

Recommender systems have become important components of portals, especially in e-commerce portals. An indicator for this argument is the one million dollar price that Netflix had offered for a recommender concept which can improve its own algorithms (Bell and Koren, 2007).

[1] Examples for scientific recommender system projects are MovieLens (http://www.movielens.org), Tapestry (Goldberg et al., 1992), GroupLens (Konstan et al., 1997).
[2] http://www.last.fm
[3] http://www.netflix.com
[4] http://www.facebook.com

The wide range of applications for recommender systems is reflected in multiple concepts tailored for the specific requirements and constraints for each field of operation. Choosing the appropriate concept is a crucial task before implementing a new recommender system. The selection process is thereby strongly dependent on the goal of the system and on the environment the system is to be implemented in (Adomavicius and Tuzhilin, 2005). For example, a recommender system that recommends items based on content comparison is appropriate for news portals but will not benefit a friend recommender system such as Facebook's. Here, a recommender system that analyzes relationships between users is more suitable.

Several factors have to be considered for the development of a recommender system for corporate portals. Some of them are different to common recommender systems as discussed in scientific literature. These factors concern the components that are available for the computation of recommendations and also the mode, how the recommendations are communicated to the users.

This chapter discusses the special requirements to a corporate portal recommender system. Based on the results an appropriate concept for the computation of recommendations is selected from the wide range of technologies. This selection process is discussed step by step to serve as guide for the development of such systems. Finally, a methodology is presented to test the quality of a system before its final activation.

## 4.1. Setting the Framework for a Corporate Portal Recommender System

The core demands to a corporate portal recommender system are (cf. Section 2.4.4):

- to inform users about the portal's service portfolio,
- to identify and suggest the most relevant services for each user,
- to require little to no input from users,
- to run automatically without requiring constant attention by portal management,
- to cope with heterogeneous users and services.

To fulfill these demands the concept for the recommender system has to be developed by discussing the overall process to generate recommendations for users. As first step, it has to be determined which data is available as basis. For the discussed concept this is the data about users and services, which will be analyzed with regard to their differences to recommender systems being used in public portals.

### 4.1.1. User Data Considerations

E-Commerce applications such as Netflix and Amazon have a user base of millions. Corporate portals, in contrast, have *significantly less users*. Here, the numbers is a few thousands users (cf. Section 1.3). According to Zanker and Gordea (2006) this can cause problems with the computation of recommendations because not enough information might be available to generate recommendations. However, they discuss corporate portals with less than thousand users which are smaller than the corporate portals being discussed in the previous chapters. Still, this matter might be one of the reasons for the lack of scientific sources on this topic.

In a corporate portal environment with a relatively static number of users, who are employed at the company, it can be assumed that more accurate *profile information* is available compared to public portals.[5] In addition to usage patterns, the department and position of each user is available (Pazzani, 1999). Using single sign-on (cf. Section 1.2.1), each portal visitor can be identified. Combined with further information about the security roles of each user (cf. Section 1.4.4) more explicit data is available that can be used to compute recommendations compared to public portals that depend on each user to voluntarily provide this information. However, privacy considerations have to be considered (cf. Section 2.3.1) which might lower the information quality.

An additional aspect is the *heterogeneity of the users* which is higher in corporate portals where users access the portal to fulfill different work tasks while visitors of public portals usually follow a similar goal such as buying a book at Amazon or

[5]The user base of corporate portals might change due to corporate growth or reorganization. However, unexpected significant changes in the user base will not occur.

choosing the next movie to watch at Netflix (cf. Section 1.3.2). Still, grouping users by job tasks provides a set of smaller user groups with similar users (cf. Section 2.3.1).

### 4.1.2. Service Data Considerations

In a recommender system, items are recommended to users. In the context of this work, items are services of a corporate portal and users the employees who use these services. Similar to the number of users, the *number of recommendable items* is less in a corporate portal environment with (a lot) less than 1000 items compared to public portals with millions of books, movies or other items. Also, the items are rather *heterogeneous* compared to other environments (cf. Section 2.1.1 and 2.1.2).

#### Interpreting Services as Recommendable Items

The definition of a *service* in a corporate portal depends on the granularity level that the recommender system should support. On the one hand, all the functionality assembled on a page of the portal can be summarized as one service; on the other hand, each functionality can count as a single service. E.g. a service could be a site offering to read and edit data but the functionalities to read and to edit could also each be regarded as single service. Another example is a site that offers multiple reports. The site itself can be regarded as a service or each different report individually.

The decision what to recommend as service depends on the goal for the implementation of a recommender system. The broader definition is suitable to familiarize users with the overall portal's functionalities. Users have to discover by themselves which functionality of each recommended service is of relevance to them. The granular definition, in contrast, can present every offered feature. Recommendations in the form of "do you know this report?" directly communicate the benefit of the recommendation while the recommendation "do you know the reporting service" leaves the user with the task of further exploration of a service's features. However, the number of recommendations will rise when using granular definitions. This can confuse users and prevent them from further following recommendations.

If the overall knowledge of the target portal among its users is already at a high level, granular definitions are more suitable since they help in exploring the portal in detail while broader definitions leads to a better overall knowledge. Thus, the decision depends on the data about portal usage and knowledge. Also, a combination of both concepts is possible by differentiating between *service* and *sub-service*. Recommendations could take the form of "do you know the reporting service? It offers the functionality to request the report X which might be of relevance to you".

Another option is to only show recommendations for sub-services to users while they are using the parent service.

#### Service Access Restrictions

The access to several services might be restricted to certain users or user-groups (cf. Section 1.4.4). Thus, not all services are visible to all users. While, in theory, every user who should have access to a service will already have the required permissions, this is not the case in reality. Often, portal management will simply not know that a protected service could be of relevance to a certain user and thus do not grant access. Three cases exist regarding the service visibility for a user:

- the service is visible,
- the service is not visible and not of relevance to the user,
- the service is not visible but could be of relevance while not breaking any clearance levels.

This circumstance introduces a complexity for the computation and display of recommendations which is not inherent to common recommender system applications. For example, Amazon's recommender system does not have to consider whether a book cannot be recommended to a user due to access restrictions. In a corporate portal it has to be decided how to handle services that are not visible for a user.

The easiest solution is not to show a recommendation for the service. However, this would prevent the detection of cases in which access is not granted only because it is not known that it is of relevance to the user. Another solution is not to display such recommendations to a user but to inform portal management that the system has

detected a service which might be of relevance to the user. The responsible persons can decide whether to grant access or not, thus manually confirming or rejecting the accuracy of the recommendation. Also, each service's attributes can be enhanced with a flag that indicates whether the service should be recommended to users who do not (yet) have access. In such a scenario an information text should be added to the recommendation that informs the user about the services main features. This message could even provide information about the process how the user can request access to this service.

A further limitation can arise for services that should not be recommended at all. This can be portal management services, highly specialized ones or those that provide highly sensitive data to a fixed group of people.

These restrictions require an additional attribute for each service that indicates how to handle its visibility. The following settings are thereby possible:

- recommend to all users, regardless of their access level,
- recommend to users, who have access, or
- do not recommend.

### 4.1.3. Overall Considerations

Several other factors have to be taken into account for the selection of a suitable recommender system besides the attributes of users and items. These factors concern the overall objective for the implementation of such a system and minor issues in which such a system differentiates itself from other recommender systems. The factors are summarized in Table 4.1.

The *objective for the implementation* of recommender systems in e-commerce applications is to increase sales or popularity of items. Therefore, the recommender system can play an important role in generating profit. In a corporate portal environment, the role of the recommender system is to increase the usage of services that have the potential to further support users thus contributing to their performance. However, it is assumed that all users already know the primary services they require for their work. Therefore, the role of the recommender system is likely to be considered less

| | *Attribute* | *Common Context* | *Corporate Portal Context* |
|---|---|---|---|
| *Users* | *- Number* | High (millions) | Few thousand |
| | *- Heterogeneity* | Low | High |
| | *- Information* | Usage data, voluntary profile | + Department, position |
| *Items* | *- Number* | High (thousands to millions) | < Thousand |
| | *- Heterogeneity* | Low | High |
| | *- Item visibility* | All items public | Role-based access |
| | *- New items* | Frequently | Seldom |
| | *- Information* | Description, ratings, comments | Description |
| *Overall* | *- Scalability* | Important | Barely relevant |
| | *- Manipulation issues* | Critical | Not relevant |
| | *- Role of the system* | Increase sales/visits | Supporting concept |
| | *- Quality* | Find good items | Find all good items |

**Table 4.1.:** Differences between recommender systems in corporate portals and in common context.

important. Accordingly, the system should not require users to provide any input to receive recommendations.

The *addition of new items* to the system in e-commerce applications will be significantly more frequent than in corporate portals where new services are, at most, introduced every month. Amazon, for example, has thousands of new items per week if not per day (Linden et al., 2003).

Due to the usually high number of users and items, the *scalability* is an important factor for many recommender systems (Linden et al., 2003). Since the number of items and users is significantly lower in corporate portals and unplanned increases are not to be expected (cf. Section 1.3) scalability is no major issue for corporate portals.

Another more general aspect is the topic of *manipulations* to recommender systems. A single user can create multiple accounts to manipulate ratings or descriptions of items (Mobasher et al., 2006). Corporate portals will not suffer from this problem for two reasons: According to Mobasher et al. (2006) a user requires several accounts to manipulate the system. This is not possible in a corporate portal where accounts

are created per employee. Furthermore, there is no benefit for users to manipulate a recommender system in a corporate portal while, for example, sellers in public portals will benefit from higher rankings. Therefore, the topic of manipulation is not further discussed in this work. Relevant literature on manipulation detection is provided by Resnick and Sami (2008), Mobasher et al. (2006) and van Roy and Yan (2009).

The *quality of the recommendations*[6] is strongly related to the number of recommendable items and the size of the user base and also to the purpose of the recommender system within the portal. In large recommender systems with many users and items, the recommender algorithms have to be optimized to handle large amount of data, thus possibly affecting the quality of the recommendations (Berkovsky et al., 2007). However, if the recommendation of a relevant item is missed in a system with many items, there are still enough other items with similar quality left that are recommended. The miss of the one item will not have a significant impact (Hill et al., 1995). Shardanand and Maes (1995) flag this scenario with the term *find good items*. In corporate portals with comparably low numbers, quality will be more important than scalability. Missing to recommend one relevant item does have a huge impact since the overall number of relevant items is usually low. Therefore, the aim of a corporate portal recommender system is to *find all good items* (Herlocker et al., 2004).

## 4.2. Choosing a Suitable Recommender Concept

In the context of this work, a recommender system is defined as a system that provides recommendations for items $I$ to the users $U$ of the system. Recommendations can be created automatically or manually. Usually the target group consists of individual users but could also consist of user groups.

Several classifications for recommender systems exist. They mainly differ in the number of applied criteria for the classifications and sometimes in the order of the classification hierarchy. Runte (2000) identifies four dimensions sorted in a hierarchy while Adomavicius and Tuzhilin (2005) only propose two dimensions without a hi-

[6] *Quality* in this context stands for the ability of the recommender system to predict relevant items for a user.

erarchy (cf. Appendix E). Similarly, Schafer et al. (2007) differentiates by attributes of the system which are not sorted.

The classification approach used for this work is based on a combination of the classifications by Adomavicius and Tuzhilin and Runte. Figure 4.1 shows the selected hierarchy with four dimensions.

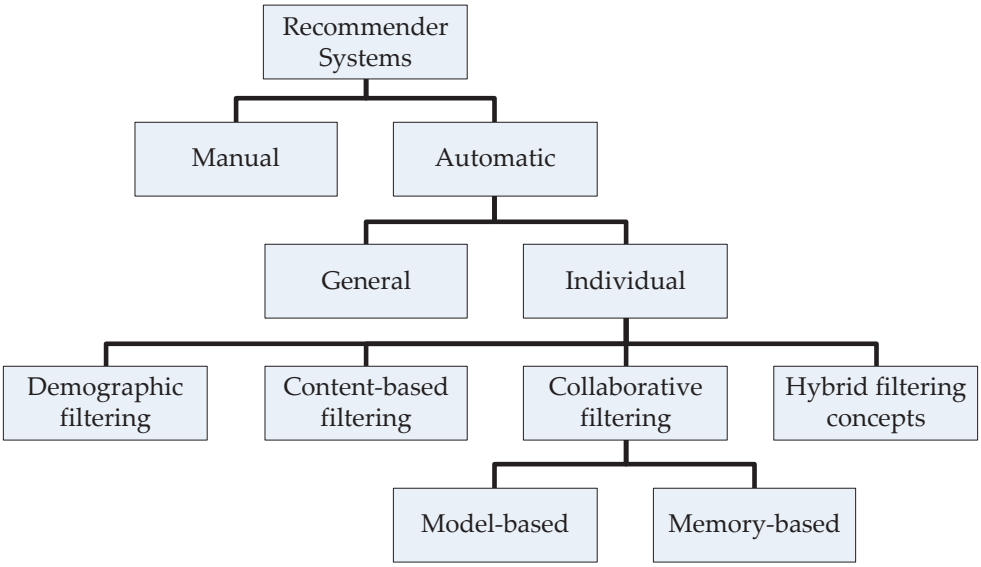


**Figure 4.1.:** Classifications of recommender systems based on Adomavicius and Tuzhilin (2005) and Runte (2000).

The first two dimensions are not discussed in detail. *Manual* recommender systems require users to create recommendations for other users of the portal (Maltz and Ehrlich, 1995). *Automatic systems*, in contrast, derive their knowledge from ratings, usage patterns or META-information about each item. They do not necessarily require manual input by users. Furthermore, they cover all users and items which manual systems might not do. The concept is comparable to the concept described as *internal marketing* in Section 2.4.3. The here required recommender system however is supposed to *automatically* compute recommendation without as little manual effort as possible.

At the next dimension that distinguishes between individual and general systems, the decision for individual system is again given by the requirements to the system. The goal is to create *individually* matching recommendations for users in an environ-

ment with heterogeneous items. *General* recommender concepts, however, are only able to recommend the most popular items based on usage or ratings (Schafer et al., 2001). The disadvantage of general recommendations is (i) that they only recommend the most popular items in the system and (ii) that they do not take user preferences into account. Accordingly, it has been shown that the prediction quality of individual system is significantly better (Konstan et al., 1997; Breese et al., 1998; Herlocker et al., 1999).

## 4.3. Comparison of Filtering Techniques

An individual recommender system filters all available items (here, the services offered by a corporate portal) for those that might be of relevance to a user. The most popular filtering techniques are *content-based filtering* and *collaborative filtering* (Breese et al., 1998; Runte, 2000; Adomavicius and Tuzhilin, 2005).

### 4.3.1. Content-Based Filtering

Content-based filtering is based on the assumption that items with similar attributes will be equally valued by users. They recommend items to users if they match their user profile (Buono et al., 2002; Schafer et al., 2007; Pazzani and Billsus, 2007). The basic concept consists of three steps:

1. Determine and collect descriptions for each item. If only unstructured textual descriptions are available, methods from information retrieval are used to convert this data to structured information.

2. Determine attributes that describe the preferences of each user. These preferences can be requested explicitly or be determined implicitly by analyzing usage behavior.

3. Compare item attributes with user preferences to identify recommendable items.

Popular applications for content-based filtering recommender systems are news-portals, restaurant-guides, TV-guides and online-shops (Pazzani and Billsus, 2007). The concept is based on the availability of descriptive attributes for each item. Often, such information is already available. E.g. in movie-databases, book- and other stores there are certain META-data such as genre, author/actors, description already available for each product or movie (Basu et al., 1998). If such data is not available, it needs to be added manually or be derived by using information retrieval techniques (Marchionini, 1995; Adomavicius and Tuzhilin, 2005). An example is to describe objects with the Ontology Language (OWL) which makes it possible to compare items (Hornung et al., 2007).

**Advantages**

The content-based filtering technique has four major advantages: (i) The process for the generation of recommendations is transparent and easy to understand for and communicate to users. For example, it is obvious to users of a news-portal that the system will recommend articles about topics that are comparable to those they have read before. The technique is (ii) independent of the number of items and users in the system. This factor is especially positive for new users. As long as they provide their preferences, the system will be able to find relevant items. The recommendations are (iii) independent from other users in the system and the system can work without the availability of a usage history. Similarly, new items can be recommended directly if the META-data is available. Furthermore, (iv) for systems that already have the required META-data set for the items, the implementation effort is low (Runte, 2000).

**Disadvantages**

Content-based filtering requires information about users and items. If such data is not available it has to be created with significant effort (Shardanand and Maes, 1995). In a corporate portal context, users would have to be motivated to provide their preferences such as job requirements. It is even questionable whether they can provide so detailed data that would allow the identification of matching services. The same

problem arises for the description of items which has to be created by portal management with high effort (Balabanović and Shoham, 1997).

These challenges are amplified by the heterogeneity of users and services (cf. Section 2.1.1 and 2.1.2). The detail level of each description has to be granular enough so that the system can match them with the requirements of different target groups. The characteristic of the content-based filtering technique to recommend items to users with similar attributes will prevent the recommendation of items that do not share enough attributes. In a heterogeneous environment, this will reduce the number of recommended items because only similar items are recommended. Unexpected recommendations are very unlikely; the system is "overspecialized" (Montaner et al., 2003; Adomavicius and Tuzhilin, 2005). Also, it is questionable whether the attributes are always able to represent an item so that it can be compared to others (Balabanović and Shoham, 1997). This is especially the case if the items are of different types (Burke, 2002).

Furthermore, the recommended item's quality is neglected. Recommendations are created based on their attributes and not on popularity or other quality metrics (Shardanand and Maes, 1995; Balabanović and Shoham, 1997; Schafer et al., 2007).

Although the system can directly create recommendations for new users who have provided their preferences, the opposite effect arises for users who do not provide any information (Runte, 2000). The system cannot create recommendations without the explicit input of user preferences.

### 4.3.2. Collaborative Filtering

The term *collaborative filtering* was introduced by Goldberg et al. (1992). They implemented the technique for the experimental mailing system *Tapestry* that recommended interesting mails based on ratings by other users.

The technology is based on the assumption that users with similar preferences will comparably value and rate items and uses statistical correlation between historic ratings of items from different users (Shardanand and Maes, 1995; Montaner et al., 2003; Schafer et al., 2007). Sarwar et al. (2001) describes the process as follows:

"The goal of a collaborative filtering algorithm is to suggest new items or to predict the utility of a certain item for a particular user based on the user's previous likings and the opinions of other like-minded users."

In an environment with $m$ users $U = \{u_1, u_2, ..., u_m\}$ and $n$ items $I = \{i_1, i_2, ..., i_n\}$ each user $u$ has expressed an opinion for several items $I_u$. Opinions are represented by a rating $r$, usually on a numeric scale. Ratings can be stated explicitly or implicitly derived from historic purchase data, from analysis of page views or from other mining techniques (Konstan et al., 1997; Terveen et al., 1997). For $I_u$ the following rules apply: $I_u \subseteq I$ and $I_u$ can be empty. Thus, each user has a mean rating of

$$\overline{r}_u = \frac{1}{|I_u|} \sum_{i \in I_u} r_{u,i} \tag{4.1}$$

Likewise the mean rating given to each item $i$ is

$$\overline{r}_i = \frac{1}{|U_i|} \sum_{u \in U_i} r_{u,i} \tag{4.2}$$

where $U_i$ is the set of users who have rated the item $i$ on a given scale.

The collaborative filtering technique predicts for an active user $a \in U$ the likeliness of an item $j \notin I_a$. The predicted value $p_{a,j}$ is calculated with

$$p_{a,j} = \overline{r}_a + \kappa \sum_{u \in U_j} sim(a,u)(r_{u,j} - \overline{r}_u) \tag{4.3}$$

The factor $sim(u,a)$ represents the similarity between the active user $a$ and all users $u \subset U_j$ that have rated item $j$. $\kappa$ is a normalization factor such that the absolute values of the similarities sum to unity (Breese et al., 1998).

The prediction $p_{a,j}$ is calculated for each item. Sorting the predictions delivers a list of items $I_p \subset I$, that the active user will probably like the most where $I_p \cap I_{u,a} = \Phi$. Figure 4.2 displays the basic concept. The collaborative filtering (CF) algorithm calculates predictions for all items $j$ that the active user $a$ has not rated yet. The resulting *top-N* predictions $p_{a,j}$ are shown in a list as recommendations.

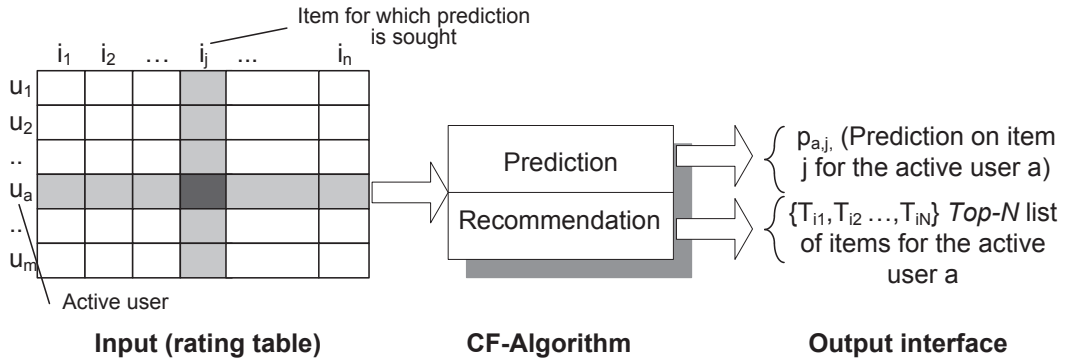


**Figure 4.2.:** The collaborative filtering process (Sarwar et al., 2001).

### Preconditions

Collaborative filtering is based on the availability of ratings for the items in the system. These ratings can be expressed explicitly by users or be derived from usage data. All ratings have to be limited or converted to a comparable scale (e.g. a five-point Likert scale - cf. Section 4.4.1). Since the computation of recommendations is based on comparing ratings, the system will only function properly if enough ratings are available. Users who have not provided any ratings (and also new users who have no history in the system) cannot be recommended any items and no recommendations can be computed for items that have not been rated. Furthermore, in order to compute similarities between users, overlapping ratings for items are required (Billsus and Pazzani, 1998). If no overlapping ratings exist, the collaborative filtering algorithms will not work.

### Advantages

The main advantages of the collaborative filtering technique are (i) the independence of any descriptions for items and, thus, (ii) the ability to handle all types of items regardless of their attributes - no domain knowledge is required (Bell and Koren, 2007) - and (iii) the inclusion of subjective user opinions in the computation of recommendations.

Using opinions about items rather than their description (cf. content-based filtering in Section 4.3.1) saves the effort of manual input or automatic identification of any

such attributes. Furthermore, the technology considers likes and dislikes of users. If an item is not popular among users, it will receive low ratings and thus will either not be recommended or receive a low prediction value. Popular items will be recommended, even if they objectively do not appear to be relevant to some users. Such *unexpected effects* are one of the main strengths of the collaborative filtering technique compared to content-based filtering because they allow identifying cross-genre items (Burke, 2002; Bell and Koren, 2007). This behavior is especially helpful to provide recommendations in heterogeneous environments such as corporate portals.

Focusing on the input demands to users, the technique only requires ratings for some items in the system. If such ratings are derived implicitly, users are not required to provide any input such as personal preferences.

**Disadvantages**

The dependency on ratings leads to problems with starting a new recommender system. Without any ratings, no recommendations can be calculated. This matter also applies for new items and users. Neither items that have not been rated nor users who have not provided any ratings are included in the recommendation process. Literature calls these problems the *new-item* and *new-user problem* (Balabanović and Shoham, 1997; Baudisch, 1999; Buono et al., 2002; Zanker et al., 2007; Schafer et al., 2007). Hybrid filtering techniques are suggested to overcome them (cf. Section 4.3.4).

In contrast to content-based filtering, the recommendation creation process is not transparent to users. Unexpected recommendations might not seem logical to the recipient. This lack of transparency leads to the perception of a *black-box* system and lower the trust into the system (Goldberg et al., 1992; Resnick et al., 1994; Hill et al., 1995; Sinha and Swearingen, 2001; Massa and Avesani, 2004).

The collaborative filtering technique is based on subjective criteria and does not include objective ones. Thus, any available attributes are ignored (Runte, 2000).

A further disadvantage is the *sparsity* and *coverage* problem. The terms describe the problems related to missing ratings, especially in systems with thousands of items. If an item is not or seldom rated, it will not be included in any recommendations or,

if only few ratings exist, the predicted will be low. This matter is strongly related to the *new-item problem*. The term *sparsity* describes the percentage of items that a user has not rated. In large e-commerce applications with millions of items the sparsity is usually high (Billsus and Pazzani, 2000). The term *coverage* is used to specify the percentage of items that are covered by the recommender system. Sparsity and coverage problems occur in systems with many more items than users (Balabanović and Shoham, 1997; Sarwar et al., 2001; Wang et al., 2006).

A last disadvantage occurs for users with very special preferences. If few or no overlapping item ratings exist between them and other users, they will not receive any recommendations.

### 4.3.3. Content-based vs. Collaborative Filtering

The advantages and disadvantages of content-based and collaborative filtering methods are summarized in Table 4.2. The comparison shows that the methods are based on fundamentally opposing concepts. Content-based filtering is entirely based on objective criteria and collaborative filtering on subjective ones. Accordingly, the content-based filtering concept requires a description of each item and preferences for each user while collaborative filtering ignores any of such information. Therefore, it is possible with collaborative filtering to compare items regardless of their attributes also across different genres.

| | *Content-based filtering* | *Collaborative filtering* |
|---|---|---|
| *Advantage* | - Transparent<br>- Supports systems with few users and items | - "Unexpected" recommendations<br>- Supports heterogeneous systems (cross-genre)<br>- Uses subjective criteria |
| *Disadvantage* | - Overspecialization<br>- Ignores subjective opinions<br>- Requires META data<br>- Overspecialization<br>- no cross-genre recommendations | - Intransparent<br>- Ignores objective criteria<br>- New-item problem<br>- Sparsity<br>- New-user problem |

**Table 4.2.:** Comparison: Content-based and collaborative filtering.

The collaborative filtering concept is more suitable for the application in corporate portals. The main reasons are its ability to handle heterogeneous users and items (cf. Section 2.5) and the independence from user input such as preferences or item descriptions. Content-based filtering, in contrast, requires every service to be described in details so that the potential relevance for users from different departments can be identified.[7] Furthermore, each user would need to provide his preferences. Also, some of the disadvantages of collaborative filtering do not apply for corporate portals: New users are initially told which *primary* services to use (cf. Section 2.1.4). Thus, the new-user problem will not arise. New services are introduced upon request and therefore have an initial set of users. Regarding sparsity and coverage: the number of services is less than the number of users and it can be assumed that each service is used by several users. Therefore, these problems should not arise in a corporate portal recommender system.

