To appear in: Journal of Computer Supported Cooperative Work (JCSCW)

Expert Recommender: Designing for a Network Organization

Tim Reichling, Michael Veith, Volker Wulf
Institute for Information Systems and New Media, University of Siegen
{reichling; veith; wulf}@fb5.uni-siegen.de

Abstract. Recent knowledge management initiatives focus on expertise sharing within formal organizational units and informal communities of practice. Expert recommender systems seem to be a promising tool in support of these initiatives. This paper presents experiences in designing an expert recommender system for a knowledge-intensive organization, namely the National Industry Association (NIA). Field study results provide a set of specific design requirements. Based on these requirements, we have designed an expert recommender system which is integrated into the specific software infrastructure of the organizational setting. The organizational setting is, as we will show, specific for historical, political, and economic reasons. These particularities influence the employees' organizational and (inter-)personal needs within this setting. The paper connects empirical findings of a long-term case study with design experiences of an expertise recommender system.

Introduction

Approaches to knowledge management (KM) have attracted both practitioners and scholars in the field of organizations and IT. The basic assumption underlying this trend is that knowledge creation and distribution will become core processes in a complex, fast-changing world. However, the focus of research in knowledge management has shifted. Ackerman et al. [2003] have proposed the concept of "expertise sharing" as a step beyond traditional KM approaches which mainly focus on the externalization of codified knowledge. To label these developmental stages, Huysman and de Wit [2004] recognize two distinct waves of knowledge management. The 'first wave' of knowledge management concentrated on the analysis, modeling, storing, and retrieval of codified knowledge. These initiatives focused either on the knowledge of individuals or on knowledge management from a managerial perspective. IT tools were mainly designed to support acquisition and retrieval of codified knowledge in order to improve knowledge bases, i.e. also the externalization of knowledge itself. By contrast, the ‘second wave’ of knowledge management focuses on the tacit and emergent aspects of knowledge (on those forms which cannot be easily externalized), and on knowledge sharing within communities.

In a socio-technical understanding of KM, the importance of IT tools is contested [e.g. Cohen and Prusak 2001]. The primary function of these tools is not to represent codified knowledge but to support processes of informal interaction within communities (e.g. communication tools, such as email or instant messaging, and tools which help to find partners for communication). Compared to first generation knowledge management tools, little research on expertise sharing and expert recommendation has been conducted so far. Only few case studies distill technical requirements from real world settings [e.g. McDonald 2000] or evaluate the impact of these technologies on organizational performance. Given our understanding of codified knowledge as just one relevant dimension of knowledge, it still needs to be investigated how knowledge sharing as a part of mundane work practices can be supported by appropriately designed technologies.

There are several challenges in investigating the fit of KM technologies designed for such purpose [Ackerman et al. 2003; Huysman and Wulf 2004; Lindgren et al. 2004]. Employees often have different skills, goals, or cultural backgrounds which can lead to failures of IT systems [Pipek and Wulf 2003; Normark and Randall 2005]. Even the successful application of new technologies can have unexpected individual or organizational outcomes that are contrary to the initial goals as Orlikowski [1996] and Pipek and Wulf [1999] describe.

Research in this area needs to explore specific organizational settings, identify requirements, introduce dedicated KM systems, and observe actors appropriating the system. The connection between a specific field of application and its design outcomes is another point of interest. The research area of KM is not yet mature enough to come up with general concepts. Instead KM can be seen as a kind of mosaic that still needs to be further completed before one may hope to be able to identify ‘general patterns’ in applying KM strategies.

In this paper, we present yet another mosaic stone – experiences when designing an expertise recommender system for the specific needs of a major European national industry association (NIA[1]). The design is based on a long-term field study. We outline basic requirements for KM derived from the field study and present a system which deals with the particularities of the identified problems. We conclude by discussing our design experiences.

---

1 “NIA” is actually not the real name but a pseudonym.

Technical Support for Second Generation Knowledge Management

There are several computer applications that have the potential to play a role within initiatives of second generation knowledge management. Most of these applications are designed to overcome spatial or temporal boundaries by making users aware of each other or of artifacts others have created. Among the systems that bridge spatial and temporal boundaries, topic- and member-centered communication spaces are common examples. While member-centered communication spaces, such as the Babble or Loops system presented by Ackerman and Halverson [2004], foster social ties in existing communities, topic-centered communication spaces, such as news groups, allow members (or guests) to exchange ideas and find solutions to given problems[2]. An important motivational factor to participate in topic-centered communication spaces seems to be the gain of personal reputation. Beyond pure communication, applications may foster knowledge exchange by offering virtual spaces which allow creating, developing, and storing of topic-centered materials. These repositories of materials are typically augmented with communication and annotation functionality [cf. Buckingham Shum 1997; Stahl and Herrmann 1999; Pipek and Won 2000]. Editing tools support the development of materials and may have additional functionality to distill content out of communication spaces [Ackerman et al 2003]. The Answer Garden [cf. Ackerman and Malone 1990; Ackerman 1998] is one of the most influential approaches in integrating shared repositories within communication spaces. It was mainly built to encourage learning within organizations. While the general functionality of these systems may be similar, their concrete implementation is specific with regard to the topic under discussion and the application domain [e.g. Chapman 2004]. The systems discussed so far offer places in virtual space where human actors can direct themselves to and exchange knowledge through intended but informal communication.

By contrast, awareness features capture selected activities of individual actors and make them visible to their cooperation partners. Won and Pipek [2003] suggest the application of specific awareness services in order to make knowledge creation perceivable for others within an organization. Furthermore, they propose to collect users' data on these computer supported activities, which are regarded as indications for their personal expertise. After different steps of aggregation, their Expertise Awareness mechanism supports finding human participants who possess a required skill profile which is dynamically updated.

We will now have a closer look at a case study of another class of applications in the field of second generation knowledge management: *expert recommendation systems*. In this class of applications, the system suggests knowledgeable actors. Contrary to first generation KM tools, these applications do not focus on externalizing and representing

2 Member-centered communication spaces may include some topic-centred functionality. However, their design assumes a rather stable group of users basically known to each other.

knowledge but on calculating the participants' level of expertise. These applications therefore require personal data of all users and domain-specific algorithms that can match actors appropriately.

Considering the latest approaches of KM applications, five *sources of personal data* have been applied to generate user profiles:

(1) *Content of textual documents*: Documents that were written, read or reviewed by a user appear to be an indicative source of data for users' expertise. Documents can include work reports, www-pages as well as emails or newsgroups postings. Systems such as *Who Knows* [Streeter and Lochbaum 1988], *Yenta* [Foner 1997], *Lotus Discovery Server* [Pohs et al. 2001], *MII Expert Finder* and *XperNet* [Maybury et al. 2003], *HALe* [McArthur and Bruza 2003], or the *ExpertFinding Framework* [Reichling et al. 2005] automatically extract personal data about human interests or knowledge from text-based documents.

(2) *Yellow pages (YP)/Directories*: Documents created with the specific intent to portray a human actor's interests and expertise can be a highly relevant source. In an E-Learning platform, Becks et al. [2004] ground their matching algorithm partly upon a self-classification of the learners concerning their educational and professional background.

(3) *Inscriptions into software artifacts*: Software – like many other artifacts – can contain inscriptions which allow referring to the expertise of its creator. Concepts of a programming language applied when realizing an application can indicate a programmer's skill level [Vivacque and Lieberman 2000]. Comments within the documentation of the source code can indicate who was responsible for the specific module [McDonald 2000].

(4) *Data on Interaction history*: Structured data about a user's history in interacting with specific applications can be an indicator for his interests or expertise. Vignollet et al. [2005] proposes to make use of similarities in browsing histories for expert recommendation. In a similar way, Becks et al. [2004] presented an approach which is integrated into an E-learning platform. Interacting with certain learning modules out of a well structured hierarchy of learning materials is interpreted as an indication of interest in that domain.

(5) *Social network analysis:* Analyzing the user's social environment is another method of user assessment. The basic idea is that users who have collaborated in the past are likely to successfully collaborate in the future. Referral Web tries to support scientific communities based on an analysis of co-authorship in publications [Kautz et al. 1997a and 1997b]. McArthur and Bruza [2003] as well as Tyler et al. [2003] analyze email traffic to learn about social networks within organizations. The latter approach focuses on discovering Communities of Practice (CoPs) that coexist with the formal organizational structure. McDonald [2003] suggests integrating data of organizational structure and social networks into the algorithms for ranking potential experts.

Interestingly, almost all approaches on expert recommender systems are technology-driven – which means that the authors do not refer to any empirical study that grounds their design to real world practice. The major exception is McDonald's [2000] work. He analyzed in depth the related needs of a smaller sized medical software company. His tool's functionality is therefore, very specific for this software company. However, he also abstracts some of his experiences into a software architecture for expert recommender systems (ER Arch). Groth and Bowers [2001] have challenged some assumptions underlying this architecture by means of a second case study conducted within a Swedish consultancy. Consultants showed patterns of behavior when searching expertise, which were not considered in ER Arch, e.g. choosing accessibility and availability for prioritizing which expert to contact. ER Arch suggests prioritizing experts solely according to data concerning expertise and personal/organizational relations between the information seeker and potential experts.

Due to the lack of empirical grounding, the discourse on expert recommender systems has not yet developed a sufficient understanding on the requirements in real world settings. It is also unclear which type of user profiles and matching algorithms will lead to acceptable matching results. Most of the expert recommenders stand alone in the sense that they collect the required user data from just one application. A more practice-oriented perspective could investigate how to integrate expert recommender systems into a given software infrastructure on specific fields of application.

In the following sections, we will present NIA as a case to explore the potentials of expert recommendation systems. Our research with NIA is still ongoing. We discuss findings from the first two stages of our longitudinal study. Up to this point, we have completed the design of the system based on a comprehensive field study and have implemented a first version of an expertise recommender system. Hence, the software design is based on results of an empirical study which put light on the requirements of different subjects within the underlying heterogeneous organization.

Setting

With almost 3000 member companies from a large variety of different branches, NIA is one of the biggest confederations of industries in Europe. The association is divided horizontally into 37 sections, each dedicated to companies from a certain industry sector[3] (e.g. "agricultural technology", "lifts and escalators" or "pumps and systems" but all related to machine and plant construction) and vertically into general departments (e.g. "business administration", "law" or "taxes"). In addition, NIA consists of several spin offs and other specific organizational units such as forums, projects and regional offices. At the NIA headquarters, about 450 employees work in one of the organizations' sections or departments. Member companies pay for their membership according to their size. These

3 In the following we will use the term "sector" in the sense of branch of industry, whereas "section" will refer to NIA's sectoral departments.

fees are received directly by the corresponding sections. The sections pay a certain percentage of their fees for organizational overhead to finance the vertical departments (among other parts of the organization). The payments of member companies are the main source of income for NIA.

