Gothenburg Studies in Informatics, Report 18, October 2000, ISSN 1400-741X

Mobile Knowledge

Henrik Fagrell

Department of Informatics,
Göteborg University, Sweden

fagrell@informatics.gu.se, www.informatics.gu.se

Doctoral Dissertation

Abstract

Both the issues of “knowledge management” and “mobility” have received much attention recently. The interest in these issues is often motivated by the fact that work, in many organisations, has become “knowledge intensive” and “mobile.” However, so far these issues have been explored separately. The thesis is a collection of six papers that address these issues from an Informatics perspective. Informatics can be described as a theory and design oriented study of information technology use, an artificial science which focuses on the intertwined complexity of people and information technology as its subject matter. The scope of this thesis is delimited to the design of co-operative technologies. The overall research question is: *how can we provide mobile workers with timely knowledge*? Here, timely knowledge means knowledge that is relevant for the task at hand. The thesis contains empirical studies of mobile work, technologies for knowledge systems, and the design and validation of prototype systems. The studies mainly consist of observations of mobile service electricians and mobile news journalists. They show that current perspectives on knowledge management do not adequately accommodate mobile work. Furthermore, implications for design are derived. Technologies to capture and measure text and hyperlink data are developed and refined to be used in knowledge systems. The design implications and the technologies serve as base for the design of three knowledge systems. One of them has been validated in several workshops and under real working conditions of mobile news journalists. The final contribution is a generalised technological architecture, designed to be easily adapted to several mobile work settings and emerging mobile technologies. The architecture is derived from the summation of previous results and the first practical implementation is in the domain of mobile news journalism.

Keywords

Mobile CSCW, Knowledge management, Organisational memory, Ethnography, Empirically informed systems design, Handheld devices

Language

English

Gothenburg Studies in Informatics,
Report 18, October 2000, ISSN 1400-741X

No. of pages 198

Preface

This thesis is a collection of six papers reporting a research effort that started in the summer of 1996. The theme of the thesis is issues of "knowledge management" and "mobility." The interest in these issues is often motivated by the fact that work in many organisations has become "knowledge intensive" and "mobile." However, these issues have to date been explored separately. Here, these issues are considered together and approached from the Informatics perspective.

The scope of the thesis is delimited to design of co-operative technologies and the overall research question is: *how can we provide mobile workers with timely knowledge*? Here, timely knowledge means knowledge that is relevant for the task at hand.

Sponsor acknowledgements

The research is funded by the Swedish Transport and Communications Research Board (Kommunikationsforskningsberedningen) grants for the Internet Project and the Swedish Research Institute for Information Technology (SITI) grants for the Mobile Informatics program.

Acknowledgements

First, a big thanks to Fredrik Ljungberg, who recruited me in the first place. Fredrik has always been a great support and a friend. Also, thanks to Steinar Kristoffersen who has helped me much the last two years. Working with Steinar and Fredrik has been a great experience and I will always look up to them as most admirable professionals.

Of course, I owe much to my supervisor Bo Dahlbom. It has been a pleasure to partake in the establishment of the Viktoria Institute and the Mobile Informatics group. Bo has been a great support during my entire Ph.D. studies.

A special thanks to my friend and colleague Carsten Sørensen, who helped me much in the beginning of my Ph.D. studies. Also, thanks to Maxine Robertson for help with extensive proof reading.

Thanks to the colleagues that I have been working closely with the past years: Kerstin Forsberg, Magnus Bergquist, Peter Ljungstrand, Johan Sanneblad, Per Dahlberg, Dick Stenmark, Erik Johanneson and Jens Bergqvist.

Thanks to Dave Randall – a great fellow who took his time to inspire me to conduct fieldwork.

During the spring of 1999 I visited Mark Ackerman's group at the University of California, Irvine. I like to give a big thanks to Mark and his group including my favourite Californian: Yuzo J. Kanomata.

Thanks to my colleagues and friends: Antonio Cordella, Staffan Björk, Senja Edvardsson, Mattias Esbjörnsson, Jennica Falk, Henrik Gater, Leif Grönqvist, Christian Hardless, Lars Erik Holmquist, Patricija Jaksetic, Oskar Juhlin, Mattias Klang, Jonas Larsson, Rikard Lindgren, Nina Lundberg, Jan Ljungberg, Andreas Nilsson, Daniel Normark, Urban Nuldén, Daniel Olsson, Stefan Olsson, Kalevi Pessi, Agneta Ranerup, Johan Redström, Kai Simon, Alexandra Weilenmann and Karl-Petter Åkesson. Also, all the people in the Internet Project - thanks to you all.

Thanks to Premilla Bäck for the cover design (the Ghost beat), Paula Rosell, Karin Svensson and Ann Andreasson for administrative help and Niklas Paulsson, Lennart Pettersson and Christer Falk for technical support.

I would also like to thank the empirical partners, especially Radio Sweden Göteborg, Volvo Group, Göteborg Energi. Thanks to the industrial partner Adera (www.aderagroup.com). Also, thanks to everyone working at Newmad Technologies (www.newmad.com).

I wish all the best to ICTech (www.instantcontext.com) - a start-up company with the objective to commercialise ideas reported in this thesis.

Finally, I wish to acknowledge with love and gratitude my girl friend Sofia.

Henrik Fagrell, Göteborg, August 2000.

Contents

Mobile Knowledge

1. Introduction

Knowledge as a major means for securing social and economic results has become the mainstay of the "new economy." This trend can be found both in theories and in practice. Researchers from a wide range of disciplines identify knowledge as an increasingly critical issue for organisations (Scarbrough 1998) on both macro (Castells 1996) and micro (Nardi et al 2000) levels.

Knowledge about customers is very important for all organisations. Organisations that can respond to customers' needs effectively will secure competitive advantage. Accordingly, the ability for organisational members to share knowledge about customers is of strategic importance.

Closely related to this discourse is the emerging theme of "knowledge management." The concept of knowledge management is increasingly used to cover many phenomena. For example, Hedlund (1994) claims that knowledge management addresses the generation, representation, storage, transfer, transformation, application, embedding, and protection of organisational knowledge. Needless to say, there is no general agreement on a definition of this contemporary concept and most attempts are very broad.

In 1996 when I started the research reported in the thesis the Web had just become accepted as a tool for sharing knowledge both within and across organisations. A main problem that soon emerged however was how to find and present relevant knowledge for different users. As a consequence many researchers became concerned with techniques for information retrieval, filtering and visualisation. The

first two papers of the thesis add to our understanding of these issues. They form a technical platform for further development which is deployed in the succeeding papers.

During the latter part of the 90s, mobile IT became increasingly important. In particular, the diffusion of mobile phones globally has been extensive. The current number of cellular telephone subscribers (GSM) in Europe for example, is approximately 200 million. The forecast for the year 2003 is 400 million (EMC 2000)! In addition to mobile phones, the diffusion of Personal Digital Assistants (PDAs) and other mobile computers is also predicted to be very high the next couple of years. For example, the International Data Corporation (IDC) predicts that nearly 62 million people will use Personal Digital Assistants (PDA) to access the Internet by 2003 (Bruno 2000, p. 280).

The trend of mobile IT is part of an overall trend in society towards mobility. People are more mobile today than they were ten, twenty and thirty years ago. At the same time, the investments in technology to *reduce* travelling are higher than ever. Paradoxically, these investments seem to correlate with increased mobility, not vice versa. According to predictions, mobility will increase even more in the future. Against this background, the focus of my research started to drift towards the use of knowledge in *mobile* situations.

Considering the IT oriented research on knowledge management, the focus has almost exclusively been on the PC and the Web. Even though I do acknowledge the importance of this research, I also find it interesting to notice that the issue of mobility has largely been overlooked so far. In the context of my research, this observation opened up new possibilities and challenges: How to provide *mobile* workers with knowledge?

Mobile phones and PDAs support mobile work, but PDAs have mainly been used as personal electronic calendars and contact directories. A simple solution would be to give mobile access to existing knowledge systems. Unfortunately, it is not that easy. Recent studies in the area of Computer Supported Co-operative Work (CSCW) and Human Computer Interaction (HCI) have shown that mobile environments pose different problems and possibilities than the stationary setting (Bellotti and Bly 1996; Kristoffersen and Ljungberg 1999, 2000; Bellotti and Rogers 1997). The current tendency to localise activities to an individual and stationary user has ment that flexible

and unexpected interaction that often emerges out of a mobile work situation has remained generally unsupported (cf., Luff and Heath 1998). The assumptions upon which stationary systems are based are not fully aligned with how work is carried out in mobile settings. The need to understand mobile work and to design knowledge systems that really accommodate mobile settings is thus growing.

In my research on mobile knowledge, I have chosen to concentrate on the issue of "timely knowledge," i.e., knowledge that is relevant, and pertinent for the task at hand. Previous research on mobility has shown the situated and changing nature of many mobile work situations (e.g., Luff and Heath 1998). Even though knowledge support, of course, always should be "timely," this seems even more relevant in a mobile work setting. For example, when meeting a client in the field it is good to be reminded of the latest interaction and the current agreements with the client, and thereby being able to offer good service.

Against this background, I formulated the following research question for the remaining four papers (contribution 3 – 6): *How can we provide mobile workers with timely knowledge*? In these mobile contributions of the thesis I have deployed and refined techniques developed in paper one and two.

The aim of this introduction of the thesis is to link together the six individual papers, found in chapter two to seven. In addition, the introduction offers a technical literature review, a discussion of the research methodology and theory and a summary of the contributions.