### 4.3.4. Hybrid and Other Filtering Concepts

*Demographic filtering* is further technique that uses demographic features representing users to learn relationships between an item and the type of people who like it (Kobsa et al., 2001; Montaner et al., 2003; Zanker et al., 2007). This information is used to cluster users and derive recommendations for each group. The technique is not suitable for a corporate recommender system for two reasons: it generalizes user interests and thus tends to generate rather impersonalized recommendations (Montaner et al., 2003). Furthermore, it would require highly detailed information about each user's interests to be able to match heterogeneous users and services. This contradicts the requirements specified in Section 4.1.1. Therefore, the demographic filtering technique is not further discussed.

Other techniques are presented as novel approaches such as *utility-based filtering* (O'Mahony et al., 2004) and *knowledge-based filtering* (Schmitt and Bergmann, 1999;

[7] Different departments often have different wordings for comparably issues which will be reflected in service descriptions and user preferences. Thus, a service description would need to include terms from both sides (cf. Section 2.1.2).

Towle and Quinn, 2000). However, these are merely sub-types of the content-based filtering technique (Huang, 2008).

The application of one filtering technique does not necessarily exclude the others. Several *hybrid techniques* exist which combine the strengths of each technique. An early example is the experimental mailing system *Tapestry* which combines collaborative and content-based filtering techniques (Goldberg et al., 1992).

The goal of such systems is to gain better performance while minimizing the drawbacks of the individual technologies. Most commonly, collaborative filtering is combined with other techniques (Burke, 2002). Different types of hybrid systems exist which are summarized in Table 4.3. Burke (2002) provides an excellent detailed discussion of the various hybrid concepts. The general conclusion is that hybridization is especially helpful to avoid startup problems and to eliminate problems related to coverage and sparsity. Hybrid systems are not further evaluated in this work since their main strength is to help with problems regarding new users or items. However, as already discussed for the collaborative filtering, these problems can be neglected for corporate portals.

## 4.4. Adapting Collaborative Filtering for Corporate Portals

The process for the generation of recommendations using the collaborative filtering technique consists of four process steps:

1. representation of items through implicit or explicit ratings $r$,
2. computation of the similarity between users $sim(u,a)$,
3. determination of the neighborhood for each user $a$, which specifies all users with a high similarity to the active user, and
4. prediction of the relevance $p_{a,j}$ of all items $j$ that user $a$ has not rated.

Several concepts exist for each process step that need to be analyzed for their suitability for the corporate portal context.

| *Hybridization* | *Description* | *Authors* |
|---|---|---|
| *Weighted* | The scores (or votes) of several recommendation techniques are combined together to produce a single recommendation. | Pazzani (1999); Towle and Quinn (2000) |
| *Switching* | The system switches between recommendation techniques depending on the current situation. | Tran and Cohen (2000); Billsus and Pazzani (2000) |
| *Mixed* | Recommendations from several different recommenders are presented at the same time | Wasfi (1999); Burke et al. (1997); Burke (2000) |
| *Feature combination* | Features from different recommendation data sources are thrown together into a single recommendation algorithm. | Basu et al. (1998); Condliff et al. (1999); Baudisch (1999) |
| *Cascade* | One recommender refines the recommendations given by another. | Burke (2002); Balabanović and Shoham (1997) |
| *Feature augmentation* | Output from one technique is used as an input feature to another. | Condliff et al. (1999); Sarwar et al. (1998); Mooney and Roy (2000) |
| *Meta-level* | The model learned by one recommender is used as input to another. | Condliff et al. (1999); Balabanović and Shoham (1997) |

**Table 4.3.:** Hybrid recommender systems overview. Adapted from Burke (2002).

### 4.4.1. Representing Service Relevance through Ratings

In the context of this work ratings for services represent the relevance that a service has for a user. High ratings thereby represent a high relevance. Ratings can be derived implicitly by analyzing usage patterns or explicitly by directly asking users to provide them. The key challenge of finding to best approach is strongly related to the required effort and the reliability of the resulting ratings. Combining implicit and explicit ratings is also possible (Adomavicius and Tuzhilin, 2005).

### Rating Scale

Whether ratings are provided by users explicitly or derived implicitly, a rating scale has to be defined that allows the distinction of each user's preferences. The finer the scale is grained, the more distinguishable information is available. However, too many steps on the scale will reduce the numbers of overlapping ratings.

Schafer et al. (2007) identify three types of rating scales:

- unary: Good or "don't know",
- binary: Good or bad,
- Likert scale: Integers (1-5, 1-7, ...).

Unary scales do not provide enough information. They could only reveal whether a user knows or does not know a service but not whether it is of relevance. Binary scales can provide this information by providing the options (i) know and of relevance, (ii) know and not of relevance and (iii) does not know. However, this scale does not reveal the relevance level. A service that is used with a high frequency will have the same rating as one that is barely used. For e-commerce portals such as Amazon, a binary scale is useful since it represents the possible cases for shopping scenarios: (i) known and bought, (ii) known and not bought and (iii) not known. For the corporate portal context, however, it does not provide enough information.

Therefore, a *Likert scale* is the appropriate concept to represent ratings in corporate portals. Usually the range of each scale starts with one, thus allowing the zero to serve as indicator for "don't know". The range of the scale defines the granularity of the information. Wide ranges (1-10 or more) provide very detailed information about each user's preferences. Herlocker et al. (2004) point out that wide ranges lead to an overspecialization of the recommender system because they lower the number of users who have provided similar ratings, thus limiting the prediction quality of the collaborative filtering algorithms. Even with implicit computation of ratings the results can vary if temporary usage fluctuations slightly change the input parameters. On a 1-5 scale tiny fluctuations will have no impact, but on a 1-10 scale the fluctuation might lead to lower or higher ratings thus influencing the similarity weights between users. Basu et al. (1998) overcomes this problem by grouping high ratings as relevant

and low ones as irrelevant, which basically leads to a lower range. If possible it seems logical to choose a smaller range in such cases.

The choice depends on the type of portal and rating concept. Wide ranges can deliver better results when using implicit rating approaches while they should not be used when explicitly requesting ratings from users. A scale from 1-5 or 1-7 is appropriate independent of the rating determination concept. It leaves enough options to distinguish between e.g. between "relevant" and "highly relevant" but is not overspecialized.

To determine the perceived quality of items, users are often enabled to provide negative ratings. Such cases are excluded for the ratings in corporate portals because the ratings should represent the relevance of an item with low ratings representing low relevance. A negative relevance is not applicable in this context.

#### Explicit Ratings

The obvious method to determine ratings from users is to explicitly request them. Many websites do this by showing the rating scale in form of stars that the user can give to an item. Other common modes to request ratings are color-lists, radio-buttons, drop down-lists or sliders. The display of such rating-forms varies. Some portals allow to always rate an item while others randomly ask for it. A recent rating feature has been introduced with Facebook's "like-button". Every portal can implement this button that only allows giving a positive rating to an item. The rating is then published on the user's personal page at Facebook and on the rated item's portal.[8]

Explicitly requesting the subjective opinion of an item delivers a high data quality (Schafer et al., 2007) but also incorporates some challenges. Users are required to participate in the process by rating items. Avery and Zeckhauser (1997) describe the problem of *freeriding*. The recommender system will only work if enough users participate. However, there might be some users who want to benefit from the recommender system but are not willing to invest the time for rating items themselves.

[8] A dislike button has recently been announced by Facebook (http://abcnews.go.com/WNT/video/zuckerberg-dislike-button-works-11221793). This feature would change the scale from unary to binary.

If too many *freeriders* exist within a community, the recommender system will fail due to lack of ratings. Avery and Zeckhauser further point to the problem of diminishing motivation to participate in the process. It is not only difficult to provide incentives to rate (Stenmark, 2004; Kaplan and Henderson, 2005; Schafer et al., 2007) but also to keep up the motivation to further rate items. They also warn that too high incentives will motivate users to rate as many items as possible which can lead to ratings that do not necessarily reflect the user's attitude. Finally, requesting the user, to provide an explicit rating, interrupts his current working process (Buono et al., 2002).

For the explicit rating of services another difficulty has to be considered. Usually people are asked how much they like an item. For the corporate portal recommender system the requested value is not the subjective opinion about the quality but the relevance of the item. Users have to be made aware of this difference. If a service is of high relevance for a user but the design or functionality is not according to the user's demands, he still would have to give a high rating. Since users are accustomed to questions about the quality rather the relevance, they might give the wrong rating in such cases.

The best results for explicit ratings can be obtained by interviewing all users or by performing a survey. In both cases the questionnaire would ask the user to rate the relevance of all services. While providing the best possible results, this method requires the highest efforts.

Asking users to actively participate in the process can have some additional positive side-effects (Harper et al., 2005). By participating users will have an increased feeling of having contributed to the community and might be grateful for having an option to provide feedback.

Explicit ratings can provide a high information quality but are prone to insufficient numbers of participating users and wrong replies. Furthermore, they require an extra effort from users. Especially in the corporate portal context where users are supposed to operate as efficiently as possible the questions of effort and reliability have to been seen as critical factors against explicit ratings.

### Implicit Ratings

In an analysis of rating concepts, Oard and Kim (1998) identify three categories for implicit ratings. The first category is based on the number of references (hyperlinks, comments) that point to an item. The second one analyzed retention patterns. The more users save, print or forward an item, the higher the rating will be. The third category examines usage which includes factors such as visits, duration of usage and repeated usage. This third category is best suited for corporate portals because it uses data that can be obtained without much effort and best represents the relevance of items (Zanker and Gordea, 2006). The overall advantage compare to explicit ratings is that it gathers information without any costs to the users (Oard and Kim, 1998; Buono et al., 2002).

Implicit ratings, which are based on the conversion from usage-data to ratings, require appropriate data such as the information about the purchase of an item, visits of sites or clicks on a site. Converting this data to ratings on a predefined scale is necessary to obtain comparable values. In case of online shops the transformation is made to a simple binary scale. A user gives a positive rating to an item by acquiring it, a negative by not buying it and none if he has not viewed the yet. Other typical conversions, e.g. for newsgroups, are measuring the time a user spends for looking at an item. Such conversions are widely discussed (Caglayan et al., 1997; Konstan et al., 1997; Oard and Kim, 1998; Schein et al., 2002; Middleton et al., 2004). Adomavicius and Tuzhilin (2005) warn that such nonintrusive ratings are often inaccurate and therefore see the need to further explore "the problem of minimizing intrusiveness while maintaining certain levels of accuracy of recommendations".

For the corporate portal recommender system, the conversion takes place based on service usage data. The data is determined by logging every click (every action) the user executes. As described in Section 2.3.1, the accumulated click-data is converted to visits to services. The system can hereby differentiate between multiple clicks on a service which indicates that the visitor has *used* the service and a one click visit which merely indicates a visit without usage. With this logging concept, a matrix is filled with the number of service visits for each user. This matrix can include the

number of visits per interval as third dimension. As later shown, this information can be relevant for the conversion to rating values (cf. Section 4.4.2). The matrix also reveals the total number of visits per service by all users. For this data conversion rules have to be defined to be used to fill the user-item matrix which is required for the collaborative filtering algorithms to compute recommendations.

This approach has several advantages compared to explicit ratings but also to other nonintrusive concepts. It uses every interaction with the system and thus fills the user-item matrix with all available data (Buono et al., 2002). Compared to explicit ratings, the problem of sparse ratings does not exist (Delgado and Ishii Nahoiro, 1999) and, further, users are not required to provide any input, thus avoiding freeriding, incentive challenges and decay in interest.

Explicit and implicit rating concepts can also be combined to profit of the strength of each approach (Buono et al., 2002). Here the role of an explicit rating can be to confirm or contradict implicit derived ones to measure the accuracy of the implicit rules. Also they can replace or strengthen the weight of an item compared to those that have not been explicitly rated by a user.

### 4.4.2. Implicit Rating Conversion Rules

The computation process to retrieve implicit ratings from usage data consists of three steps:

- collection of usage data (cf. Section 2.3.1),
- consolidation to service visits per user in a service-visit matrix,
- transformation of service visits to ratings (user-item matrix).

The relevant usage information in context of a corporate portal is the number of active visits per service. A visit is called *active* when the user not only opened a service but used it by executing an action. For example, if a user opens a service and requests a report, enters a trade or downloads a document, the visit was active. If he just opens it, e.g. to navigate to another service which is located in a sub-branch of the opened one, the visit is not considered relevant.

The following parameters have to be specified for the transformation from the service-visit matrix to the user-item matrix :

- the observed time frame, optional with
  - division into intervals or
  - decay algorithms,
- translation rules from number of visits to ratings (relative or absolute)

The length of the observed time frame depends on the type of portal. If all portal services are frequently used, this interval can be set to two or three months. This will allow compensating for temporary usage fluctuations due to illness, holidays or other circumstances that can influence the usage pattern. If some of the services are seldom used or are only relevant for certain time periods the time frame should be longer. The decision depends on the available information. If it is known that some services are only used once per month or quarter, the interval should encompass at least two of these usage periods to be able to determine average usage patterns based on at least two periods.

With a specified time frame it is further possible to favor recent over older data. For example, using a six month time frame, if a user increased the usage of a service in the past two months compared to the previous four months, the resulting rating value should reflect this matter by favoring the recent visit numbers. This can be achieved by (i) dividing the time frame into intervals and weighting more recent intervals higher or by (ii) using a decay factor.

When using intervals the rating $r_i$ for a service $i$ is computed with

$$r_i = \sum_{t=1}^{T} w_t r_{i,t} \tag{4.4}$$

where $T$ represents the number of intervals in the observed time frame, $w_t$ the weight that each interval is given and $\sum_{t=1}^{T} w_t = 1$. The weight for each interval and the number of intervals has to be determined by testing different values and comparing the resulting ratings. The goal thereby is to find a setting that generates high and low ratings instead of values in the middle of the rating scale. If most of the

ratings are located in the middle of the scale, the algorithm to compute recommendations cannot differentiate between items because the differences between the input variables (the ratings) are insignificant. Therefore, the final setting should be the one that returned the highest standard deviation. By increasing the number of intervals, a liner decay function can be derived (Zanker and Gordea, 2006).

*Conversion rules* have to be defined that specify how visit numbers are converted to a value on the rating scale. In the following it is distinguished between rule based on absolute and ones based on relative values. *Absolute transformation rules* define a range of visits for each available rating value. E.g. one to four visits per interval are represented with a rating value of one, five to ten with two and so on. The selection of the range for each value again depends on experience and testing.

*Relative rules* take different usage patterns per user or per service into account. This is relevant to better rate visits for less frequently used services or for rather inactive users. Table 4.4 shows an example for a service-visit matrix with three users and three services. User U3 is comparably inactive and service S3 is less frequently visited by all users compared to the other two. Absolute rules would return low ratings for all services U3 is using. However, the services might have a high relevance for this user since he uses all equally. Thus, the low ratings given by absolute rules will not represent the real relevance. Similarly the rating for S3 would be low since the average usage of 2.67 is lower compared to S1 and S2 although all users use the service with equal frequencies. This could indicate that S3 is relevant to all users but does not require high usage to be of relevance. Thus, a higher rating could be adequate. A simple method to derive rules based on average usage is to compute the relative distance to the average visit numbers with

$$dist = \frac{v_{u,i} - \overline{v}_i}{\overline{v}_i} \tag{4.5}$$

Here, $v_{u,i}$ represents the number of visits to service $i$ by user $u$ and $\overline{v}_i$ the average number of visits either for the service or the user. Distance ranges have to be defined that represent the rating for a specific distance range.

| | *S1* | *S2* | *S3* | **Total** | **(avg)** |
|---|---|---|---|---|---|
| *U1* | 20 | 15 | 2 | **37** | **(12.33)** |
| *U2* | 18 | 15 | 3 | **36** | **(12.0)** |
| *U3* | 3 | 2 | 3 | **8** | **(2.67)** |
| ***Total (avg)*** | **41 (13.67)** | **32 (10.67)** | **8 (2.67)** | | |

**Table 4.4.:** Service-visit matrix example with users U and services S. Each field in the matrix represents the number of visits for each service user combination in an interval.

Relative and absolute rules can be combined by adding or reducing a rating that was derived by an absolute rule if the distance surpasses a limit. In the given example in Table 4.4, the ratings for S3 would be raised for all users because the distance of the visits of each user to the average usage of S3 is close to zero. More sophisticated approaches are possible. It is, however, questionable whether too special rules diffuse the reliability of the ratings. If the services for user U3 in the given example are indeed not highly relevant for him, relative rules would lead to invalid ratings. Also a comparably low usage average for a service can truthfully indicate that it is not highly relevant for all users. Therefore, relative rules have to be used carefully and more information about the recommended services should be integrated in the rules. For example, if a service is known to require few visits to be of high relevance, relative rules can be useful. The collaborative filtering technique also recommends services to a user if many similar users are using it independent of its ratings. The resulting relevance of the recommendation will be lower for such services, but it will be on the list. Too low ratings can therefore delay the appearance of a recommendation at the top of the list but will not prevent it.

### 4.4.3. Model-based vs Memory-Based Algorithms

The user-item matrix, which contains the service ratings for each user, builds the foundation for the computation of recommendations. Several algorithms exist to compute the predicted relevance of a service for each user using this data. Breese et al.

(1998) group these algorithms in *memory-based* and *model-based* algorithms. Memory-based algorithms operate over the entire data set of users and items while model-based algorithms only use historic data sets to train a model which is then applied for the computation of recommendations. This widely accepted distinction (Calderón-Benavides et al., 2004; Adomavicius and Tuzhilin, 2005) serves as basis for the remainder of this section.

*Model-based algorithms* do not use the entire available set of data.[9] Instead, historic data is taken to develop and train a model that is then used to compute the recommendations. The common process is to split the historic data into a training and a test set. The training set is used to develop the model which is then tested on the test set. Since a model is based on historic data, the model should be regularly refreshed to react to changes to the system. Breese et al. (1998) distinguish between cluster models such as the expectation maximum algorithm and Bayesian networks which use decision trees. Other concepts like neuronal networks or machine-learning also fall into this category (Adomavicius and Tuzhilin, 2005).

*Memory-based algorithms* are using the available set of data to compute the predictions *online* with the latest available information. One main challenge of this concept is the required effort for the computation of large data sets (Breese et al., 1998). The entire user-item matrix of existing ratings is used to compute similarities between users and, based on these similarities, the predicted rating for each item that a user has not rated yet. The computational effort for this approach is therefore higher compared to the model-based approach (Sarwar et al., 2001).

The *advantages* of memory-based algorithms are the usage of real-time data which allows to directly incorporate new users and items and to immediately react to changes in the environment (Breese et al., 1998). In such cases the model-based algorithm needs to train a new model. In comparison, the main advantage of model-based algorithms is its scalability (Breese et al., 1998; Ziegler and Lausen, 2004; Schafer et al., 2007). Since the models are trained offline, the computation of recommendations is comparably fast. However, the prediction quality is significantly lower

[9] The entire set of data consists of all historic usage information and ratings in addition to the latest ones.

compared to collaborative filtering (Breese et al., 1998; Runte, 2000; Adomavicius and Tuzhilin, 2005). Therefore, and because scalability is no major issue in corporate portals, model-based algorithms are not further discussed in detail.

### 4.4.4. User Similarity Algorithms

In order to compute the prediction $p_{a,j}$ of the rating that the active user $a$ would give to an item $j$, Equation 4.3 requires a value that expresses the similarity $sim(a,u)$ between $a$ and all other users $u$. The similarity is based on the similarity of ratings which two users have given to other items. Thus, for users who have no overlapping ratings, the similarity is zero. The similarity between two users can be described as *weight* that reflects distance or correlation between two users.

Several algorithms to compute user similarities are presented in recommender literature. Following Breese et al. (1998) the following three algorithms are further discussed in this work:

- Pearson's correlation coefficient,
- vector similarity (also cosine-based similarity), and
- inverse user frequency.

Among other popular algorithms is the *mean squared difference* (Shardanand and Maes, 1995; Pennock et al., 2000) measures the degree of dissimilarity between two users. The similarity is than computed by taking the users with the lowest dissimilarity compared to the active user. The algorithm is almost identical to the vector similarity and thus not further discussed.

A *constrained* version of Pearson's correlation algorithm (Shardanand and Maes, 1995) exists to support positive and negative ratings. This algorithms is also excluded from further discussion because negative ratings are not considered in the corporate portal recommender concept.

*Default voting* (Kamishima, 2003) is designed for situations in which few items are rated. The method ranks all rated items per user and inserts unrated items in the middle of the resulting list. Thus, a default neutral rank is guessed for these items.

This method is useful for cases with explicit ratings. However, in the corporate portal context with relatively few items and a strong preference for implicit ratings, the concept is rather irrelevant.

#### Pearson's Correlation Coefficient

The basic approach to compare two sets or ratings from different users $a$ and $u$ is *Pearson's correlation coefficient*

$$sim(a,u) = \frac{\sum_j (r_{a,j} - \overline{r}_a)(r_{u,j} - \overline{r}_u)}{\sqrt{\sum_j (r_{a,j} - \overline{r}_a)^2 \sum_j (r_{u,j} - \overline{r}_u)^2}} \tag{4.6}$$

where the summation over $j$ is over the items that both users have rated. Billsus and Pazzani (2000) criticize this matter as the main disadvantage because all other ratings are neglected when comparing two users. Resnick et al. (1994) initially applied it for the *GroupLens* project. For systems with heterogeneous users and items the correlation is likely to be lower than in more homogeneous system. Therefore, its applicability for corporate portals with such heterogeneous properties is questionable.

#### Vector Similarity

In information retrieval, similarities between documents are often determined by treating each document as a vector of word frequencies and computing the cosine of the angle between two vectors (Salton and McGill, 1986). For collaborative filtering this approach is adopted to treat users as documents, items as words and ratings as word frequency (Breese et al., 1998; Sarwar et al., 2001; Linden et al., 2003). The algorithm is defined as

$$sim(a,u) = \sum_j \frac{r_{a,j}}{\sqrt{\sum_{(k \in I_a)} r_{a,k}^2}} \frac{r_{u,j}}{\sqrt{\sum_{(k \in I_u)} r_{u,k}^2}} \tag{4.7}$$

The equation does not permit negative ratings. If a user has not rated an item, a rating of zero is taken which removes the item from the algorithm. The squared

denominator normalizes the ratings so that users who have rated many items will not automatically have higher weights than others.

**Inverse User Frequency**

The *inverse user frequency* algorithm is also taken from the field of information retrieval. The idea is to reduce the weight of popular words based on the assumption that a high frequency of such words does not express the relevance of a document. Applied for collaborative filtering, this concept will reduce the impact of universally liked items (Breese et al., 1998). The algorithm uses $f_j = \log \frac{n}{n_j}$ where $n_j$ is the number of users who have rated the item $j$. If everyone has rated item $j$ the value $f_j$ is zero. Applying this technique for the correlation algorithm leads to the similarity algorithm:

$$\begin{aligned} sim(a,u) &= \frac{\sum_j f_j \sum_j f_j r_{a,j} r_{u,j} - (\sum_j f_j r_{a,j})(\sum_j f_j r_{u,j})}{\sqrt{AU}} \\ A &= \sum_j f_j (\sum_j f_j r_{a,j}^2 - (\sum_j f_j r_{a,j})^2) \\ U &= \sum_j f_j (\sum_j f_j r_{u,j}^2 - (\sum_j f_j r_{u,j})^2) \end{aligned} \tag{4.8}$$

The main differences between the three algorithms are that the first allows negative ratings, which is irrelevant in corporate portals as there are only positive ratings. The latter two consider all rated items while the correlation coefficient uses only corated items.

Empirical tests by Breese et al. (1998) have identified Pearson's correlation coefficient as the algorithm with the best performance. Also, Konstan et al. (1997) have shown that the algorithm is best suited for continuous rating scales which are used in this context. According to (Calderón-Benavides et al., 2004) the algorithm is best suited as it best represents the basic collaborative filtering concept. It is questionable whether the special factors in a corporate portal environment will support these

findings. Few overlapping ratings might influence the results. Therefore, it has to be tested which algorithm delivers the best results.

## 4.5. Neighborhood Selection Concepts

The computation of similarities between users defines a *neighborhood* $U_a$ for each user $a$ that consist of other users who use at least one similar service as $a$. In systems with many users, the size of the neighborhood can reach levels that the quality of the predictions or the effort for the computation of recommendations could suffer. Therefore, three concepts exist for the selection of the neighborhood.

### 4.5.1. All Users

This concept does not filter any users from the neighborhood. All users, who have any similarity with the active user $a$, are member of his neighborhood. Many recommender concepts use this approach (Herlocker et al., 1999; Montaner et al., 2003). It is useful if the potential size for the neighborhood is so small that any filters would diminish it in a way that no useful recommendations could be computed.

For systems with many users, this approach can cause scalability problems. Large neighborhoods increase the computational effort. The determination whether a neighborhood is too large is case dependent (Sarwar et al., 2001).

### 4.5.2. Good Neighbors

Shardanand and Maes (1995) limit the neighborhood to users who surpass a predefined similarity level. Only users who fulfill this criteria are incorporated in the computation of recommendations for the active user $a$. This approach ensures that only *good* neighbors that share a common set of attributes are considered. Eliminating all other potential neighbors with a low similarity value prevents the system from recommending items of little relevance. However, if the active user has several neighbors with a strong similarity, their *weights* will already ensure that items from the *weak*

neighbors will have little relevance. On the other side, if the number of *weak* neighbors is significant enough, their impact could cause *weak* recommendations. Montaner et al. (2003) have determined that in some cases to large neighborhoods can reduce recommendation quality. Therefore, this approach is especially well suited for environments with many users. It also limits the required computational effort by reducing the neighborhood size.

Reducing the neighborhood size will become a disadvantage if the resulting neighborhood for a user becomes too small (Montaner et al., 2003). Therefore, the radius for the selection has to be tested for each system.

### 4.5.3. Fixed Neighborhood

Similar to the *good neighbor* concept, limiting the size of the neighborhood to a fixed limit $N$ ensures that users with low similarities will not be considered. Here, all potential neighbors are sorted by weight and the resulting list is cut after $N$ users (Herlocker et al., 2004). The value for $N$ has to be tested. A too small value will reduce the quality of the recommendations and too large values might reduce the quality in case of many *weak* neighbors. Again, this approach also reduces the computational effort.

For the selection of the appropriate neighborhood concept for a corporate portal recommender system, the simple concept to take *all users* seems suitable. The potential neighborhood size in such heterogeneous systems with comparably few users is already low so that further reductions could prevent the generation of useful neighborhoods. The definition of *low* thereby depends on the size and properties of the target system. Therefore, testing is required for each implementation. If the average neighborhood size becomes too large, one of the other concepts should be used.

## 4.6. Overview of Prediction Algorithms

When the neighborhood $U_a$ for each user $a$ and the weights for the similarity $sim(a,u)$ for each user $u$ within the neighborhood have been determined, the predictions for each item $j$ that the user $a$ has not rated can be calculated. Equation 4.3 introduced the common approach, however, more exist (Breese et al., 1998).