Member companies in turn are welcome to request NIA's services, when needed. NIA defines its core competencies as:

- *Networking* (Introducing member companies to each other for business transactions)
- *Technical* or *professional support*
- *Representation* (lobbying at governmental (or other important) institutions – this kind of service is offered by NIA exclusively[4])

Knowledge management initiatives are regarded as a means to deal with a slight decline in membership registrations. While membership in an industrial association was given to companies in technical sectors during the past decades, companies today are expected to justify their expenses by claiming a "return on investment". This can be difficult to calculate for the membership in an organization 'dealing' with *support*, *network* and *representation*. So one of the projects goals is to better define and present its services to the members and to make NIA and its members 'move closer to each other'. This should be done by improving the mutual awareness of one another; i.e. the awareness of NIA's services on the members' side and the awareness of the members' needs on NIA's side.

An illustrative example should sharpen the project's vision. In short: When developing a new agricultural tractor, one member company got into trouble as they realized (after the design phase was over) that their machine did not conform to certain regulations related to its physical dimensions. This was a very costly error that could have been avoided by turning to NIA where this information was available. In turn, NIA was not aware of the company's intention of developing this kind of agricultural machine and thus was unable to inform the company. KM strategies are now expected to connect both NIA and its members more efficiently to each other in order to avoid situations like the one described. It is especially the communicative level of knowledge sharing between member companies and NIA that needs to be improved or at least supported.

However, there are also serious intra-organizational challenges in communication. As mentioned above, NIA consists of 37 sections which serve 37 sectors of national machine and plant construction industries. Earlier, many of these sections were autonomous associations and in the 1990s these former rivals were merging into a single organization. There is still a considerable level of rivalry remaining since many member companies are hardly associable with just one section of NIA[5]. Still, quite a number of sections consider

---

4 Since about 10 years there are additional organizations which offer networking and knowledge services to – mainly small and medium sized – companies. NIA, however, is the only officially recognized industry representation in these branches. Thus, it can provide exclusive links to political committees and policy-makers.

5 As an example, some agricultural companies need support from the '*agricultural technology'* and from the '*pumps and system'* section since many agricultural machines and tractors are additionally equipped with systems and machines that are developed by member companies of other sections.

members which they had enlisted long before the creation of NIA as their exclusive customers.

The members' obvious advantage of being served by just a single section becomes a problem when deficits in communication within NIA lead to suboptimal services. So, in our field of application it seems that two distinguishable challenges in knowledge management and especially in communication behaviour exist. On the one hand, NIA has considerable problems in providing qualitative and quick help for their member companies. On the other hand, there are also KM problems among NIA sections which seem to be the main reason for the communication problem between the member companies and NIA.

Research Methods

The methodological approach taken in the field follows the theoretical framework of *Integrated Organization and Technology Development* (OTD). Wulf & Rohde [1995] and Rohde [2006] describe OTD as an evolutionary concept which tries to take technological, organizational, and human factors into consideration. As commonly known, the introduction and establishment of novel technology influences not only work processes but also organizational structures and human habits. Keeping that in mind, human and organizational needs have to be taken into account when new software solutions are supposed to support work practices. One particular characteristic of the NIA project is to implement technical opportunities for additional collaborative work processes. NIA existing work processes often seemed to fail when employees looked for expertise beyond their own section, e.g. when answering a request of a member company. With regard to this general problem statement, our empirical study tried to deal with three specific questions. First, what are the relevant work processes within NIA? Second, why do they often seem to fail? And third, which alternatives could be envisioned and which type of technical support would be required? Through our field study we tried to understand the formal organization, the employees' *interpretation* of it, and the informal patterns employees apply in order to deal with these types of problems.

With regard to the empirical case study, we applied the OTD framework by employing the ethnomethodological concept of "Studies of Work" to focus on personal and interpersonal issues [cf. Flick 2002: 39ff; Bergmann 2003: 129ff; Harper et al. 2000]. Beyond workplace observations, investigations into the technical infrastructure, and workshops on specific topics, our research included three cycles of 16 semi-structured interviews conducted with 14 different employees and two managers of NIA. The majority of the employees worked in the agricultural section; the others worked in several vertical units such as the staff-, IT departments or standardization committees. The two managers headed the agricultural section and the IT department, respectively.

In addition to the interviews, we conducted continuous field observations which we documented in written form. Central to our observations were KM related work practices and the use of corresponding software applications. We also looked at the employees' (electronic) documents when conducting interviews and observations. We observed mainly employees who took part in the interviews. Most of the “formal” observations took place in conjunction with the interviews or were carried out after the interviews were conducted. Since a researcher was present in the field at least twice a month for almost three years, observations happened additionally in an informal way. These sources provided further empirical material for analysis.

For each interview cycle, we created a particular interview guideline. This allowed us to modify the guidelines evolutionarily as we assimilated the experiences gained before. The guidelines included issues such as “everyday’s life on the job”, “working history within NIA” (both were main topics in cycle 1 which included six interviews), “communication and cooperation with others”, and “knowledge management and expertise sharing” (treated already in cycle 1 but became the main focus in cycle 2 which included seven additional interviews). In cycle 3, technical and socio-technical issues, such as problem solving and dealing with IT, were brought up in three more interviews.

The interviews lasted between 60 and 120 minutes. We set up the interviews in a relatively open and talkative manner since guided questions were accompanied by narrative elements. By doing so, we had the opportunity to reflect upon the interviews with regard to the organizational and cultural environment. We recognize and situate our findings in an interpersonal, organizational context [cf. Randall and Bentley 1994]. Therefore, we tried to steer the interviews as little as possible, i.e. we followed an open-ended protocol. In order to guarantee solid and valid results, we used a tape recorder to avoid note taking during the interviews which would have influenced the fluency of the conversations negatively.

To facilitate data reduction and to spend time efficiently, we decided to split the analysis phase into five specific steps [cf. Schmidt 2003]:

1. Based on the written materials, we constructed ‘ex-post’ categories for analysis. This categorization was mainly in line with the three aspects of OTD: technology, organization, and human factors, however, the specific subcategories resulted mainly from the interviews.
2. We put together the analytical categories to create a coding guideline which helped us to cluster the data in terms of meaningful units. Each unit focused on a specific problem.
3. We coded the material in order to make critical data anonymous and to generalize the data into meaningful patterns.
4. We constructed nodes of correlating units which provided a quantitative overview of the material. This also gave us a clue which questions and problems might be most prominent and urgent, and had to be reconsidered in later steps of the project.
5. Finally, we thought about possible hypotheses which had been derived from the previous steps of analysis. Based on those, we modified the guideline for the next interview cycle.

We employed the five steps mentioned above for each interview cycle. By doing so, we believe to have satisfied two important requirements: (1) participation of NIA's employees and (2) consideration for processes and practises.
Following the OTD approach, we also looked at the empirical findings from a design perspective. A part of the interviews was dedicated to 'brainstorming and discussing KM solutions'. In the beginning, we had only a very coarse vision of which KM strategy could be applied within NIA. In later interview sessions, we focussed on KM concepts which potentially offered a solution to problems already been identified. As a result of the empirical analysis, we came up with a set of specific design requirements for NIA. To evaluate their validity, we created mock-ups and early prototypes and discussed them with employees of NIA[6].

Empirical Findings

After analyzing the interviews certain issues appeared to be central in terms of KM. These issues can be roughly assigned to the domains of work processes, organizational transparency and knowledge management. Here, we will focus mainly on those issues which have technical implications for the design of the expert recommender system. More detailed insights of the study are provided by Reichling and Veith [2005]. In the following section, we will have a closer look at the results which form a starting point for deriving design requirements.

Working for member companies

The work of NIAs' employees (especially those of the agricultural sector) is dominated by a high and varying workload as targeted events (e.g. internal work groups, standards committee meetings, exhibitions, trade fairs) structure their work activities seasonally. The services directly delivered to member companies such as technical support and representation, strongly depend on the members' requests and can hardly be anticipated in advance.

Handling customers' requests for help is an important task in the agricultural section. Responses in the interviews indicate that handling inquiries can often be reduced to "finding the right expert". In a first instance, employees try to find experts within their section. In particular cases, even contacts to member companies or ministries are utilized to answer certain questions. A few requests demand further inquiries or are delegated to

6 About 20 users participated in evaluating the expertise recommender. We conducted another 21 more interviews and workplace observations with users (some of them twice at different stages) during the evaluation period. These participants overlapped partly with those of the early empirical study.

other work units within the association. A common way to deal with costumer requests – as indicated in the interviews – is to sort them according to their "importance" for NIA.[7]

Especially newcomers find it hard to deal with this dynamic organization of knowledge work. Moreover, new employees are often overwhelmed by NIA's institutional complexity. In this domain experienced employees have an advantage over novices because they are already experienced in dealing with this structural labyrinth. Some of the interviewees claimed that the internal complexity is a result of NIA being created out of the merger of several independent associations.

Particularly in situations of urgent requests from member companies, knowledge about who has expertise in a certain field is crucially important. Participants described their expectations this way:

> "If we had a rough idea of what everyone is doing [within NIA] – which of course is unmanageable for 450 people – then for us this would be a giant step forward."

> "The goal is to create transparency. Responsibilities must be clearly defined and assigned unambiguously."

Due to the complex organizational structure and the missing transparency, it is not surprising that requests from member companies were redirected without being answered appropriately. Situations occur, as interviewees reported, where requests were handed over from one colleague to the next without being finally answered. It was considered to be difficult to identify actors possessing specific competencies, responsibilities or skills within the organization

Building and maintaining social networks is a task of central importance for NIA's employees. It was reported to take years to have established appropriate networks externally within the industrial sector and towards ministries as well as internally within the different sections of NIA. Moreover, once established social networks cannot be simply transferred among NIA's employees.

Tools for Expertise Sharing

For internal networking purposes NIA provides two tools which allow employees to seek for experts and maintain social networks: First, a printed catalogue lists internal experts for specific topics. In response to the question whether this booklet eases the search for experts within NIA[8] appropriately, some interviewees stated that it is often pure chance to find an expert within the booklet. The information was often outdated or imprecise.

Similar problems were identified when discussing the second tool, the central *Address Information Management System* (AIM). The AIM system suffered from being designed in order to meet diverse requirements defined by different organizational units. For instance, as one participant stated, she would like AIM to be capable for seeking persons

---

7 This importance is defined by two major criteria: first, the actual size of a member company, as the membership subscription is based on size, and second, by micro-political considerations regarding the engagement of a member company within NIA's' policies.

8 These experts may work in certain horizontal sections (e.g. the agricultural section) or in vertical sections (e.g. the IT or standardization department).

within the organization carrying certain competencies or responsibilities. However, AIM's functionality did not allow this type of inquiries.