The six individual papers of the thesis are listed below:

1. Fagrell, H. and C. Sørensen (2000) "Surveying the World Wide Web," *Scandinavian Journal of Information Systems*, 11 (1), pp. 25-50.

2. Fagrell, H. (1999) "IntraNews: A News Recommending Service for Corporate Intranets," In *Proceedings of Computer Supported Cooperative Work in Design*, pp. 323-328, Compiègne, France: Université de Technologie de Compiègne.

3. Fagrell, H., S. Kristoffersen and F. Ljungberg (1999) "Exploring Support for Knowledge Management in Mobile Work," In *Proceedings of the Sixth European Conference on Computer-*

Supported Cooperative Work, pp. 259-275, Copenhagen, Denmark: Kluwer Academic Publishers.

4. Fagrell, H. and F. Ljungberg (2000) "A Field Study of News Journalism: Implications for Knowledge Management Systems," Accepted for publication in *Proceedings of the Sixth Biennial Participatory Design Conference*, New York, NY: CPSR and ACM Press.

5. Fagrell, H. (2000) "NewsMate: Providing Timely Knowledge to Mobile and Distributed News Journalists," Revised version accepted for publication in *Beyond Knowledge Management: Managing Expertise*, M. Ackerman et al., forthcoming. The paper is a further development of Fagrell, H. and E. Johanneson (2000) "NewsMate: Knowledge Management for Mobile Journalists," In *Proceedings of IRIS 23*, Uddevalla, Sweden.

6. Fagrell, H., K. Forsberg and J. Sanneblad (2000) "Mobile Knowledge Management: Model, Architecture and Applications," Accepted for publication in *Proceedings of ACM 2000 Conference on Computer Supported Cooperative Work*, Philadelphia, PA: ACM Press.

Other publications of mine that are related to the thesis include, two book chapters, seven peer-reviewed conference publications and a peer-reviewed workshop paper:

1. Bergquist, M., H. Fagrell, P. Ljungstrand and M. Storsten (1998) "Designing for Informal Mobile Cooperation," At the Workshop *Handheld CSCW* at the ACM 1998 Conference on Computer Supported Cooperative Work, Seattle, MA, USA.

2. Bergqvist, J., P. Dahlberg, H. Fagrell and J. Redström (1999) "Location Awarness and Local Mobility: Exploring Proximity Awarness," In *Proceedings of IRIS 22*, Vol. I, pp. 103-110, Keuruu, Finland.

3. Fagrell, H. (1999) "An Alternative Perspective on Knowledge Management," In *Informatics in the next millennium*, F. Ljungberg (ed.), pp. 213-224, Studentlitteratur, Lund, Sweden.

4. Fagrell, H. and C. Sørensen (1997) "It's life Jim, but not as we know it," In *Proceedings of WebNet 97*, Toronto, Canada.

5. Fagrell, H. and E. Johanneson (2000) “NewsMate: Knowledge Management for Mobile Journalists,” In *Proceedings of IRIS 23*, Uddevalla, Sweden.

6. Fagrell, H. and P. Ljungstrand (1998) “Make an Agent and you shall find,” In *Proceedings of IRIS 21*, pp. 197-206, Sæby, Denmark.

7. Forsberg, K. and H. Fagrell (1999) “Let’s talk about News,” In *Proceedings of IRIS 22*, Vol. I, pp. 341-350, Keuruu, Finland.

8. Holmquist, L.E., H. Fagrell and R. Busso (1998) “Navigating Cyberspace with CyberGeo Maps,” In *Proceedings of IRIS 21*, pp. 391-399, Sæby, Denmark.

9. Ljungberg, F., B. Dahlbom, H. Fagrell, M. Bergquist and P. Ljungstrand (1998) “Innovation of new IT use: Combining approaches and perspectives in R&D projects,” In *Proceedings of the Fifth Biennial Participatory Design Conference*, pp. 203-209, Seattle, WA: CPSR and ACM Press.

10. Sørensen, C., H. Fagrell and P. Ljungstrand (2000) “Traces: From order to chaos,” In *Planet Internet*, K. Braa et al (eds.), pp. 157-183, Studentlitteratur, Lund, Sweden.

2. Mobile Work and Knowledge

Current IT services for mobile users are very much based on experiences and ideas from office computing. There is therefore a "stationary bias" in the use of mobile computers (Kristoffersen and Ljungberg 2000). But, what is mobile work anyway? And, are there any specific characteristics of the management of knowledge in mobile work?

It is virtually impossible to define mobile work in a meaningful way. Since mobile work is a human activity and not machinery it is difficult to make general statements that are far-reaching, precise and complete. On one hand you can say that all work is mobile. People move within the workplace, they visit customers and partners, and they travel between workplaces. At the same time, all work has a "stationary" component. Even the construction worker and the nurse are still at some point, in a working day. However, their jobs differ in important ways from for example, a secretary or a programmer when it comes to *degree* of mobility. Although it is hard to explain in general terms we all know that the postman and the travelling salesman are mobile, the secretary and the chef, in relative terms, are not. Thus, we are able to conceive of typical situations in which people are mobile and when they are not.

We can distinguish between situations that are typical instances of a type of mobility. Kristoffersen and Ljungberg's (2000) way of conceiving mobility is to distinguish between travelling, visiting and wandering modalities (figure 1).

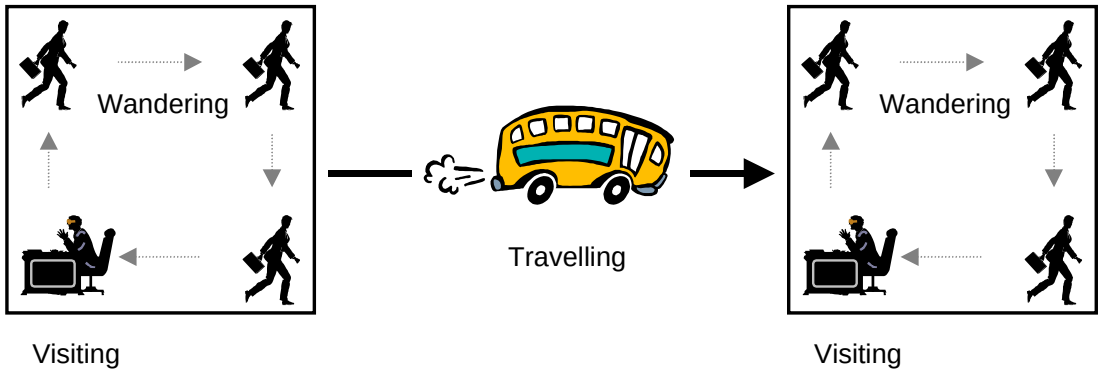

Figure 1: Types of mobile modalities.

The three modalities are conceptualised from a perspective of design of IT use. Let us consider them in more detail.

Travelling is the process of going from one place to another in a vehicle. For example, commuters are travelling when they go by train from their homes to the work place; the travelling representative is travelling when she or he goes by car from one client organisation to another. The travelling modality seeks to capture the mobility of people who go in vehicles. For example, the train commuter can use a laptop while travelling whereas a commuter who drives a car can only use a mobile phone. Given the vast amounts of time spent on, e.g., car commuting, there is little research on IT support for travelling.

Visiting is spending time in one place for a prolonged period of time before moving on to another place. For example, a consultant is visiting when spending time in a client organisation. Visiting seeks to capture the mobility involved when people spend time in a place on a temporary basis before moving on to another place. The visitor can either bring IT to the place they visit, e.g., a laptop, or they can use PCs that are already there. In addition, the Web has offered new possibilities for supporting visiting, since it now is widely used and solves many interoperability problems (cf., Brown 1999). A concept, which is related to visiting is *remote mobility*, i.e., when remote users interact with each other using technology (Luff and Heath 1998). The example they give is construction foremen visiting teams of construction workers.

Wandering is extensive local mobility in a building or local bounded area. A wandering person spends considerable time walking around. For example, IT support personnel in some organisations spend time wandering around helping people across functions and

buildings. Due to the high degree of personal mobility when wandering, the IT people use is necessarily, typically easy to carry. Wandering or local mobility has been described in several field studies of, e.g., product designers at a consulting firm (Bellotti and Bly 1996), personnel at London underground (Luff and Heath 1998), bank officers at a customer service centre (Kristoffersen and Rodden 1996), and software developers at an IT company (Bergqvist et al 1999).

Luff and Heath (1998) propose yet another category of mobility, called micro-mobility. They describe micro-mobility as the way in which the artefact may by mobilised and manipulated for various purposes around a relatively circumscribed, or "at hand" domain. An example is the medical record that is used for both synchronous and asynchronous collaboration between both doctors and other professionals, and between the patient and the doctor.

Not only does the modalities need to be considered but also the issue of local knowledge. Local knowledge is conceived as specific to place and procedures. Importantly, the ways in which it appears in mobile work settings has not yet been investigated. Clifford Geertz (1983, p. 167), who coined the term, views local knowledge from the perspective of "to-know-a-city-is-to-know-its-streets." Here, local knowledge could be similar in many different geographically dispersed workplaces, but it is created through local experiences and used locally by a limited group of people that are working more or less together. In CSCW, local knowledge has been described as: knowledge of the particularities (Rouncefield et al 1994), what makes work run smoothly (Randall et al 1995), to know who knows what, who is busy, who is worth asking about "x" (Bowers et al 1996).