### 4.6.1. Simple Average

The most simple and "naive" approach (Schafer et al., 2007) is to take the average of all ratings for an item which is computed with

$$p_{a,j} = \frac{1}{m} \sum_{u \in U_j} r_{u,j} \tag{4.9}$$

where $U_j$ specifies all users who have rated the item $j$. This algorithm ignores the similarities between users and thus does not belong to the group of collaborative filtering algorithms. However, it can serve as a benchmark to evaluate the quality of other algorithms. If the average calculation predicts the relevance of a recommendation with a comparable quality to other algorithms, it should be used because the computational effort is lower.

The computation of the average rating can also be limited to the ratings of users in the neighborhood of the active user.

### 4.6.2. Weighted Sum

The *weighted sum* includes the similarities between the users. Ratings of users with a higher *weight* have a higher impact on the prediction than others.

$$p_{a,j} = \frac{\sum_{u \in U_j} sim(a,u) r_{u,j}}{\sum_{u \in U_j} |sim(a,u)|} \tag{4.10}$$

Dividing the added weighted ratings by the number of neighbors serves as normalizing factor.

### 4.6.3. Adapted Weighted Sum

The *weighted sum* algorithm does not incorporate different rating behaviors. When ratings for items are determined by explicit ratings, the resulting rating values might not be directly comparable. Some users will generally give higher or lower rating values[10] than others (Resnick et al., 1994; Buono et al., 2002). The adapted weighted sum algorithm considers this with

$$p_{a,j} = \bar{r}_a + \frac{\sum_{u \in U_j} sim(a,u)(r_{u,j} - \bar{r}_j)}{\sum_{u \in U_j} |sim(a,u)|} \tag{4.11}$$

Each rating for an item is subtracted by the user's average rating. Thus, the fraction delivers the mean deviation between each user's average rating and the one given for the item $j$. By adding the average rating of the active user $a$ the algorithm computes the prediction under the assumption that the user will deviate in his rating of this item similarly to all his neighbors. This approach is similar to Equation 4.3 when $\kappa$ is replaced by $\frac{1}{\sum_{u \in U_j} |sim(a,u)|}$.

The *adapted weighted sum* has been applied in many systems (e.g. Resnick et al. (1994), Herlocker et al. (1999), Buono et al. (2002), Adomavicius and Tuzhilin (2005)). Its main advantage is the "normalization" of ratings which is of high relevance for systems that are based on explicit ratings. The effectiveness compared to the *weighted sum* in systems with implicit ratings is questionable (cf. Section 4.4.1).

The simple average algorithm does not really belong to the collaborative filtering algorithms but might serve as benchmark to the other two. The adapted weighted sum is well suited if the item ratings are derived explicitly. For implicitly computed ratings, the normalization of the data is not necessary. However, in case that the usage patterns are very different in a form that some users are highly active (which implicitly returns high ratings) and others barely, the adapted weighted sum might lead to better results as it considers these differences.

[10] In systems using a rating scale, e.g. a five-point Likert scale.

## 4.7. Choosing the right Algorithms

The previous sections introduced several concepts for each step in the process of computing recommendations with the collaborative filtering technique. They are listed in Table 4.5. Before a recommender system can be implemented in an existing portal, it has to be decided which combination of these concepts is applied. This decision has to be made by testing different compositions. Deciding individually for each step is difficult because each process step depends on the outcomes of the previous one. E.g. the user similarities are computed based on the ratings that were retrieved with the selected rating concept. Different rating concepts will return different values which will affect the user similarity values for two users. Therefore, a test of all possible combinations is required to determine the best set.

| *Process step* | *Concept* |
|---|---|
| *Rating determination* | - Explicitly<br>- Implicitly from usage data |
| *User similarity* | - Pearson's correlation coefficient<br>- Vector similarity<br>- Inverse user frequency |
| *Neighborhood selection* | - Take all<br>- Good neighbors<br>- Top-N |
| *Prediction algorithm* | - Simple average<br>- Weighted sum<br>- Adapted weighted sum |

**Table 4.5.:** Collaborative filtering algorithms and concepts.

The foundation for the testing process is laid by the comparison of recommendation algorithms. Several authors have compared the quality of recommender algorithms with empirical tests (Breese et al., 1998; Calderón-Benavides et al., 2004; Herlocker et al., 2004; Zanker et al., 2007). For this purpose they have used rating data of existing sites. Their testing concepts can also be applied for testing of new implementations.

### 4.7.1. Testing Methods: All But X and Given X

The methodology to compare algorithms is based on concepts called *All but X* and *Given X*. Both concepts are based on predicting the relevance of an item whose rating is known. Thus, the quality of a recommender system can be measured by comparing the value of the predicted relevance with the original rating value.[11]

For the *All but 1* testing method, one random is user is selected from the data set who has at least rated two items. In the next step, the rating for one item is randomly removed. Based on the new set of rated items for this user, the prediction for this one removed item is computed with all algorithms that are subject to the comparison (Breese et al., 1998).

The concept can be extended to *All but X* by removing more than one item from the set of rated items. Herlocker et al. (2004), for example, withhold five ratings for their tests.

The *Given X* concept is quite similar. The difference is that instead of removing one or more rated items, All but X items are removed. E.g., with *Given 5* and a user with 11 ratings, six ratings would be removed and the prediction for all of these six items would be calculated.

To prepare such tests, all possible compositions of algorithms are specified in a list. For example, if two implicit rating methods, three user similarity algorithms and three prediction algorithms are available, 18 compositions are possible. As basis for the test runs, historic usage data is taken for a time frame that has to cover a sufficient range in case of implicit rating computation (cf. Section 4.4.2).

For each compositions Given X and All but X tests are executed with different values for X. Examples in literature use All but 1, Given 2, 5, 10 and 20 (Breese et al., 1998; Calderón-Benavides et al., 2004; Zanker et al., 2007). For the context of corporate portals with few services Given 10, 20 and maybe even Given 5 might not result

[11] The case study in the following chapter presents an example for comparing different test runs in Section 5.1.3.

in usable data sets in case the average number of services per user is too low. The decision depends on this value.[12]

### 4.7.2. Quality Metrics

Once the predictions for the excluded items have been calculated, the predicted relevance values can be compared with the original ratings. Several quality metrics are suitable for this process step. The results of these metrics reflect the ability of the chosen algorithm to predict the correct relevance of an item for a user. The most popular quality metrics exist are grouped by coverage, accuracy, precision and recall (Herlocker et al., 2004).

#### Coverage

Coverage measures the percentage of items that the recommender system can provide predictions for (cf. Section 4.3.2). Coverage is computed with

$$coverage = \frac{I_{recommended}}{I} * 100 \tag{4.12}$$

$I$ is the number of all recommendable items in the system. Coverage values should be as high as possible. Calderón-Benavides et al. (2004), for example, have mainly coverage values of 99%. System with few users, many more items than users or barely overlapping ratings usually have a smaller coverage (Herlocker et al., 1999). When a recommender system has a bad coverage, many items will not be recommended. The best prediction algorithms are worthless with a bad coverage.

#### Accuracy

Accuracy metrics compare predictions and ratings as, for example, given by the Given X or All but 1 concepts. Common metrics are the mean absolute error (MAE) (Shardanand and Maes, 1995; Sarwar et al., 1998) or the root mean squared error (MSE) (Sarwar et al., 1998). Appendix F lists the most common metrics. Only a few

[12] In the case study presented in Chapter 5 the tests are run with a maximum of Given 5 because less than 5% use more than five services on average.

of these will be further discussed as the others do not fit the requirements of this work. For example, the *variance MAE* delivers information about the mean variance of the results. *Swet's a measure* metric displays the area below the *receiver operation characteristic (ROC) curve* which is not relevant if the recommender should display all relevant items (Herlocker et al., 2004). Therefore, only five metrics are presented in detail:

- Mean Absolute Error MAE,
- Mean Squared Error MSE,
- Pearson's Correlation Coefficient,
- Half-Life Utility Metric,
- Top-N Evaluation,
- Precision and Recall.

**Mean Absolute Error:** This metric is often applied in the analysis of time series. In the recommender context, it is a measure for the accuracy of the predictions. Each prediction $p_j$ for an item $j$ is compared to the original rating $r_j$ with

$$e_{mae} = \frac{1}{M} \sum_{j=1}^{M} |p_j - r_j| \tag{4.13}$$

where $M$ is the number of predictions.

**Mean Squared Error:** The mean squared error is defined as

$$e_{mse} = \frac{1}{M} \sum_{j=1}^{M} (p_j - r_j)^2 \tag{4.14}$$

In comparison to the MAE metric, higher differences between prediction and rating are emphasized. A small value indicates few deviations between prediction and rating and thus a good result. Compared to the MAE, the MSE serves as indicator for the *reliability* of the algorithm while the other measures *accuracy*.

**Pearson's Correlation Coefficient:** The Pearson Correlation Coefficient which also serves as user similarity algorithm (cf. Equation 4.6) measures the strength of linear dependence between two variables. The higher the dependency, the closer the variables are together. Thus, for recommender testing, a high positive correlation between predicted and removed ratings is an indicator for good results.

However, a careful interpretation of the results is required. The values are not likely to get close to 1, the best possible value. Deviations between predicted and removed rating in systems with a rating scale of 5 or 7 values are to be expected. Therefore, a positive correlation result already can already serve as good indicator for the quality of the system.

**Half-Life Utility Metric:** Breese et al. (1998) use the MAE metric only for the evaluation of single predictions. To measure the accuracy of a list of recommendations, they use the *Half-Life Utility Metric*. They argue that the relevance of a recommendation in a ranked list (sorted by index $j$ in order of declining $r_{a,j}$) will decline for items that are lower in the list. The metric takes this into account by assuming that each successive item in the list is less likely to be viewed by the user with an exponential decay. The expected utility of a ranked list of items is calculated with the formulas

$$\begin{aligned} R_a &= \sum_j \frac{max(r_{a,j} - d, 0)}{2^{(j-1)/(\alpha-1)}} \\ R &= 100 \frac{\sum_a R_a}{\sum_a^{max} R_a} \end{aligned} \tag{4.15}$$

where $d$ is the neutral vote and $\alpha$ is the viewing half-life. The half-life is the number of the item on the list such that there is a 50-50 chance that the user will review that item (Breese et al., 1998)[13]. $R$ defines the final score with $R^{max}$ as the maximum achievable utility if all items had been at the top of the list.

[13]Breese et al. (1998) used a half-life of five items in their experiments.

The half-life utility is a good method to cope with longer list of rankings. For environments with few items which accordingly have a shorter list of recommendations it might not be necessary to apply the metric.

**Top-N Evaluation:** Comparable to the half-life utility metric, Lam (2004) proposes to limit the list of recommended items, in case of long lists, to the *top-N* results because the others will not receive any attention from the users. Measuring the percentage of relevant items in the list for each tested algorithm allows comparing their quality. This concepts is quite similar to precision and recall as described in the following.

#### Precision and Recall

Precision and recall are popular metrics for the evaluation of information retrieval systems (Cleverdon and Keen, 1968) and are equally popular for the evaluation recommender systems (Billsus and Pazzani, 1998; Sarwar et al., 2001; Basu et al., 1998; Herlocker et al., 2004; Zanker and Gordea, 2006; Schafer et al., 2007).

The input factors for both equations are all relevant[14] items $R$ and all predicted items $P$. $P_r$ represents all predicted items that are indeed relevant. Precision and recall are then defined as

$$precision = \frac{P_r}{P} \tag{4.16}$$

$$recall = \frac{P_r}{R} \tag{4.17}$$

*Precision* represents the probability that a predicted item is relevant, *recall*, on the other hand, the probability that a relevant item will be predicted. Both values have to be evaluated together. To achieve a high recall value, the system would only need to recommend all items. This would, however, lead to a low precision value.

[14] *Relevant* items are those that should be recommended.

### 4.7.3. Test Evaluation

All introduced quality measurement metrics are summarized in Table 4.6. By executing Given X and All but X test runs for every composition of available algorithms, a list of quality values is created for each composition. Based on these values the compositions can be compared to each other. Thereby each types of metric focus on different aspects of quality.

| *Metric* | *Indicator for* |
|---|---|
| Coverage | Ability to cover all items in the system |
| Precision | Probability that recommended items are relevant |
| Recall | Probability that relevant items are recommended |
| Mean Absolute Error (MAE) | Accuracy of the predictions |
| Mean Squared Error (MSE) | Reliability of the system |
| Pearson's Correlation Coefficient | Linear Correlation |

**Table 4.6.:** Test evaluation metrics.

Coverage, precision and recall indicate the overall ability of the system to efficiently detect all recommendable items while the other metrics measure the quality of the predictions. Bad results for the first metrics make further testing irrelevant because the best prediction quality is worthless if not all relevant items are detected - especially with the specified goal to *find all good items* (cf. Section 4.1.3).

MAE and MSE reveal the accuracy and reliability of the system. When comparing different test runs with these metrics, the ones with the lowest values indicate the better system. The result can be further evaluated by comparing the linear correlation between the predicted and observed ratings.

## 4.8. Communicating Recommendations to User

The final step in the implementation of a recommender step is the decision how to communicate the computed recommendations to the users. Recommendations can be delivered to the users via different *channels*. Most common is the display of the

recommendations in the portal but other external channels (outside the portal) such as emails or trainings are possible. This decision does not only affect the channel and design but also the number of recommendations and the information level being communicated to the users. Further considerations have to weigh the benefits of recommendations versus the *annoyance factor* of the system.

### 4.8.1. Displaying Recommendations in the Portal

The integration of the recommendation display into the portal's interface can follow several patterns. Each pattern has its advantages and disadvantages regarding visibility, detail level and obtrusiveness. Table 4.7 lists all display patterns with these attributes. The *visibility* indicates the likelihood that a user will perceive the recommendation. For example, the display in the footer of the portal has less visibility than opening a pop up window. When showing a recommendation, different *detail levels* are possible which range from simple links to the recommended service up to a detailed explanation of the recommendation computation or a description of the recommended service. High detail levels can increase the likelihood that a user will look at a recommended item but might also prevent him from reading the recommendation if too much information is communicated. A critical matter is the *obtrusiveness* of the recommendation display. Highly visible recommendations such as pop-ups or messages in the center of the portal also have a high potential of annoying users. Pop ups interrupt the flow of working with the portal because they have to be closed in order to continue to work with the portal. Centrally located messages force users to scroll to further continue using a service.[15] The obtrusiveness factor is the more important in corporate portals since users are supposed to work efficiently. Therefore, the challenge is to find the right balance between visibility and obtrusiveness so that recommendations are perceived but do not prevent users from working without interruptions.

For corporate portals the display in content and as pop up appears too annoying. More suitable are the display in footer, sidebar or header. Here, the sidebar is

[15] News portals often place advertisements in the middle of an article. In the flow of reading, users will perceive those but are interrupted from reading the article and have to scroll to continue reading.

| *Pattern* | *Comment* | *Visibility* | *Obtrusiveness* |
|---|---|---|---|
| *Pop up* | Open a new window with one or more recommendations. Possible to close the pop up when the user has not activated it within a time frame. | high | high |
| *Header* | Display in the header of the portal as message or advertisement. | middle | low |
| *In content* | Message in the content (E.g. advertisements in news articles). | high | high |
| *Footer* | Display in the footer. (Similar to header). | low | low |
| *Sidebar* | Message in the sidebar e.g. above or below the menu (E.g. used by Facebook to recommend possible friends). | middle | low |
| *Mailbox* | Show as message in the mailbox of the portal (if available). | middle | middle |

**Table 4.7.:** Display patterns for recommendations in portals.

preferable because users are more likely to perceive it when the recommendations are located next to the menu. Using a portal mailbox can be effective. Such mailboxes are used to deliver messages from the portal to the users using an internal system rather than sending an email to the user's mailing account. Such systems can be used to inform the user about the finish of a service's process, the delivery of a report or other issues. Using this system to deliver recommendations will lead to good results, if the portal frequently delivers other relevant messages. Otherwise the user might ignore the entire mailbox because this will not affect any of his job tasks.

### 4.8.2. External Communication of Recommendations

Besides the display within the portal, the recommendations can be delivered outside the portal. Similar to the portal mailbox, which was introduced as internal channel, is the delivery of personalized mails to the user's mailing account. Such mails can be sent in with a fixed interval or when new recommendations have been computed. The advantage of this approach is that users will notice them when checking their regular

mails. However, this channel might also annoy users as it delivers information from the portal while they are not using it.

An alternative concept is to use recommendations in portal workshops or personal trainings. The recommendations can be used to further discover the portal during such events.

Furthermore, the recommendations can be sent in a report to the user's supervisor. The supervisor can analyze the recommendations and forward relevant ones to his subordinate. This process ensures that only relevant recommendations are forwarded. However, it delivers a report about usage patterns to managements which might contradict privacy issues.

Communication of recommendations outside the portal context is possible but has to be considered critical as it requires additional effort and could lead to irritations since the recommendations are not delivered while the recipients are using the portal.

### 4.8.3. Number of Displayed Recommendations

Recommendations are computed with a relevance that represents the prediction how the user would rate the service. By sorting all recommendations for each user descending by this value, a ranked list is created with the most relevant recommendations at the top. Lam and Riedl (2004) suggest that users will only look at the top of such lists and ignore lower rated ones. Users are not willing or interested in looking through a long list of recommendations.

The length of a ranked list can be restricted by defining a *barrier value* which prevents the display of recommendations whose relevance value is below the barrier. This concept is comparable with the good neighbors approach, which was introduced in Section 4.5.2. An alternative or addition is to set a limit for the *top N* items in the list (Karypis, 2001; Deshpande and Karypis, 2004).

When limiting the number of shown recommendations, it has to be decided how to handle the recommendations that are not displayed. If a user visits a recommended service the recommendation will disappear from the list and allow the next best one to appear. The question whether to show this one immediately or delayed is discussed

in the following section. The recommender system could also allow users to look at the next recommendations in the list even if they are below a barrier value. Thus, interested users can explore the recommendations by themselves while the others are not bothered with long lists. Implementing such a system requires more effort but can be helpful for curious users. It might also help them understand the recommender system when they see recommendations in context of a longer list because it might reveal the type of services that are recommended to them.

Limiting the list of recommendations is required to not overwhelm or irritate users with too many recommendations. For a corporate recommender system a maximum of five or less services can be useful as the user has to explore each service on its usefulness. Other recommender systems such as Amazon's book recommender can show more recommendations because the title of each recommended book allows the user to quickly decide whether it is of interest or not. If services are recommended, the title can seldom sufficiently describe the functionality. Therefore, users will have to explore it. Showing too many recommendations will prevent them from starting because they first have to choose which service to look at. Showing five or less recommendations allows quickly deciding which service to further explore.

### 4.8.4. Frequency and Detail Level of Recommendations

When users visit a service that they have not used before, based on a recommendation or not, their service portfolio changes and thus their similarities to other users might change. This can lead to new recommendations or changed relevance values of existing ones. Thus, the list of recommendations can change. Furthermore, users might choose to decline some recommendations,[16] thus removing them from the list.

For these cases it has to be decided whether the new list of recommendations is displayed at once (when a user has visited or declined a recommendations or when new once are created) or based on a predefined schedule. Such a schedule could specify that the list is refreshed every morning, once per week or at any other given point of time. This applies only for a new order in the list or new recommendations.

[16]The recommender should allow users to decline recommendations either because they do not see any relevance in the recommended item's description or because they already know the item.

Declined or visited recommendations have to be removed from the list immediately to not confuse the user.

The advantage of showing the refreshed list at once is the possibility for users to directly explore the new recommendations. If a user has started looking at recommendations he might be interested in continue looking at further ones even if he has looked through the entire initial list. However, if the system provides the opportunity to further explore the list as described in the previous section, the user can already do that even if the initial list remains unchanged.

Against the immediate refreshing of the list stands the feeling of arbitrariness. If ever new items are added to the list when one is removed, users can get the impression that the recommendations are just created randomly. The motivation to further look at the recommendations will decline. The user will have the feeling that he cannot finish the list because new items are always added. This argument strongly speaks for refreshing the list on a fixed schedule. When a user receives new recommendations only once per week or even more seldom, his interest might be higher. The new list can create the feeling of fresh information compared to always seeing a list of N recommendations.

An important factor beside the number of items is the detail level of the recommendation. A short list of service titles might not carry enough information to motivate users to look at the services. Better is to provide a short description of the service that summarizes its key benefits.

Since every recommendation is computed with a predicted relevance it is possible to show this relevance. Here it can be chosen between showing the real value or scales with “highly relevant” for the highest and “possibly relevant” for the lowest values. When showing the relevance level, a short explanation should be available that explains the different relevance levels. Also, if the recommendations are shown without relevance values, such a description of the recommendation computation process can help in motivating the users to use the system. If they are put in a position to comprehend the system, they might gain more trust into the recommendations even if they encounter irrelevant ones.

Choosing the best patterns for the display and refresh rates of recommendations is a difficult tasks. Rather invisible recommendations have no effect and obtrusive channels and immediate refreshes will annoy users. The best solution would be to allow each user to personalize the recommender to his preferences. Thus, each user could decide whether to receive emails with recommendations, get pop ups or have a list in the sidebar. However, the effort for implementing such a system is quite high.

In all cases it is relevant to explain the recommendations. Whether to use a short text such as "users who have visited this service also visited the following" or a detailed process diagram depends on the target group but also the effort that management is willing to put into the service and recommender system description. Best would be to test different approaches with a test group to determine the one which is accepted by most users.[17]

## 4.9. Chapter Summary

Various recommendation system concepts have been introduced in this chapter and were evaluated for their qualification to be applied as a recommender system for services in a corporate portal. The literature review has shown that no comparable case studies exist that concern smaller systems with highly heterogeneous items and users as presented in this work. Often quoted sources such as the broad analysis by Breese et al. (1998) or Herlocker et al. (2004), implementations like GroupLens (Konstan et al., 1997) or Tapestry (Goldberg et al., 1992) use systems with several thousand items and users. Well-known recommender system such as Amazon's book recommendation (Linden et al., 2003) deal with numbers of millions. While Zanker and Gordea (2006) shortly introduces issues arising in smaller systems, no further literature with focus on corporate portals exist. The analysis in this chapter laid the foundation for the implementation of such a system by selecting appropriate concepts from wide spectrum of recommender systems.

[17] The case study presented in the following chapter does not compare different methods but compares a treatment group that sees the recommender system and a control group that does not.

While the decision for *automatic* systems, which provide *individual* recommendations, is quite simple, the subsequent selection process requires careful consideration of the nature and the requirements of corporate portals.

*Content-based filtering* concepts, that are based on comparing user and item attributes (personal preferences, item descriptions, job descriptions, etc.), have a high potential to identify related items. However, they require significant effort from users and administrators to provide these attributes. The concept of *collaborative filtering*, in contrast, is based on subjective parameters. Grouping users by similar ratings, which they implicitly or explicitly give to items, leads to recommendations that are based on similar likes or dislikes. The required effort to users is low or non-existing, the latter if ratings are derived implicitly. *Hybrid systems* exist that try to combine the advantages of different recommender concepts. However, this work is focused on the *collaborative filtering* concept (with memory-based algorithms) because it best fits the requirements that were stated in Chapter 2.[18]

The four steps for the computation of the predicted relevance of a service in collaborative filtering process were introduced and each discussed for its suitability for the context of this work. For each of the process steps, (i) the determination of ratings, (ii) the computation of similarities between users, (iii) the selection of a neighborhood for each user and (iv) the computation of predictions, several concepts exist.

For the first step, the determination of service ratings, which represent the relevance of services, a concept was presented to implicitly derive ratings, thus removing the need to explicitly require users to provide these. While other authors (Lieberman, 1995; Konstan et al., 1997; Billsus and Pazzani, 2000) present approaches to implicitly derive ratings from shopping data, number of references or reading time, the approach at hand converts information about repeated usage into ratings.

Taking together the various options from each of the four collaborative filtering steps, several concept compositions can be created. A testing process was introduced to compare the outcome of such compositions in order to find the one with the best prediction quality.

[18]The collaborative filtering is part of the most common hybrid systems (Burke, 2002). Therefore, all findings from this chapter can be applied for further analysis of such systems.

A final aspect for the implementation of a recommender system, which is barely discussed in literature, is the channel through which recommendations are communicated to the users. Several options were presented and discussed with focus on their visibility and obtrusiveness.

In summary, this chapter serves as a manual that can be followed when implementing a recommender system into a corporate portal. It does not only discuss various algorithms but also describes in detail how to test the quality of different concepts and how to communicate the recommendations.

# Chapter 5.

# Case Study: CoFiPot Recommender System

In Chapter 3 Bayer's Corporate Financial Portal (CoFiPot) was presented and challenges for the management of the portal were discussed. The users' lack of knowledge about the service portfolio that the portal offers was identified as a key problem. It was shown that users barely browse the portal to learn about other services. Therefore, it was decided to implement a recommender system for portal services. This chapter describes the attributes of the CoFiPot recommender system and evaluates its effects that were observed during a four month test period. During this period the users of the portal were divided into two groups: a treatment group, which received recommendations for services, and a control group, which did not receive any recommendations. This division was made to subsequently determine whether the recommender system has any influence on the usage patterns. Afterwards a survey was conducted to determine the users' experiences with the recommender system and their overall satisfaction with the portal.

The overall aim of the case study is to analyze whether the recommender system can positively influence the user's knowledge about the portal. The effects are determined by analyzing the number of visits to services that a user has previously not visited. Accordingly, the research questions for the case study were:

- R1: Does the recommender system influence the number of first-time visits to services?
- R2: Can the recommender system increase the number of actively used services?

The first section of this chapter presents the initial requirements and constraints to the recommender system and applies the selection process as proposed in the previous chapter (cf. Section 4.7). In the second section, the design of the recommender system and the setup of the case study are presented. The third section summarizes the effects of the system on the usage of the portal. The findings are further enhanced with the results of the survey in the fourth section.

## 5.1. System Characteristics

For the implementation of the recommender system into the portal, the services to be recommended, the rating system and the algorithms for user similarity and prediction have to be defined. Thereby, the following constraints must be considered.

### 5.1.1. Requirements and Constraints

The information about the number of services being offered by CoFiPot and the number of its users was introduced in Section 3.2.3. As described in Section 4.1.2, some of the 72 services are not supposed to be recommended. 8 services were excluded from the recommender system because they provide functionalities that only a fixed and known group of users is allowed to access. 26 services were flagged so that only those users will receive recommendations for these services who are member of a role that is authorized to access them. Therefore, 38 services remained to be recommended to all users. 16 of them are protected and require authorization; the other 22 are accessible for all users. For all these services the usage data of the four month period was collected before the start of the recommender system.

Subsequently, the decision between explicit or implicit service ratings has to be made. Based on the result of the expert interviews and discussions with the portal

management, only implicit ratings were used. The decision was based on the requirement that the recommender system should not disturb the users by requesting information and on the estimation that users would seldom provide explicit ratings.[1] Further restrictions were imposed regarding the display of the recommendations. These are discussed in Section 5.2.