To compensate for the problems of AIM and the paper catalogue, we found an interesting approach. One employee described his way to manage external requests by means of a personal archive of contacts. He had built up this archive continuously over time by adding contacts and skills sets every time a human actor proved to be 'useful'. His archive allows seeking for skills and experts, but it was not commonly shared since the employee considers it to be his private property.

When discussing technical functionalities to overcome the given problems, the most illustrative suggestion was described as "Google for NIA", an extended *yellow pages system* (YP) that supports finding people based on an extended set of personal attributes, such as *activities*, *interests*, *experiences,* or *responsibilities*. One of the participants explained his vision in the following way:

> "It would be heaven, just to enter a keyword and then to get back exactly those ten experts [that I am seeking for]."

Yet another interviewee had a vision:

> "… that it becomes transparent, who's responsible for what, would be surely helpful. And this might be some kind of compliment for certain people; they can see themselves in a leading position within the organization (…). This can even be motivating."

In the latter quotation, another issue is addressed, namely the interviewee gives a subtle reasons for employees to take part in the system: People might feel some kind of 'glory' when being assigned as an expert for a certain topic area. Such a feeling, as some interviewees stated, may improve the employees' motivation to keep their profiles up to date and to share their knowledge with others. As we see in the following subsection the willingness to share expertise must not be assumed to be naturally given at NIA.

A controversial question with regards to YP systems in general is how to create and update the user profiles [cf. Ehrlich 2003; Pipek et al. 2003]. Since most of the interviewees typically worked under time pressure and high work load, it seems to be unlikely that they would update their profiles periodically. On the other hand it became clear from the interviews that a YP system would not be used in case the stored user profiles appeared to be outdated or simply erroneous.

An alternative suggestion was made by one of the interviewees outlining the idea of a 'virtual notice board'. In her assessment NIA was too complex to be effectively covered by a YP system. Her idea was to simply change from a 'pull' to a 'push' concept. By having such a virtual notice board, the system would no longer be required to seek people – but have the users themselves responding to the requests they feel qualified for. Even though the 'virtual notice board' approach has obvious advantages over the Google idea, it is expected to lead to problems when urgent requests were to be handled. As another participant stated, it was impossible for him to wait for "someone who is willing to accept a request".

Obstacles to Expertise Sharing

When discussing KM strategies during the interview sessions, we primarily focussed on the management of human resources rather than of content. The concepts of a YP system ("Google for NIA") require the employees to actively take part in, i.e. share their knowledge/expertise which they must be willing and ready to do. Several interviewees found it perfectly natural to share their expertise with others and expected others to follow suit. Others however, expressed good reasons not to do so. Additionally, some participants stated that the 'potential of knowledge sharing' was highly overestimated.

As one participant assumed, KM was primarily required by the younger employees with limited experiences. These are the ones who could benefit most from 'sharing' expertise. In contrast, some of the more experienced employees seemed to be more short-spoken with regard to expertise sharing. An interviewee stated that expertise sharing would endanger his status by 'making his unique knowledge accessible for others'. So he suggested that it could be a good idea to create incentives for sharing knowledge in the sense of monetary rewards. According to his assessment:

> "[…] everything inside my head is mine. And I must keep it to myself […] to increase or to keep up my market value."

Besides this obvious reason not to share his knowledge, other impeding factors of organizational or cultural nature were brought up by the same participant. As he stated, it was a typical behaviour of some of his colleagues to "strut in borrowed plumes". They would solve certain problems with the help of colleagues within NIA, and thereafter, declare it to be their own work. Therefore, colleagues with lower qualification appeared as 'experts' who would be requested exclusively in future. Such behaviour would reduce the colleagues' willingness to share their knowledge. In his eyes the only way to make colleagues share their knowledge would be to pay them money (see above). Asked whether standardizations or guidelines (mandating colleagues to share expertise as part of their work) were capable of increasing the colleagues' willingness to share knowledge, he spontaneously responded that "guidelines are to be avoided".

Assuming that a YP system such as 'Google for NIA' would be introduced, employees are expected to spend time on helping each other. Taking the workers' heavy workload into account, cooperation among colleagues could be fostered by creating a balance sheet to bill internal services. Some of the interviewees argued that such a balance sheet would make it easier for them to legitimate the amount of time spent on helping or cooperating with colleagues. Otherwise these efforts had to be labelled as "lost time".

Requirements for Technical Support

Based on the findings from the analysis phase, we discuss requirements for a KM tool. We will first summarize the requirements and then link them to appropriate system features. The subsequent section will describe the actual implementation of the expert recommender system for NIA which was designed to meet these requirements.

Basic requirements

The empirical findings can be summarized as follows: NIA is a large organization with a decentralized internal structure consisting of 37 sections and several general departments. The boundaries among the sections are strongly developed. This is an effect of the organization's history since the sections were formerly independent associations. The mode of dividing the member fees within the organization further strengthens the independence of the different sections. Especially the section heads act in a highly self-determined manner.

The decentralized organizational structure fosters the dedication of NIA towards the individual branches. However, the strict borders between the sections lead to losses in potential synergies. Due to 37 sections and several other departments, the formal organizational structure within NIA is very heterogeneous and the relevant social networks – formal as well as informal – are very complex. Several interviewees stated that increasing mutual awareness about skills, expertise, experiences, and knowledge among the employees would be a "giant step forward"[9]. Based on such an awareness, occasional sharing of expertise may happen which many (but not all) of the interviewees assumed to be highly promising. Our study shows that this awareness information is not yet available within the existing organizational and IT infrastructures. It is neither provided by the AIM database nor by the employees' catalogue. Employees requested an efficient and reliable search function (that is what the Google metaphor was used for) for appropriate experts in a given domain. However, we cannot assume that a sufficient amount of employees are willing to continuously or periodically update their profile data which 'traditional' Yellow Pages systems would require. Hence information about the employees' expertise needs to be extracted from well chosen, indicative sources of existing data.

Indicators for Expertise

When considering expert recommender technology, we had to find appropriate indicators for human actors' expertise. With regard to the indicators found in the literature, our field of application offered the following opportunities:

(1) *Content of textual documents*: Textual documents were a major means to report inside the sections and towards member companies. So most of the employees created considerable amounts of textual documents as part of their daily work practice. For instance, project documentation, protocols, correspondence or articles for the organization's website could be widely found. Almost all textual documents were given in PDF or MS Office file formats such as Word, PowerPoint or Excel. The textual documents were typically organized in a directory system hosted on the department's file server. Documents stored on

[9] It becomes obvious from the empirical findings that not only high level expert knowledge is requested. Even lower degrees and different aspects of 'expertise' (namely interests, experiences, activities or abilities) belong to the requested properties (cf. [Hinds and Pfeffer 2003] or [Ackerman, Pipek and Wulf 2003]).

local hard drives, though a rare phenomenon, do still exist.
Email was an important means of communication inside as well as outside NIA. Lotus Notes was commonly used as an email client and time scheduler.

(2) *Yellow pages*: Except for the limited information AIM provided, NIA did not run any personal profile system. Personal profiles were updated, e.g. in case an employee moved to another office. However, employees found it hard to use AIM efficiently for matching people. AIM is – as its name indicates – a data base to administrate address information. It did neither function as a YP-system nor an expertise recommender system.

(3) *Social network analysis*: The organizational structure given by the sections plays a major role for expertise sharing inside NIA. However, the sections seem to be small enough that expert recommendation is not needed inside. Social network analysis might play an important role in sharing knowledge beyond the individual sections. For privacy reasons, we did not access email information (see below). Beyond email, there was not any other comprehensive base for social network analysis available.

Taking the given sources for data on the human actors' expertise into account, we decided to base a first version of the expert recommender system on textual documents. We assumed that a collection of textual documents would be a rather good indication of an employee's interests and expertise. Based on a choice of the employees' text documents a keyword vector was derived using methods of automatic text analysis (see below). We also assumed that the given folder structure would make it easy for the employees to quickly and reliably select relevant and indicative documents from their working environment.

Since NIA did not have its own Yellow Page system, we decided to integrate a YP component into the expert recommender application. Employees could use it to create a personal expertise profile based on their self-assessments. We extracted the personal data as another indicator of an employee's expertise. Moreover, we presented this information when visualizing the list of experts found by the system. Beyond 'ordinary' contact information, the YP component allows users to enter information describing their educational background, professional experiences, interests, and organizational involvement – properties which are relevant for expertise matching, as well.

Hence the users' expertise profile consists of two components: A YP component containing self-assessments that can be directly edited by the users and the keyword vector which is created automatically from a personal selection of the employees' text documents. Both profile components are distinct with respect to their content and the level of accessibility by the (other) users. So, these two components are likely to complement each other. The keyword vector which is supposed to contain several thousands of keywords adds to the self-assessments provided by the YP component. Its creation requires much less effort since the employees just need to specify the folders from which the textual documents should be extracted.

Privacy Issues

Further requirements emerged from the legal situation in the German context. According to German law, the storage and exploitation of personal data which can be used to monitor employees' behaviour and performance is a matter of workers' co-determination. Therefore, we designed a system architecture which allows the users to determine which of their documents or document folders should be taken as an indicator of their expertise (instead of an automatic selection of an employee's entire hard disk or the "My Documents" folder).[10] When documents or folders were selected to be an indicator of expertise, the system's client just extracts the keyword vector but does not upload the whole documents to the server. The system also provides a function to let the user manually delete specific items from the keyword vector in case of a misleading or privacy-threatening extraction. Hence, none of the user's documents is uploaded to the server or could be inspected by other members anyway. As a result, our system – unlike the original Google – does not actually support detecting documents, which makes the usage of the Google metaphor appear somewhat misleading at this point (an example of the systems output is given in the subsequent section). Finally, the keyword profile is uploaded at the explicit request of a user (when s/he has pressed the upload button).

For privacy reasons, we decided not to include email folders or time schedulers as indicators for expertise. While emails and scheduling information may be highly promising indicators for a person's interests and expertise, we assumed that their integration would be perceived as a violation of the actors' privacy. The question of how to preserve privacy without violating legal constraints or employees' attitudes needs to be further investigated within NIA after the systems roll-out.

With regard to the YP component, similar methods of privacy protection are included: For each item that can be entered, its visibility can be set up to be: 'not visible for others', 'internally visible', or 'externally visible'. The last option is relevant in case member companies are connected to the system. This feature is described in more detail in the subsequent section.

Feedback Component

In order to guarantee reliable and trustworthy recommendations, the tool should provide some kind of 'social control' in order to prevent negative or opportunistic behaviour. For instance, users may handle other users' requests for help carelessly or let themselves inappropriately appear as experts in certain domains – as reported by employees during the analysis phase. The latter might be done by entering wrong or misleading personal information to the system or release misleading text documents. In order to achieve some kind of social control, we include a feedback mechanism that allows users to evaluate the support they received from an expert.