Orr (1996) describes a classical example of how mobile photocopier repairmen share knowledge. Orr's suggests that the staff use "war stories" to assist their community's learning. This is related to the notion of "community of practice," which is a unified view that emphasises the collective, group-oriented nature of workplace activity (see Brown and Duguid 1991; Wenger 1998). The photocopier repairmen in Orr's study conducted the repairs at the customers' sites and they typically met during lunch breaks to socialise and tell stories of how to solve and diagnose problems with the photocopiers. Hence, the community's knowledge was codified and shared in the stories. Story telling is one way of providing mobile workers with knowledge.

However, when a worker is confronting a new situation in the field, a computerised support connected to the task at hand may be more useful.

In many field studies researchers have documented the importance of peripheral awareness as means for timely knowledge sharing. Peripheral awareness is constituted by actions like “talking out loud” and “overhearing,” i.e., inviting colleagues to give feedback (e.g., Heath and Luff 1991). This approach is not practicably appropriate when people are individually mobile, because the local and unique circumstances that each actor confronts are difficult to mediate remotely. However, there are studies of knowledge in work on micro-level, but they do not consider mobility. Instead, the empirical results have been used to criticise theoretical frameworks or to describe how knowledge is shared in a local setting (e.g., Randall et al 1996).

There is also an IT oriented strand of research with the main focus of developing knowledge management systems. This research has rarely been based on empirical investigations of how work really takes place in practice (see, e.g., Ackerman 1994; Conklin and Begeman 1988). Accordingly, the assumptions on which the systems are based are not informed by empirical work. IT oriented research does not usually express a perspective on organisational knowledge and what aspect of it, it seeks to support. The systems are assumed to speak for themselves so to say.

So far, there has been little research on the specific characteristics of the management of knowledge in mobile work. Most research on mobile work can be characterised as empirical studies (cf., Bellotti and Bly 1996; Kristoffersen and Ljungberg 1999, 2000; Bellotti and Rogers 1997). None of them have focused on the roles of knowledge in organisations.

The view of organisational knowledge that will be emphasised in this thesis is aligned with the view of Anderson and Sharrock (1993). They argue:

> “For social science knowledge is not to be considered apart from the situations and courses of action within which it is deployed. Knowledge and action are conjoined. […] The connection between knowledge and action is defined in constitutive terms. ” (ibid., p. 149)

Against this background, the design of mobile knowledge systems can be informed by observations of organisational knowledge in action in mobile settings. So, let us now look into how well existing knowledge systems provide mobile workers with *timely* knowledge.

3. Knowledge systems

The review of knowledge systems in this section is based on the systems described in the scientific literature of the 1990s. The review excludes expert systems and systems designed to simulate the problem-solving behaviour of a human expert in a narrow domain or discipline. Furthermore, the review mainly covers the CSCW and HCI literature. Platforms for knowledge systems development, e.g., Microsoft Exchange and Lotus Notes, are not considered.

In the literature survey I have found three categories of systems that relate to knowledge management. Some of these systems are typical knowledge management systems focusing on building organisational memories. Other systems, however, rather demonstrate the typical engineering community interest in its own professional work practice. I call them design rational systems. Yet a third category of systems express the interest of this community in the latest technology, i.e., the Internet and the World Wide Web.

- Organisational memory systems. The rationale of organisational memory systems is to help people to find solutions to problems. It is assumed that the user has a problem, which is not entirely unique. The system is then used to identify an already solved, similar problem. The solution for the old problem is then applied to the new problem.

- Design rationale systems. The objective of design rationale systems is to give an overview of complex design problems. The idea is to make the rationale for a design decision explicit by capturing the argumentation in a hypermedia system. Although the objectives of design rationale systems and organisational memory systems are related they are typically considered to be distinguished from each other.

- Collaborative filtering systems. Collaborative filtering systems are based on the assumption that people who are looking for information should be able to make use of what others have already found and evaluated. These systems aggregate ratings of people and

use them for filtering. The ratings can be explicit, i.e., the users grade information, or implicit, e.g., the system mines some existing interaction log.

Fifteen systems have been identified that can be categorised in this way - apart from the three systems that are reported in this thesis. A short description of each system follows.

3.1 Organisational memory systems

In 1990 Mark Ackerman did some seminal work on organisational memory with the *Answer Garden* system. The aim was to help organisations capture and retrieve the experiences of their employees (Ackerman and Malone 1990; Ackerman 1994). The users could browse a network of diagnostic questions to find the answers they required. If the answer could not be found the question was routed to an expert that inserted the answer (along with the question) into the network. Experts could also modify the network, and thus organisational knowledge developed. This system has mainly been tested on software developers where the mobility of each user was limited.

The second version of Answer Garden introduced features that routed the user directly to the expert if the solution was not found in the network (Ackerman and McDonald 1996). A collaborative help feature was added to make the interpretations of de-contextulized information easier. Both versions of Answer Garden actively require a substantial amount of authoring to have utility for users.

Figure 2 illustrates the Web-based interface of Answer Garden 2. The Web version may be useful for troubleshooting in a mobile setting.

Figure 2: A screen dump of Answer Garden 2 (Ackerman and McDonald 1996, p. 99).

Karduck (1994) described a system called *TeamBuilder*. The objective of TeamBuilder was to support the identification of expertise when forming teams. TeamBuilder was a Dexter hypertext system and the model of how teams were formed is by linking information units of people and documents. The success of the system seemed to rely on each user explicitly storing what they know in the system. This can be difficult to achieve because it requires considerable work and making knowledge of this kind explicit is very difficult. It is also hard to keep the knowledge up to date manually.

If the system could be made to work appropriately given the considerations highlighted above mobile access would be useful for mobile personnel, e.g., project leaders and sales people who need to have an overview of the available resources.

Anther related system is *Bubble-up* that has been used at a customer support department of a company selling high-capacity backup products (El Sawy and Bowler 1997). The support personnel used the system as an integrated part of the support work as old problems and solutions were accessed. The system had a Web interface where customers could log on and suggest updates and report bugs. The system was proven useful on two levels, i.e., the customer support department could give better support, but also the interaction with the

system was seen as very valuable feedback for the developers of new product versions. In this way Bubble-up facilitated considerable indirect collaboration and knowledge sharing. Since Bubble-up is tailored for product support it is not certain that access to the system is beneficial for a mobile worker. However, mobile access is probably useful when the personnel are locally mobile, i.e., *wandering*.

A system that does provide mobile access is *Darwin*. The objective of Darwin is to assist task distribution and exchange of lessons learned within a dispersed IT-support group (Kristoffersen and Ljungberg 1998). The design was informed by an empirical study and the objective of the system was to help the group offer high quality IT-support at the users' desks (as opposed to via the telephone). Darwin supports the co-ordination of work and the mobility of the group as the system runs on the mobile device Apple Newton. The users are however required to enter their work tasks for the system to be useful.

Darwin therefore does consider mobility and the exclusive focus is on task distribution in a tightly coupled dispersed work group. Figure 3 shows two screen shuts of Darwin.

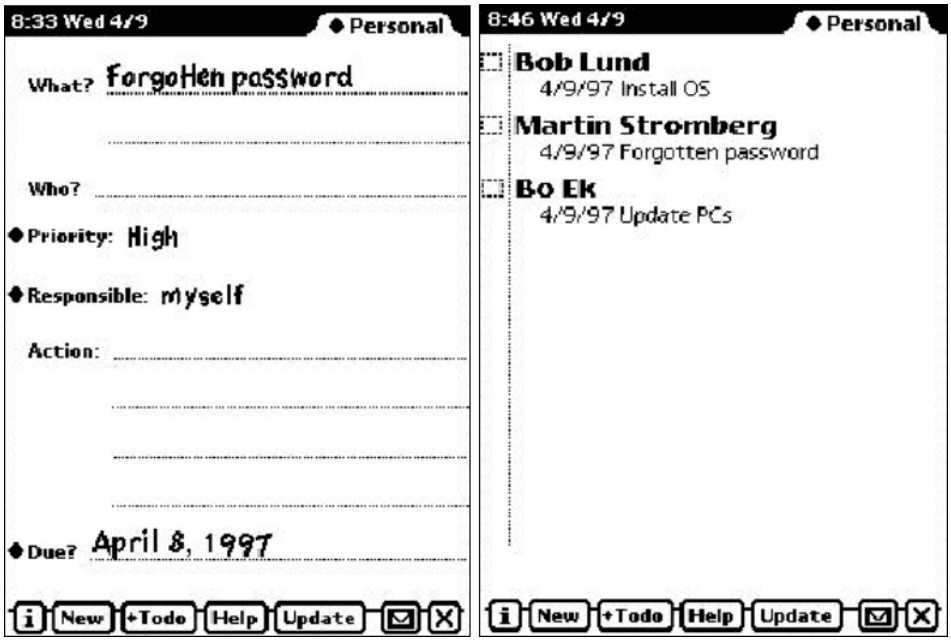

Figure 3: Screen dumps of the Darwin system. Add new task (left) and view public tasks (right).

3.2 Design Rationale systems

One of the first design rationale systems was *gIBIS*. The design of the system was based on an approach introduced by Horst Rittel, called Issue Based Information System (Conklin and Begeman 1988; Yakemovic and Conklin 1990). The approach encourages team members to actively discuss problems by raising Issues that need to be addressed; Positions in response to those Issues, and Arguments to support or object to Positions. The aim of gIBIS was to make the rationale for a design decision explicit by capturing this argumentation in a hypermedia structure. Information, e.g., design sketches and code, were incorporated with hyperlinks. There was little evidence that knowledge was re-used with the system. Rather, it seems as if gIBIS mainly supported brainstorming of co-located users.