### 5.1.2. Implicit Rating Rules

The basic concept of implicit rating rules is to derive the relevance of a service for a single user by analyzing his usage patterns (cf. Section 4.4.2). The more a service is used by a user, the higher its relevance for this user is assumed. The challenge in defining an implicit rating rule is to define a rule that returns a set of ratings with high standard deviation. If a rule returns rather homogeneous ratings, it is difficult to determine which ones are potentially more relevant when computing recommendations. A second challenge lies in defining the time frame from which the usage data is taken. E.g. a rule derives ratings based on the weekly usage. An alternative rule derives ratings based on monthly usage. Assuming that a user uses a service only in one week of a month, the first rule will return a high rating, if it considers this one week and a low one for the other weeks. The second rule, however, is less sensitive for such fluctuations and will return the same rating based on the monthly average. The length of the considered time frame depends on the usage patterns of the portal. Since the implicitly derived ratings are part of the overall process to compute recommendations, the suitability of a rule also depends on the combination of selected recommendation algorithms (cf. Section 4.7). Therefore, it is best to define multiple rules and to test them with the potential recommendation algorithms.

For the CoFiPot recommender system, four rating rules for a five-point Likert scale were defined; they are listed in Table 5.1. The basis for the determination of the rules was usage data from a four month period. They were defined in multiple steps. Based on a first version, all ratings and the resulting standard deviation were computed. Afterwards, each rule was adapted and the new standard deviation was compared to

[1] This assumption was later confirmed by the low rate of explicit ratings for visited recommendations as described in Section 5.2.2

the previous one until the results of the adapted rule became less than the previous version. In that case, the previous version was selected.

| *Rating* | *Weekly average 1 (WA1)* | *Weekly average 2 (WA2)* |
|---|---|---|
| *5* | > 2 visits per week | > 4 visits in 2 weeks |
| *4* | 1-2 visits per week | 3-4 visits in 2 weeks |
| *3* | > 2 visit in 2 weeks | 2 visits in 2 weeks |
| *2* | 1-2 visits in 2 weeks | 1 visit in 2 weeks |
| *1* | > 1 visits in 2 months | > 1 visits in 2 months |
| | *Monthly average 1 (MA1)* | *Monthly average 2 (MA2)* |
| *5* | > 8 visits per month | > 24 visits in 3 months |
| *4* | 5-7 visits per month | 13-23 visits in 3 months |
| *3* | 3-4 visits per month | 7-12 visits in 3 months |
| *2* | 2 visits per month | 4-8 visits in 3 months |
| *1* | > 1 visits in 3 months | > 1 visits in 3 months |

**Table 5.1.:** The four CoFiPot implicit rating rules used for the further selection process.

All rules are using weighted interval ratings with different interval lengths as specified in Equation 4.4.[2] The four rules were added as parameters for selection process for the recommender's algorithm composition.

### 5.1.3. Algorithm Composition

Following the rating rules, the method for the neighborhood selection had to be chosen. Although CoFiPot has several thousand registered users, only a few more than 300 users actively use more than two services (cf. Table 3.2). Therefore, the neighborhood for each user was not restricted any further to ensure a large enough neighborhood to provide good predictions. Hence, the *take all users* approach was selected (cf. Section 4.5).

The selection of four rating rules, one neighborhood composition method combined with the user similarity and prediction algorithms, which were introduced in

[2]Rules with relative ratings (cf. Equation 4.5) were tested but returned ratings with a very low standard deviation and were consequently not further tested.

the previous chapter, left three process steps with at least two choices as listed in Table 5.2. Combining all options lead to 36 possible compositions.

| *Rating Rules* | *User Similarity Algorithms* | *Prediction Algorithms* |
|---|---|---|
| Weekly Average 1 (**WA1**) | Pearson's Correlation (**Corr**) | Average Rating (**AR**) |
| Weekly Average 2 (**WA2**) | Inverse User Frequency (**IUF**) | Weighted Sum (**WS**) |
| Monthly Average 1 (**MA1**) | Vector Similarity (**VS**) | Adjusted Weighted Sum (**AWS**) |
| Monthly Average 2 (**MA2**) | | |

**Table 5.2.:** Rules and algorithms used for the CoFiPot recommender test runs.

As specified in Section 4.7, four test methods, All but 1, All but 2, Given 2 and Given 5, were run for all these 36 compositions. The MAE, MSE, the correlation and the percentage of missed predictions were computed for each test run. The latter values showed that all test-runs with *Pearson's correlation coefficient* had a missing rate of over 14% with an average of 31%. This result means that less than 86%, at maximum, of the removed ratings were actually predicted (cf. Section 4.7.2). This contradicts the requirement to find all relevant items. Consequently, all test-runs with *Pearson's correlation coefficient* were excluded from further analysis. The remaining results are listed in Table 5.3 with the top three results for each computed value set emphasized in bold. E.g. the best MAE results for the All But 1 test run were 0.512, 0.522 and 0.522.

The comparison of the values for MAE and MSE showed the best results for the MA1 rating rule, the WS and AR prediction algorithms with little differences between the two remaining user-similarity algorithms VS and IUF. Taking into account the results for the missing rate, the combinations using the VS algorithm lead to better results than the ones using IUF. The best missing rates are achieved with both weekly average algorithms in combination with VS. The results of the correlation analysis do not reveal a dominant combination but confirm a positive relationship for all values. The best correlation results occur for the VS algorithm but do not deliver clear result regarding the rating rule or prediction algorithm.

As next step, the results were further evaluated to determine which of the factors in the computation of the predictions influence the result and whether it also depends on the combination of factors. Factors in this context are the three steps in the calcu-

| Rating | Sim. | Pred. | Missing rate | | | | MAE | | | | MSE | | | | Corr | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | AB1 | AB2 | G2 | G5 | AB1 | AB2 | G2 | G5 | AB1 | AB2 | G2 | G5 | AB1 | AB2 | G2 | G5 |
| M1 | IUF | AR | 6,23% | 1,67% | 5,86% | 1,03% | 0,564 | 0,600 | **0.773** | **0.738** | 0.727 | 0.809 | **1.203** | **1.139** | 0.393 | 0.356 | 0.386 | 0.475 |
| | | AWS | 6.23% | 1.67% | 6.42% | 1.26% | 0.796 | 0.664 | 0.947 | 0.853 | 1.176 | 0.861 | 1.702 | 1.419 | 0.262 | 0.348 | 0.292 | 0.379 |
| | | WS | 6.23% | 1.67% | 5.86% | 1.03% | **0.512** | **0.533** | **0.729** | **0.704** | **0.685** | **0.720** | 1.244 | 1.173 | 0.406 | 0.395 | 0.380 | 0.477 |
| | VS | AR | 1.95% | 0.48% | 4.16% | 0.67% | **0.552** | **0.590** | 0.775 | 0.747 | **0.702** | **0.807** | **1.197** | **1.158** | 0.411 | 0.446 | 0.407 | 0.474 |
| | | AWS | 1.95% | 0.48% | 4.72% | 0.89% | 0.833 | 0.697 | 0.949 | 0.855 | 1.229 | 0.885 | 1.690 | 1.395 | 0.257 | 0.451 | 0.314 | 0.398 |
| | | WS | 1.95% | 0.48% | 4.16% | 0.67% | **0.522** | **0.554** | **0.750** | **0.733** | **0.675** | **0.746** | **1.213** | **1.162** | 0.423 | 0.487 | 0.400 | 0.478 |
| M2 | IUF | AR | 4.14% | 2.22% | 14.35% | 1.63% | 0.702 | 0.784 | 0.932 | 0.882 | 0.967 | 1.239 | 1.523 | 1.405 | 0.397 | 0.352 | 0.344 | 0.497 |
| | | AWS | 4.14% | 2.22% | 15.64% | 2.21% | 0.874 | 0.790 | 1.173 | 0.995 | 1.414 | 1.161 | 2.397 | 1.784 | 0.265 | 0.402 | 0.261 | 0.423 |
| | | WS | 4.14% | 2.22% | 14.35% | 1.63% | 0.623 | 0.713 | 0.910 | 0.863 | 0.856 | 1.112 | 1.593 | 1.451 | 0.417 | 0.399 | 0.337 | **0.501** |
| | VS | AR | 2.07% | 1.33% | 9.54% | 1.05% | 0.715 | 0.802 | 0.932 | 0.892 | 0.975 | 1.284 | 1.508 | 1.411 | **0.483** | 0.438 | 0.368 | **0.502** |
| | | AWS | 2.07% | 1.78% | 10.82% | 1.52% | 0.924 | 0.796 | 1.201 | 0.978 | 1.478 | 1.160 | 2.499 | 1.673 | 0.341 | 0.524 | 0.262 | 0.459 |
| | | WS | 2.07% | 1.33% | 9.54% | 1.05% | 0.683 | 0.769 | 0.921 | 0.879 | 0.916 | 1.202 | 1.522 | 1.402 | **0.499** | 0.473 | 0.363 | **0.513** |
| W1 | IUF | AR | 5.84% | 3.92% | 1.17% | 0.64% | 0.969 | 0.804 | 0.909 | 0.899 | 1.479 | 1.055 | 1.324 | 1.316 | 0.321 | 0.398 | **0.443** | 0.492 |
| | | AWS | 5.84% | 3.92% | 1.46% | 0.71% | 0.945 | 0.770 | 0.989 | 0.883 | 1.710 | 1.102 | 1.710 | 1.438 | 0.338 | 0.420 | 0.379 | 0.472 |
| | | WS | 5.84% | 3.92% | 1.17% | 0.64% | 0.919 | 0.783 | 0.898 | 0.893 | 1.375 | 1.036 | 1.378 | 1.353 | 0.379 | 0.416 | 0.427 | 0.495 |
| | VS | AR | **0.73%** | **0.30%** | **0.88%** | **0.50%** | 0.949 | 0.813 | 0.910 | 0.906 | 1.456 | 1.076 | 1.312 | 1.329 | 0.374 | 0.508 | **0.449** | 0.483 |
| | | AWS | **0.73%** | **0.30%** | 1.17% | 0.57% | 0.993 | 0.770 | 0.985 | 0.878 | 1.800 | 1.065 | 1.671 | 1.391 | 0.327 | **0.541** | 0.392 | 0.486 |
| | | WS | **0.73%** | **0.30%** | **0.88%** | **0.50%** | 0.930 | 0.790 | 0.908 | 0.904 | 1.397 | 1.035 | 1.337 | 1.320 | 0.413 | **0.540** | 0.435 | 0.493 |
| W2 | IUF | AR | 5.84% | 3.92% | 1.05% | 0.71% | 0.944 | 0.787 | 0.897 | 0.882 | 1.447 | 1.081 | 1.347 | 1.318 | 0.341 | 0.427 | 0.438 | 0.482 |
| | | AWS | 5.84% | 3.92% | 1.40% | 0.85% | 0.899 | 0.766 | 0.973 | 0.872 | 1.586 | 1.161 | 1.694 | 1.427 | 0.380 | 0.429 | 0.381 | 0.465 |
| | | WS | 5.84% | 3.92% | 1.05% | 0.71% | 0.885 | 0.763 | 0.872 | 0.864 | 1.350 | 1.079 | 1.394 | 1.341 | 0.408 | 0.440 | 0.427 | 0.494 |
| | VS | AR | **0.73%** | **0.30%** | **0.88%** | **0.50%** | 0.917 | 0.794 | 0.895 | 0.891 | 1.405 | 1.097 | 1.321 | 1.341 | 0.386 | 0.517 | **0.450** | 0.477 |
| | | AWS | **0.73%** | **0.30%** | 1.17% | 0.64% | 0.935 | 0.765 | 0.967 | 0.871 | 1.627 | 1.110 | 1.657 | 1.398 | 0.369 | 0.534 | 0.394 | 0.477 |
| | | WS | **0.73%** | **0.30%** | **0.88%** | **0.50%** | 0.888 | 0.769 | 0.883 | 0.884 | 1.352 | 1.057 | 1.344 | 1.328 | **0.429** | **0.550** | 0.439 | 0.489 |

**Table 5.3.:** Results of the recommender algorithm test runs for All but 1, All but 2, Given 2 and Given 5 test runs.

lation, rating rule, user similarity and prediction algorithm. Following the example from Breese et al. (1998) the univariate analysis method (ANOVA) was used for this purpose. The results are listed in Table 5.4. It shows that rating rule and prediction algorithm have a significant impact for all test methods while the influence of the user similarity algorithm is only measurable for G2 and G5. The combination of rating rule and prediction algorithm shows significant results with all other combinations without significant influence.

The result of the ANOVA tested the influence of each computation step. To evaluate the differences within each step, e.g. between the different rating rules, the values were analyzed using the Bonferroni method based on observed means (cf. Breese et al., 1998). The results revealed significant differences between some of the factors as summarized in Table 5.5. The detailed results are listed in Appendix G.

The analysis of Table 5.3 had shown the best results for the MA1 rating rule. Combining this information with the values from Table 5.5, showing that MA1 has significant different outcomes compared to all other rating rules, lead to the decision, to chose the *Monthly Average Rating Rule 1 (MA1)* for the CoFiPot recommender system.

The results of the ANOVA results had shown that the combination of rating rule and prediction algorithm had a significant influence on the results. For the test runs with MA1, the IUV and VS results did not have any significant differences; in fact, there were no differences between these two algorithms for all test runs as Table 5.5 shows. According to these results, any of the user similarity algorithms could have been taken. However, comparing the missing rates in Table 5.3 left the *Vector Similarity (VS)* as best choice for the system because the missing rates for this algorithm were lower compared to the results for combinations using IUF.

For the prediction algorithm, the values of the Bonferroni comparison showed significant differences between the WS and AWS and also between AR and AWS with exception of the AB2 test run. No differences were identified between WS and AR. Again, the values in Table 5.3 had shown that AWS had worse results than the other two algorithms. Therefore AWS was excluded. The final choice between AR and WS was made for *weighted sum*, because the comparison of the MAE and MSE values showed better results for the AB1 and AB2 test runs.

| *Test Run* | *Source* | *Type III Sum of Squares* | *df* | *Mean Square* | *F* | *Sig.* |
|---|---|---|---|---|---|---|
| *AB1* | Rating | 88.131 | 3 | 29.377 | 55.122 | 0.000 |
| | UserSim | 0.487 | 2 | 0.244 | 0.457 | 0.633 |
| | Prediction | 12.672 | 2 | 6.336 | 11.888 | 0.000 |
| | Rating x UserSim | 0.379 | 6 | 0.063 | 0.118 | 0.994 |
| | Rating x Prediction | 13.610 | 6 | 2.268 | 4.256 | 0.000 |
| | UsersSim x Prediction | 1.839 | 4 | 0.460 | 0.863 | 0.486 |
| *AB2* | Rating | 67.470 | 3 | 22.490 | 47.537 | .000 |
| | UserSim | 1.345 | 2 | .673 | 1.422 | .241 |
| | Prediction | 3.996 | 2 | 1.998 | 4.223 | .015 |
| | Rating x UserSim | .216 | 6 | .036 | .076 | .998 |
| | Rating x Prediction | 8.442 | 6 | 1.407 | 2.974 | .007 |
| | UsersSim x Prediction | .975 | 4 | .244 | .515 | .725 |
| *G2* | Rating | 639.534 | 3 | 213.178 | 354.186 | .000 |
| | UserSim | 17.310 | 2 | 8.655 | 14.380 | .000 |
| | Prediction | 244.331 | 2 | 122.166 | 202.973 | .000 |
| | Rating x UserSim | 2.354 | 6 | .392 | .652 | .689 |
| | Rating x Prediction | 23.586 | 6 | 3.931 | 6.531 | .000 |
| | UsersSim x Prediction | 4.619 | 4 | 1.155 | 1.919 | .104 |
| *G5* | Rating | 548.408 | 3 | 182.803 | 316.510 | .000 |
| | UserSim | 46.911 | 2 | 23.456 | 40.612 | .000 |
| | Prediction | 28.330 | 2 | 14.165 | 24.525 | .000 |
| | Rating x UserSim | 6.515 | 6 | 1.086 | 1.880 | .080 |
| | Rating x Prediction | 33.146 | 6 | 5.524 | 9.565 | .000 |
| | UsersSim x Prediction | 1.806 | 4 | .451 | .782 | .537 |

**Table 5.4.:** ANOVA results for the test of between-subjects effects for MAE values for the test runs All but 1, All but 2, Given 2 and Given 5. The sources are the factors, rating rule, user similarity algorithm (UserSim) and prediction algorithm and the combinations of these.

| *Compared Algorithms* | | | *AB1* | *AB2* | *G2* | *G5* |
|---|---|---|---|---|---|---|
| WA1 | x | WA2 | 1.000 | 1.000 | .121 | 1.000 |
| | x | MA1 | **.000** | **.000** | **.000** | **.000** |
| | x | MA2 | .000 | 1.000 | .000 | .000 |
| WA2 | x | MA1 | **.000** | **.000** | **.000** | **.000** |
| | x | MA2 | .000 | 1.000 | .000 | .000 |
| MA1 | x | MA2 | **.000** | **.000** | **.000** | **.000** |
| Corr | x | IUF | 1.000 | .291 | .000 | .000 |
| | x | VS | 1.000 | 1.000 | .000 | .000 |
| IUF | x | VS | 1.000 | 1.000 | 1.000 | 1.000 |
| AR | x | WS | .425 | .252 | .210 | .652 |
| | x | AWS | .000 | .330 | .000 | .000 |
| WS | x | AWS | **.000** | **.003** | **.000** | **.000** |

**Table 5.5.:** Multiple comparisons using the Bonferroni method for the All but 1, 2 and Given 2, 5 test runs. The results are based on observed means. E.g. the combination of WA2 and MA2 have shown significant differences for the test runs AB1, G2 and G5 but not for AB2.

Thus, in summary, the specifications for the CoFiPot recommender system were set to:

- Rating rule: Monthly average 1,
- User similarity algorithm: Vector similarity,
- Prediction algorithm: Weighted sum.

With these settings, the recommendations for each user were automatically computed every night and sorted by predicted relevance. Finally, it had to be decided whether to show all recommendations independent from their predicted relevance to the users. This approach would allow the display of recommendations with potentially little relevance. It was arbitrarily decided to set the filter value to 1.5 so that all recommendations with a lower predicted relevance are not displayed.

## 5.2. Recommendation Display and Case Study Design

### 5.2.1. Notifications and Recommendations

With the settings for the daily recommendation computation process specified, the portal interface had to be adapted to integrate the recommender system. Several methods to communicate recommendations to the users had been introduced and discussed in Section 4.8.1. Based on them, the following five were presented and rated in the expert interview (cf. Section 3.3):

- Random pop-up display with a maximum of one appearance per week,
- display in the footer comparable to the recommendations at Amazon,
- a balloon-tip appearing in the lower right corner of the portal,
- in combination with the portal's notifier service (The notifier informs each user about important messages from services), and
- a regular automatically sent email.

The preferences of the interviewed managers differentiated strongly. Figure 5.1 shows the answers. The footer display got the highest ratings followed by notifier and balloon-tip. Concerning the option to open a pop-up window, the comments revealed a strong opposition against any display pattern that requires users to manually discard the message. Similarly, some managers opposed the automatic sending of emails because it would distract users from working and further fill their mailboxes with "irrelevant" mails. The high ratings for the footer-display were commented with the advantage of the approach that it does not disturb users, because it does not block any content and still is more visible than the notifier.

The first three options were further discussed with portal management. Although the footer had received the best ratings, the final decision by management was to combine the recommendation display with the portal's notification system. The argumentation against the display in the footer was that it would (i) irritate the users when such messages were displayed below the services and (ii) require them to scroll

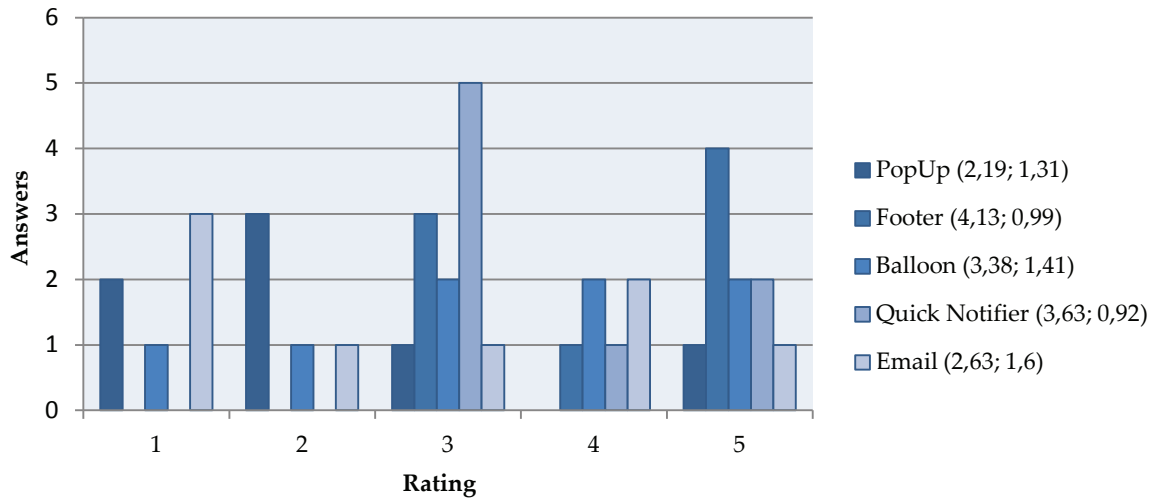


**Figure 5.1.:** Expert interview results: Rating of recommendation display methods on a five-point scale with five as highest rating. Each Option with average and standard deviation in brackets.

down to see them. The final design decision as shown in Figure 5.2 was to add an icon next to the notification system. Clicking on the icon shows the top three recommendations.

The decision to show the top three recommendations was based on the idea that three items were likely to attract a user's attention but would not appear overwhelming or randomly selected as more items might have done.

This decision was critically discussed with management, because the chosen approach would not guarantee a high visibility. The first objection was that only users of certain services receive messages from the notification system. Therefore, all others users are not likely to check for any messages by looking at the notification icon next to the recommendation icon. Secondly, the separation between notification messages and recommendations further minimizes the likelihood of users looking at them. Mixing the notification messages with recommendations would at least have caused users to see them when they check for new notifications. But with the chosen approach the users were required to actively click the notification icon to see any message. Without this action, they would not be able to detect them. The results of the survey, which was conducted to evaluate the system, confirmed these concerns and will be discussed in more detail in Section 5.4.

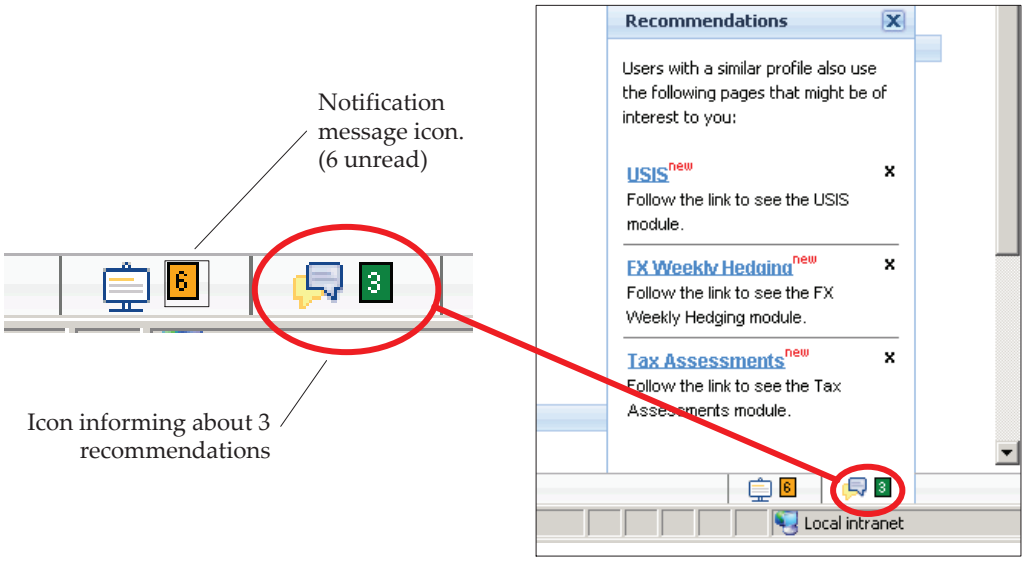


**Figure 5.2.:** Screenshot of the CoFiPot recommender. The icon and message box are located in the lower right corner of the portal. To the left of the recommender icon is the icon for the notification messages. Clicking on the icon opens the recommendations as shown in the right image.

### 5.2.2. Recommendation Visit Process

When a user clicks on the recommender icon, a list with the top three recommendations is shown (cf. Figure 5.2). Each recommendation provides a link to the recommended service. When a service is visited through this link, the system shows a short description of the service's main functionalities, if available. In addition a feedback form is displayed that allows the user to rate the quality of the recommendation. Figure 5.3 shows this feedback form.

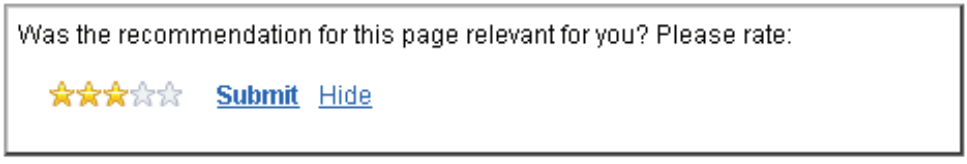


**Figure 5.3.:** Explicit feedback form for recommendations.

Once a user has visited a recommended service through the provided link, the recommendation is immediately removed from the list. If the service is visited through the menu or another link, the recommendation will remain visible until the next nightly update-process. This process will detect that the user has visited a recom-

mended service and therefore not generate a new recommendation for it. Users can also manually remove recommendations by clicking the "x" next to the recommendation message as shown in Figure 5.3.

In both cases, the visit of a recommended service or the manual removal of a recommendation, the next recommendation with the highest predicted relevance is at once added to the recommendation list. Thus, each user will see three recommendations as long as there are that many available.

### 5.2.3. Case Study Setup

Once the recommender system was implemented, a four month observation period started (March - June). In this period users were separated in a treatment and a control group. Members of the latter did not see any recommendations. During the observation period no changes were made to the portal's service structure. Furthermore, data had already been collected for the four months previous to the observation period during which the service portfolio had not been changed. Thus, the number of services is consistent between the two periods so that usage patterns from both periods are comparable. In the remainder of this chapter, the periods are named *comparison* and *treatment period*, respectively.

#### Treatment and Control Group

The assignment of users to the treatment or control group was performed randomly with regard to the user activities. To receive comparable groups, all users were split in groups according to their activity level as introduced in Section 3.2.1. In each activity group the users were randomly assigned to one of the two groups. The resulting groups with 2713 users in the treatment and 2568 users in the control group are shown in Figure 5.4. Every new user, who visited the portal during the treatment period for the first time, was randomly assigned to one of the groups.