---

10 We believed that such a design approach would also improve the quality of the keyword vector since we included only relevant documents for matching.

Different 'methods' of providing feedback can be implemented: Scores (i.e. from 1 to 5), tendencies (i.e. "+", "-" or "neutral") or comments (i.e. free-style text of defined maximum length) are popular elements of user feedback in groupware systems. Another design issue concerning feedback mechanisms is the 'degrees of freedom': Will users be able to provide negative feedback as well as positive? Or will 'negative feedback be indicated by lacking feedback or the lowest number of positive scores? And will other users have access to the feedback details (i.e. the comments)? In case they have: Will they be provided with the overall number of feedback statements, such as the feedback ratio (positive feedback divided by the overall amount of feedback)?

Another problem may come up by different 'feedback cultures'. While some people are used to give honest and straightforward responses, others may use a euphemism to describe poor cooperation. Others in turn may intentionally return undeserved bad feedback in order to keep up their own "market value" compared to others (see section above).

In order to avoid this kind of behavior and eventual resulting conflicts, our feedback system does not permit negative scores. In detail, after a request is answered the requestor is given the opportunity to score by assigning between one to five points (where one means "thanks for trying" and five means "great assistance") to each actor who was involved in answering. Assigning scores, however, is not mandatory. When feedback information is presented, the members' overall number of received feedback statements (and thus, the feedback ratio) is also displayed. We believe that only the combination of both values provides an impression of the actors' actual performance.

Bringing It All Together

Table 1 gives an overview of the central requirements mainly taken from our empirically derived scenario and describes the resulting recommender functionalities. The application can be described as a *text based recommender system for expertise* inside NIA (and its member companies). Each user can provide personal information about himself, including typical yellow page information, such as contact information, a photo, and work-related information. We assume that most of the basic functionalities, i.e. text-based indication of expertise, would be relevant for the second application scenario, as well. However, the privacy-related functionality and the feedback mechanisms will probably need design adaptations with regard to the inter-organizational setting.

The architecture of the system is designed in an extendible manner to allow more sophisticated recommendation strategies which include additional indicators of expertise. It allows integrating modules which connect a variety of different applications to access further types of personal data. For instance, in future system versions email content, interaction histories, internet bookmarks or newsgroup postings may provide additional insights into the employees' expertise or interest. Appropriate modules capable of accessing these data sources can extend the system's ability to create indicative user profiles.

In order to enable interoperability with and integration into NIAs IT infrastructure, we included a *Web Service* interface into the system, allowing clients access to the system via standardized interaction technologies and protocols like HTTP, XML, SOAP and WSDL [Alonso et al. 2003]. This technology allows for embedding expert recommender functionality into other applications such as AIM or NIA's intranet website. When analysing the KM needs within NIA, many interviewees pointed to the fact that they did not want to get yet another additional application. They felt already overwhelmed by the number of applications they were using. Web Service technology thus enables (limited) access to the system without installing the ExpertFinding's client application.[11]

Issues and Requirements	**Technical Functionalities**
Creating organizational transparency and gathering expertise related information	Search for experts based on textual documents and YP data
Identifying competencies, abilities and responsibilities	
Reflecting the potentially rapid developments in expertise, skills, and interests	Semi-automatic generation of user-profiles
Preventing users from time consuming profile updates	
Encouraging users' motivation and reliability	Feedback mechanisms
Handling sensible, expertise-indicating data in a trustworthy manner	Local creation and editing of expertise profiles; upload on users explicit request
Ease of use and learning	Pure expertise recommender engine, no undesired extra functionality
Extendibility in terms of expertise-indicators and matching functionality	Server *and* client rely on an open, modular architecture, allowing for extensions and adjustments
Interoperability with the existing IT infrastructure	*Web Service* interface for connecting other systems and databases

Table 1: Requirements and potential according KM technology

11 The Web Service interface allows only for expertise recommendations. Since local creation of the users' keyword profiles is a central part of our approach and a Web Service would require uploading entire files, the client system still needs to be installed locally in order to generate and eventually edit keyword profiles before uploading them.

Expert Recommender for NIA

We will now have a closer look at the implementation of the expert recommender system for NIA. It is a specific instance of a more widely applicable software framework which we call the *ExpertFinding framework* (EFF) [cf. Becks et al. 2004; Reichling et al. 2005]. We will first outline the system's basic attributes. Afterwards we will describe the most important and distinctive features in more detail. Finally, the application on the client side is presented.

ExpertFinding Framework

EFF is implemented in Java using the Java 1.5 language specification. We rely on a client-server architecture (rather than on peer-to-peer architecture) since user profiles need to be stored on a central server instance in order to guarantee for continuous and quick access to expertise profiles even when clients are offline. Hence, the server-side application manages storage, update and eventual deletion or renaming of user profiles. Furthermore, requests for expertise are handled by the server. In turn the client-side application offers a user interface for editing the personal data, generating a keyword profile from text documents, generating requests for expertise, accordingly displaying the results (which may be seen as the realization of the Google metaphor), and managing emerging requests for help which are facilitated by the system. Both, server and client side applications allow for extensions ('plug-ins') in order to enrich the systems functionalities. This way further data sources serving as indicators for expertise may be included.

The creation of keyword profiles from text documents is done by using statistical methods of keyword analysis [Heyer et al. 2002]. Since the text documents generally are present in some proprietary file format, plain text needs to be extracted from the files in a first step. Namely we are able to extract plain text from MS Word documents, MS PowerPoint presentations, PDF and HTML files. For this purpose we rely on open source projects capable of reading these file formats [Apache POI 2005; PDFBox 2005]. So, we cover the most common text file formats used also as standard formats within NIA.[12]

The steps to extract keywords and create keyword profiles are explained in more detail in [Reichling et al. 2005]. In short: plain text is extracted from documents of diverse file formats. Afterward stop words (terms such as 'the', 'a', 'me' or 'with') are filtered using stop word listings. The remaining terms are regarded as meaningful terms – i.e. keywords. Together with their frequency within all the recognized documents these keywords are stored within a large keyword vector – the 'keyword profile'. It has to be pointed out that

[12] MS Excel files are widely used as well at NIA. However, Excel files are not yet supported since we assumed these files to include numbers rather than meaningful keywords. Our latest investigations show that this is actually not true for NIA since some Excel sheets appeared to include meaningful keywords. Hence, support for Excel files will be included in the next version of the system.

the documents that users release for their keyword profile are not shared with other users. Only the keyword vector is used to match human actors.

The use of personal data is highly restricted by the users' privacy concerns. Careless handling with these data would threaten the users' acceptance of the system. In order to avoid this, none of the users' personal data is released without the users' explicit agreement. By default personal data (after uploading to the server) still remains invisible to other users unless they are explicitly released by the owner of these data. None of the text documents used for automatically generating the user's keyword profile is uploaded to the server or can be inspected by other users. Instead, the keyword profile is generated locally and then sent to the server on the users' demand.

Software Architecture

Figure 1 presents the software architecture of the expert recommender. It shows nine system components, realized as plug-ins. Furthermore interactions among modules are depicted as dotted lines. Such interaction occurs when matching components access corresponding storage components for user data or when the feedback collection component 'feeds' the feedback storage component. As described above, a WebService interface acts as a standardized interface for interaction with the 'world outside'.

The users are represented by three types of profiles. They are stored in different modules (storage modules). We will refer to single components of the user profile as 'profile components, whereas the collection of each user's profile components is referred to as the 'user profile'. The first profile component consists of the yellow page information that users provide about themselves. This component is strongly motivated by our empirical findings (see above). It includes contact information (name, telephone number, email, etc.), information about their role within the organization (position, tasks, job description) and information about skills, experiences and expertises (apprenticeship, study, publications, presentations, etc.). We assume this information to be rather static since it's only infrequently changed.

The second profile component is given by the keywords derived from the users' text documents which are created, manipulated, or read in the context of the actors' actual work. The extraction is carried out semi-automatically as described above. The third profile component is created from feedback events that occur during continued usage of the system (see above). For each user the overall score is stored in the *feedback* component. The average value of scores gained by others is likely to reflect properties like the participants' '*willingness to help*' or their '*engagement in users' requests*'. This component is not a part of the first rolled out version of the implementation of the EFF. It is a plug-in which will be further developed in a second cycle.

In order to match the user profiles against requests and in order to give recommendations, we draw from a set of five matching algorithms ('matching modules'), rather than from just one. We do so in order to deal with possible failures of single matching modules (that may occur as a result of missing, obsolete or insufficient profile data). We further believe that over time certain matching components will prove to be useful while others may emerge as rather useless. So, eventual adjustments of the systems configuration over time (e.g. removal of useless components) are part of our concept.

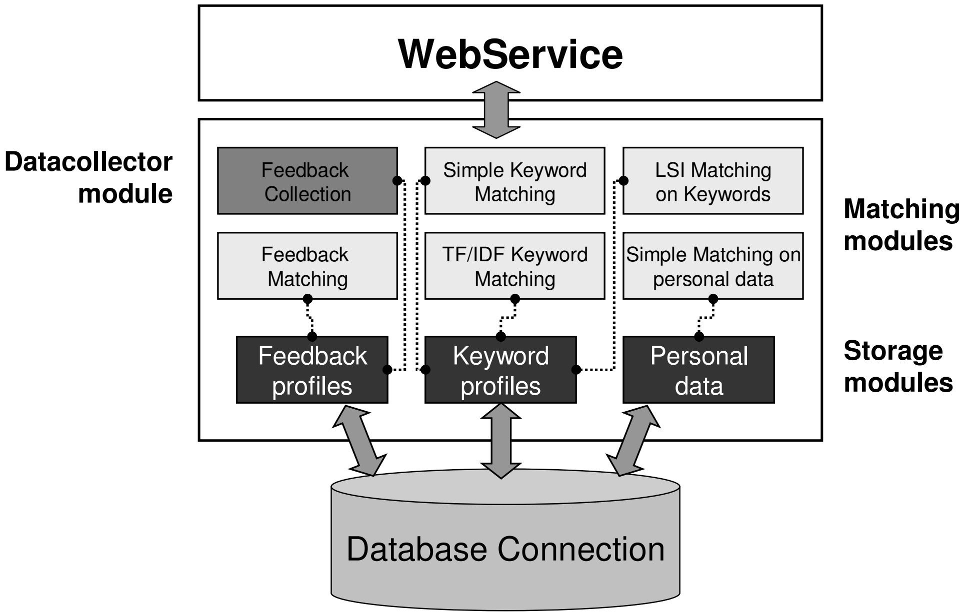

Figure 1: The ExpertFinding framework

The overall matching result is given by the average value of theses five modules (we will later see that users can adjust the weight of each module according to their individual needs). In figure 1 these five different matching modules are depicted as light grey boxes, each implementing one single matching strategy. Five strategies might be a surprising number since only three different types of profile data sets are stored. However, by no means is it given that only one matching strategy can be applied to a specific profile component. In this case, three different matching strategies can apply to the *keyword profile*: *Simple Keyword Matching*, *TF/IDF Keyword Matching* and *LSI Matching on Keywords*. The *Simple Matching on Personal Data* and the *Feedback Matching* apply to the remaining two profile components (*personal data* and *feedback profiles*).