The gIBIS system is a commercial product today, called Questmap (figure 4).

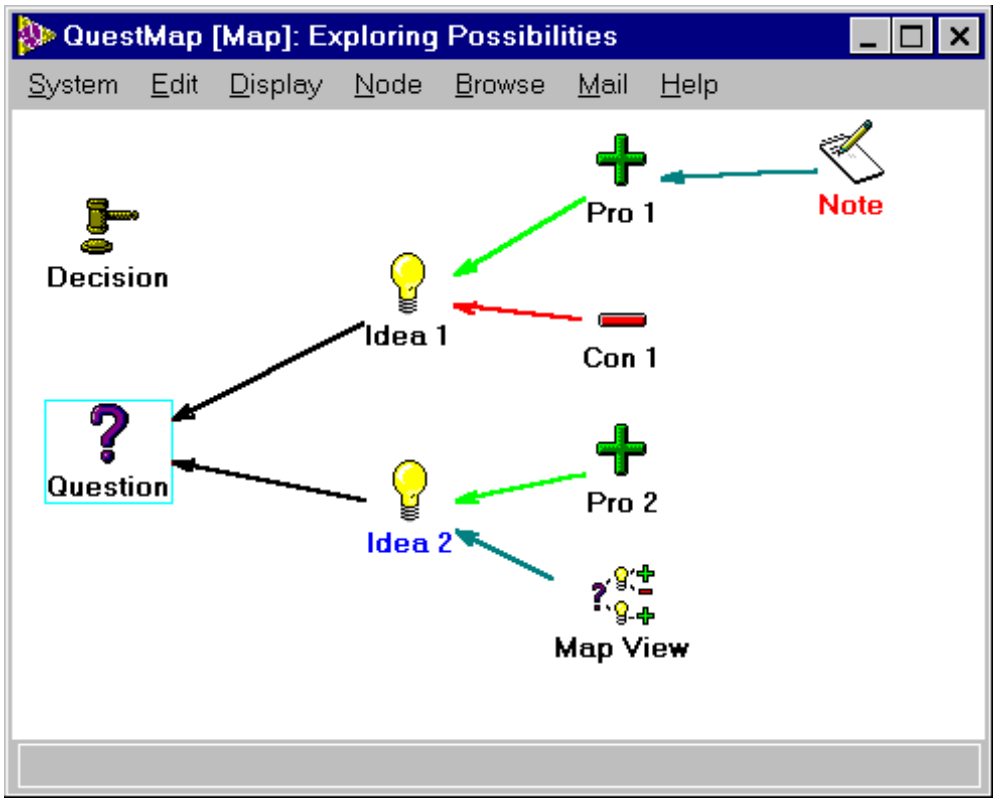

Figure 4: A screen dump of the QuestMaps system. Source: www.gdss.com

Ideas similar to gIBIS can be found in a recent system, the *RepTool* (Jordan et al 1998). The system supports the collection, analysis, and presentation of data about "workscapes." A workscape is defined as a collection of relevant information about a workplace and work practice in a central data structure centred on a map. The objective of the system was to support systematic data collection and collaborative data analysis that should be utilised at all levels of an organisation to build a shared view of the formal and informal work processes.

It is unclear how well the system works in practice, but to use maps to attach information to places where it is relevant could be interesting in mobile work.

Like gIBIS and RepTool, the Project memory system offers team members a central repository and access point for project communications and documents (Weiser and Morrison 1998). The information is linked with hypertext and classified into five discrete entities: projects, users, events, meetings, and documents. The Project memory system tries to integrate all kinds of knowledge while still emphasising that it can be accomplished with a minimal amount of extra user effort and in a manner that motivates use.

A laboratory study demonstrated that the system performed quite well. However, some difficulties with the capturing of relevant information were reported. Such problems could increase if the system was used in mobile settings.

Another related design rationale system is the *Designer Assistant* (Terveen et al 1993, 1995). The objective of the system was to create a "living design memory" for software developers. This was accomplished by introducing an interface between the developers and the product in the form of hyperlinks associating information units with the pertaining parts of the products. The system was integrated with the software development process and it is claimed that relevant knowledge exists in the form of "folklore" and that it should be incorporated in the system. The authors underline that understanding the organisational contexts and work practices is necessary for the system to succeed.

Perhaps the Designer Assistant could be useful for mobile service technicians as they may see comments by the designers of the product. On the other hand, it may be problematic since a service

technician does not have the same organisational context as a designer.

3.3 Collaborative filtering systems

One of the first collaborative filtering systems for knowledge support in the literature is the seminal system *Information Lens* (Malone et al 1987). The objective was to help people share and filter information communicated by computer-based messaging systems. Information Lens had different kinds of filters, but the most interesting one was the collaborative aspect. The systems could learn from how a group of users treated messages and thereby assist the filtering.

Another system for collaborative filtering is *Fab,* a recommendation service for Web documents supporting the users long-term interests (Balabanovíc 1997, 1998). The system relies on users explicitly rating Web documents in combination with content analysis of the text. The novel idea is to combine the specialisation (ratings), with the serendipitous character of collaborative filtering. The author shows that the combinations of the two techniques are best suited for simple single-topic long-term interests.

It may be useful to access single-topic long-term interests in mobile settings. However, to read a Web document that is aligned with a single-topic long-term interest is done individually and disconnected from the task at hand.

SOAP (Social Agent Platform) sends queries to Web search engines and utilises collaborative filtering to support the long-term interests of its users (Voss and Kreifelts 1997). A user can start agents and with little overhead exploit bookmark collections as sources of describing the long-term interest. Feedback on the results was used to achieve good precision.

The results of an agent can also be shared among users and groups of users known to each other (see figure 6).

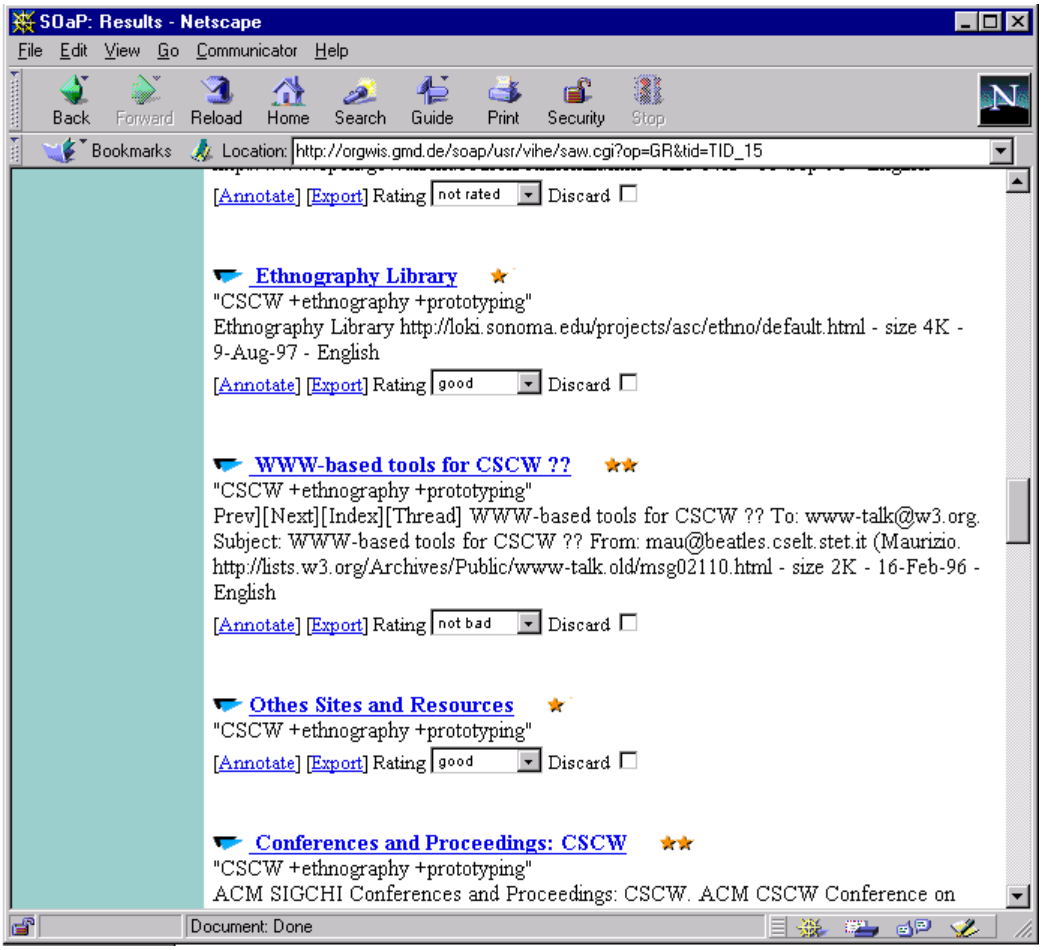

Figure 6: A screen shot of the result dialog of the SOAP system. The stars are recommendations by other users.

The system seemed to perform well, but no empirical results were reported. Access to the system for mobile users could be beneficial, but not for the task at hand, but rather to fill a free moment with an interesting activity.