During the treatment period 3434 users of the treatment group and 2073 of the control group actively used the portal. The differences to the previous period are explicable with the fact that (i) several users of the control group did not access the

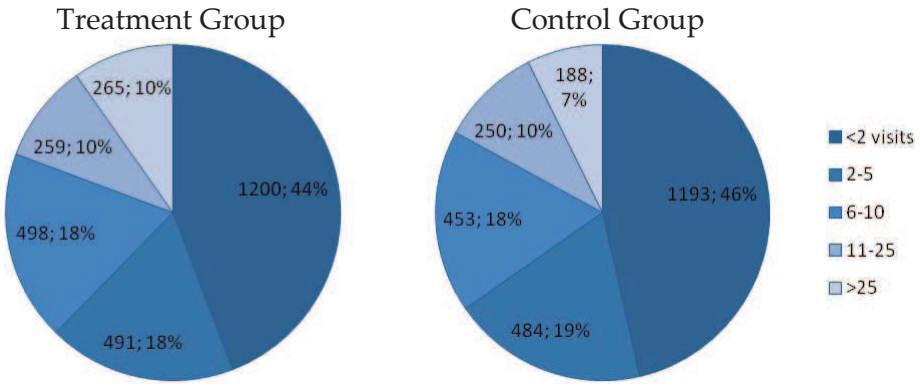


**Figure 5.4.:** Treatment and control group clustered by activity level at the beginning of the treatment period. The numbers show the absolute size of each group and the relation to the overall group size.

portal during the treatment period and (ii) that multiple users of the treatment group had not been active during the comparison period but were active again in the treatment period. Restricting the perspective on users who actively used the portal in both periods leaves 1408 users in the treatment group and 1287 in the control group. Table 5.6 summarizes the specified groups and shows the average number of visits to the portal per user.

| *Group* | *Users* | *Total Visits* | |
|---|---|---|---|
| | | *Avg* | *StdDev* |
| *T1* | 2871 | 23.982 | 90.123 |
| *T2* | 3546 | 24.388 | 87.039 |
| *C1* | 2709 | 17.659 | 62.818 |
| *C2* | 2073 | 26.040 | 74.296 |
| *T1** | 1408 | 45.744 | 124.848 |
| *T2** | 1408 | 53.668 | 130.642 |
| *C1** | 1287 | 34.103 | 88.119 |
| *C2** | 1287 | 39.041 | 91.104 |

**Table 5.6.:** Users split in treatment (T) and control (C) group for comparison (1) and treatment period (2). The values for total visits represent the average number of visits per user per period.
* Users who have actively used the portal in both periods

In the analysis in Chapter 3, it was found that the majority of users had only visited a small number of services. Thus, for each user a multitude of services remained to

be discovered. Accordingly, a check of the computed recommendations showed more than three feasible recommendations for each user.[3]

## 5.3. Case Study Evaluation

To increase the visibility of the recommender system, the list of recommendations was set to open automatically[4] for each member of the treatment group during his first visit after the start of the treatment period. This procedure was later repeated several times to increase the visibility of the recommendations. In order to detect the recommendations, each user either had to notice this short opening of the recommendation list or the new icon next to the notification messages icon (cf. Figure 5.2).

### 5.3.1. Recommendation Visits

The usage of the recommender system was constantly monitored during the treatment period. Figure 5.5 shows the recommendation visits. The repeated procedure of opening the recommendation list is emphasized with red lines. It shows a peak following the initialization of the system but no major changes due to the repeated procedure. The timeline also covers a short period after the end of the treatment period showing a slight usage increase. This increase is probably related to the user survey that asked users whether they have seen the recommendation system. As the subsequent analysis of the survey will show, many users did not detect the recommender system. Therefore, it can be assumed that a few of them tested it after they had learned about it in the survey.

A total of 226 recommendations were visited by 128 users during the treatment period (3.73% of the treatment group). This leaves 3306 users in the treatment group who either did not see the recommendations or chose not to follow them. These numbers indicate that the system was not detected by most users as predicted during the discussions about its design (cf. Section 5.2.1). Table 5.7 extends Table 5.6 with the

[3] Excluding three users who were identified as portal administrator.
[4] If not activated, the message box disappeared after 10 seconds.

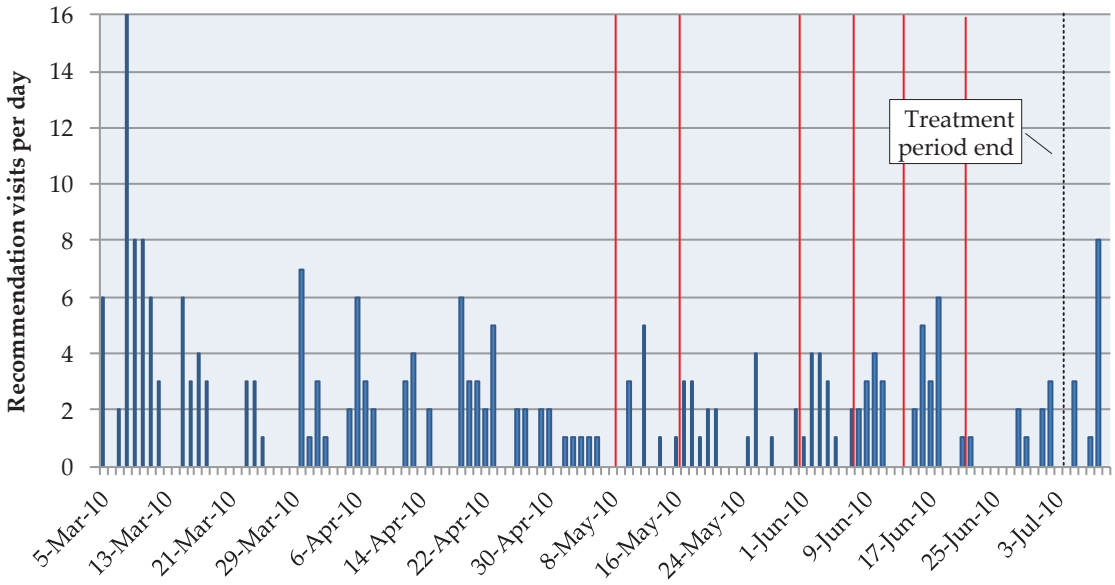


**Figure 5.5.:** Recommendation visits during treatment period. Overall daily visits.

information about the user group who used the recommender system (R) for both periods. It shows that except for nine users, all others have actively used the portal in the comparison period. Furthermore, the high average of visits in comparison to the values in Table 5.6 indicates that rather active users have visited the recommendations.

| *Group* | *Users* | *Total Visits* | |
|---|---|---|---|
| | | *Avg* | *StdDev* |
| *R1* | 121 | 159.719 | 321.778 |
| *R2* | 128 | 180.762 | 305.552 |

**Table 5.7.:** Users who visited at least one recommendation for comparison (1) and treatment (2) period with number of total visits in each period.

An indicator for the quality of the recommendation besides the explicit ratings is the amount of services that users continued to use after they had visited them based on a recommendation. Of the 226 visited services, 31 were visited again by the user, half of them more than once. It is difficult to further interpret these numbers, because some of the follow-up visits took place a few weeks after the initial visit and others within days. This can partially be explained with the fact that some services are only relevant during specific time periods, e.g. the end of a month (cf. Section 3.2.4). To

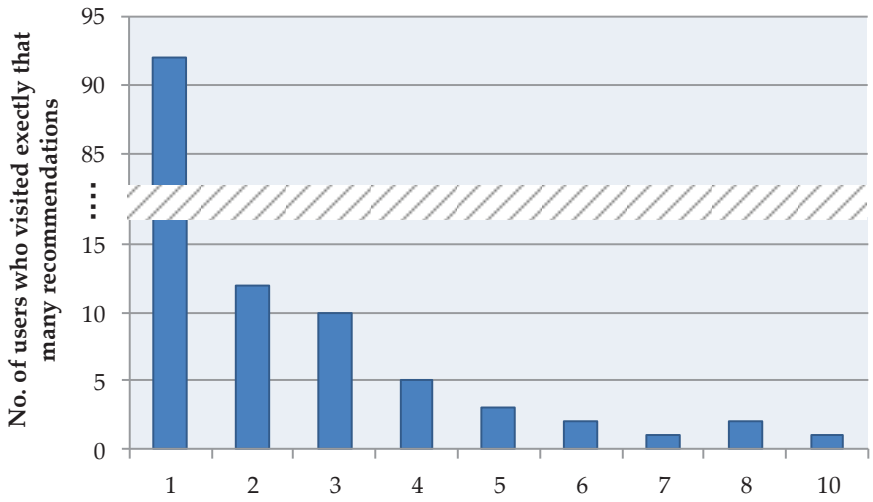


**Figure 5.6.:** Visited recommendations per user (Avg: 1.766; StdDev: 1.624) restricted on all users who visited at least one recommendation. The groups are exclusive, e.g. users who visited more than one recommendation are not included in the first group.

create a founded analysis, the follow-up usage patterns for each visited recommendation would need to be analyzed for at several months after each visit. Thus, it can be ensured that even services that are only relevant at the end of a quarter are covered. Unfortunately, the data of the treatment period could not be analyzed for such a time period. Therefore, the analysis has to remain superficial.

### 5.3.2. Recommendation Ratings

Only 11 explicit ratings were given by six users which constitutes a feedback rate of 4.86%. The ratings in combination with the predicted relevance and the number of follow-up visits are shown in Table 5.8. The recommendations which lead to follow-up visits all received high ratings of four or five. The majority of the remaining ratings were at a similar level with only two below three. However, a detailed analysis would require more ratings and, as described above, more information about the long term usage of the recommended services.

| *User* | *Follow-up Visits* | *Rating* | *Predicted Relevance* |
|---|---|---|---|
| A | 1 | 4 | 2.3 |
| | 0 | 4 | 1.82 |
| B | 0 | 2 | 2.3 |
| C | 0 | 4 | 2.46 |
| | 0 | 1 | 3.22 |
| D | 8 | 5 | 2.96 |
| | 3 | 5 | 2.5 |
| | 3 | 4 | 2.65 |
| | 3 | 5 | 2.74 |
| E | 8 | 4 | 3 |
| F | 0 | 5 | 2.56 |

**Table 5.8.:** Explicit ratings for recommendations with predicted relevance and number of follow-up visits grouped by users.

### 5.3.3. Service Usage Evaluation

The analysis of the recommendation visits so far showed that some users used the recommender system to explore new services. However, it has to be critically stated that only 3.73% users of the treatment group have visited at least one recommended service. However, it is unclear whether some users of the treatment group visited any new services motivated by the recommendation message without using the provided link. Instead they could have visited a recommended service through the portal's menu. In addition, some users might have been motivated by seeing the recommendations to browse the portal for new services.

The aim of this section is to determine whether differences exist in the usage patterns between the treatment and control group. The dependent variables are the number of new visits (NV) and the number of new visits that lead to follow-up visits (NV*). The first variable serves as an indicator for the browsing patterns of the user base. High values indicate a higher motivation to search for potentially relevant services. The second variable serves as success rate of the browsing.

As first analysis, it was tested whether differences exist between the treatment and the control group during the treatment period. Accordingly, the hypotheses were:

- $H_1$: The number of newly visited services is significantly higher for the treatment group than for the control group in the treatment period.
- $H_2$: The number of newly visited services that were subsequently used is significantly higher for the treatment group than for the control group in the treatment period.

To ensure that the groups were not biased, it was also required to test whether the two groups had similar usage patterns in the comparison period. Figure 5.7 shows the available data and the comparisons that were made. The comparison between the two groups during the comparison period is marked with (a) and the comparison required to test the hypotheses with (b).

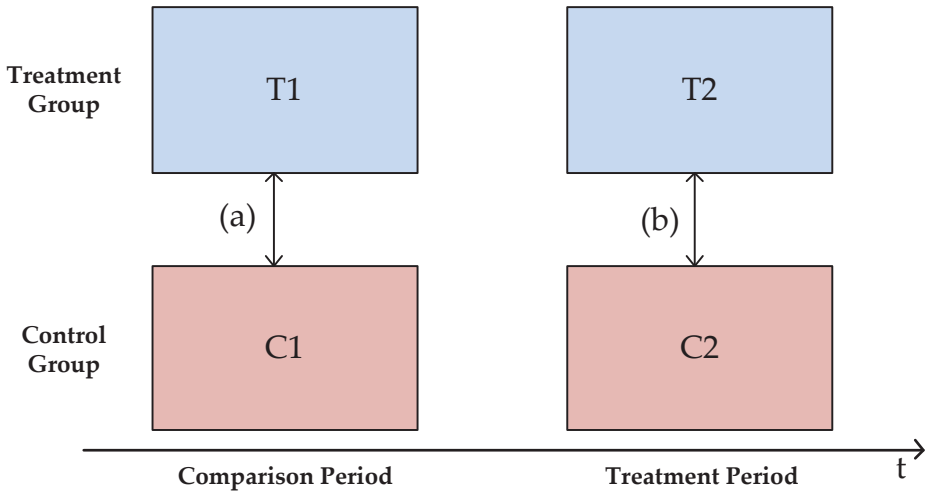


**Figure 5.7.:** Available data for treatment (T) and control group (C) for the two periods.

The data was compared using an independent two-sample two-sided *t*-test. The results are listed in Table 5.9. They show significant differences at a 1% level for the newly visited services (NV) in the treatment period between the two groups (a). Thus, $H_1$ can be confirmed. The result for the comparison of the newly visited services that were subsequently used is similar; $H_2$ was also confirmed with a higly significant result. The validation wether any differences existed between the groups during the comparison period (b) shows no significant differences between the means can be identified for neither NV nor NV*. The basic analysis indicates a positive effect of the

recommender system because significant differences between the two groups during the treatment period were identified while there were none during the previous comparison period. This is also recognizable in Figure 5.8, which shows the number of new visits per week for the two groups.

| | *Groups* | | | *NV* | | *NV** | |
|---|---|---|---|---|---|---|---|
| | | | | *t* | *Sig.* | *t* | *Sig.* |
| *(a)* | T1 | x | C1 | 0.790 | 0.430 | 1.474 | 0.141 |
| *(b)* | T2 | x | C2 | 3.648 | **0.000** | 3.132 | **0.002** |

**Table 5.9.:** Result of the *t*-tests for new visits (NV) and new visits that were revisited (NV*).

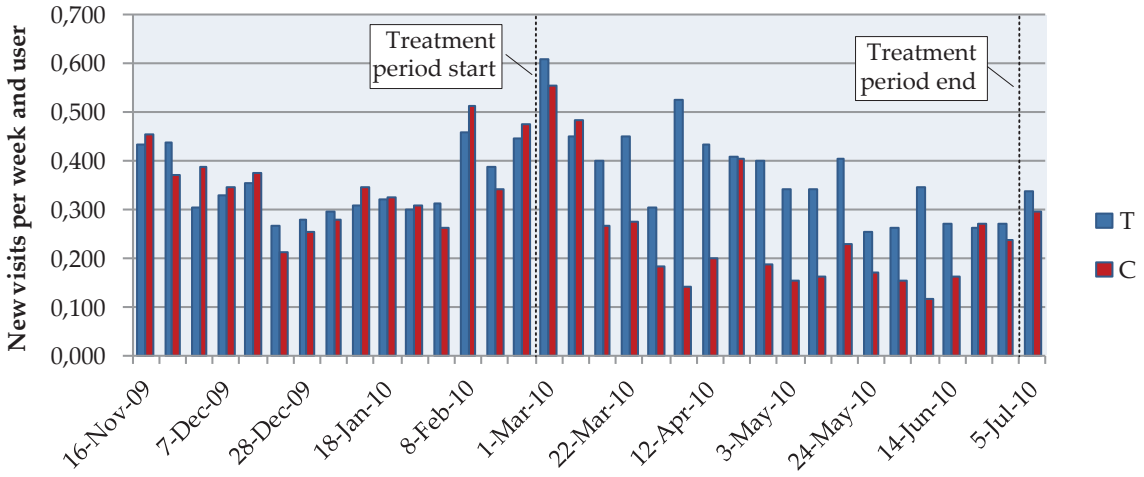


**Figure 5.8.:** Comparison: New service visits per week for treatment (T) and control group (C).

To measure the effect of the recommender system, the data was analyzed using the difference in difference method (Ashenfelter and Card, 1985). This method compares the two groups over the two periods and removes biases that could be the results of a trend. The model for this method is given by

$$y = \beta_0 + \beta_1 TG + \delta_0 TP + \delta_1 TP * TG \tag{5.1}$$

where $y$ is the outcome of NV or NV* and $TG$ is a dummy variable that captures possible differences between the two groups. $TP$ is a dummy variable for the treatment period that captures aggregate factors that would cause changes in y caused by trends independent of any treatment. The final variable is the interaction of $TP * TG$,

which is a dummy variable equal to one for the observations in the treatment group in the treatment period.

The results of the regression analysis are summarized in Table 5.10. For the newly visited services (NV) the values indicate a significant effect of the treatment. The highly significant result for $\delta_0$ show an overall negative trend; the number of newly visited services declined between the two periods. This is also recognizable in Figure 5.8. The positive significant result for $\delta_1$ reveals a positive effect of 15.9% of the treatment compared to the control group. It shows that the users from this group, who were supported with the recommender system, did visit that many more new services. This confirms the initial results of the *t*-tests.

| *Factor* | *Unstandardized Coefficients* | | *Standardized Coefficients* | *t* | *Sig.* | $R^2$ |
|---|---|---|---|---|---|---|
| | *B* | *Std. Error* | *Beta* | | | |
| **Dependent Variable: New Visits (NV)** | | | | | | |
| *(Constant)* | 1.055 | 0.38 | | 27.460 | .000 | .002 |
| $\beta_1$ | .042 | .054 | .010 | .788 | .431 | |
| $\delta_0$ | -.219 | .058 | -.055 | -3.748 | .000 | |
| $\delta_1$ | .159 | .077 | .037 | 2.070 | 0.038 | |
| **Dependent Variable: New Visits with follow-up visits (NV*)** | | | | | | |
| *(Constant)* | .340 | .016 | | 21.360 | .000 | .016 |
| $\beta_1$ | .038 | .022 | .022 | 1.707 | .088 | |
| $\delta_0$ | -.223 | .024 | -.133 | -9.200 | .000 | |
| $\delta_1$ | .019 | .032 | .011 | .593 | .553 | |

**Table 5.10.:** Results for the regression for difference-in-difference method with the dependent variables NV and NV*.

The results for the newly visited services that were subsequently used (NV*), however, do not further strengthen the initial findings. A negative trend similar to to the newly visited services exists but no significant effects of the treatment are observable.

To evaluate the effects of other influence factors, two more were added to the regression. The model is given with

$$y = \beta_0 + \beta_1 TG + \delta_0 TP + \delta_1 TP * TG + \kappa_0 TV + \kappa_1 LT \tag{5.2}$$

The variable *TV* represents the total number of visits over the two periods, which serves as indicator for the influence of the user's activity level. *LT* is a dummy variable whether a user belongs to the group of long term users who have been registered in the portal for more than a year. The results that are listed in Table 5.11 show highly significant influences for both factors. Obviously, the activity level positively influences the number of new visits (NV). The more a user visits the portal, the more new services he is likely to visit. More interesting is the result for long term users. A long term membership in the portal positively influences the number of new visits. This is surprising because a negative influence would have seemed reasonable: new users would seem more likely to be the ones who look for potentially relevant services. However, the results indicate that in CoFiPot the more experienced users are more likely to look for new services.

| *Factor* | *Unstandardized Coefficients* | | *Standardized Coefficients* | *t* | *Sig.* | $R^2$ |
|---|---|---|---|---|---|---|
| | *B* | *Std. Error* | *Beta* | | | |
| **Dependent Variable: New Visits (NV)** | | | | | | |
| *(Constant)* | .930 | .038 | | 24.361 | .000 | .047 |
| $\beta_1$ | .004 | .052 | .001 | .080 | .936 | |
| $\delta_0$ | -.269 | .057 | -.067 | -4.706 | .000 | |
| $\delta_1$ | .219 | .075 | .051 | 2.904 | .004 | |
| $\kappa_0$ | .005 | .000 | .182 | 18.687 | .000 | |
| $\kappa_1$ | .374 | .058 | .063 | 6.445 | .000 | |
| **Dependent Variable: New Visits with follow-up visits (NV*)** | | | | | | |
| *(Constant)* | .302 | .016 | | 19.056 | .000 | .057 |
| $\beta_1$ | .024 | .022 | .014 | 1.121 | .262 | |
| $\delta_0$ | -.241 | .024 | -.144 | -10.147 | .000 | |
| $\delta_1$ | .036 | .031 | .020 | 1.159 | .246 | |
| $\kappa_0$ | .002 | .000 | .203 | 20.879 | .000 | |
| $\kappa_1$ | .008 | .024 | .003 | .352 | .725 | |

**Table 5.11.:** Results for the regression for difference-in-difference method with the dependent variables NV and NV* extended with variables for total visits per user and an indicator long term membership.

The second analysis for NV* also shows that the number of total visits has a highly significant influence; with 2% it is, however, smaller than the 5% for NV. Still, more

active users are more likely to start using a new service. The second variable for long term users does not reveal any significant inlfuence. This shows that the lenght of membership in the portl does not influence the decision to start using additional services.

In summary, the results from the *t*-tests and the regression analyses indicate a positive influence of the recommender system. Significant differences between the control group and the treatment group were identified. The result for the difference-in-difference variable with 15.9% gives an estimate for the impact of the treatment. The additional inclusion of factors for activity and length of membership revealed additional influence factors on the number of newly visited services.

It has to be stated critically that less than 5% of the treatment group visited at least one recommendation. This leaves the question why the others did not use the recommender system. It is even more surprising that despite the small percentage of recommender users within the treatment group the effects compared to the control group were significant. An open question, which is addressed in the following section, is whether some users from the treatment group visited recommended services without clicking on the provided link and thus had not been detected.

## 5.4. User Survey

A survey among all users was conducted at the end of the treatment period. Several objectives were pursued with the survey. The key questions for the research were whether the users consciously noticed the recommender system and, in case they did, how they evaluate its value for the portal. The remaining questions focused on the overall satisfaction with the portal. The answers to the anonymous survey were stored together with the usage patterns of each participant. Thus, it is possible to compare the actual usage of the recommender system with the given answers.

### 5.4.1. Participants

The survey was conducted by showing a message to all users while they were visiting the portal. 272 users participated in the survey. 175 (64.34%) of those were member of the treatment group and thus had access to the recommender system. Similar to previous evaluations, the users were grouped by their activity level as shown in Figure 5.9. The values show that mainly very active users participated and only 7 (3%) who seldom used the portal. The distribution of users did not vary between the treatment and control group. 30 users (11.03%) among all survey participants had visited at least one recommendation.

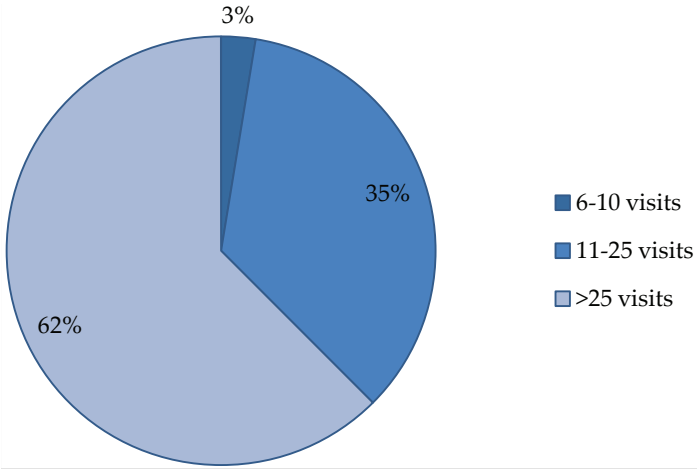


**Figure 5.9.:** Survey participants grouped by their monthly activity level. The activity level is based on the data collected during the treatment period.

### 5.4.2. Overall Perception of the Recommender System

The first set of questions focused on the overall perception of the system. The first question asked whether they had noticed the recommendations. To avoid any misunderstandings, a screenshot of the recommendation messages was displayed next to this question. 50 users positively answered this question as shown in Table 5.12; all of them were member of the treatment group.[5] The 50 positive answers show that

[5]The question whether the user had noticed the recommender system was also shown to the control group but none answered with "yes". Such cases could have indicated possible misunderstandings with the question.

28% of the users from the treatment group, who answered the survey, had noticed the recommender system. The usage analysis had resulted in a rate of less than 5%. This difference can have several reasons. It could indicate that many users have seen the recommendations but did not further look at them. On the other hand, the usage analysis had shown that most of the recommender users were part of the rather active user group. The participants of the survey were also mainly part of this group. Thus, the higher rate of 28% can also be explained with a higher perception rate of the system within the group of active users.

A critical factor in the answers to the first question is the fact that seven of the users who answered with "no" actually had visited a recommendation. It is difficult to explain this result other than with careless answer patterns or with misconceptions of the question despite the display of the screenshot.

| | *All users* | | *Rec-User** | |
|---|---|---|---|---|
| *yes* | 50 | (28.57%) | 23 | (13.14%) |
| *no* | 125 | (71.43%) | 7 | (4.00%) |
| ***Total*** | **175** | | **30** | **(17.14%)** |

**Table 5.12.:** Number of participants who have seen the recommendations.
* Users who followed a recommendation according to the usage log

The remaining questions were only shown to the 50 users who had seen the recommender. Therefore, all following answers are in reference to this number. The second question asked whether the users had looked at any of the recommended services. 26 (52%) answered with "yes" with 15 who had visited recommendations according to the usage analysis. Eight users who also had visited recommendations answered with "no". Six of them had visited the recommendations within the first week of the treatment period. Thus, a possible explanantion is that they forgot that they had visited them. Otherwise this matter can only be explained with careless answers.

The fact that 11 of the 26 users (42.3%) who answered that they had looked at recommendations were not among the recommender users identified in the usage log analysis underlines the assumption that some users did not follow the provided link in the recommendation message but rather visited the recommended service by se-

lecting it in the menu. Thus, these visits were not detected by the system. Considering this fact, it can be assumed that more users of the treatment group followed a recommendation as the 3.73% that were identified in Section 5.3.3.

All 26 users who had stated that they had visited at least one recommended service were asked whether any of these services was of interest for. 17 (65.38%) of them gave positive answers. The answers of the 15 users whose recommendation visits were recorded by the system were further compared with the usage data. The values are listed in Table 5.13 and 5.14. The six users who answered that no recommendation was of interest had visited a total of 20 recommended services (3.33 per user). Two users actually have visited one of the recommended services two more times after the first visit. A long term analysis might deliver better insights since it would show whether the regularly use the service although they declined this. The nine users who positively valued the recommendations together visited 40 services (4.44 per user). They revisited eight of them with four services being visited more than seven which indicates that they are of relevance to them.