Matching Keyword Profiles

The *Simple Keyword Matching* compares a given request in a fairly simple way: It detects exact or partial matches between the keywords in the request and those in each user profile. For all exact matches, the matching result is increased by the weight of the matching term within the user profile.[13] *TF/IDF Keyword Matching* works in a very

[13] Partial matches (truncations) are considered as matches of 'lower quality'. Accordingly, in cases of partial matches a smaller value is added to the matching result, due to the number of matching letters.

similar way: However, it differs since the matching result is increased by the weight of a keyword divided by the overall number of user profiles containing this keyword. [14]

A more sophisticated method of comparing user profiles is implemented in the *LSI algorithm.* LSI (*Latent Semantic Indexing*) applies to a set of documents that reveal hidden ('latent') similarities between documents.[15]. In the case of a 'proper' set of documents (some of the keywords should occur in more than one document), similarities are discovered even if there are not any exact matches of terms to be detected. Again, in our case it is the user`s profile rather than the documents that is being considered.

Matching Personal Data and User Feedback

Another matching strategy fairly similar to the *Simple Keyword Matching* is the *Simple Matching on personal data*. Again, exact or partial matches between a given request and the user profile are detected. In case of the Matching on Personal Data, the users' personal data are used instead of keyword profiles. The matching result is computed in the same way as with the *Simple Keyword Matching* method.

Finally, *Feedback Matching* is applied to the *Feedback Profiles*. The matching result is computed independently to the actual request since no matching needs to be performed. It computes an average value describing the users' engagement in answering requests (see above) based on all received feedback items. For privacy reasons no detailed information about these data is presented. The feedback is included as part of the result list that is displayed on demand. The results are colour-coded: green indicates a positive feedback whereas white indicates negative or missing feedback.

Since feedback needs to be gathered, a *Feedback Collection* module is part of the software architecture and depicted in the upper left part of figure 1. This component gathers user feedback after a cooperation among users was initiated and communication took place. Collected feedback is stored in a profile (*Feedback Profile*).

User Interface

Now we describe the client-side component ('user client') of the system. The design rationale was to provide a gentle slope of increased complexity to allow users to manage the system and its inherent flexibility [cf. McLean et al. 1991]. On the lowest level of complexity, the client provides 'preconfigured parameters' for non-experienced users to apply the five different matching strategies. However, users can adjust these parameters to fit their individual needs. Some users may want a configuration of the matching modules, which takes the personal profiles to be the most important data to base the matching on. Others might be interested in the users' willingness to help rather than their pure

---

14 TF/IDF stands for *Term Frequency Inversed Document Frequency*. Compared to simple term frequency, this method provides a better understanding of the keyword's importance in the context of the other user profiles (cf. Salton and McGill 1983).

15 Presenting this method in detail would exceed the scope of this paper. Therefore, we refer to Berry et al. [1995] for a detailed description.

qualification. The flexibility in the weighting mechanism is an attempt to deal with partly controversial empirical findings in which participants' motivated technical strategies strongly grounded in their actual work contexts. These contexts again, partly differ tremendously due to NIA's organizational structure.

Searching for Experts

Figure 2 shows the system's search interface. In figure 2a, users can select which type of search is required. The options for search include *keyword-based search, document driven* search, and the *search for similar users*. To set parameters for individual search strategies, a fourth button needs to be pressed.

The keyword-based inquiry can be seen as 'straight forward' search functionality. It approximates most closely to the aspired *Google for NIA* idea, which is motivated by the empirical findings. However, the two other search functions need further explanation: First, the document driven search allows users to select one or more text files from their file system. When starting the search, a vector of weighted keywords is derived from these documents in the same way as it is done when creating keyword profiles. In this case, the inquiry is conducted by matching the resulting keyword vector with the user profiles stored in the system. The inquiry for similar users compares the requestor's own profile with those of other users. The goal of this type of search is to find user profiles that are similar to the requestor's profile. Since the user profiles consist of keywords (in cases of the keyword profile and the personal data), similarities between user profiles are detected in the same way as it is done with keyword-based requests. These search methods are meant to offer additional access to the expertise indicators. While they were not explicitly required by the NIA interviewees, we assumed that specifically employees in member companies would benefit from these additional search options.

In figure 2b a list of experts for an exemplary search inquiry for the terms "market", "agriculture", and "china" is shown. From this list, users can directly create requests to experts or add the contact data to their personal address book. The result list shows the expert's name and a photo as long as the data are released to be displayed. By doing so, we dealt with legal concerns as well as with NIA-specific privacy issues (see above). The overall matching result (defined as the average of the results returned by all the involved matching modules) is indicated by the vertical bar chart (figure 2b). Finally, the right part of each row shows details of the matching results in natural language, explaining *why* the results are as they are and which are the actual matching terms. Since three matching modules (the *Simple Keyword Matching*, the *LSI Matching* and the *Simple Matching on Personal Data*) were involved in answering this search inquiry, three comments are displayed, each describing the result of one matching module (depicted in the box in the lower right of figure 2b, where details are show in increased size and translated to English).[16]

16 The current implementation offers information on the quality of the automatically generated recommendations. However, in case member companies deploy the system to a large extend, we may need to reconsider the display

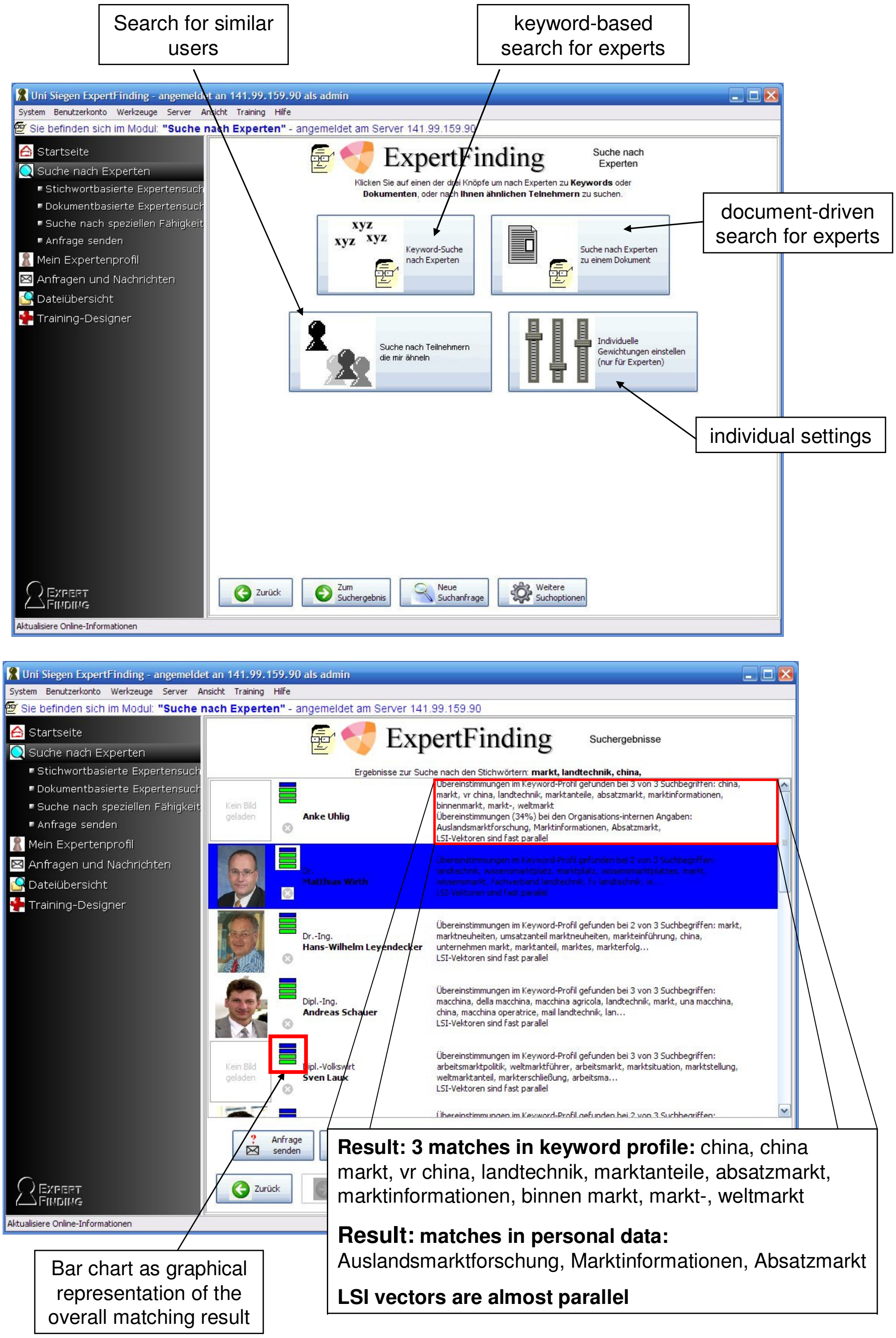

Figure 2: a) Interface to select the preferred type of search b) Interface presenting the list of experts

---

of too many details at this point. For example, NIA may not like to let its member companies know that the best match with regard to a 'hot topic' has a score of only 26%.

When creating a request to an identified expert via the expert recommender a blank email-like window appears. The urgency of the request (maximum number of days to work on it) can be adjusted using a slider element. When moving the slider to the left, the urgency increases which is indicated by the background colour that synchronously turns orange (signalling urgency). The receiver(s) of that request will see an indication of an incoming request. The receivers are provided with the following options: They can accept the request, delegate the request to someone else (contained in the personal address book), decline the request for reasons of lacking time, or decline the request for missing expertise to handle it. In each case, an automatic notification message is sent to the requestor providing the expert's decision. In case of accepting the request, users can specify the expected time it takes them to respond using another slider element.

For two reasons we did not draw on existing email infrastructure for messaging. First, traditional email does not provide the speech-act type of functionality described before [cf. Winograd 1987/88]. Second, though most users are willing to show their name and photo, for privacy reasons users may not wish their email address to become visible to others. This may specifically be the case when the system reaches out into member organizations. However, the next version of the system will include email notification in order to keep the users aware of incoming requests even if the user client is not running.