Similar to SOAP is the *Knowledge Pump* system, which gives support for long-term interests by recommending Web pages and getting people with similar interests together (Glance et al 1998). The system integrates and refines ideas from intelligent information sharing (Malone et al 1987) and collaborative filtering. The aim seems to be long-term competence development rather than support for the task at hand. The authors' vision is that the Knowledge Pump should be an integrated and seamless part of the users' work practice. However, again there are no empirical results reported.

Another collaborative filtering system is the *Expert Finder*, an agent tailored for Java programmers (Lieberman and Vivacqua 2000).

The system suggests co-workers with expertise on the modules that the programmer is currently using. The expertise is automatically identified based on what source code the co-workers have written. The authors show that the system performed well compared to manual approaches to determine the expertise of co-workers. Another advantage is that the expertise dictionary is automatically updated.

The nature of Java programming is quite stationary, and thus there has been little need to support mobility. On the other hand, the authors argue that the systems could be useful in other domains of work.

Another way of locating experts is offered by the *Referral Web* system (Kautz et al 1997ab). Referral Web helps people to find experts based on their relationship in a social network. An assumption is that there is a topical expertise among all co-authors that can be identified by their participation in co-author relationships. An index is built up based on text content analysis of bibliography lists on Web pages. The index can then be queried for a particular topic.

For example, figure 5 shows a search for experts on “situated action.” The researcher Lucy Suchman is ranked highly in this search. The user can also investigate the relationship between people in the index. The figure 5 also shows the relationship between Lucy Suchman, Herbert Simon and Paul Luff.

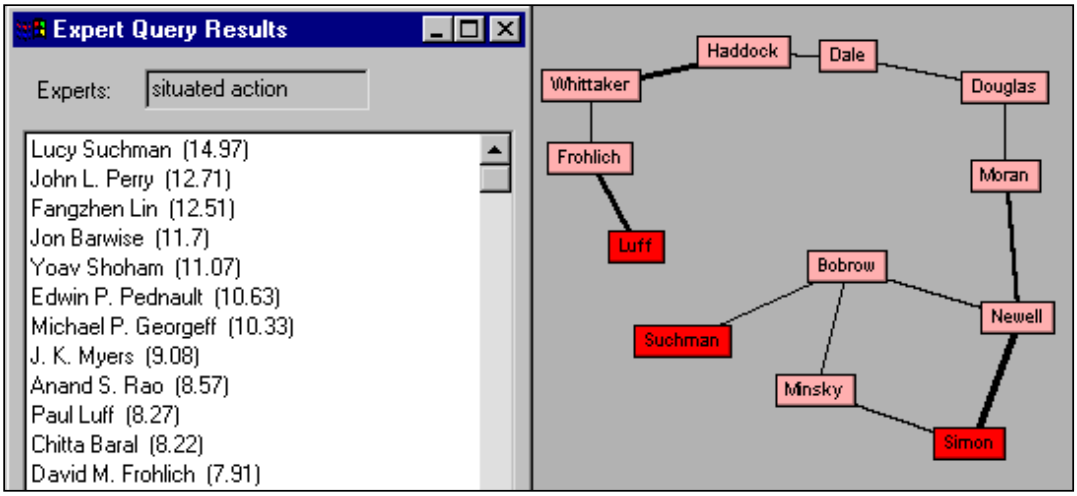

Figure 5: A screen shot of Referral Web. The figure is generated with a demo version of Referral Web found at http://foraker.research.att.com/refweb/version2/RefWeb.html.

The authors' report very promising technological results, but it is unclear how it could be useful in mobile work situations. To be aware of who is an established researcher on a topic is useful on a long-term basis rather than for the task at hand in the field.

Leonard Foner's *Yenta* is a de-centralised matchmaking agent architecture supporting distributed and automatic generation of clusters of users who are interested in similar topics (Foner 1997, 1999). The users can send messages to each other or to everyone in the cluster. Every user runs their own copy of Yenta, and a large numbers of agents form coalitions of users across the Internet who shares common interests through text content analysis. Figure 7 shows how a user gives feedback on interest profiles.

The author reports very promising technological results, but similar to most of the other systems discussed, in this section, it again mainly supports long-term interests. Thus, it is unclear how it can provide timely knowledge in mobile work.

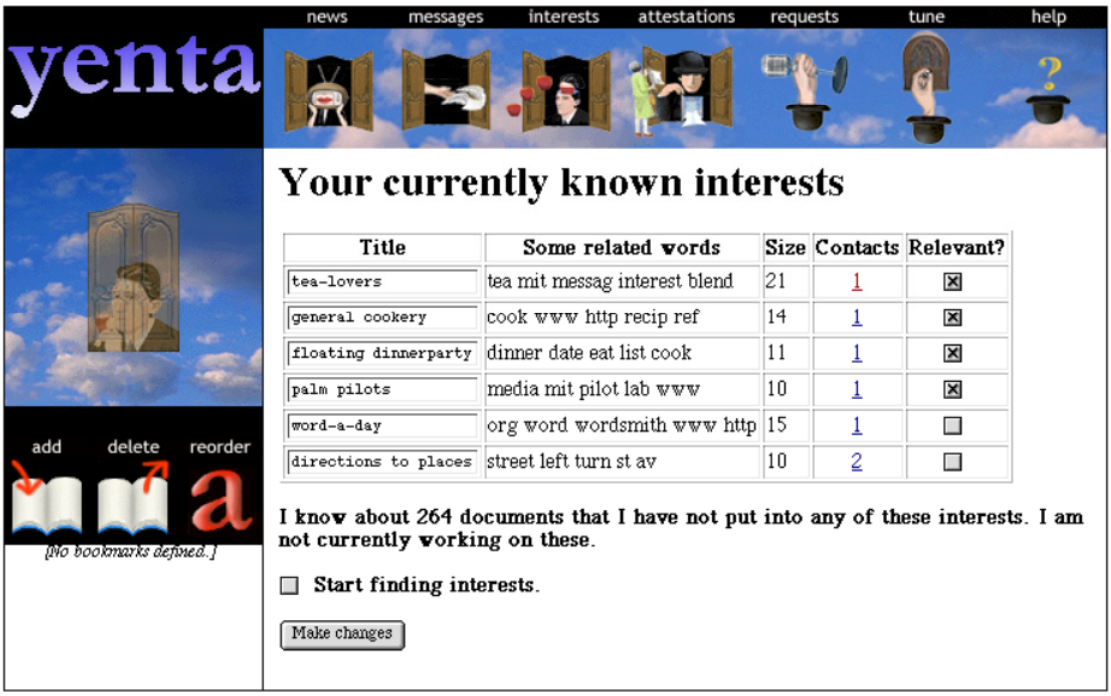

Figure 7: A Screen dump of the Yenta system (Foner 1999, p. 79).

3.4 Discussion of knowledge systems

Concepts like knowledge management and organisational memory have been the subject of much critique. Randall et al (1996), for example, have criticised the metaphor of organisational memory, suggesting that it can only support a very constrained set of organisational problems. Similarly, Bannon and Kuutti (1996) argue that if organisational memory is a useful notion at all, it is only to the extent that it refers to active remembering in its own context. Their conclusion is that a model of active remembering should replace the passive "repository" model of memory. The implication is, they argue, that designers of knowledge systems should not rely on idealised models of human activity but take into account how work is actually conducted (see also, Brown and Duguid 1991).

Clearly, the systems described here are useful in some domains. However, few of them are designed to support mobile users. Inasmuch as mobile work differs from stationary work (which previous research suggests) it may not be easy to redesign these systems for mobility. However, some aspects of mobile work do in fact share features with stationary work, which make stationary systems useful in some mobile situations. For example, a person "travelling" by train could logistically use a networked laptop to access systems like Fab, Yenta and SOAP. However, in other mobile situations, the stationary system would not offer proper support (which they of course were not designed to provide).

Systems with Web access could also be used in some forms of mobile work, e.g., when "visiting" another organisation. In this case, a stationary PC with web access could be the access point. "Wandering" on the other hand can only be supported by systems with mobile access. Some PDAs have Web browsers. However, Web access via a mobile device is problematic today because mobile browsers have very limited display size and support for Java.

Summarizing the above review of the literature on knowledge systems from the perspective of mobile work, we may conclude:

- most knowledge systems are mainly aimed at supporting stationary work and not mobile work;
- making stationary knowledge systems mobile could be useful, but only for some organisational problems;

- there are to date no systems that offer timely knowledge to mobile personnel.

4. Method

As noted in the previous section most knowledge systems do not translate well to mobile work. Therefore, we may ask: how do we design systems that make timely knowledge available for mobile worker? Below I describe the research approach used in this thesis, i.e., how an answer for the question was found.

4.1 Research Approach

The research approach adopted in this thesis has several roots, which together has provided both a foundation and inspiration for the research reported here:

- One root is that of Informatics, which according to Dahlbom (1996, p. 29) can be described as "… a theory and design oriented study of information technology use, an artificial science with the intertwined complex of people and information technology as its subject matter." The objective of Informatics is to improve the use of IT by contributing to the development of both the use and the technology itself.

- Another root is computer supported co-operative work (CSCW), which is an inter-disciplinary field concerned with issues ranging from highly technological aspects of groupware technology to sociological studies of collaborative work. According to Dix et al (1997) CSCW is about "groups of users – how to design systems to support their work as a group and how to understand the effect of technology on their work patterns."

- A third tradition that has inspired my research is that of human-computer interaction (HCI), which is the study of how people interact with computers and to what extent computers are or are not developed for successful interaction (see e.g., Preece et al 1994). As opposed to CSCW, HCI has traditionally been inspired by

psychological theory and focused on the individual user. This is perhaps not as relevant today as most systems are collaborative.