The numbers confirm the given answers. The users have found relevant services through the recommender system. Furthermore, the higher rate of visited recommendations per user among the users who found relevant services (5.33 visits per user) compared to the other users (3.33) shows that positive experience with the system motivate users to further look at recommendations.

| | *All users (50)* | | *Rec-User* (23)* | |
|---|---|---|---|---|
| *yes* | **(a)** | 26 (52.0%) | **(b)** | 15 (65.22%) |
| *no* | | 24 (48.0%) | | 8 (34.78%) |

**Table 5.13.:** Survey question: Have you looked at any recommendation?
* Users who followed a recommendation according to the usage log

The 24 users who had answered that they had not visited any recommendation were asked to choose among several reasons why they did not do so. Multiple selections were possible. Two third (16) agreed with the choice that they did not have the time to do so, a quarter (6) that the recommendations did not look interesting, four that they already knew the recommended service and three that they did not have

| | *Group a*** | *Group b*** | *Visited rec. (Avg; StdDev)* | *Revisited rec.**** *(Avg; StdDev)* |
|---|---|---|---|---|
| *yes* | 17 (65.38%) | 9 (60.0%) | 40 (4.44; 3.36) | 8 (2.0; 1.41) |
| *no* | 9 (34.62%) | 6 (40.0%) | 20 (3.67; 2.58) | 2 (1.0; 0) |

**Table 5.14.:** Survey questions concerning the overall perception of the recommender system: Were any of the recommendations of interest to you?
** Groups marked in Table 5.13
*** Services, a user has at least visited three more times

access to it (cf. Figure 5.10). The indication of lack of time shows one of the key problems, which was already discussed in the first part of this work (cf. Section 3.2.4). Obviously even the prospect of learning about a possibly interesting service does not sufficiently motivate users to interrupt their current work process. This leads to the question for further research whether users would rather prefer of receiving a summarized report of recommendations via mail. Thus, they could choose when to visit the recommendations. The disadvantages of this approach had already been discussed in Section 4.8. The answer that four users already known one of the recommended services are most likely explicable with the fact that it was formerly[6] possible to visit some of the services without being logged in to the portal. In such cases, the recommender system could not know whether users already knew the service. Other explanations can be that the users had learned about the services from other users or that they had mistaken them for another service.

With final question of this set users were asked to rate the statement "I started using new services based on the recommendations" on a scale from 1-5 with 5 representing total agreement. This question adds on to the previous one who only demanded whether any of the recommendations was of interest. The answers shown in Figure 5.11, with an average of 2.76 and a standard deviation of 1.12, reveal no obvious preference. Similar to the previous question, the answers were compared to the usage data, if available. Two of the users who answered with a four had started using a new service and two of those who rated the statement with a three. One of the users who only rated with a two had even revisited one of the recommended services 14

[6] Before October 2010.

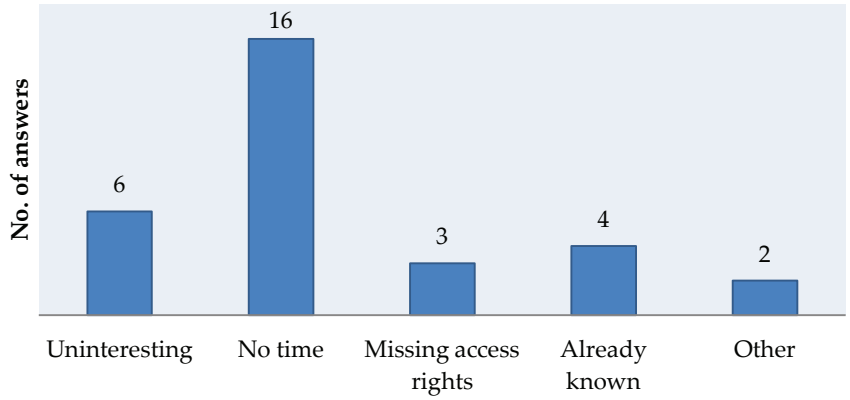


**Figure 5.10.:** Answers to the question why users did not visit any recommendation. Multiple selections were possible with 24 users providing answers.

times. The answers to this and the previous question show that some of the recommendations were of relevance to the users and in some cases lead to constant usage. It remains unclear why some users denied this answer although usage data showed it.

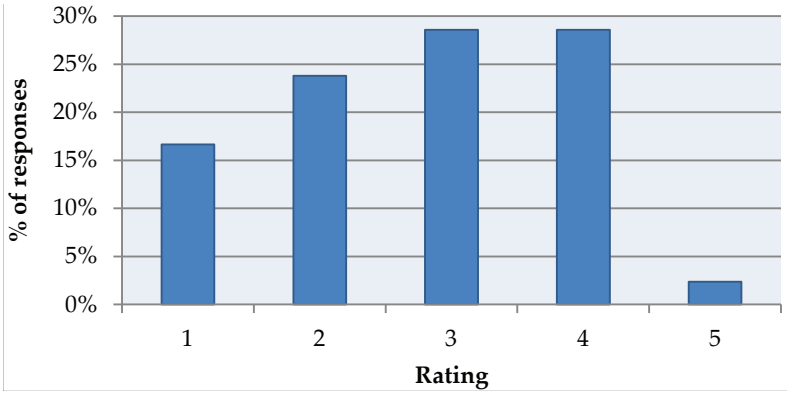


**Figure 5.11.:** Did users start using new services based on recommendations. Rating on a five-point scale with 5 being the highest.
Valid Answers: 42, Avg: 2.76, StdDev: 1.12

In summary, the answers show a positive effect of the recommender system among the users who detected it. Half of the users tested it while the remaining ones did not due to lack of time or because the recommended services did not appear interesting. The majority of the users found the recommendations interesting with some having found new services for constant use.

### 5.4.3. Satisfaction with the Recommender Systems

The second set of questions focused on the satisfaction with the recommender system. All questions in this set requested users to rate statements on a five-point scale with 5 being total agreement. Two questions pointed at the usefulness of the recommender system. The first asked whether the recommender system was a useful addition to the portal and the second whether any of the recommendations were useful. The answers are shown in Figure 5.12 and reveal a positive perceived usefulness of the individual recommendations (a) with an average of 3.3 (StdDev: 0.86) and also of the entire system (b) with 3.67 (0.82). More than half the users who have noticed the recommender system see it as a positive addition to the portal.

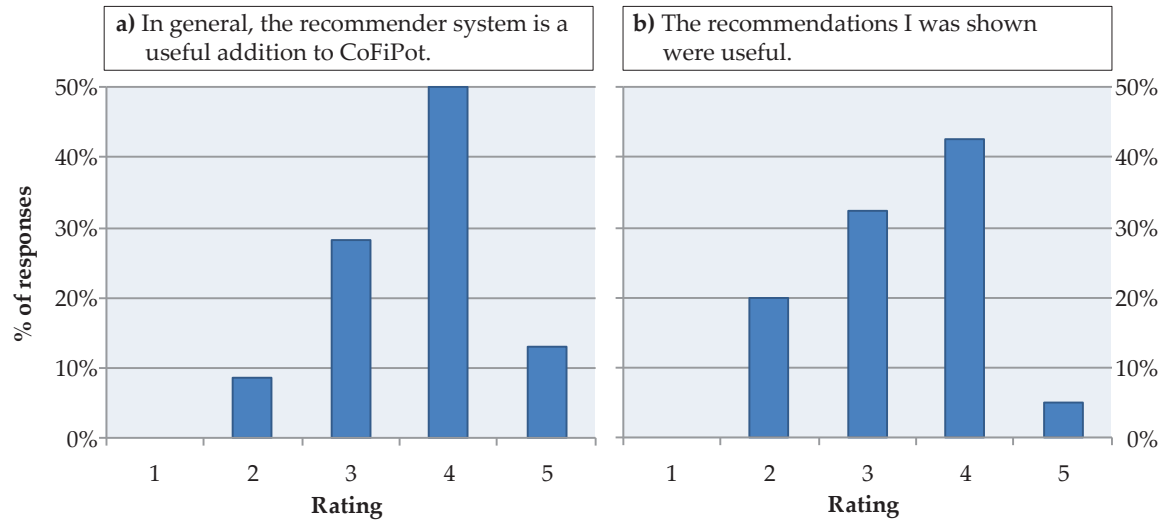


**Figure 5.12.:** Survey: Perceived usefulness of the recommender system.
a) Valid answers: 46, Avg: 3.67, StdDev: 0.82
b) Valid answers: 40, Avg: 3.33, StdDev: 0.86

With regard to the desired effects of the recommender system, users were asked whether the system helped them in better getting to know the portal. The results shown in Figure 5.13 with an average of 3.03 (1.05) reveal no main agreement or disagreement. Still, 32.5% of the users have a rather positive opinion of the system.

During the implementation phase the design of the recommender system and the availability of information about each recommended service have been critically discussed (cf. Section 5.2). Therefore, users were asked about their perception of the

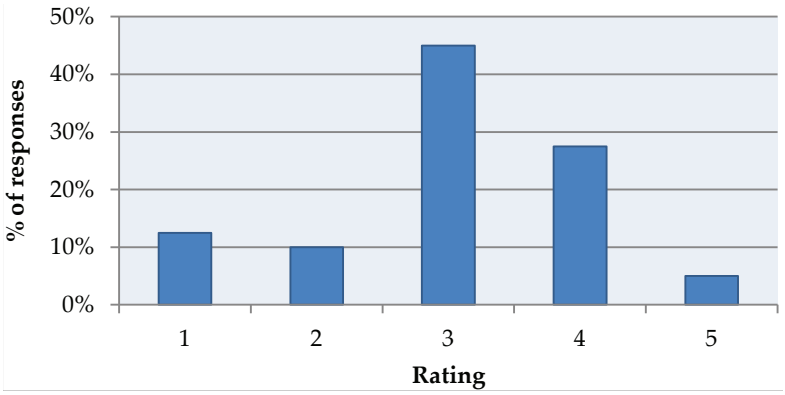


**Figure 5.13.:** Replies to learning-effect question: "The recommendations helped me getting to know CoFiPot better".
Valid answers: 40, Avg: 3.03, StdDev: 1.05

information availability when using the recommender system. The two questions shown in Figure 5.14 were designed to be comparable. The first asked whether sufficient information was available and the other whether there was not enough information. It was expected that users would provide similar answers for both questions with one side being negative and the other positive. The correlation for the answers with one side transformed[7] resulted in a negative correlation of -0.08. This result questions the reliability of the answer. However, as the graphic and the averages of 3.02 and 2.84 show, the users do not regard the availability of information as a considerable problem. Still, the values demonstrate that there is potential for improvements.

In summary, the majority of the survey participants regard the recommender system as positive addition to CoFiPot. However, only half of them took the time to look at a recommended service while the remaining lacked the time to do so or did not consider their personal recommendations as relevant. Besides the information about the lack of time to look at the recommendations, the rate of 28% of the users from the treatment group who perceived the recommender system is one of the most critical aspects that the survey revealed. Obviously, the visibility of the recommendation messages is too low which confirms the concerns that were stated during the discussion about the interface design. Improving the visibility requires therefore the

[7]Ratings of 5 were converted to 1 and 4 to 2.


**Figure 5.14.:** Questions about the availability of information on a five-point scale (with 5: totally agree).
a) Valid answers: 40, Avg: 3.03, StdDev: 1.00
b) Valid answers: 37, Avg: 2.84, StdDev: 0.93

highest attention. The detail level of the recommendation description had been another concern. However, the answers to the related questions show that users have a neutral opinion.

## 5.5. Chapter Summary

This chapter demonstrated how a recommender system was implementation into CoFiPot based on the concepts that had been presented in Chapter 4.

The comparison of different algorithms demonstrated that the decision needs to incorporate multiple factors such as the missing rate of each composition, the MAE and MSE values. In the presented case, the results for compositions with the inverse user frequency algorithm were similar to the ones based on the vector similarity algorithm. However, by comparing number of missed predictions, it was shown that the vector similarity algorithm delivered better results with a lower missing rate. Therefore, this algorithm was chosen for the computations of recommendations in CoFiPot. The results show that the testing process is required for every new implementation

of a recommender system. The different compositions of algorithms and rating rules have to be tested with real data in order to find the best matching concept.

The effects of the system were analyzed by comparing the number of newly visited services between users who were shown recommendation and a control group. Usage data from a comparison period previous to the implementation of the recommender system was included in the analysis. The research questions were accordingly whether the recommender system leads to significant differences between the two groups. The results show such a positive effect of the recommender system. Users from the treatment group visited more services that they had not visited before and also started using more services than the users from the control group. These effects occurred even though less than 5% of the treatment group's users did actually visit a recommendation according to the usage logs. However, it can be assumed that a higher percentage actually followed the recommendations but were not detected by the system because they did not follow the provided links. This assumption is underlined by the survey; several participants stated that they had visited recommendation but were not detected through the usage analysis.

Despite the significant effects of the system, the low visibility of the system has to be seen critical. Only 28% of the participants of the survey detected it. Taking into account that the majority of the participants were among the most active portal users, the detection by less active users is likely to be even lower. This assumption is confirmed by the analysis of the group of users who have visited recommendation according to the usage analysis: less than a third of these users visit the portal less than 11 times per month. Therefore, the focus for the improvement of the recommender system should be laid on the visibility of the system. Whether a different design or one of the alternative communication pattern as discussed in Section 4.8 will lead to a higher visibility has to be tested.

Another relevant result of the survey is that several users did not explore the recommendations, because they lacked the time to do so. These answers should also be considered when improving or changing the communication pattern for the recommendations. A simple solution might be to add a "remind me later" option to a

recommendation, but it is also possible that a different mode of informing the users about recommendations, e.g. via mail, could lead to a higher usage rate.

Taking together the visibility and the lack of time problem emphasized the requirement to support concepts such as the recommender system, to not request any user input. If users do not even have the time to follow a recommendation, they are unlikely to provide input such as explicit ratings. This is underlined by the six users who provided explicit ratings after having followed a recommendation with 248 other users who did not.

The case study showed that a concept that does not require any user input has the potential to increase the users' knowledge of the service portfolio of a corporate portal. This fulfills the key demands to the concept that were specified in Section 2.4.4.

# Chapter 6.

# Concluding Summary & Outlook

In this work, two main topics were covered: effects of corporate portal evolution and the concept for a corporate portal recommender system. The effects of constant corporate portal evolution were identified from the users' perspective and concepts to support portal management in handling with the resulting challenges were introduced. By means of the CoFiPot case study, the findings were further evaluated. One of the concepts, a recommender system for corporate portal services, was selected for a closer evaluation. In the second part, a model for the implementation of such a system was developed and applied for the CoFiPot case study.

## 6.1. Contributions

Basis for the analysis of corporate portal evolution was new a classification for corporate portals that was defined in Chapter 1. The literature overview had shown that existing classifications did not reflect the advancements in (web) technology. Especially the possibility to integrate all kinds of applications into a portal is not covered; existing scientific papers are rather focused on the knowledge management aspects of corporate portals. The classification that was defined in this work distinguishes between three main portal types:

- business application portal,
- application integration portal,
- employee portal

Moreover, for each type three evolutionary stages were introduced that allow to further differentiate the requirements of existing or planned portals.

With the classification as basis, the focus was shifted to the topic of corporate portal evolution in the second chapter. Thereby, the reasons for portal evolution and effects of this process were analyzed concentrating on the user perspective. It was shown, that changing user requirements and a users base with employees from different departments put high demands on the ongoing portal development process. Likewise, the implementation of new services and upgrades of existing ones require users to continuously adapt. These challenges were evaluated in detail and core tasks for portal users and management were derived and grouped by two core missions:

- to help users in their search for services and
- to provide a high usability.

Furthermore, the tasks were discussed in relation to the different corporate portal types and evolutionary stages.

As next step, several concepts to meet the identified challenges were presented. These concepts were taken from different areas and put very different demands to either portal users, management or developers. E.g., internal marketing and the organization of trainings are tasks that have to be executed by management while a recommender system is a portal service that will run automatically without any input requirements. Usability engineering and portal design are tasks for portal developers and documentation involves multiple stakeholders. Each concept was discussed for its applicability for the different portal types and evolutionary stages.

With the identified tasks and support concepts, which are moreover discussed with reference to the portal types and evolutionary stages, a concept was presented that facilitates the planning process for portals. The concept helps in identifying the needs of users for the given portal type and stage.

The case study of CoFiPot, which was introduced in Chapter 3, presents a typical application integration portal at the highest evolutionary stage. According to the specifications that were made in the second chapter, challenges related to the growth

of the portal since its implementation in 2001 were identified by analyzing usage data and further performing an expert interview. Two key challenges for CoFiPot were identified: (i) users barely know portal's service portfolio. Even very active users only visited a few of the services they had access to. Moreover, the expert interview indicated that several of the managers do not either know the range of services being offered by the portal. With regard to the lack of knowledge, it was shown (ii) that users do not browse the portal for other services by themselves. Therefore, with help of the findings from the second chapter, the concept of a recommender system for portal services was chosen to increase the users' portal knowledge.

The second part of the work was focused on finding an appropriate recommender system for implementation into a corporate portal. Accordingly, as first research step, the requirements for a corporate portal recommender system compared to common recommender systems were identified. Key findings in this context were the size of the targeted system which is rather low for a corporate portal compared to e-commerce applications. Also, the heterogeneity of the services and users and furthermore aspects of scalability and access control were identified as differences to common recommender systems. Based on this findings, the following research pointed at the task of developing a concept for the computation of recommendations with the research question: *Which recommendation technique is best suited for a corporate portal recommender system?*

For this purpose, different concepts for recommender systems were briefly discussed. The collaborative filtering technique was identified as best match for the requirements of corporate portals. This technique, which provides a variety of algorithms for each step in the process of computing recommendation, has to be adapted to the needs of each targeted portal. Therefore, a model to select and test an appropriate set of algorithms was defined. Among these algorithms, the method to derive implicit ratings for services from usage data was developed in this work. The other algorithms were taken from scientific literature with regard to the predefined criteria for corporate portals. Also, further considerations regarding the communication channels were presented.

Chapter 5 demonstrated how to apply the model from Chapter 4 by describing the implementation of a recommender system into the Corporate Financial Portal. Besides the testing of the implementation model, two research questions were pursued by analyzing the usage of the recommender system and performing a survey:

- *Does the recommender system influence the number of first-time visits to services?*
- *Can the recommender system increase the number of actively used services?*

For this purpose, only half of the users were shown the recommender system. Their usage patterns were compared to the other group and, furthermore, to their previous usage patterns. The findings showed that the users of the treatment group visit significantly more new[1] services and also started using more new services. This is even more remarkable because the usage statistics showed that most users did not noticed the recommender system. While this revealed major disadvantages with the chosen communication channel for the recommendations, it also demonstrates the high potential of the selected concept. If a small percentage of users who follow recommendations already causes measurable differences in usage, a higher visibility of the recommender system will surely benefit the overall usage of a corporate portal. The findings from the survey backed these assumptions. It showed that users appreciate the recommender system and see it as a valuable addition to CoFiPot.

## 6.2. Outlook

### 6.2.1. Corporate Portal Engineering

The classification model for corporate portals that was developed in Chapter 1 will help in structuring existing and new research on corporate portals. While business intelligence portals have been widely discussed in scientific literature, especially in the field of knowledge management, application integration and employee portals require more attention. The potential of application integration will continue to increase the importance of the portal technology. In this context, the aspect of the inte-

[1] "New" in this context means a service that a user has not visited before.

gration of internal and external (web) services will raise the range of processes that can be provided through a portal to the employees of a company.

As discussed in the second chapter, the process of continuous portal engineering plays an important part in the ongoing evolution a portal. This aspect should be further discussed and backed with more case studies on the subject. Thereby, the focus should be laid on the "soft" challenges related to keeping users informed about the potential that the portal offers to support their work. The introduced distinction between primary and secondary services will help in analyzing these challenges. The existence and potential relevance of secondary services can be extended to other areas of software engineering. While this work is limited on a rather broad view of service, which regards a portal service as a tool to execute a task, this view can be extended to single features. In CoFiPot, i.e., a typical service is the one to look up exchange rates. The detailed view would differentiate between the different reports that are provided by this service. While the broad view is especially useful to increase the awareness of new users, the detailed view would also provide new information to users who already have a good overview of the portal services.

The change management aspect of the portal engineering process also requires further studies. Especially in large portals, this aspects needs to become part of the ongoing development process. The findings from the analysis of CoFiPot have shown, that the overall user perspective is not considered when implementing new or changing existing services. Only the key users of each service are informed, if even. An interesting research question would be to determine how other corporate portal providers perceive and prioritize this task.

### 6.2.2. Corporate Portal Recommender Systems

The findings from the recommender system case study have shown the potential that the technology offers to support the process to inform users about a portal's service portfolio. The case study also proofed the applicability of the model for the implementation of such a recommender system, which was developed in Chapter 4. A long term study of the effects of such a system would support the recommender research.

The interesting aspect compared to other scientific literature is the focus on recommending services for long term usage compared to, e.g., recommending products or news articles for "one-time" usage. A long term study should also evaluate the effects on new users and the implementation of new services. It would be of special interest to provide a recommender system directly at the start of a new corporate portal.

The channel how recommendations are communicated to the user should also be reviewed in detail. The implementation in CoFiPot has revealed drawbacks regarding the visibility of the messages. The question is whether a redesign can ensure a higher visibility or if alternative channels as presented in Section 4.8 will have lead to better results.

### 6.2.3. Alternative Support Concepts for Corporate Portals

This work focused on the application of a recommender system to increase the users' knowledge about a portal's service portfolio. Alternative concepts were introduced and discussed in Section 2.4. Comparing the effects of a well planned documentation concept, internal marketing or any of the other concepts will support the selection of the best fitting concept for a portal.

None of the support concepts are exclusive, therefore, combinations might lead to even better results. E.g. the recommender system could be extended to not only recommend relevant services but also documentations about services the user is already using. Also, the availability of documentation could be considered when improving the usability of services. Links to relevant documents or short explanations could be displayed next to related buttons or fields in a form.

Combining the advantages of the different concepts could lead to an overall concept for the continuous portal engineering process. This concept could be used to integrate the aspects of evolution in the development of new corporate portals and help in improving the support to existing implementations.

# Appendices

# A. Corporate Portal Definitions

| *Author(s)* | *Definition* |
|---|---|
| Collins (2000) | A browser-based application that allows knowledge workers to gain access to, collaborate with, make decisions, and take action on a wide variety of business-related information regardless of the employee's virtual location or departmental affiliations, the location of the information, or the formal in which the information is stored. |
| Daniel and Ward (2005) | Enterprise portals are secure web locations that can be customized or personalized, that allow staff and business partners access to, and interaction with a range of internal and external applications and information sources. Uses of the portal may include: improved access to information, increased collaboration, greater use of existing applications and effective integration between applications. |
| Davydov (2001) | [...] The key focus of portals is the integrated access to both information and application services, and greater levels of integration between the two. Corporate portals are extremely valuable to business users because they simplify complex information, provide context-specific and useful application services, and foster collaboration and community building across the extended enterprise. These characteristics represent the rapid evolution of corporate portals from mere corporate intranet sites to mission-critical business tools. [...] |
| Delphi Group (2001) | A personalized workspace that integrates our most relevant sources of information and the underlying connections that make this information valuable to us in a single point of access. |
| Detlor (2000) | Corporate portals are single-point Web browser interfaces used within organizations to promote the gathering, sharing, and dis- |

| *Author(s)* | *Definition* |
|---|---|
| | semination of information throughout the enterprise. As such, these tools offer corporations a means by which to manage and access information from disparate sources across the firm. |
| Eckerson (1999a) | Wayne Eckerson defines a Business Portal as an application that "provides business users one-stop shopping for any information object they need inside or outside the corporation." He therefore emphasizes the gateway aspect of business portal applications as fundamental to the concept. He also emphasizes the importance of shared services such as "security, metadata repository, personalization, search, publish/subscribe," etc., as well as a common look and feel to the gateway. |
| Gartner Group (1998) | A portal is a web site targeted at a specific audience that provides: content aggregation and delivery of information relevant to the audience, collaboration and community services, and application access for the target audience, delivered in a highly personalized manner. |
| Hall and Heffner (2001) | A portal leads somewhere or to something, it is a doorway. In the enterprise, the doorway leads to content, data and services within and beyond the organization. The value of the enterprise portal lies in its ability to provide a single access point to disparate information. |
| Murray (1999) | Portals that focus only on content are inadequate for the corporate market. Corporate portals must connect us not only with everything we need, but with everyone we need, and provide all the tools we need to work together. This means that groupware, e-mail, workflow, and desktop applications-even critical business applications-must all be accessible through the portal. Thus the portal is the desktop, and your commute is just a phone call. |

| *Author(s)* | *Definition* |
|---|---|
| Raol et al. (2003) | An enterprise portal can be defined as a single point of access (SPOA) for the pooling, organizing, interacting, and distributing of organizational knowledge. |
| Ryley (2001) | A corporate portal should put in place the tools and procedures necessary to allow departments to quickly and simply create, maintain, and update their own Websites, and provide appropriate navigation that ties them all together. |
| Shilakes and Tylman (1998) | Enterprise information portals are applications that enable companies to unlock internally and externally stored information, and provide users a single gateway to personalized information needed to make informed business decisions. |
| Smith (2004) | I define portal as an infrastructure providing secure, customizable, personalizable, integrated access to dynamic content from a variety of sources, in a variety of source formats, wherever it is needed. |
| Tatnall (2005) | Enterprise Information Portals. The term enterprise (or corporate) information portals (EIP) is now often being applied to the gateways to the corporate intranets that are used to manage the knowledge within an organization. These are designed primarily for business-to-employee (B2E) processes and offer employees the means to access and share data and information within the enterprise. An EIP may include facilities such as: a categorization of information available on the intranet, a search engine covering the entire intranet, organizational news, access to e-mail, access to common software applications, document management, links to internal sites and popular external Web sites, and the ability to personalize the page. Variations on EIPs include business intelligence portals that are designed to act as gateways to decision-making processes and to provide competitive intelligence, business area portals that support specific busi- |

| *Author(s)* | *Definition* |
|---|---|
| | ness processes such as personnel or supply chain management, and facilities designed to support the field sales forces. |
| Ryley (2001) | A corporate portal should put in place the tools and procedures necessary to allow departments to quickly and simply create, maintain, and update their own Websites, and provide appropriate navigation that ties them all together. |
| Vo (2007) | A corporate portal is a web-based application that serves as a personalized and customizable and single gateway to all the information, application, processes and people that are of relevance for conducting business operations. |
| White (2003) | Application Server Suite: This suite brings together the four key integration infrastructure technologies (i.e., user interface, business process, application and data) in a single package that is combined with an application server and collaboration services. This suite places emphasis on application integration and on an infrastructure for building an integrated business environment. It is similar to what Gartner calls an Application Platform Suite (APS). |

# B. Eckerson's 15 Rules for Enterprise Portals

| | |
|---|---|
| *1)* | Ease of use |
| *2)* | Intuitive classification and searching |
| *3)* | Collaborative information sharing |
| *4)* | Universal connectivity to information resources |
| *5)* | Dynamic access to information resources |
| *6)* | Intelligent routing |
| *7)* | Integrated business intelligence tool |
| *8)* | Server-based architecture |
| *9)* | Distributed services |
| *10)* | Flexible permission granting |
| *11)* | External interfaces |
| *12)* | Programmatic interfaces |
| *13)* | Security |
| *14)* | Easy deployment and maintenance |
| *15)* | Customization and personalization |

*Eckerson, 1999a*

# C. List of CoFiPot Services

*About Us*
- Org Chart

*Controlling*
- Compliance Tool
- BackOffice Tools
- CFC Tools

*Demo*
- Customer
- Invoices
- Payment Analysis
- FI Contracts
- FX Exposure Monitor
- Payment Terms

*Knowledge Base*
- Business in Brief
- CFI Documents
- CRM Documents
- CTR-CM Documents
- F&A Community
- F&A Manager Meetings
- FI Guidelines
- FX Market Reports
- Help
- Planning Guides
- Reporting Templates
- Research
- Search for Affiliates
- Taxes
- Workshops & Seminars
- Worldwide Contacts

*Management*
- BO200 Dates Editor
- Cash Planning Admin
- Cofiplan Setup
- Control Center
- Error Log Viewer
- Issue Tracker
- LDAP Browser
- Module Management
- Page Visits
- Report Monitor
- User Management
- User Status Management
- Validation Center

*Market Focus*
- Aspire
- Finance Charts
- Intercompany IR Rates
- Market Data Monitor
- Official Bayer Rates

*My Cofipot*
- My Documents
- My Reports

*Planning*
- Daily Cash Planning
- Dividend Planning
- Monthly Cash Planning
- Planning Monitor
- WACC Monitor

*Tax*
- GAIS
- GewSt
- Tax Assessments
- Tax Audit
- Tax Computation
- Tax Declaration
- Tax Entity Mgmt
- Tax Planning
- USIS

*Treasury*
- Accounts
- BIC Monitor
- Cash Hedging
- Commodity Monitor
- FX Auto Hedging
- FX Exposure Charts
- FX Exposure Input
- FX Hedge Accounting
- FX Planned Exposure
- FX Trades
- FX Weekly Hedging
- NewKred Monitor

*as of April 2009*

# D. CoFiPot Expert Interview

## Overview

- The interviews were conducted in German. The interviews took around 60 minutes and were recorded.
- Nine interviews were conducted with six managers from the Finance Departments at the Bayer Holding and three regional managers.
- The respondents were free to skip those questions that they did not want to or could not answer.