Another motivation to implement a messaging system internally is the documentation of feedback. It would be much harder to capture feedback from a traditional email system.

Creating the User Profile

We now turn from searching to profiling which is a central element in our approach: In figure 3 and 4 screenshots show the user interfaces for creating, checking, and manipulating keyword profiles (figure 3) and personal data (figure 4). Figure 3a shows a list of folders taken from the file system. These folders are selected by a user to indicate his expertise and interests. Figure 3a also indicates an ongoing 'document scan' which can be observed by the user: The frame in the upper right corner shows the progress on document scanning. The lower right frame shows a list of documents that have not yet been scanned. In Figure 3b a keyword profile is shown after accomplishing the document scan. Experience shows that even a small amount of documents generally results in a large set of keywords. In the left two columns the keywords and corresponding frequencies are displayed. The right column shows the number of documents in which the term has been found. The latter value is provided as additional information for the user. It is not relevant for the matching algorithm and is not transmitted to the server when uploading the keyword profile. Since not all of these terms truly reflect the users' current expertise or interests, they have the opportunity to edit their keyword profiles. Namely they can delete those keywords which could be misleading. The corresponding user interface is depicted in figure 3c. However, editing the keyword vector is meant to be an exceptional intervention.

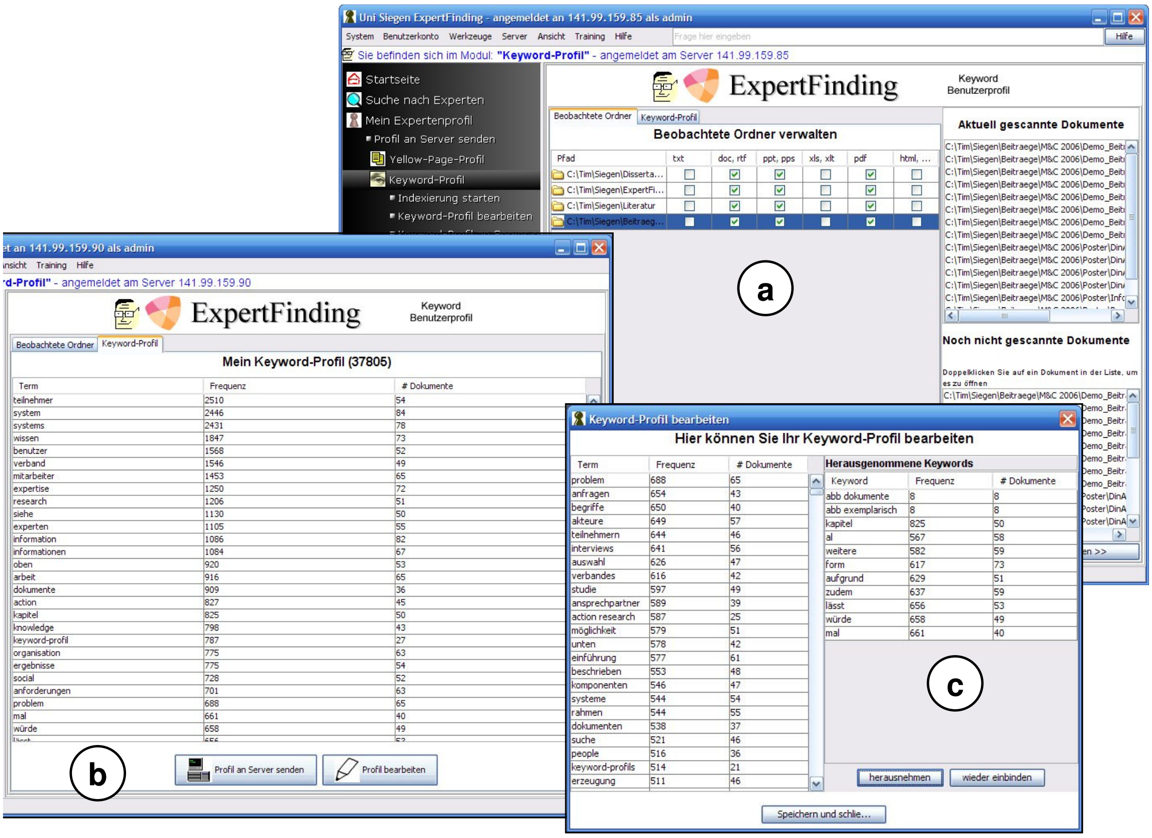

Figure 3: Screenshots of a) the selected folders, b) the generated keyword profile and c) the interface for editing the keyword profile

In figure 4 the user interface for editing the personal data is displayed. Those data are divided into contact information (4a), organizational involvement (4b) and qualification information (4c). The information gathered in 4a is also represented within AIM. This is not the case for the information displayed in 4b and 4c. To protect the users' privacy, the visibility can be configured individually for each item by using the combo box elements next to the entry fields. Visibility can be configured to either 'not visible', 'internally visible' or 'externally visible'. It should be pointed out that the information entered in these fields is used to match expertise, regardless of the item's visibility setting. Visibility restrictions only affect the information that is displayed to other users (for instance in the result lists).

The information about organizational involvement (4b) includes information about the recent position, a description of the actual tasks, and the number of years spent in the organization. The value of these data may vary according to the level of abstraction that is used. More concrete information about the users' actual expertise is gathered in 4c, where typical indicators of expertise are inquired. These include education, apprenticeship, university degrees, language skills, international experience, authored publications, given talks, project experiences, and specialized knowledge. The structure of the Yellow Page form was partly derived from the interviewees and was further motivated by AIM which provides basic information about employees' contact data.[17]

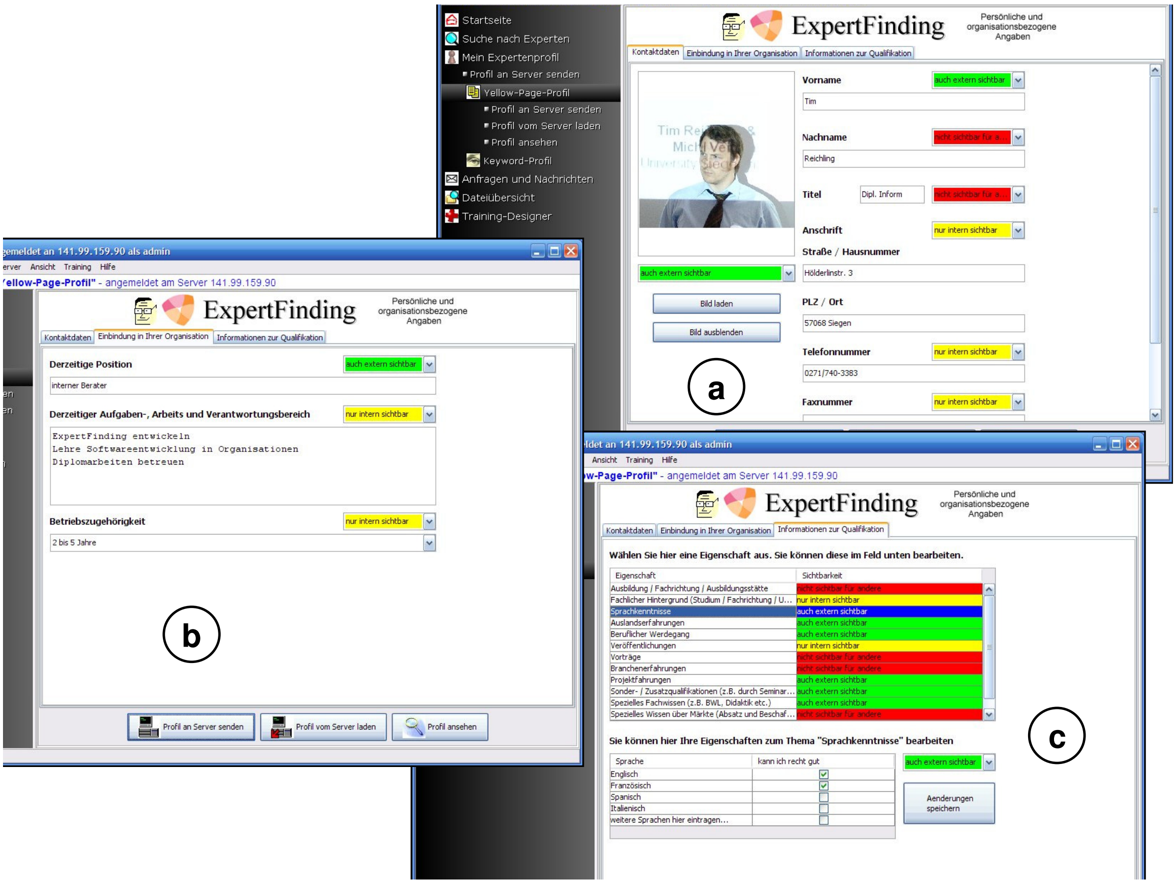

Figure 4: User interface for editing the personal profile: a) contact information, b) organizational involvement and c) qualification information

Conclusion

Second wave KM initiatives require specific technological support. Expert recommender systems are one of the promising technological options offered in that field. Due to the nature of knowledge and its intertwinedness with work practises, research in the field of expertise recommender systems needs empirical case studies to ground and evaluate the design space.

17 NIA's employees are used to work with these sets of data and do not want to maintain additional information.

In this paper, we have explored the design space with regards to the specific needs of NIA – a highly decentralized organization which is very complex and heterogeneous in its structure. We have conducted an investigation into the knowledge-related work practises of NIA. Contrary to earlier field studies [e.g. Pipek and Wulf 2003], NIA had considerable problems in locating appropriate human experts beyond the boundaries of specific sections. Work activities and the related division of labour remained hidden within the individual sections. This pattern seems to result from the organization's history and its internal reward system. These structural patterns seemed to have shaped a culture of competition among the different sections. The size of the organization and its regional distribution only add to the problems.

However, at least on the operative level, the inefficiency of the current mode of expertise sharing is strongly felt. During the empirical case study, it became clear that the services provided to member companies and employees' job satisfaction would benefit from better means of expertise sharing. New employees and actors from member companies can be expected to benefit the most from better organizational transparency within NIA. However, we also experienced considerable reservations from the side of other users. Some of the experienced employees and section heads felt threatened by the potentials of a wider culture of expertise sharing. Such a multi perspective study allows us to better understand potentials and obstacles to the appropriation of expert recommender systems. Moreover, it helps us to conceptualize an appropriate strategy for the roll-out of expert recommender systems.