- Finally, mobile computing, being mainly a technological discipline that designs, develops and evaluates mobile computer devices and wireless communication protocols (Satyanarayanan 1996), has provided me with a stronger technological focus.

A typical research process, using this combined approach is can be described in the following four steps (figure 8):

1. The approach begins with two concurrent activities, i.e., a study of an empirical setting, and experiments with technologies. The activity seeks to explore the possibilities of new IT use inspired by empirical and technological explorations.

2. The empirical finding and the technological ideas are then formulated into design proposals. The design proposals are discussed in workshops with typical users. A prototype is then developed to see what is technologically possible.

3. The prototype is validated under actual work conditions so that the users can give the researchers feedback on its design qualities.

4. The implications derived from the empirical work and the experiences with the prototype are further analysed with the objective to generalise the ideas to other domains of mobile work. Finally, the generalised ideas are used to design a technological architecture.

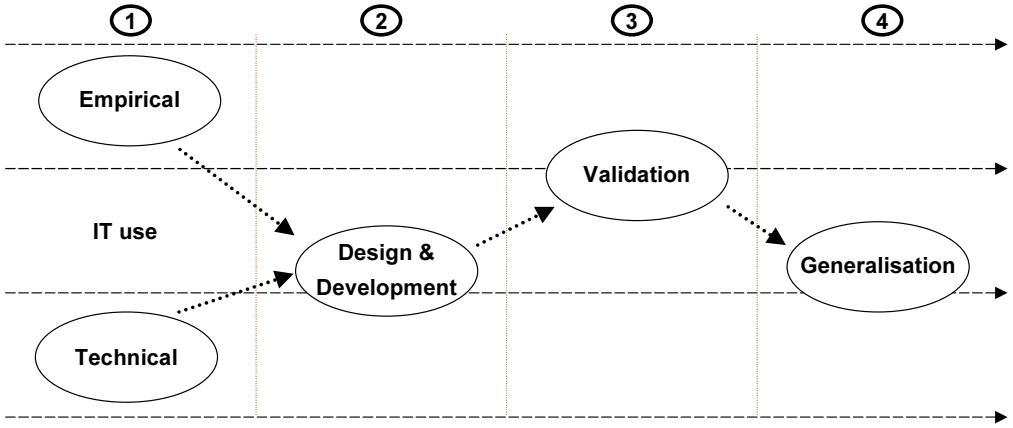

Figure 8: The research approach.

A more general discussion of research methods and this research approach can be found in Ljungberg et al (1998) and Dahlbom and Ljungberg (2000). The following sections address how I have used the approach.

4.2 Fieldwork

An ethnographic approach was used in the fieldwork and has been used frequently in Informatics. There are two trends that strongly motivate the use of ethnography in Informatics. Firstly, many systems have failed because the design pays insufficient attention to the social context of work, which existing methods for requirement elicitation mainly fail to capture (Hughes et al 1994). Secondly, mobile and collaborative technologies pose new problems for design, which require the development of new methods to analyze practice.

The research in this thesis has been conducted in three companies (see table 1).

Company	*Description*	*Data collection*	*Prototype system*
Volvo Group (Chapter 2)	Vehicle producer	Interviews, observation	IntraNews, also described in chapter 3.
Göteborg Energi (Chapter 3)	Energy provider	Observation, Workshops	SIMON, Reported in Bergquist et al (1998).
Radio Sweden (Chapter 4-5)	Public service Radio station	Observation, Workshops	NewsMate, reported in chapter 5.

Table 1: A description of the studied companies.

It is sometimes claimed that ethnography is just a data collection method and that the data analysis must rely on a theoretical perspective. This idea is rejected by, for example, Silverman (1985) who views all research as ethnography if it observes events and actions in natural situations and acknowledges the mutual dependencies between theory and practices.

The main data collection method in my research was observation. The amount of interview data was reduced as much as possible. I believe that interview data is not suitable as a basis for an analysis intended to inform design of IT use (cf., Silverman 1998). An interview, especially if open-ended, is itself a form of social control,

which shapes what people say (cf., Hammersley and Atkinsson 1993). (Interview data can be used, however, when evaluating design). Furthermore, since peoples' use of knowledge in organisations is social and dynamic it must be observed and documented as it happens and not in retrospect. The time required to obtain sufficient empirical data in this case is not several years as some anthropologically oriented approaches suggest, but usually less than a month (see, e.g., "quick and dirty ethnography" discussed by Hughes et al 1994).

An ethnographic analysis does not seek to impose a framework on the setting, but instead discover the social organisational properties of that setting as they are naturally occurring (cf., Hughes et al 1993). However, most post-positivistic approaches to philosophy of science agree that all facts are in one way or an other theory-loaded, hence we all have some theoretical preconceptions that shape our analysis. It is not a problem to analyse the empirical data inspired by the theoretical perspectives as long as the analysis is not constrained by the concepts of those theories. Randall et al (1996) and Anderson's and Sharock's (1993) discussions regarding "organisational knowledge" and "organisational remembering" have influenced me, but since my interests lies in the design of new IT artefacts the analysis of the empirical data was carried out in the following manner:

1. Start to develop categories, which illuminate the data
2. Find appropriate examples of each category to demonstrate its relevance
3. Develop the categories into implications that are relevant for design.

This approach is partly inspired by "grounded theory" method (Glaser and Strauss 1967).

4.3 Technological elaboration

In order to develop innovative designs you must know what is technologically possible. In Informatics, technological elaboration is usually about combining existing technologies, platforms and protocols that are commercially available or easy to implement based on recent

research. In some situations the combination can pose new problems and possibilities that improve the technology itself.

Technological elaboration is related to the “hunch-and-hack” approach described by Holmquist (2000) or the so-called “demo-or-die” approach from MIT Media Lab (Brandt 1987).

4.4 Design and Development

Design proposals that are based on the technological and empirical findings are discussed in workshops with potential users. The workshops are focused on the design proposals and the users can exercise their expertise in how well the design matches the possibilities they see of how to support the work practice. This is partly inspired by participatory design, but with a focus on letting the user give their view about novel features rather than engage in design (cf., Ljungberg et al 1998).

A prototype is then developed based on one of the design proposals. It is important that the novelties of the prototype are easy to demonstrate and that it is based on technology that is currently available. It is always difficult to decide when a design is good enough, but it is usefully determined by the budget of the project.

4.5 Validation

When a prototype is validated it is done both in workshops and under real working conditions.

The workshops consist of about five to ten users. A researcher moderates the workshop where issues and concerns about the features of a system are discussed (cf., Nielsen 1993, pp. 214-216). It is also easy and useful to arrange workshops for people from another work domain than the ones in the field study.

Validation in work practice means that the researcher follows the users and maps their current task with the use of the prototype. When users have a few moments free they are confronted with the results of the prototype and asked to comment. This method makes it possible for the user to see if the system is of any help in the actual

work situation and how well the design matched the work practice (see Hughes et al (1993) for a discussion of “Evaluative ethnography”).

4.6 Generalisation

After the prototype has been validated the experiences are further analysed with the objective to generalise the ideas to other domains of mobile work. In this way one meets the classical scientific objective to offer results with general applicability. One also makes the results more generally useful and interesting and, of course, more commercially appealing. One way to do this generalising is to set up requirements on a technological architecture, thus refining and reflecting on both the design and the empirical findings.

5. Results

In this section I summarise the results of the thesis. Let me first summarise the individual papers and map them to the research approach.

5.1 The papers

It is important to note that the papers do not always match the steps in the research approach, e.g., two steps can be reported in one paper (see figure 9).

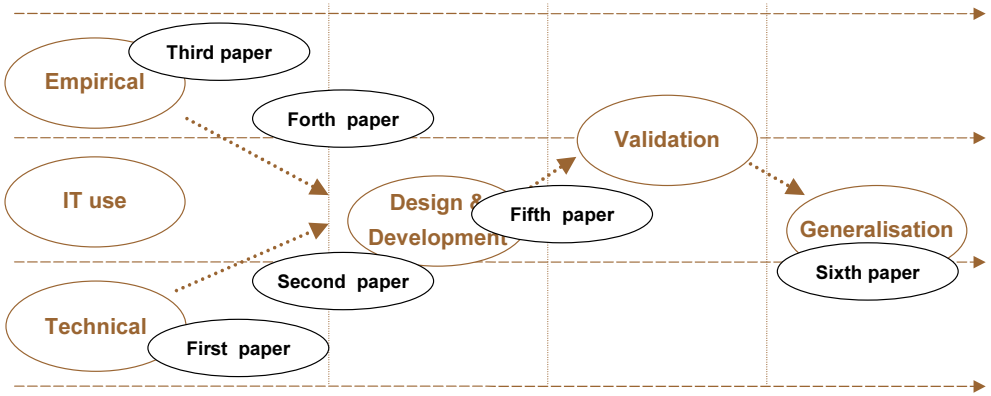

Figure 9: The papers matched with the research approach.

The main findings of the six papers are summarised in the following sections.

5.1.1 Surveying the World Wide Web

The first paper aims to demonstrate the principles of using semi-automatic measurement instruments to advance our understanding of the Web as a body of textual traces of human activity. The measurements illustrate that it is possible to detect general topics, complexity, change rate and vocabulary richness of a Web site. The results of the technological elaboration are discussed related to the

management of knowledge. Some aspects of the major technical findings are used in the succeeding papers two, five and six.