## Interview Handout

### Corporate Financial Portal - Erfahrungen und Erkenntnisse

### Leitfaden Interview

Vielen Dank für Ihre Teilnahme an diesem Interview.

- Das Ziel dieses Interviews einen Eindruck zu gewinnen, wie vertraut sie mit dem Corporate Financial Portal (CoFiPot) sind und inwieweit ihre Mitarbeiter von der Nutzung dessen profitieren. Zum Abschluss des Interviews wird eine neue Funktionalität vorgestellt, die die Nutzbarkeit des Portals erhöhen soll. Hierzu bitten wir um eine Bewertung.
- Fragen, die Sie nicht beantworten wollen, werden ausgelassen
- Wesentliche Erkenntnisse aus diesem Gespräch werden ein wichtiger Teil meiner Dissertation zum Thema “Corporate Portal Management - Enhancing Service usage” sein.
- Dieses Interview wird ca. 60 Minuten in Anspruch nehmen
- Ich möchte Sie hiermit um Erlaubnis bitten, Teile des Gespräches in meiner Arbeit anonym zu zitieren.

Name: _____

Positionsbeschreibung: _____

Datum: _____

## Teil 1: Vorstellung des Gesprächspartners

### Frage 1.1

Bitte stellen Sie sich kurz vor und skizzieren Sie Ihre Tätigkeit bei Bayer Finanzen.

- Wie lange arbeiten Sie schon für Bayer Finanzen?
- Für welche Abteilung sind Sie verantwortlich?
- Welche Kernaufgaben hat Ihr Bereich zu erfüllen?
- Wie viele Mitarbeiter arbeiten in Ihrem Bereich?
- Wie lange arbeiten Sie schon mit CoFiPot?

### Frage 1.2

Wählen Sie bitte aus welche der folgenden zwei Kategorien auf die Tätigkeiten Ihrer Abteilung zutreffen

**Abteilung**

- □ Accounting/Valuation
- □ Administration
- □ Auditing (inc. PWC)
- □ Cash Management
- □ Controlling
- □ FX Management
- □ Financing
- □ M & A
- □ Planning
- □ Reporting
- □ Risk Management
- □ Tax
- □ Andere:

**Prozesse**

- □ Accounting/Billtrolling
- □ Controlling
- □ Corporate Auditing
- □ Credit Management
- □ Law & Patents
- □ Local/Regional Management
- □ M & A
- □ IT/O&I
- □ Marketing/Sales
- □ PWC
- □ Procurement
- □ Tax
- □ Treasury Finance
- □ Andere:

## Teil 2: Persönliche Erfahrung mit CoFiPot und den angebotenen Modulen

**Frage 2.1**

Wie sind Sie auf CoFiPot aufmerksam geworden?

**Frage 2.2**

Schildern Sie bitte kurz die Bedeutung des Portals in Ihrem Tagesablauf bzw. Tagesgeschäft.

- Wie häufig nutzen Sie das Portal? (mehrmals täglich, täglich, mehrmals die Woche, wöchentlich, etc.)
- Unterliegt die Nutzung saisonaler Schwankung? (z.B. Monatsende, Quartalsende, Jahresende)
- Zu welchem Zweck nutzen Sie das Portal?

**Frage 2.3**

Wie gut kennen Sie das Portal in Bezug auf das Angebot von Modulen und die Möglichkeit, auf gesicherte Module zuzugreifen?

- Mir ist bekannt, dass es gesicherte Module gibt.
- Der Prozess zum Freischalten von Modulen ist mir bekannt.

**Frage 2.4**

Bitte schätzen Sie, wie viele Module das Portal derzeit anbietet?

**Frage 2.5**

In Bezug auf die vom Portal angebotenen Module:

- Auf wie viele Module haben sie derzeit Zugriff (inklusive der allgemein sichtbaren)?
- Wie viele Module nutzen Sie regelmäßig?

- Wie viele Module könnten Sie Ihren Mitarbeitern in ihrer Funktion kurz beschreiben?

**Frage 2.6**

Wie erfahren Sie von neuen Modulen bzw. von signifikanten Änderungen an bestehenden? Bitte beurteilen Sie anhand der folgenden Aussagen auf einer Skala von 1 (trifft nicht zu) bis 5 (trifft in hohem Maße zu).

- Ich werde regelmäßig über neue Module im Portal informiert.
- Ich werde über signifikante Änderungen an Modulen informiert.
- Es ist wichtig für mich, über neue Module informiert zu werden.
- Es ist wichtig für mich über signifikante Änderungen an Modulen informiert zu werden.
- Ich schaue mir regelmäßig das Portal an um Änderungen/Erweiterungen zu finden.

## Teil 3: Erfahrung der Mitarbeiter mit CoFiPot

**Frage 3.1**

Wie viele Ihrer Mitarbeiter nutzen CoFiPot? (Allgemein / Mindestens wöchentlich?)

**Frage 3.2**

Wie lernen Ihre Mitarbeiter das Portal kenne?

- Gibt einen Einweisungsprozess für neue Mitarbeiter?
- Wenn ja, wie sieht dieser aus? (Gibt es ein Standard-Profil)
- Wenn nein, wie lernen die Mitarbeiter das Portal kennen?
- Werden neue Module bekannt gegeben? Wenn ja, wie?

**Frage 3.3**

- Auf wie viele verschiedene Module haben Ihre Mitarbeiter in Summe Zugriff?
- Wie viele dieser Module sind geschützt?

- Wie viele verschiedene Module werden in Summe von Ihren Mitarbeitern genutzt?
- Schwankt die Nutzung dieser Module?
  (z.B. am Monatsende/Quartalsende/Jahresende)
- Welche Module werden besonders häufig genutzt bzw. sollten häufig genutzt werden.

**Frage 3.5**

Gibt es Module, die Ihre Mitarbeiter nicht nutzen sollen? Falls ja, um welche handelt es sich und warum sollen sie nicht genutzt werden?

**Frage 3.6**

Gibt es Prozesse, in denen mehrere Module zusammen genutzt werden?

- Gibt es Prozesse die das Aufrufen mehrerer Module in einer bestimmten Reihenfolge erfordern?
- Gibt es Cluster von Modulen, die immer gemeinsam genutzt werden?

**Frage 3.7**

Sind Ihnen Module bekannt, die den gleichen Zweck erfüllen und somit austauschbar sind?

**Frage 3.8**

Wie erfolgt die Freischaltung von Mitarbeitern auf gesicherte Module?

- Beantragen Sie die Freischaltung?
- Können die Mitarbeiter die Freischaltung selber beantragen?
- Von wem geht im Normalfall die Initiative für die Freischaltung aus?

## Teil 4: Relevanz von CoFiPot

**Frage 4.1**

Wenn man CoFiPot abschalten würde, dann würde dies Sie und Ihre Mitarbeiter bei der Erfüllung Ihrer Aufgaben...

- □ Nicht beeinflussen
- □ Kaum beeinflussen
- □ Beeinflussen
- □ Erschweren
- □ Sehr erschweren

**Frage 4.2**

Bitte bewerten Sie folgende Aussagen auf einer Skala von 1 (Trifft nicht zu) bis 5 (Trifft in hohem Maße zu)

- CoFiPot unterstützt mich und meine Mitarbeiter optimal
- Das Portal könnte von meinen Mitarbeitern häufiger genutzt werden
- Das Portal sollte von meinen Mitarbeitern häufiger genutzt werden
- Ich habe einen guten Überblick über die Module, die von CoFiPot angeboten werden.
- Meine Mitarbeiter haben einen guten Überblick über die Module, die von CoFiPot angeboten werden.

**Frage 4.3**

Wie würden Sie die Relevanz des Portals für Sie und Ihre Mitarbeiter beschreiben? Könnten Sie ein Maßdefinieren, anhand dessen man die Relevanz des Portals messen könnte? (Z.B. Dauer eines Prozess mit und ohne Portal, Qualität der Ergebnisse, etc.)

**Frage 4.4**

Können Sie Alternativen für Dienste nennen, die die Module des Portals bieten.

- Gibt es alternative Quellen zur Abfrage von Reports

- Gibt es Alternativen zur Dateneingabe?
- Gibt es Prozesse, die komplett ohne CoFiPot ausgeführt werden könnten?
- Wenn ja, wäre es aus Ihrer Sicht wünschenswert, wenn dies ohne CoFiPot geschehen würde?
- Wenn nein, welche Rolle spielt CoFiPot in diesen Prozessen?
- Gibt es Prozesse in Ihrer Abteilung die durch die Nutzung bestehender CoFiPot Module besser ausgeführt werden könnten?
- Bitte nennen sie ein Beispiel.
- Warum wäre die Nutzung von CoFiPot Modulen hier sinnvoll?
- Welche Rolle würde der Dienst spielen? (Information, Prozessablauf, etc.)
- Warum wird der Dienst nicht genutzt?
- Welche Art von Dienst würde ersetzt? (Excel, Adaptiv, Mail, etc.)

## Teil 5: System zur automatischen Empfehlung von Modulen

Um Mitarbeiter mit dem Modulangebot des Portals besser vertraut zu machen, planen wir die Einführung eines Empfehlungssystems. Dies stellt dem aktuellen Nutzer Module vor die von verwandten Nutzern genutzt werden. Die Methode ist vergleichbar mit der Buchempfehlung des Online-Buchversands Amazon.

Um diese Empfehlungen anzuzeigen, stehen uns folgende Kanäle zur Verfügung:

- Als Pop Up
- Unter dem aktiven Modul (ähnlich zu Amazon)
- Als Balloon-Tip (- siehe Demo)
- Im Quick Notifier
- Als personalisierte Email (beispielsweise Monatlich)

Zusätzlich erwägen wir die Erstellung von Empfehlungslisten nach Abteilungen, die dem jeweiligen Management zur Verfügung gestellt werden könnte.

**Frage 5.1**

Wie schätzen sie das Potential des Recommender Systems auf einer Skala von 1 (Trifft nicht zu) bis 5 (Trifft in hohem Maße zu) ein:

- Das System hat Potential, Mitarbeiter bei der Arbeit zu unterstützen
- Neuen Mitarbeitern wird der Einstieg in das Portal erleichtert
- Das Management kann von einer Übersicht der Empfehlungen profitieren
- Das System erhöht die Nutzbarkeit/Verständlichkeit von CoFiPot.

**Frage 5.2**

Bitte bewerten sie die verschiedenen Optionen zur Anzeige von Empfehlungen anhand der Optionen: Sehr störend, störend, neutral, effektiv, sehr effektiv:

- PopUp
- Unterhalb des aktiven Modules
- Balloon-Tip
- Quick Notifier
- Personalisierte Email

**Frage 5.3**

Haben weitere Sie Anmerkungen/Anregungen zu dem System.

**Frage 5.4**

Haben sie weitere Anregungen, wie man die Wahrnehmung von CoFiPot in Bezug auf die Angebotenen Module verbessern könnte.

Vielen Dank für das Gespräch!

# E. Recommender System Classification

Recommender system classification by Adomavicius and Tuzhilin (2005)

| *Recommendation Approach* | *Recommendation Technique* | |
|---|---|---|
| | *Heuristic-based* | *Model-based* |
| *Content-based* | - TF-IDF (information retrieval<br>- Clustering | - Bayesian classifiers<br>- Clustering<br>- Artificial neural networks |
| *Collaborative* | - Nearest neighbor (cosine, correlation)<br>- Clustering<br>- Graph theory | - Bayesian networks<br>- Clustering<br>- Artificial neural networks<br>- Linear regression<br>- Probabilistic models |
| *Hybrid* | - Linear combination of predicted ratings<br>- Various voting schemes<br>- Incorporating one component as a part of the heuristic for the other | - Incorporating one component as a part of the model for the other<br>- Building one unifying model |

# F. Statistical Accuracy Metrics

| *Metric* | *Used for* | *References* |
|---|---|---|
| Mean Absolute Error (MAE) | Accuracy | Shardanand and Maes (1995)<br>Breese et al. (1998)<br>Herlocker et al. (1999), (2004)<br>Fisher et al. (2000)<br>Calderón-Benavides et al. (2004)<br>Adomavicius and Tuzhilin (2005) |
| Variance MAE | Reliability | Fisher et al. (2000)<br>Calderón-Benavides et al. (2004) |
| Mean Squared Error (MSE) | Reliability | Shardanand and Maes (1995)<br>Herlocker et al. (2004) |
| Root MSE | Reliability | Herlocker et al. (1999), (2004)<br>Adomavicius and Tuzhilin (2005) |
| Mean Error | Distortion | Fisher et al. (2000) |
| Normalized MAE | Accuracy | Herlocker et al. (2004) |
| MAE of the Extremes | Accuracy | Herlocker et al. (2004) |
| Half-Life Utility Metric | Accuracy (in lists) | Breese et al. (1998)<br>Herlocker et al. (2004) |
| Top-N Evaluation | Accuracy (in lists) | Lam (2004) |
| Kappa-Index | Approval | Cohen et al. (1972)<br>Calderón-Benavides et al. (2004) |
| Person's Correlation Coefficient | Correlation (Complexity) | Shardanand and Maes (1995)<br>Hill et al. (1995)<br>Herlocker et al. (1999), (2004)<br>Adomavicius and Tuzhilin (2005) |

| *Metric* | *Used for* | *References* |
|---|---|---|
| Spearman's Rank Correlation Coefficient | Rank Correlation | Mooney and Roy (2000)<br>Herlocker et al. (1999), (2004) |
| Kendall's Tau | Rank Correlation | Herlocker et al. (2004) |
| Weighted Mean | Complexity | Fisher et al. (2000) |
| Receiver Operation Characteristic (ROC) Curves | | Herlocker et al. (1999), (2004)<br>Adomavicius and Tuzhilin (2005) |
| C ROC | | Herlocker et al. (2004) |
| Swet's A Measure | | Herlocker et al. (2004) |
| Precision,<br>Recall,<br>F1-Measure | | Mooney and Roy (2000)<br>Sarwar et al. (2001)<br>Herlocker et al. (2004)<br>Adomavicius and Tuzhilin (2005) |
| Standard Deviation of Errors | | Sarwar et al. (2001) |
| Mean Average Precision | Accuracy | Herlocker et al. (2004) |
| Normalized Distance-Based Performance | | Herlocker et al. (2004) |

# G. Statistical Results of Recommender Test Runs

Averages of deviations between predicted and original ratings using the Bonferroni method based on observed means. For test runs All but 1/2 (AB1/2), Given 25 (G2/5). Tests for rating rules Weekly and Monthly Average 1/2 (WA1/2, MA1/2), user similarity algorithms Pearson's Correlation (Corr), Inverse User Frequency (IUF) and Vector Similarity (VS), prediction algorithms Average Rating (AR), Weighted Sum (WS) and Adapted Weighted Sum (AWS).

| *Test* | | | | *Mean* | | | *95% Conf. Interval* | |
|---|---|---|---|---|---|---|---|---|
| *Method* | *Algorithms* | | | *Difference* | *Std. Error* | *Sig.* | *Lower* | *Upper* |
| *AB1* | WA1 | x | WA2 | .0327 | .03179 | 1.000 | -.0512 | .1166 |
| | | x | MA1 | .3138 | .02777 | .000 | .2405 | .3870 |
| | | x | MA2 | .1866 | .03140 | .000 | .1037 | .2695 |
| | WA2 | x | MA1 | .2810 | .02779 | .000 | .2077 | .3544 |
| | | x | MA2 | .1539 | .03143 | .000 | .0709 | .2368 |
| | MA1 | x | MA2 | -.1272 | .02735 | .000 | -.1993 | -.0550 |
| | Corr | x | IUF | .0204 | .02621 | 1.000 | -.0423 | .0832 |
| | | x | VS | .0029 | .02598 | 1.000 | -.0593 | .0652 |
| | IUF | x | VS | -.0175 | .02335 | 1.000 | -.0734 | .0384 |
| | AR | x | WS | .0364 | .02476 | .425 | -.0229 | .0957 |
| | | x | AWS | -.1225 | .02476 | .000 | -.1818 | -.0632 |
| | WS | x | AWS | -.1588 | .02476 | .000 | -.2181 | -.0996 |
| *AB2* | WA1 | x | WA2 | .0140 | .01874 | 1.000 | -.0354 | .0635 |
| | | x | MA1 | .1777 | .01762 | .000 | .1312 | .2242 |
| | | x | MA2 | .0093 | .02068 | 1.000 | -.0453 | .0638 |
| | WA2 | x | MA1 | .1637 | .01765 | .000 | .1171 | .2102 |
| | | x | MA2 | -.0048 | .02070 | 1.000 | -.0594 | .0499 |
| | MA1 | x | MA2 | -.1684 | .01969 | .000 | -.2204 | -.1165 |

| Test Method | Algorithms | | | Mean Difference | Std. Error | Sig. | 95% Conf. Interval Lower | Upper |
|---|---|---|---|---|---|---|---|---|
| | Corr | x | IUF | .0277 | .01672 | .291 | -.0123 | .0678 |
| | | x | VS | .0155 | .01663 | 1.000 | -.0243 | .0553 |
| | IUF | x | VS | -.0123 | .01567 | 1.000 | -.0498 | .0252 |
| | AR | x | WS | .0281 | .01624 | .252 | -.0108 | .0670 |
| | | x | AWS | -.0260 | .01624 | .330 | -.0649 | .0129 |
| | WS | x | AWS | -.0540 | .01624 | .003 | -.0929 | -.0151 |
| *G2* | WA1 | x | WA2 | .0221 | .00953 | .121 | -.0030 | .0473 |
| | | x | MA1 | .1200 | .00996 | .000 | .0937 | .1463 |
| | | x | MA2 | .2603 | .00871 | .000 | .2373 | .2833 |
| | WA2 | x | MA1 | .0979 | .00994 | .000 | .0716 | .1241 |
| | | x | MA2 | .2381 | .00870 | .000 | .2152 | .2611 |
| | MA1 | x | MA2 | .1403 | .00916 | .000 | .1161 | .1644 |
| | Corr | x | IUF | .0623 | .00879 | .000 | .0413 | .0834 |
| | | x | VS | .0557 | .00877 | .000 | .0347 | .0767 |
| | IUF | x | VS | -.0067 | .00726 | 1.000 | -.0241 | .0107 |
| | AR | x | WS | .0143 | .00792 | .210 | -.0046 | .0333 |
| | | x | AWS | -.1380 | .00793 | .000 | -.1570 | -.1190 |
| | WS | x | AWS | -.1523 | .00793 | .000 | -.1713 | -.1334 |
| *G5* | WA1 | x | WA2 | .0124 | .00980 | 1.000 | -.0134 | .0383 |
| | | x | MA1 | .1125 | .01001 | .000 | .0861 | .1389 |
| | | x | MA2 | .2450 | .00894 | .000 | .2214 | .2686 |
| | WA2 | x | MA1 | .1001 | .01002 | .000 | .0737 | .1265 |
| | | x | MA2 | .2326 | .00895 | .000 | .2090 | .2562 |
| | MA1 | x | MA2 | .1325 | .00917 | .000 | .1083 | .1567 |
| | Corr | x | IUF | .0831 | .00831 | .000 | .0632 | .1030 |
| | | x | VS | .0757 | .00830 | .000 | .0558 | .0956 |
| | IUF | x | VS | -.0074 | .00776 | 1.000 | -.0260 | .0112 |
| | AR | x | WS | .0100 | .00807 | .652 | -.0094 | .0293 |

| *Test Method* | *Algorithms* | | | *Mean Difference* | *Std. Error* | *Sig.* | *95% Conf. Interval* *Lower* | *Upper* |
|---|---|---|---|---|---|---|---|---|
| | | x | AWS | -.0461 | .00807 | .000 | -.0654 | -.0268 |
| | WS | x | AWS | -.0561 | .00807 | .000 | -.0754 | -.0367 |

# References


Abdinnour-Helm, S. F., B. S. Chaparro, and S. M. Farmer (2005). Using the end-user computing satisfaction (EUCS) instrument to measure satisfaction with a web site. *Decision Science 36*(2), 341–364.

Adomavicius, G. and A. Tuzhilin (2001). Expert-driven validation of rule-based user models in personalization applications. *Data Mining and Knowledge Discovery 5*(1-2), 33–58.

Adomavicius, G. and A. Tuzhilin (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering 17*(6), 734–749.

Aneja, A., C. Rowan, and B. Brooksby (2000). Corporate portal framework for transforming content chaos on intranets. *Intel Technology Journal 4*(1), 1–7.

Arbin, K. and U. Essler (2002). Emerging industrial eMarkets: The case of Covisint in Europe. In *Proceedings of the 15th International Electronic Commerce Conference*, Bled, Slovenia, pp. 279–292.

Arroyo, E., T. Selker, and W. Wei (2006). Usability tool for analysis of web designs using mouse tracks. In *CHI '06 extended abstracts on Human factors in computing systems*, New York, NY, USA, pp. 484–489. ACM.

Ashenfelter, O. and D. Card (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics 67*(4), p 648–660.

Avery, C. and R. Zeckhauser (1997). Recommender systems for evaluating computer messages. *Communications of the ACM 40*(3), 88–89.

Balabanović, M. and Y. Shoham (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM 40*(3), 66–72.

Basu, C., H. Hirsh, and W. Cohen (1998). Recommendation as classification: Using social and content-based information in recommendation. In *Proceedings of the 15th National Conference on Artificial Intelligence*, pp. 714–720. Madison, WI, USA.

Baudisch, P. (1999). Joining collaborative and content-based filtering. In *Online Proceedings of Interacting with Recommender Systems, CHI '99 Workshop*.

Baumard, P. and P. Baumard (1999). *Tacit knowledge in organizations*. London, UK: Sage Publications.

Bell, D. E. (2005). Looking back at the Bell-La Padula model. In *ACSAC '05: Proceedings of the 21st Annual Computer Security Applications Conference*, Washington, DC, USA, pp. 337–351. IEEE Computer Society.

Bell, R. M. and Y. Koren (2007). Improved neighborhood-based collaborative filtering. In *KDDCup'07*, San Jose, CA, USA.

Benantar, M. (2006). *Access control systems: Security, identity management and trust models*. Boston, MA, USA: Springer Science+Business Media Inc.

Benbya, H., G. Passiante, and N. A. Belbaly (2004). Corporate portal: A tool for knowledge management synchronization. *International Journal of Information Management 24*(3), 201–220.

Berkovsky, S., Y. Eytani, and L. M. Manevitz (2007). Efficient collaborative filtering in content-addressable spaces. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) 21*(2), 265–289.

Billsus, D. and M. J. Pazzani (1998). Learning collaborative information filters. In *Proceedings of the 15th International Conference on Machine Learning*, San Francisco, CA, USA, pp. 46–54. Morgan Kaufmann.

Billsus, D. and M. J. Pazzani (2000). User modeling for adaptive news access. *User Modeling and User-Adapted Interaction 10*(2-3), 147–180.

Braun, S. and A. Schmidt (2007). Wikis as a technology fostering knowledge maturing: What we can learn from Wikipedia. In *Proceedings of the 7th International Conference on Knowledge Management (I-KNOW 2007), Special Track on Integrating Working and Learning in Business (IWL), Graz, 2007*.

Breese, J. S., D. Heckerman, and C. Kadie (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence (UAI-98)*, San Francisco, CA, USA, pp. 43–52. Morgan Kaufmann.

Buffa, M. (2006). Intranet wikis. In *WWW Conference 2006*, Edinburgh, Scotland.

Bughin, J., M. Chui, and A. Miller (2009). How companies are benefitting from Web 2.0: A McKinsey Global Survey. *The McKinsey Quarterly* (3).

Buono, P., M. F. Costabile, S. Guida, and A. Piccinno (2002). Integrating user data and collaborative filtering in a web recommendation system. In S. Reich (Ed.), *Hypermedia: openness, structural awareness, and adaptivity*, Volume 2266 of *Lecture notes in computer science*, Berlin, Germany, pp. 192–196. Springer-Verlag.

Burke, R. (2000). Knowledge-based recommender systems. In A. Kent (Ed.), *Encyclopedia of Library and Information Systems*, Volume 69, Supplement 32.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction 12*(4), 331–370.

Burke, R. D., K. J. Hammond, and B. C. Young (1997). The FindMe approach to assisted browsing. *IEEE Expert 12*(4), 32–40.

Caglayan, A., M. Snorrason, J. Jacoby, R. Jones, and K. Kumar (1997). Learn Sesame - A learning agent engine. *Applied Artificial Intelligence: An International Journal 11*(5), 393–412.

Calderón-Benavides, M. L., C. N. González-Caro, J. J. de Pérez-Alcázar, J. C. Garc\iaz, and J. Delgado (2004). A comparison of several predictive algorithms for collabo-

rative filtering on multi-valued ratings. In *SAC '04: Proceedings of the 2004 ACM symposium on Applied computing*, New York, NY, USA, pp. 1033–1039. ACM.

Card, S. K., P. Pirolli, M. van der Wege, J. B. Morrison, R. W. Reeder, P. K. Schraedley, and J. Boshart (2001). Information scent as a driver of Web behavior graphs: Results of a protocol analysis method for Web usability. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 498–505. ACM.

Chui, M., A. Miller, and R. P. Roberts (2009). Six ways to make Web 2.0 work. *The McKinsey Quarterly* (1).