Based on the results of the empirical study, we designed an expert recommender system. Compared to most other implementations of expert recommendation systems [e.g. Maybury, et al. 2003; Streeter and Lochbaum 1988; Foner 1997], we have implemented an application which is specific for an organizational setting. While we were able to draw on a software framework that provided modules for different matching strategies and their parameterization [cf. Becks et al. 2004; Reichling et al. 2005], many design decisions were grounded in the specific requirements of NIA.

Expert recommenders often rely on structured data such as source code or histories of interaction [Vivacque and Lieberman 2000, McDonald 2000, Becks et al. 2004]. Structured data typically allow extracting and interpreting expertise related information in a more straight forward way than plain text. However, this kind of data was not easily available inside NIA and its member companies or problematic for privacy reasons. Since dealing with textual documents was an important aspect of the work practices, we took these documents as an indicator of expertise and complemented them with Yellow Page information. Our approach is conceptually close to the "Who Knows" application [Streeter and Lochbaum 1988] since it builds on a similar strategy to generate expertise profiles from textual documents. However, our system architecture allows dealing with different sources of expertise indicators in a specific manner. Moreover, our architecture and implementation consider privacy issues explicitly.

The Lotus Discovery Server (LDS) collects a wide variety of textual documents automatically, analysis and clusters them. Human actors are associated with certain document clusters. Search queries display documents as well as related human actors. Users can decide whether to be associated with a certain document cluster [Pohs et al. 2001; Schirmer 2003]. While LDS combines search for documents and experts, our approach focuses solely on finding expertise. Therefore, indicative documents are selected manually by the users and text analysis happens in a decentralized manner. Such an approach is more selective with regard to indicative documents and sensitive towards privacy concerns. Enterprise Content Management systems (ECM) contain typically some sort of expert recommender functionality, as well. However, this functionality is not designed for the particularities of a specific organizational setting and does not offer sufficient flexibility. So, matching results are probably suboptimal and privacy concerns are not dealt with in accordance to the legal and cultural particularities.

Alternatively, one could think of applying (commercial or non-commercial) pull technologies in order to locate and engage experts. For instance, Usenet technology or Virtual Notice Boards meet some of the requirements. Once a participant engages in answering a request, it is a clear indicator of his readiness to share expertise in this domain. However, the empirical pre-study made clear that actors often cannot wait for "someone who is willing to accept a request". We assume that (specifically in non-cooperative environments) a "push" mechanism of directly asking a potential expert would have a positive impact on the way requests are received: Dedicated requests which are exclusively directed towards a potential expert may be – due to their stronger social and communicative character – much harder to decline than a requests launched in a pull-medium.

We are not aware of any existing technology covering the functionality that NIA (and its member companies) require. While our approach is mainly based on textual documents, the quality of recommendation could be improved by taking parts of an employee's communication, such as email messages or phone calls, additionally into account. Further indicators, such as Usenet postings, Internet bookmarks or browser histories, could be considered, as well. For three reasons we decided to base our matching algorithms mainly on textual documents. First, there are still considerable technical and algorithmic problems with regards to voice recognition, which prevent the integration of spoken language. Textual documents are easily accessible and there exists an elaborated set of matching algorithms. Second, privacy issues kept us from utilizing email messages or browser histories. The integration of employees' content of communication or detailed activity protocols seemed to be more delicate than dedicated textual documents. Third, we could not rely on a significant number of users applying other text-based communication channels which could become additional indicators of expertise. For instance, Usenet or similar technologies are not widely used at NIA.

Another important issue when designing expert recommender systems is to find a balance between the control a user needs to have over his profile and the effort which is involved in keeping it updated. We believe that the solution developed for NIA, which draws on textual documents stored in the users' file system, is a promising approach to deal with this design issue. Therefore, changes in a user's document base can be tracked and used to update his profile. This approach can be applied to other document collections which are organized by means of a container structure, as well.

From a privacy perspective, a high level of control on a user's profiles is essential for any expertise recommender system. The German legal setting, with its high demands on privacy, has a great influence on our design considerations. However, the importance of controllability has also been emphasised outside the German setting, as well [e.g. Belotti and Sellen 1993]. Indeed at this point, design requirements for appropriate expert matching and the requirements for the protection of the users' privacy seem to point in the same direction.

A third lesson we learned involves integrating expert recommender systems into the given IT infrastructure. McDonald [2000] and Becks et al. [2004] already pointed out the fact that important aspects of the users' profile data need to be taken out of existing IT systems. Our study confirms this finding. However, in our case the data source was less obvious and the quality of data probably less precise for matching than in the case of McDonald's [2000] work. Our case study shows that "fitting into the given infrastructure" can have a broader meaning. On the one hand, NIA employees requested an integration of the expert finder functionality into the user interface of already given applications, e.g. their intranet portal. Since the use of expert recommender systems will be deeply embedded in a range of different work practises, this fact has to be reflected by the interface design. The software architecture of expert recommenders should allow the integration into a range of different applications with little efforts. Finally, the extent of the functionality of an expert recommender system needs to be designed in relation to the already existing software functionality. In NIA, a Yellow Page system was missing. To gain additional profile data and to offer potentially relevant functionality to the organization, we decided to integrate a Yellow Page component into our application. So our design case indicates that fit with an existing IT infrastructure is a more complex construct than just allowing importing external data.

There are also lessons to be learnt with regard to the design of the software architecture of an expert recommender system. To integrate expert recommender technologies into given IT infrastructures, a flexible software architecture and standardized software interfaces are essential. Flexibility allows to access additional sources of expertise indicators and to extend the functionality, e.g. by adding feedback mechanisms or a Yellow Page component. Standardized software interfaces allow for integration of the application into the user interfaces of another application. While the need for flexibility with regard to expertise indicators has already been discussed in the literature [McDonald 2000; Becks et al. 2004], our study provides a broader set of architectural requirements. Some of them are also the result of privacy concerns which have not yet been widely explored with regard to expertise recommender systems.

Groth and Bowers [2001] challenge expert recommender systems which base their algorithms mainly on indicators of expertise or interest. They found situational factors such as accessibility and availability of human actors more relevant when employees asked for help. To add to this discussion, our empirical study revealed the importance of social and organizational factors when employees select potential experts. These factors include the level of competition across sectors, the level of competition among NIA peers, and the importance of a member company. Groth and Bowers [2001] concluded that the situatedness of human strategies in finding expertise prevented algorithms for expertise location from being effective in case they were solely based on indicators of expertise or interest. Based on our material, one could argue that matching algorithms which do not take social and organizational factors into account would hardly be effective.

However, our interpretation of the empirical findings is rather different. NIA (and the network of its member companies) is an organizational setting in which potential experts are often unknown to the internal and external actors in need of support. So the current version of the recommender system offers information which has not yet been made available to the human actors. The system architecture allows taking further types of data into account when recommendations are computing, e.g. data dealing with situational, social, or organizational factors. However, our design experience indicates that these data may be hardly available and highly sensitive. Data such as the actual workload of an expert is currently not available within NIA's software infrastructure. Other types of data may even be too sensitive that employees would not state them openly or make them available to a computer system. This refers in our case to data such as the level of competition between sections or peers.

However, the lacking availability of these types of data may not be a too serious threat to the usefulness of expert recommender systems. These applications need to be seen as tools supplementing rather than replacing human activities in expert finding [see Pipek and Wulf 2003]. When selecting an actor out of the list of potential experts, the users can 'manually' bring into play all those factors which have not been considered within the algorithms. The results of matching algorithms which take too many different factors into account may be difficult to interpret by users. Therefore, a design approach grounding its recommendations in data sources which are semantically homogeneous, i.e. indicators of expertise and interests, has its merits. Further information helpful to choose among potential experts could be displayed as additional items in the output window which listed the matching results (see figure 2b). Our study indicates that it is still a challenge to find homogeneous data sources and to configurate the matching algorithms appropriately for specific organizational settings.

While the design of the expert recommender is grounded in the empirical case, further studies on the system's appropriation by NIA's employees are needed. NIA's organizational structures and culture does not guarantee the experts' willingness to cooperate or to use the system at all (see section "Obstacles to Expert(ise) Finding"). So the question is whether the recommender is doomed to fail if certain high-level experts will not take part. It is not clear whether those actors will become users of the system as long as they are not intrinsically motivated to contribute to expert sharing in the organizational network.

Obviously, the introduction of the expert recommender system should be best accompanied by measures of organizational and personal development [e.g. Wulf and

Rohde 1995]. For instance, a modification of NIA's internal reward system, which would take efforts spent on expertise sharing more into account, would be very beneficial for the appropriation of the expert recommender system and, most likely, for NIA's overall organizational performance. Such a reward system could be partly based on data derived from the feedback component of the expert recommender system.

However, NIA's decentralized nature seems to make changes in the overall organizational structure more difficult than the introduction of a technical system. Orlikowski (1996) as well as Pipek and Wulf (1999) show empirically that the appropriation of groupware may change organizational practises and processes even beyond the originally intended level. So, it will be an interesting issue to investigate to which extent the introduction of the expert recommender system will have an impact on the pattern of expertise sharing.

While expert recommender technologies have been discussed for quite some time, the focus was so far rather on conceptual and algorithmic issues [Maybury et al. 2003; Streeter and Lochbaum 1988; Foner 1997]. We presented a case where we evaluated its fit into a complex organizational setting. Based on this finding, we designed and implemented a specific functionality and came up with new requirements for the software architecture of expert recommender systems. We believe that further studies of this type are needed to better understand the design space for expert recommender systems and make this promising technology useful in real world settings.

Acknowledgments

The research has been funded by the German Federal Ministry for Economics and Technology. During the time of writing, Volker Wulf was supported by a Fulbright Scholarship. We would like to thank April Bowen, Andrea Bernards, and Aditya Johri for proof reading the manuscript.

References

Ackerman, M.S. (1998): 'Augmenting Organizational Memory: A Field Study of Answer Garden', *ACM Transactions on Information Systems (TOIS)*, vol. 16, no. 3, 1998, pp. 203-224.

Ackerman, M.S. and T.W. Malone (1990) Answer Garden: A tool for growing Organizational Memory Proceedings of the ACM Conference on Office Information Systems, pp. 31-39.

Ackerman, M.S, A. Swenson, S. Cotterill, C. DeMaagd (2003): I-DIAG: From Community Discussion to Knowledge Distillation, in: Huysman, M.; Wenger, E.; Wulf, V. (eds): Communities and Technologies, Kluwer, Dordrecht, pp. 307-326.

Ackerman, M.S. and Ch. Halverson (2004) Sharing Expertise: The Next Step for Knowledge Management, In: Huysman, M.H. and V. Wulf (eds) Social Capital and Information Technology, Cambridge: Cambridge: MIT Press, pp. 273-299.

Ackerman, M.S.; Pipek, V., and Wulf, V. (eds) (2003): Beyond Knowledge Management: sharing expertise. Cambridge MA, MIT Press.