5.1.2 IntraNews: A News Recommending Service for Corporate Intranets

The second paper was developed from a field study that found that it is often problematic for employees in large companies to keep up to date with what is going on within the organisation. Corporate intranets can be used to ease the problem, but existing techniques, i.e., search engines, portals, and retrieval agents, all have shortcomings when attempting to address the problem.

The contribution in the paper is a complementary approach to address the problem, which is illustrated with a prototype called IntraNews. The approach is to view the problem from a *news consumption* perspective rather than from an *information retrieval* perspective. IntraNews implicitly captures collaborative filtering information from the log files of each intranet server. The information is used to identify relevant documents that will then keep people up to date. Recommendations are issued to both stationary and mobile users. Mobile access is obtained via the Wireless Application Protocol (WAP) or an HTML compatible mobile Web browser.

5.1.3 Exploring Support for Knowledge Management in Mobile Work

The third paper reports fieldwork from the electrical utilities industry, examining the suitability of current knowledge management perspectives to the day-to-day work of mobile staff.

The results of the empirical study makes a distinction between four aspects of mobile knowledge management as it took place in the mobile work setting. These are:

- sharing, i.e., how several parties exchange knowledge;
- indexing, i.e., how one party explains to another what knowledge to retrieve;
- diagnosing, i.e., how two parties make sense of how to interpret a situation;

- foreseeing, i.e., how one party (or more) uses knowledge to project the future.

The paper shows that current knowledge management perspectives are not suitable in mobile work and an initial sketch of a framework for "practical knowledge management" is outlined.

5.1.4 Empirically Informed Knowledge Management Systems in Mobile Domains

The objective of the forth paper was to give detailed implications for the design of knowledge management systems based on empirical findings of the work of news journalists at a radio channel.

Based on the empirical analysis a model of knowledge management systems is suggested that: first, starts out from evolving interdependent tasks of the users; second, which has all information linked to authors and their accessibility; third, that provides filters that are task-based (as opposed to long and short-term only); fourth, that shows hits and overview; and fifth, which provides flexible access (stationary and mobile). The design implications complement the "the repository view" of knowledge management, which dominates the field, and provides suggestions that link to the following paper.

5.1.5 NewsMate: Providing Timely Knowledge to Mobile and Distributed News Journalists

The fifth paper describes the design and evaluation of the NewsMate prototype. The NewsMate prototype provides mobile and distributed news journalists with timely knowledge. The idea with the prototype was to suggest a complementary approach to IT-support for Knowledge Management promoting the need to: first, step away from the focus on problem-solving and instead incorporate knowledge management support in the task at hand; second, consider records from sources internal as well as external to the organisation; third, favour direct communication based on present conditions.

The system was designed based on empirical findings and the technological design of the prototype is based on the results from paper one and two. The prototype was also validated in practice with promising results.

5.1.6 FieldWise: A Mobile Knowledge Management Architecture

The final paper describes a mobile knowledge management architecture, called FieldWise. The paper is a further development and generalisation of the results in all of the previous papers. The novel requirements on which FieldWise is based are derived from the field studies in paper three and four. The techniques are based on the results in paper one, two and five. Finally, the new requirements are elicited from the empirical experiences with the prototype in paper five.

FieldWise provides mobile access from five client platforms, i.e., Pocket PC, Windows CE 2.11 (Palm-sized and Handheld), EPOC, and PalmOS. Furthermore, it allows flexible access for freelance employees. However, it has been developed to accommodate more then mobile access to stationary systems. Several novel features have been incorporated that have not been previously reported in the literature. Namely support for:

- evolving and interdependent tasks
- overview of internal and external information sources
- location of available expertise
- filtering based on task and long-term interests
- dynamic configuration of the mobile service

The first implementation of FieldWise is in the domain of news journalism, but the architecture can be easily adapted to other mobile work domains, e.g. sales.

5.2 Summary conclusions

In this section, I describe the overall contributions of the thesis to the use of knowledge systems in mobile work. The thesis answers the question *how can we provide mobile workers with timely knowledge* by offering the following significant contributions.

The main points of the introductory chapter and the following six papers can be summarised as follows:

- A review of knowledge systems demonstrating that mobile work has not been afforded significant attention.

- The design, definition and refinement of several semi-automatic techniques to capture and measure text and hyperlink data that can be used in knowledge systems.

- The design and implementation of the IntraNews prototype system, which assists users in finding organisational news on an intranet. A novel approach to collaborative filtering is used to capture user data implicitly. IntraNews addresses mobility to some extent as the system has can be accessed from mobile terminal running WAP.

- A detailed description of the day-to-day work of teams of mobile service electricians. The empirical results show that current knowledge management perspectives are not appropriate in mobile work. An initial sketch of a framework for “practical knowledge management” is outlined.

- A detailed description of the day-to-day work of mobile news journalists at a radio channel. The empirical findings were analysed to derive design implications for knowledge management in mobile work. Six inter-related design implications were derived forming an approach to mobile knowledge management that improves the “repository model of knowledge,” which currently dominates the field.

- The design, implementation and evaluation of the NewsMate prototype in the domain of news journalism. The prototype was based on design implications derived from fieldwork of mobile news

journalists. The prototype validation shows promising results, i.e., it matches the work practice well taking the mobility of the users into account.

- The design and implementation of FieldWise, a mobile knowledge management architecture. The design is derived from the fieldwork and experiences with the prototypes. The architecture can easily be adapted to several mobile work settings and emerging mobile technologies. The first practical implementation is in the domain of news journalism.

6. Discussion

How can we therefore provide mobile workers with timely knowledge? Some ways of doing this is addressed in this thesis. However, it really depends on what type of mobile work and knowledge is to be supported. A methodological apparatus to tackle the quest is offered in this thesis. This is valuable in the ever-changing business environment of the “new economy.”

In the face of the unavoidable transformational nature of technology and system design in working settings, it may seem that ethnography is rather powerless. It has traditionally been used to analyse practice, rather than design. Button (1993, p. 26) suggests that sociologically generated analytic categories, constructed in micro-level studies, may be of little relevance to the practical problems of designing systems. However, Bentley et al (1992) suggest that ethnographic studies are helpful in informing design and may produce insights that contradict conventional thinking in systems design. Both Button (1993) and Bentley et al (1992) seem to suggest a “strict” division of labour between the system designers and the social scientist. Although much of the work in this thesis has been collaborative I have been focal in all aspects of the research process. My view is that fieldwork can contradict conventional thinking in systems design.

Dourich and Button (1998) have developed a new model of interaction, “technomethodology,” with computer systems that is based on ethnomethodological perspectives on everyday human social action. They suggest that there is a trap in thinking that fieldwork in itself will result in novel and generally applicable systems. Instead, they suggest that the ethnomethodological position can inform, re-specify and re-conceptualise fundamental elements of systems design from an analytical point of view. However, it is difficult to “assure” the general applicability of a design and I am not certain that an ethnomethodological position can change that. Although ethnomethodology emphasises the everyday ordinary in human social action it is a very complex way of reasoning. I have found the

ethnomethdological way of reasoning appealing in many ways, but another way to find out if a design is successful is to validate it. When several users in different fields agree that a design is relevant and useful the approach can be considered justified.

Today, many companies are experimenting with outsourcing, downsizing, and automations of routine jobs. There is an increased focus on business relationships between companies as new structures are being formed between businesses, suppliers, governments and customers. However, organisations are inherently slow to change compared to individuals and yet, it is vital for their survival to adjust and respond to changing market conditions. Some experts are, from the perspective of the company, best suited to be on short-term contracts to guarantee that their knowledge is up to date and valid (cf., Albert and Bradley 1997). For example, very few organisations hire lawyers because their legal counselling needs cannot be held by a single person due to the high degree of specialisation of lawyers. Perhaps this is becoming more common for other type of jobs? More and more jobs are taking the contours of “knowledge work.” More and more employees need to combine technological expertise with the ability to work effectively in teams, form productive relationships with clients and customers (Argyris 1991). Furthermore, the massive increase of communications technologies, e.g., mobile telephones, e-mail, instant messaging, and PDAs have led to changes in established work-based communication practices (Kraut et al 1993; Whittaker et al 1994; Nardi et al 2000). Does this suggest that organisational relationships based on long-term contracts are becoming obsolete? If this is the case, what will the knowledge systems be like? The sixth paper offers some suggestions for how freelance workers can be supported. However, there inevitably will be other issues that need to be addressed if almost everyone is mobile and hired on short-term contracts. For example, who will “own” the knowledge system? Will there be any central co-ordination on technical (database) or managerial levels? These are unanswered questions that further research will have to address.

7. References

Ackerman M.S. and T.W. Malone (1990) “Answer Garden: a tool for growing organizational memory,” In *Proceedings of the conference on Office information systems*, pp. 31-39, Cambridge, MA: ACM Press

Ackerman, M.S. (1994) “Augmenting the organizational memory: a field study of answer garden,” In *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 243-252, Chapel Hill, NC: ACM Press.

Ackerman, M.S. and D.W. McDonald (1996) “Answer Garden 2: Merging organizational memory with collaborative help,” In *Proceedings of the ACM 1996 Conference on Computer Supported Cooperative Work*, pp. 97-105, Cambridge, MA: ACM Press.

Albert, S. and K. Bradley (1997) *Managing Knowledge: Experts, Agencies and Organizations*. Cambridge: Cambridge University Press.

Anderson, R. and W. Sharock. (1993) “Can Organisations Afford Knowledge?” *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, 1 (3), pp. 123-142.