Cingil, I., A. Dogac, and A. Azgin (2000). A broader approach to personalization. *Communications of the ACM 43*(8), 136–141.

Clark, D. D. and D. R. Wilson (1987). A comparison of commercial and military computer security policies. *IEEE Symposium on Security and Privacy*, 184–194.

Cleverdon, C. W. and M. Keen (1968). Factors determining the performance of indexing systems. https://dspace.lib.cranfield.ac.uk/handle/1826/863.

Cloete, M. and R. Snyman (2003). The enterprise portal – Is it knowledge management? *Aslib Proceedings 55*(4), 234–242.

Cohen, M. D., J. G. March, and J. P. Olsen (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 1–25.

Collins, H. (2000). *Corporate portals: Revolutionizing information access to increase productivity and drive the bottom line*. New York, NY, USA: American Management Assoc., Inc.

Condliff, M. K., D. D. Lewis, D. Madigan, and C. Posse (1999). Bayesian mixed-effects models for recommender systems. In *SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation*. Berkeley, CA, USA.

Cooley, R., B. Mobasher, and J. Srivastava (1999). Data preparation for mining World Wide Web browsing patterns. *Journal of Knowledge and Information Systems 1*(1).

Counsell, A. (2004). Opening the door to collaboration. *Financial Times IT Review (31.03.2004)*.

Cunningham, W. (2003). Exploring with Wiki: A conversation with Ward Cunningham: Part I by Bill Venners. http://www.artima.com/intv/wiki.html.

Daniel, E. M. and J. M. Ward (2005). Enterprise portals: Addressing the organisational and individual perspectives of information systems. In *Proceedings of the 13th European Conference on Information Systems, Information Systems in a Rapidly Changing Economy, ECIS 2005, Regensburg, Germany, May 26-28, 2005*.

Daniel, E. M. and J. M. Ward (2006). Integrated service delivery: Exploratory case studies of enterprise portal adoption in UK local government. *Business Process Management Journal 12*(1), 1463–7154.

Davenport, T. H. (2005). *Thinking for a living: How to get better performance and results from knowledge workers*. Boston, MA, USA: Harvard Business Publishing.

Davenport, T. H., D. W. de Long, and M. C. Beers (1998). Successful knowledge management projects. *Sloan Management Review 39*(2), 43–57.

Davenport, T. H. and L. Prusak (1997). *Information ecology: Mastering the information and knowledge environment*. Oxford University Press.

David, F. R. and V. Venkatesh (1996). A model of the ancedents of perceived ease of use: Development and test. *Decision Science 27*(3), 451–481.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly 13*(3), 319–340.

Davydov, M. M. (2001). *Corporate portals and e-business integration*. Emerging business technology series. New York, NY, USA: McGraw-Hill.

Delgado, J. and Ishii Nahoiro (1999). Memory-based weighted-majority prediction for recommender systems. In *SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation*. Berkeley, CA, USA.

Delphi Group (2001). Enabling the agile Enterprise: White Paper.

Denning, D. E. (1976). A lattice model of secure information flow. *Communications of the ACM 19*(5), 236–243.

Department of Defense (1985). Department of Defense trusted computer system evaluation criteria. http://handle.dtic.mil/100.2/ADA207905, Department of Defense Fort George G Meade MD Computer Security Center.

Deshpande, M. and G. Karypis (2004). Item-based top-N recommendation algorithms. *ACM Transactions on Information Systems 22*(1), 143–177.

Désilets, A., S. Paquet, and N. G. Vinson (2005). Are wikis usable? In *Conference proceedings of the 2005 International Symposium on Wikis*, pp. 3–15. New York, NY, USA: ACM.

Detlor, B. (2000). The corporate portal as information infrastructure: Towards a framework for portal design. *International Journal of Information Management 20*(2), 91–101.

Detlor, B. (2004). *Towards knowledge portals: From human issues to intelligent agents*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dias, C. (2001). Corporate portals: A literature review of a new concept in Information Management. *International Journal of Information Management 21*(4), 269–287.

Doll, W., A. Hendrickson, and D. Xiandong (1998). Using Davie's perceived usefulness and ease-of-use instruments for decision making: A confirmatory and multigroup invariance analysis. *Decision Science 29*(4), 839–869.

Eckerson, W. (1999a). 15 rules for enterprise portals. *Oracle Magazine 13*(4), 13–14.

Eckerson, W. (1999b). Plumtree blossoms: New version fulfils enterprise portal requirements: Whitepaper. Boston, MA, USA.

Edelman, R. and T. Jussila (2000). Portals help integrate interdependent applications: Corporate portals provide a single focal point for initiating processes. http://www.informationweek.com/816/portalsoftware.htm.

Eirinaki, M. and M. Vazirgiannis (2003). Web mining for web personalization. *ACM Transactions on Internet Technologies 3*(1), 1–27.

Elsner, H. and J. Krämer (2010a). Management komplexer Unternehmensportale – Herausforderungen und Lösungsansätze. In *72. Wissenschaftliche Jahrestagung des Verbandes der Hochschullehrer für Betriebswirtschaft*, Bremen, Germany.

Elsner, H. and J. Krämer (2010b). Phasen und Herausforderungen bei der Entwicklung eines Unternehmensportals. In *Tagungsband der 1. Konferenz zu Mitarbeiterportalen in der Unternehmenspraxis (MiPo'10)*, Aalen, Germany.

Elsner, H. and H. T. K. Vo (2007). Management of portal evolution - Introducing evolution management for the Corporate Financial Portal. In *Wirtschaftsinformatik Proceedings*, Volume 2 of *eOrganisation: Service-, Prozess-, Market-Engineering*, pp. 337–352.

Fagin, R., R. Kumar, K. McCurley, J. Novak, and D. Sivakumar (2003). Searching the workplace web. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, Budapest, Hungary, pp. 366–375.

Fernandez, E. B. and J. C. Hawkins (1997). Determining role rights from use cases. In *RBAC '97: Proceedings of the second ACM workshop on Role-based access control*, New York, NY, USA, pp. 121–125. ACM.

Firestone, J. M. (2000). Defining the enterprise information portal. http://www.tgc.com/dsstar/00/0822/102054.html.

Fisher, D., K. Hildrum, J. Hong, M. Newman, M. Thomas, and R. Vuduc (2000). SWAMI (Poster-Session): A framework for collaborative filtering algorithm development and evaluation. In *SIGIR 2000: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, pp. 366–368. ACM.

Gantner, J., A. Geyer-Schulz, and A. Thede (2006). A single sign-on protocol for distributed web applications based on standard internet mechanisms. In J. a. Ascenso,

L. Vasiu, C. Belo, and M. Saramago (Eds.), *e-Business and Telecommunication Networks*, pp. 176–183. Springer Netherlands.

Gold, A., A. Malhotra, and A. Segars (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems 18*(1), 185–214.

Goldberg, D., D. Nichols, B. M. Oki, and D. Terry (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM 35*(12), 61–70.

Großmann, M. and H. Koschek (2005). *Unternehmensportale: Grundlagen, Architekturen, Technologien*. Berlin, Germany: Xpert.Press.

Grudin, J. (1994). Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM 37*(1), 92–105.

Gurugé, A. (2003). *Corporate portals empowered with XML and web services*. Digital Press.

Hahn, J. and M. R. Subramani (2000). A framework of knowledge management systems: Issues and challenges for theory and practice. In *Proceedings of the twenty first international conference on Information systems*, Brisbane, Queensland, Australia, pp. 302–312. Association for Information Systems.

Hall, K. and R. Heffner (2001). Enterprise portals: Evaluation criteria. Cambridge, MA, USA: Giga Information Group, Inc.

Harper, F. M., X. Li, Y. Chen, and J. A. Konstan (2005). An economic model of user rating in an online recommender system. In *User Modeling 2005*, pp. 307–316.

Harrison, M. A., W. L. Ruzzo, and J. D. Ullman (1976). Protection in operating systems. *Communications of the ACM 19*(8), 461–471.

Hasan, H. and C. C. Pfaff (2006). The Wiki: An environment to revolutionise employees' interaction with corporate knowledge. In *Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments*, pp. 377–380. ACM.

Hazra, T. K. (2002). Building enterprise portals: Principles to practice. In *Software Engineering, 2002. ICSE 2002. Proceedings of the 24rd International Conference on*, pp. 623–633.

Herlocker, J. L., J. A. Konstan, A. Borchers, and J. Riedl (1999). An algorithmic framework for performing collaborative filtering. In *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 230–237. ACM.

Herlocker, J. L., J. A. Konstan, L. G. Terveen, and J. T. Riedl (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems 22*(1), 5–53.

Hill, W., L. Stead, M. Rosenstein, and G. Furnas (1995). Recommending and evaluating choices in a virtual community of use. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 194–201. ACM/Addison-Wesley.

Holtshouse, D. (1998). Knowledge research issues. *California Management Review 40*(3), 277–280.

Hornung, T., A. Koschmider, and A. Oberweis (2007). A recommender system for business process models. In *Seventeenth Annual Workshop on Information Technologies and Systems in conjunction with International Conference on Information Systems*, Montréal, Canada, pp. 127–132.

Huang, S.-L. (2008). Comparison of utility-based recommendation methods. In *PACIS 2008 Proceedings*.

Kamishima, T. (2003). Nantonac collaborative filtering: Recommendation based on order responses. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, pp. 583–588. ACM.

Kaplan, S. and R. Henderson (2005). Inertia and incentives: Bridging organizational economics and organizational theory. *Organization Science 16*(5), 509–552.

Karypis, G. (2001). Evaluation of item-based top-N recommendation algorithms. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, New York, NY, USA, pp. 247–254. ACM.

Knorr, E. (2004). The new enterprise portal: The browser-based portal is fast becoming the enterprise UI and the nexus for a new breed of integration and app dev. http://www.infoworld.com/d/developer-world/new-enterprise-portal-131.

Kobsa, A. (2007). Privacy-enhanced web personalization. In P. Brusilovsky (Ed.), *The Adaptive Web*, Lecture notes in computer science; 4321: State-of-the-art survey. Berlin, Germany: Springer-Verlag.

Kobsa, A., J. Koenemann, and W. Pohl (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *The Knowledge engineering review 16*(2), 111–155.

Konstan, J. A., B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. T. Riedl (1997). GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM 40*(3), 77–87.

Kosala, R. and H. Blockeel (2000). Web mining research: A survey. *ACM SIGKDD Explorations Newsletter 2*(1), 1–15.

Krämer, J., H. Elsner, and C. Weinhardt (2010). The Ricommander - Rating recommendations for corporate community portals. In *Proceedings of the 11th Group Decision and Negotiation Conference (GDN)*, Delft, The Netherlands.

Lam, A. (2000). Tacit knowledge: Organizational learning and societal institutions: An integrated framework. *Organization Studies 21*(3).

Lam, C. (2004). SNACK: Incorporating social network information in automated collaborative filtering. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce*, New York, NY, USA, pp. 254–255. ACM.

Lam, S. K. and J. Riedl (2004). Shilling recommender systems for fun and profit. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA, pp. 393–402. ACM.

Lampson, B. W. (1974). Protection. *SIGOPS Operating Systems Review 8*(1), 18–24.

Lederer, A. L., D. J. Maupin, M. P. Sena, and Y. Zhuang (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems 29*(3), 269–282.

Leonard, D. and S. Sensiper (1998). The role of tacit knowledge in group innovation. *California Management Review 4*(3), 112–132.

Leuf, B. and W. Cunningham (2001). *The wiki way: Collaboration and sharing on the Internet*. Addison-Wesley.

Lieberman, H. (1995). Letizia: An agent that assists web browsing. In Chris S. Mellish (Ed.), *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI–95)*, Montréal, Canada, pp. 924–929. Morgan Kaufmann.

Linden, G., B. Smith, and J. York (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing 7*(1), 76–80.

Luce, C. (2002). Three trends in enterprise information portals. http://news.zdnet.com/2100-10532_22-297610.html.

Maltz, D. and K. Ehrlich (1995). Pointing the way: Active collaborative filtering. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 202–209. ACM/Addison-Wesley.

Marchionini, G. (1995). *Information seeking in electronic environments*. New York, NY, USA: Cambridge University Press.

Massa, P. and P. Avesani (2004). Trust-aware collaborative filtering for recommender systems. In R. Meersman and Z. Tari (Eds.), *On the move to meaningful Internet systems 2004: CoopIS, DOA, and ODBASE*, Volume 3290 of *Lecture notes in computer science*, pp. 492–508. Berlin / Heidelberg, Germany: Springer-Verlag.

Mathieson, K., E. Peacock, and W. Chin (2001). Extending the technology acceptance model: The influence of perceived user resources. *SIGMIS Database 32*(3), 86–112.

McAfee, A. (2006). Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review 47*(3), 21–28.

Middleton, S. E., N. Shadbolt, and D. C. de Roure (2004). Ontological user profiling in recommender systems. *ACM Transactions on Information Systems 22*(1), 54–88.

Mobasher, B. (1999). WebPersonalizer: A server side recommender system based on web usage mining. In *Proceedings of the 9th Workshop on Information Technologies and Systems (WITS'99)*.

Mobasher, B., R. Burke, and J. J. Sandvig (2006). Model-based collaborative filtering as a defense against profile injection attacks. In *AAAI'06: proceedings of the 21st national conference on Artificial intelligence*, pp. 1388–1393. AAAI Press.

Mobasher, B., R. Cooley, and J. Srivastava (1999). Creating adaptive web sites through usage–based clustering of URLs. In *Workshop on Knowledge and Data Engineering Exchange. (KDEX '99) Proceedings*, pp. 19–25.

Mobasher, B., R. Cooley, and J. Srivastava (2000). Automatic personalization based on web usage mining. *Communications of the ACM 43*(8), 142–151.

Mobasher, B., H. Dai, T. Luo, and M. Nakagawa (2002). Discovery and evaluation of aggregate usage profiles for web personalization. *Data Mining and Knowledge Discovery 6*(1), 61–82.

Moffett, J. D. and M. S. Sloman (1991). Content-dependent access control. *SIGOPS Operating Systems Review 25*(2), 63–70.

Molloy, I., H. Chen, T. Li, Q. Wang, N. Li, E. Bertino, S. Calo, and J. Lobo (2008). Mining roles with semantic meanings. In *SACMAT '08: Proceedings of the 13th ACM symposium on Access control models and technologies*, New York, NY, USA, pp. 21–30. ACM.

Money, W. and A. Turner (2004). Application of the technology acceptance model to a knowledge management system. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004*, Los Alamitos, CA, USA. IEEE Computer Society.

Montaner, M., B. López, and J. L. de La Rosa (2003). A taxonomy of recommender agents on the Internet. *Artificial Intelligence Review 19*(4), 285–330.

Mooney, R. J. and L. Roy (2000). Content-based book recommending using learning for text categorization. In *DL '00: Proceedings of the fifth ACM conference on Digital libraries*, New York, NY, USA, pp. 195–204. ACM.

Mulvenna, M. D., S. S. Anand, and A. G. Buchner (2000). Personalization on the Net using web mining: Introduction. *Communications of the ACM 43*(8), 122–125.

Murray, G. (1999). The portal is the desktop. Los Altos, CA, USA: Intraspect Inc.

Neumann, G. and M. Strembeck (2002). A scenario-driven role engineering process for functional RBAC roles. In *SACMAT '02: Proceedings of the seventh ACM symposium on Access control models and technologies*, New York, NY, USA, pp. 33–42. ACM.

Nielsen, J. (1993). *Usability engineering*. Morgan Kaufmann.

Nielsen, J. (1999). *Designing web usability: The practice of simplicity*. Thousand Oaks, CA, USA: New Riders Publishing.

Oard, D. W. and J. Kim (1998). Implicit feedback for recommender systems. In *Proceedings of the AAAI Workshop on Recommender Systems*, pp. 81–83. Madison, WI, USA.

O'Mahony, M. P., N. J. Hurley, and G. C. M. Silvestre (2004). Utility-based neighbourhood formation for efficient and robust collaborative filtering. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce*, New York, NY, USA, pp. 260–261. ACM.

O'Reilly, T. (2005). What Is Web2.0: Design patterns and business models for the next generation of software. http://oreillynet.com/lpt/a/6228.

Orlikowski, W. J. (1992). Learning from Notes: Organizational issues in groupware implementation. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, New York, NY, USA, pp. 362–369. ACM.

Orlikowski, W. J. (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization Science 13*(3), 249–273.

Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review 13*(5-6), 393–408.

Pazzani, M. J. and D. Billsus (2007). Content-based recommendation systems. In *The Adaptive Web*, pp. 325–341. Berlin / Heidelberg, Germany: Springer-Verlag.

Pennock, D. M., E. Horvitz, and C. L. Giles (2000). Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, Menlo Park, CA, USA, pp. 729–734. AAAI Press.

Phifer, G. (2003). The portal is dead, long live the portal: Gartner Report.

Polanyi, M. (1997). The tacit dimension. In L. Prusak (Ed.), *Knowledge in organizations*, Resources for the knowledge-based economy, pp. 135–146. Boston, MA, USA: Butterworth-Heinemann.

Raghavan, P. (2002). Social networks: From the Web to the enterprise. *IEEE Internet Computing 6*(1), 91–94.

Raol, J. M., K. S. Koong, and C. S. Yu (2003). An identification and classification of enterprise portal functions and features. *Industrial Management & Data Systems 103*(9), 693–702.

Remus, U. (2007). Critical success factors for implementing enterprise portals: A comparison with ERP implementations. *Business Process Management Journal 13*(4), 538–552.

Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl (1994). GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94)*, New York, NY, USA, pp. 175–186. ACM.

Resnick, P. and R. Sami (2008). Manipulation-resistant recommender systems through influence limits. *SIGecom Exch. 7*(3), 1–4.

Roeckle, H., G. Schimpf, and R. Weidinger (2000). Process-oriented approach for role-finding to implement role-based security administration in a large industrial organization. In *RBAC '00: Proceedings of the fifth ACM workshop on Role-based access control*, New York, NY, USA, pp. 103–110. ACM.

Runte, M. (2000). *Personalisierung im Internet: Individualisierte Angebote mit Collaborative Filtering*. Ph. D. thesis, Christian-Albrechts-Universität zu Kiel, Kiel, Germany.

Ryley, S. (2001). Corporate portal development: A practical approach ensures real business benefits. *Business Information Review 18*(2), 28–34.

Salton, G. and M. J. McGill (1986). *Introduction to modern information retrieval*. New York, NY, USA: McGraw-Hill.

Sarwar, B., G. Karypis, J. Konstan, and J. Reidl (2001). Item-based collaborative filtering recommendation algorithms. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, New York, NY, USA, pp. 285–295. ACM.

Sarwar, B. M., J. A. Konstan, A. Borchers, J. L. Herlocker, B. Miller, and J. Riedl (1998). Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. In *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pp. 345–354. New York, NY, USA: ACM.

Schafer, J. B., D. Frankowski, J. Herlocker, and S. Sen (2007). Collaborative filtering recommender systems. In *The Adaptive Web*, pp. 291–324. Berlin / Heidelberg, Germany: Springer-Verlag.

Schafer, J. B., J. A. Konstan, and J. Riedl (2001). E-Commerce recommendation applications. *Data Mining and Knowledge Discovery 5*(1), 115–153.

Schein, A. I., A. Popescul, L. H. Ungar, and D. M. Pennock (2002). Methods and metrics for cold-start recommendations. In *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 253–260. ACM.

Schelp, J. and R. Winter (2002). Enterprise portals und enterprise application integration. In S. Meinhardt and K. Popp (Eds.), *Enterprise-Portale & Enterprise Application Integration*, Volume 225 of *HMD*, pp. 6–20. Heidelberg, Germany: dpunkt-Verl.

Schmitt, S. and R. Bergmann (1999). Applying case-based reasoning technology for product selection and customization in electronic commerce environments. In *12th Bled Electronic Commerce Conference. Bled, Slovenia, June 7-9*.

Shahabi, C., A. M. Zarkesh, J. Adibi, and V. Shah (1997). Knowledge discovery from users web–page navigation. In *Seventh International Workshop on Research Issues in Data Engineering, 1997. Proceedings.*, pp. 20–29.

Shardanand, U. and P. Maes (1995). Social information filtering: Algorithms for automating "word of mouth"'. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'95)*, pp. 210–217.

Sharp, J. (2007). Development, extension, and application: A review of the Technology Acceptance Model. *Information Systems Education Journal 5*(9).

Shilakes, C. C. and J. Tylman (1998). Enterprise information portals, Merrill Lynch In-depth Report.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction* (3 ed.). Addison-Wesley.

Sinha, R. and K. Swearingen (2001). Comparing recommendations made by online systems and friends. In *Joint DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin, Ireland*.

Smith, M. A. (2004). Portals: Toward an application framework for interoperability. *Communications of the ACM 47*, 93–97.

Spiliopoulou, M. and L. C. Faulstich (1999). WUM: A web utilization miner. In *Proceedings of the EDBT Workshop WebDB98*, Valencia, Spain. Springer-Verlag.

Srivastava, J., R. Cooley, M. Deshpande, and P.-N. Tan (2000). Web usage mining: Discovery and applications of usage patterns from Web data. *ACM SIGKDD Explorations Newsletter 1*(2), 12–23.

Srour, L., A. Kayssi, and A. Chehab (2007). Personalized web page ranking using trust and similarity. In *International Conference on Computer Systems and Applications (AICCSA '07)*, pp. 454–457. IEEE Computer Society / ACS.

Stenmark, D. (2004). Failures of reward-driven behaviour in industry: A case of systems, management and creativity. In T. Leino (Ed.), *The European IS profession in the global networking environment*.

Stenmark, D. (2005). Knowledge sharing on a corporate intranet: Effects of reinstating web authoring capabilities. In *Proceedings of ECIS 2005*, Regensburg, Germany, pp. 26–28.

Stenmark, D. and R. Lindgren (2004). Integrating knowledge management systems with everyday work: Design principles leveraging user practice. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004*, Los Alamitos, CA, USA. IEEE Computer Society.

Sullivan, D. (2003). *Proven portals: Best practices for planning, designing, and developing enterprise portals*. Boston, MA, USA: Addison-Wesley.

Summer, J. (2006). *Transforming your intranet: New techniques in strategy, technology and measurement from the world's leading organizations*. London, UK: Melcrum Publishing Limited.

Tatnall, A. (2005). Portals, portals everywhere. In *Web Portals: The New Gateways*, pp. 1–14. Idea Group Publishing.

Taylor, R. S. (1991). Information use environments. In B. Dervin and M. Voigt (Eds.), *Progress in communication sciences*, Volume 10, pp. 217–255. Norwood, NJ, USA: Ablex.

Terveen, L. G., W. Hill, B. Amento, D. McDonald, and J. Creter (1997). PHOAKS: A system for sharing recommendations. *Communications of the ACM 40*(3), 59–62.

Top, D. (2005). The smart business. http://www.intelligententerprise.com/showArticle.jhtml?articleID=163702662.

Towle, B. and C. Quinn (2000). Knowledge based recommender systems using explicit user models. In *Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, AAAI Technical Report WS-00-04*, Menlo Park, CA, USA, pp. 74–77. AAAI Press.

Tran, T. and R. Cohen (2000). Hybrid recommender systems for electronic commerce. In *Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, AAAI Technical Report WS-00-04*, Menlo Park, CA, USA, pp. 78–83. AAAI Press.

van Roy, B. and X. Yan (2009). Manipulation-resistant collaborative filtering systems. In *RecSys '09: Proceedings of the third ACM conference on Recommender systems*, New York, NY, USA, pp. 165–172. ACM.

Venkatesh, V. and F. D. Davis (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science 46*(2), 186–204.

Vo, H. T. K. (2007). *Engineering corporate portals*. Ph. D. thesis, Universität Karlsruhe (TH), Karlsruhe, Germany.

Vo, H. T. K., M. Glaum, and R. Wojciechowski (2008). Streamlining foreign exchange management operations with corporate financial portals. In D. Seese, F. Schlottmann, and C. Weinhardt (Eds.), *Handbook on Information Technology in Finance*, Springer-11643 /Dig. Serial, pp. 123–139. Berlin / Heidelberg, Germany: Springer-Verlag.

Vo, H. T. K., C. Weinhardt, and R. Wojciechowski (2005). Integration of electronic foreign exchange trading and corporate treasury systems with web services. In *Wirtschaftsinformatik*, pp. 471–488.

Wang, J., A. P. de Vries, and M. J. T. Reinders (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 501–508. ACM.

Wasfi, A. M. (1999). Collecting user access patterns for building user profiles and collaborative filtering. In *IUI '99: Proceedings of the 1999 International Conference on Intelligent User Interfaces*, pp. 57–64.

White, C. (2003). Is the portal dead? *Information Management Magazine* (July 2003).

White, M. (2000a). Corporate portals: Realizing their promise, avoiding costly failure. *Business Information Review 17*(4), 177–184.

White, M. (2000b). Enterprise information portals. *The Electronic Library 18*(5), 354–362.

Zanker, M. and S. Gordea (2006). Recommendation–based browsing assistance for corporate knowledge portals. In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, New York, NY, USA, pp. 1116–1117. ACM.

Zanker, M., D. Jannach, S. Gordea, and M. Jessenitschnig (2007). Comparing recommendation strategies in a commercial context. *IEEE Intelligent Systems 22*(3), 69–73.

Ziegler, C.-N. and G. Lausen (2004). Analyzing correlation between trust and user similarity in online communities. In *Proceedings of 2nd International Conference on Trust Management*, pp. 251–265.

Zykov, S. V. (2004). Enterprise portal: From model to implementation. In *Proceedings of the 6th International Workshop on Computer Science and Information Technologies CSIT'2004*, Volume 2, pp. 88–193. Budapest, Hungary.

# List of Abbreviations

| | |
|---|---|
| **ANOVA** | Univariate Analysis of Variance |
| **API** | Application Programming Interface |
| **AR** | Average Rating |
| **AVG** | Average |
| **AWS** | Adapted Weighted Sum |
| **B2B** | Business-to-Business |
| **B2C** | Business-to-Consumer |
| **B2E** | Business-to-Employee |
| **CFC** | Corporate Financial Controlling |
| **CoFiPot** | Corporate Financial Portal |
| **EBIT** | Earnings before Interest and Taxes |
| **FAQ** | Frequently Asked Questions |
| **FA** | Finance and Administration |
| **FX** | Foreign Exchange |
| **IP** | Internet Protocol |
| **IR** | Interest Rate |
| **IUF** | Inverse User Frequency |
| **MA** | Monthly Average Rating Rule |
| **MAE** | Mean Absolute Error |
| **MSE** | Mean Squared Error |
| **RSS** | Really Simple Syndication |
| **ROC** | Receiver Operation Characteristic |
| **SOAP** | Simple Object Access Protocol |
| **StdDev** | Standard Deviation |
| **TAM** | Technology Acceptance Model |
| **TaxIS** | Tax Information System |
| **VS** | Vector similarity |
| **WA** | Weekly Average |
| **WS** | Weighted Sum |
| **XML** | Extensible Markup Language |