Alonso G, Casati F, Kuno H, Machiraju V. (2003): Web Services - Concepts, Architectures and Applications, Berlin: Springer, 2003

Apache POI (2005): http://jakarta.apache.org/poi/, seen Nov 2005.

Becks, A., Reichling, T., Wulf, V. (2004): 'Expertise Finding: Approaches to Foster Social Capital', in M. Huysman and V. Wulf (eds.): *Social Capital and Information Technology*, MIT Press, Cambridge, 2004, pp. 333-354.

Belotti, V., Sellen, A. (1993): Design for Privacy in Ubiquitous Computing Environments, in: G. de Michelis, C. Simone, K. Schmidt (eds), Proceedings of the Third Conference on Computer-Supported Cooperative Work - ECSCW '93 (Kluwer, Dordrecht, NL), 77-92.

Berry, M.W., Dumais, S. T. and O'Brien, G. W. (1995): *Using linear algebra for intelligent information retrieval*. SIAM Review, vol. 37, no. 4, 1995, 573-595.

Bergmann, J. R. (2003): 'Ethnomethodologie', in U. Flick, E. von Kardorff, and I. Steinke (eds.): *Qualitative Forschung: Ein Handbuch*, Rowohlt Taschenbuch Verlag, Hamburg, pp. 118-135.

Buckingham Shum, S (1997): Negotiating the Construction and Reconstruction of Organisational Memories Journal of Universal Computer Science (Special Issue on IT for Knowledge Management), vol 3/8. pp. 899-928.

Chapman, R. (2004): 'Pearls of Wisdom: Social Capital building in informal leanring environments', in: Huysman, M.H. and V. Wulf (eds) Social Capital and Information Technology, MIT Press, Cambridge MA, pp. 310-333.

Cohen, D. and Prusak, L. (2001): *In Good Company: How Social Capital makes Organizations Work*, Harvard Business School Press, Boston.

Deerwester, S., Dumais, S., Furnas G., Landauer, T. and Harshman, R. (1990): Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science, vol. 41, no. 6, 391-407.

Ehrlich, K. (2003): 'Locating Expertise: Design Issues for an Expertise Locator System', in M. Ackerman, V. Pipek, and V. Wulf (eds.): *Sharing Expertise - Beyond Knowledge Management*, MIT Press, Cambridge, 2003, pp. 137-158.

Flick, U. (2002): *Qualitative Sozialforschung : Eine Einführung*, Rowohlt Taschenbuch Verlag, Hamburg.

Foner, L. N. (1997): 'Yenta: a multi-agent, referral-based matchmaking system', in *Proceedings of the first international conference on Autonomous agents*. USA, ACM Press, New York, pp. 301-307.

Groth, K.; Bowers, J. (2001): *On Finding things out: Situated organizational knowledge in CSCW*, in: Proceedings of the 7th ECSCW, Kluwer, Dordrecht, pp. 279-298.

Harper, R., Randall, D., Rouncefield, M.: (2000): *Organizational change and Retail Finance: An Ethnographic Approach*, Routledge, London.

Heyer, G., Quasthoff, U. and Wolff, Ch. (2002): Möglichkeiten und Verfahren zur automatischen Gewinnung von Fachbegriffen aus Texten. http://wortschatz.informatik.uni-leipzig.de/asv/publikationen/HeyFachbegriffe100902.pdf, 2002, seen: 20.11.2004.

Huysman, M. and de Wit, D. (2004): Practise of Managing Knowledge Sharing: Towads a Second Wave of Knowledge Management, in: Knowledge and Process Management, Vol. 11, No 2, 2004, pp. 81-92.

Huysman, M.; Wulf, V. (2004): Social Capital and Information Technology, MIT-Press, Cambridge MA, 2004.

Kautz, H. A., Selman, B., Shak, M. (1997a) Referral Web: Combining Social Networks and Collaborative filtering, in Communication of the ACM,. 40(3):, 63 – 65.

Kautz, H. A.; Selman, B.; Shak, M. (1997b) The hidden web, AI Magazine, 18(2): 27-36.

Lindgren, R.; Henfriedsson, O.; Schultze, U. (2004): Design Principles for Competence Management Systems: A Synthesis of an Action Research Study. In: MIS Quaterly, Vol. 28, 2004, pp. 435-427

Maybury, M.; D'Amore, R.; House, D.: Automated Discovery and Mapping of Expertise, in: in: Ackerman, M.; Pipek, V.; Wulf, V. (eds): Expertise Sharing: Beyond Knowledge Management, MIT-Press, Cambridge MA 2002, pp. 359–382.

McArthur, R. and Bruza P. (2003): Discovery of implicit and explicit connections between people using email utterance. In: *ECSCW 2003 – Proceedings of the Eighth Conference on Computer Supported Cooperative Work*, Helsinki, Finland, 21-40.

McDonald, D. W. (2000): *Supporting Nuance in Groupware Design: Moving from Naturalistic Expertise Location to Expertise Recommendation*, PhD-thesis, University of California, Irvine.

McDonald, D.W. (2003) Recommending Collaboration with Social Networks: A Comparative Evaluation. In: *Proceedings of the 2003 ACM Conference on Human Factors in Computer Systems*, Ft. Lauderdale, 2003, 593-600.

Normark and Randall (2005): Local Expertise at an Emergency Call Centre in: Proceedings of the Ninth European Conference on Computer Supported Cooperative Work, Springer, Dordrecht, Netherlands, 2005, pp. 347-366.

Orlikowski, W. J. (1992): 'Learning from Notes: Organizational Issues in Groupware Implementation', in Proc. ACM Conf. Computer Supported Cooperative Work (CSCW'92), pp.362-369.

Orlikowski, W. J. (1996): 'Evolving with Notes: Organizational change around groupware technology', in C. Ciborra (ed.): *Groupware & Teamwork*, J. Wiley, Chichester et al., pp. 23-60.

PDFBox (2005): http://www.pdfbox.org/, seen Dec 2005.

Pipek, V. and Wulf, V. (1999): 'A Groupware's Life', in *Proceedings of the Sixth European Conference on Computer Supported Cooperative Work, ECSCW'99*. Netherlands, 1999, Kluwer Academic Publishers, pp. 199-218.

Pipek, V. and Wulf, V. (2003): 'Pruning the Answer Garden: Knowledge Sharing in Maintenance Engineering', in *Proceedings of the Eighth European Conference on Computer Supported Cooperative Work (ECSCW 2003)*, Helsinki, Finland, 14-18 September, Kluwer Academic Publishers, pp. 1-20.

Pipek, V., Hinrichs, J., Wulf, V. (2003): 'Sharing Expertise: Challenges for Technical Support', in M. Ackerman, V. Pipek, and V. Wulf (eds.): *Sharing Expertise - Beyond Knowledge Management*, MIT Press, Cambridge, 2003, pp. 111-136.

Pipek and Won (2000): Wissenslandschaften und Kommunikation, Position Paper for the Workshop on 'Communication and Cooperation in Knowledge Communities' German Conference on Computer-Supported Cooperative Work (DCSCW), München, 2000.

Pipek, V.; Won, M. (2002): Communication-oriented Computer Support for Knowledge Management. Informatik/Informatique - Magazine of the Swiss Informatics Societies 2002(1), pp. 39-43.

Pohs, W.; Pinder, G.; Dougherty, C.; White, M. (2001): The Lotus Knowledge Discovery System: Tools and Experiences, in: IBM Systems Journal, Vol. 40, No. 4, pp. 956 - 966

Randall, D. and Bentley, R. (1994): Tutorial on *'Ethnography and collaborative systems development II: practical application in a commercial context'*, CSCW '94, Chapel Hill, North Carolina, Sept, 1994.

Reichling, T., Schubert, K., Wulf, V. (2005): Matching Human Actors based on their Texts: Design and Evaluation of an Instance of the ExpertFinding Framework, in: Proceedings of GROUP 2005, ACM-Press, New York 2005.

Reichling, T. and Veith, M. (2005): Expertise sharing in a heterogeneous organizational environment, in: Proceedings of the Ninth European Conference on Computer Supported Cooperative Work, Springer, Dordrecht, Netherlands, 2005, pp. 325-345.

Rohde, M. (2006): Integrated Organization and Technology Development (OTD) - A process model for action research, PhD-Thesis, Department of Communication, Journalism and Computer Science, Roskilde University, Roskilde 2006.

Salton, G.; McGill, M. J. (1983): Introduction to Modern Information Retrieval, McGraw-Hill 1983

Schirmer, A. L. (2003): Privacy and Knowledge Management: Challenges in the Design of the Lotus Discovery Server, in: IBM Systems Journal, Vol. 42, No. 3, pp. 519 - 531

Schmidt, C. (2003): ‘Analyse von Leitfadeninterviews’, in U. Flick, E. von Kardorff, and I. Steinke (eds.): *Qualitative Forschung: Ein Handbuch*, 2003, Rowohlt Taschenbuch Verlag, Hamburg, pp. 447-456.

Stahl, G.; Herrmann, T. (1999): *Intertwining Perspectives and Negotiation*; in: Int. Conference on Supporting Group Work (GROUP’99), ACM Press, New York, pp. 316-325.

Streeter, L. A. and Lochbaum, K. E. (1988): ‘Who Knows: A System Based on Automatic Representation of Semantic Structure’, in RIAO '88, Cambridge, MA, pp. 380-388.

Tyler, J. R., Wilkinson, D. M. and Huberman, B. A. (2003): Email as spectroscopy: automated discovery of community structure within organizations, in: Huysman, M.; Wenger, E.; Wulf, V. (eds): *Communities and technologies 2003*, Kluwer, Dordrecht, pp. 81-96.

Vignollet, L,; Plu, M.; Marty, J. C. ; Agosto, L. (2005): Regulation Mechanisms in an Open Social Media Using a Contact Recommender System, in: Proceedings of Communities and Technologies (C&T 2005), Springer, Dordrecht 2005, pp. 419-436.

Vivacque, A. and Lieberman, H. (2000): ‘Agents to assist in finding help’, in *Proceedings in the Conference on Human Computer Interaction (CHI 2000)*, ACM Press, New York, pp. 65-72.

Won, M., Pipek, V (2003) Sharing Knowledge on Knowledge - The eXact Peripheral Expertise Awareness System, Journal of Universal Computer Science, Vol. 9/12, pp. 1388-1397.

Winograd, T.: A Language/Action Perspective on the Design of Cooperative Work, in *Human-Computer Interaction* 3:1 (1987-88), pp. 3-30.

Wulf, V. and Rohde, M. (1995): ‘Towards an integrated organization and technology development’, in *Proceedings of the conference on Designing interactive systems: processes, practices, methods, & techniques*, 1995, ACM Press, New York, pp. 55-64.