Argyris, C. (1991) “Teaching Smart People How to Learn,” *Harvard Business Review*, 69 (3), pp. 99-109.

Balabanovíc, M. (1997) “An adaptive Web page recommendation service,” In *Proceedings of the first international conference on Autonomous agents,* pp. 378-385, Marina del Rey, CA: ACM Press.

Balabanovíc, M. (1998) “Exploring versus Exploiting when Learning User Models for Text Recommendation,” *User Modelling and User-Adapted Interaction: An International Journal*, 8, pp. 71-102.

Bannon, L.J. and K. Kuutti (1996) “Shifting Perspectives on Organizational Memory: from Storage to Active Remembering,” In *Proceedings of the 29th Hawaii Conference on System Sciences,* Vol. IV, pp. 156-167, Los Alamitos. IEEE Computer Society Press.

Bentley, R., J.A. Hughes, D. Randall, T. Rodden, P. Sawyer, D. Shapiro and I. Sommerville (1992) “Ethnographically-informed systems design for air traffic control,” In *Proceedings of the ACM 1992 Conference on Computer Supported Cooperative Work*, pp. 123-129, Toronto, Canada: ACM Press.

Bergquist, M., H. Fagrell, P. Ljungstrand and M. Storsten (1998) “Designing for Informal Mobile Cooperation,” At the Workshop *Handheld CSCW* at the ACM 1998 Conference on Computer Supported Cooperative Work, Seattle, MA, USA.

Bergqvist, J., P. Dahlberg, S. Kristoffersen and F. Ljungberg (1999) “Moving Out of the Meeting Room: Exploring support for mobile meetings,” In *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work*, pp. 81-98, Copenhagen, Denmark: Kluwer Academic Publishers.

Bellotti, V. and S. Bly (1996) “Walking away from the desktop computer distributed collaboration and mobility in a product design team,” In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*, pp. 209-218, Cambridge, MA: ACM Press.

Bellotti, V. and Y. Rogers (1997) “From Web Press to Web Pressure: Multimedia Representations and Multimedia Publishing,” In *Proceedings of ACM 1997 Conference on Human Factors in Computing Systems,* pp. 279-286, Atlanta, GA: ACM Press.

Bowers, J., J. O’Brien and J. Pycock (1996) “Practically accomplishing immersion: cooperation in and for virtual environments,” In *Proceedings of ACM 1996 Conference on Computer-Supported Cooperative Work,* pp. 380-389, Cambridge, MA: ACM Press.

Bowker, G.C. (1997) “Lest we remember; organizational forgetting and the production of knowledge,” *Accounting, Management and Information Technologies*, 7 (3), pp. 113-138.

Brand, S. (1987) *The Media Lab: Inventing the Future at M.I.T. New York*, NY: Viking Penguin.

Brown, J.S. (1999) “Internet technology in support of the concept of ‘communities-of-practice’: the case of Xerox,” *Accounting, Management and Information Technologies*, 8 (4), pp. 227-236.

Brown, J.S. and P. Duguid (1991) “Organizational Learning and Communities of Practise: Towards a Unified View of Working, Learning and Innovation,” *Organization Science*, 2 (1), pp. 40-57.

Bruno, L. (2000) “Central Standards Time,” *Red Herring*, May 2000.

Button, G. (1993) “The curious case of the vanishing technology,” In *Technology in Working Order,* G. Button (ed.), pp. 10-27, London: Routledge.

Castells, M. (1996) *The Rise of the Network Society*, Oxford: Blackwell.

Conklin, J. and M.L. Begeman (1988) “gIBIS: a hypertext tool for exploratory policy discussion,” *ACM Transactions on Office Information Systems*, 6 (4), pp. 303-331.

Dahlbom, B. (1996) “The New Informatics,” *Scandinavian Journal of Information Systems*, 8 (2), 29–48.

Dahlbom, B. and F. Ljungberg (1999) “Mobile Informatics,” *Scandinavian Journal of Information Systems*, 10 (1-2), pp. 227-234.

Dix, A., J. Finlay, G. Abowd and R. Beale (1997) *Human-Computer Interaction*, Prentice Hall.

Dourish, P. and G. Button (1998) "On Technomethodology: Foundational Relationships between Ethnomethodology and System Design," *Human-Computer Interaction*, 13 (4), pp 395-432.

El Sawy, O.A. and G.E. Bowles (1997) "Redesigning the customer support process for the electronic economy: Insights from storage dimensions," *MIS Quarterly*, 21 (4), pp. 457-484.

EMC (2000). EMC World Cellular Database. www.emc-database.com

Foner, L. (1997) "Yenta: A Multi-Agent, Referral Based Matchmaking System," In *Proceedings of the first international conference on Autonomous agents,* pp. 301-307, Marina del Rey, CA: ACM Press.

Foner, L. (1999) Political Artifacts and Personal Privacy: The Yenta Multi-Agent Distributed Matchmaking System. Unpublished PhD. dissertation, Program of Media Arts and Sciences, Massachusetts Institute of Technology. URL: foner.www.media.mit.edu/people/foner/PhD-Thesis/Dissertation/

Geertz, C. (1983) *Local Knowledge: Further Essays in Interpretive Anthropology*, New York: Basic Books.

Glance, N., A. Damián and M. Dardenne (1998) "Knowledge Pump: Supporting the Flow and Use of Knowledge," In *Information Technology for Knowledge management*, U.M. Borghoff and R. Pareschi (eds.). Berlin: Springer-Verlag.

Glaser, B. and A. Strauss (1967) *The Discovery of Grounded Theory*. Chicago: Aldine.

Hammersley, M. and P. Atkinson (1993) *Ethnography: Principles and practice*. Second edition, London: Routledge.

Heath C. and P. Luff (1991) "Collaborative Activity and Technological Design: Task Coordination in London Underground Control Rooms," In *Proceedings of the Second European Conference on Computer-Supported Co-operative Work,* pp. 15-22, Amsterdam, The Netherlands: Kluwer Academic Publishers.

Hedlund, G. (1994) "A model of knowledge management and the N-form corporation," *Strategic Management Journal*, 15, pp. 73-91.

Holmqvist, L.E. (2000) *Breaking the Screen Barrier,* PhD. thesis, Department of Informatics, Göteborg University, Sweden. ISSN: 1400-741X.

Hughes, J., D. Randall, and D. Shapiro (1993) "From ethnographic record to system design. Some experiences from the field," *Computer Supported Cooperative Work. An international journal*, 1 (3), pp. 123-141.

Hughes, J.A., V. King, T. Rodden and H. Andersen (1994) "Moving out from the control room: ethnography in system design," In *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 429-439, Chapel Hill, NC: ACM Press.

Jordan, B., R. Goldman and A. Eichler (1998) “A Technology for Supporting Knowledge Work: The RepTool,” In *Information Technology for Knowledge management*, U.M. Borghoff and R. Pareschi (eds.), pp. 79-96, Berlin: Springer-Verlag.

Karduck, A. (1994) “TeamBuilder: a CSCW tool for identifying experts and team formation,” *Computer Communications*, 1994, 17 (11), pp. 777-887.

Kautz, H., B. Selman and M. Shah (1997a) “Referral Web: combining social networks and collaborative filtering,” *Communications of the ACM*, 40 (3), pp. 63-65.

Kautz, H., B. Selman and M. Shah (1997b) “The Hidden Web,” *The AI Magazine*, 18 (2), pp. 27-36.

Kraut, R., R. Fish, R. Root and B. Chalfonte (1993) “Informal communication in organizations: form, function, and technology,” In *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*, R. Baecker (ed.), pp. 387-314, San Mateo, CA: Morgan Kaufmann Publishers.

Kristoffersen, S. and T. Rodden (1996) “Working by Walking Around. Requirements of flexible interaction management in video-supported collaborative work,” In *Proceedings of Human Computer Interaction,* pp. 315-329, London, UK: Springer-Verlag.

Kristoffersen, S. and F. Ljungberg (1998) “MobiCom: Networking dispersed groups,” *Interacting with Computers*, 1998, 10 (1), pp. 45-65.

Kristoffersen, S. and F. Ljungberg (1999) “‘Making place’ to make IT work: empirical explorations of HCI for mobile CSCW,” In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pp. 267-275, Phoenix, AZ: ACM Press.

Kristoffersen, S. and F. Ljungberg (2000) “Mobility: From Stationary to Mobile Work,” In *Planet Internet*, K. Braa et al (eds.), pp. 137-156, Studentlitteratur, Lund, Sweden.

Lieberman, H. and A. Vivacqua (2000) “Agents to Assist in Finding Help,” In *Proceedings of the ACM 2000 Conference on Human Factors in Computing Systems*, pp. 65-72, The Hauge, The Netherlands: ACM Press.

Ljungberg, F., B. Dahlbom, H. Fagrell, M. Bergquist and P. Ljungstrand (1998) “Innovation of new IT use: Combining approaches and perspectives in R&D projects,” In *Proceedings of the Fifth Biennial Participatory Design Conference*, pp. 203-209, Seattle, WA: CPSR and ACM Press.

Luff, P. and C. Heath (1998) “Mobility in Collaboration,” In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 305-314, Seattle, WA: ACM Press.

Malone, T.W., K.R. Grant, F.A Turback, S.A. Brobst, and M.D. Cohan (1987) "Intelligent information sharing systems," *Communications of the ACM*, 30 (5), pp. 390-402.

March, J.G. (1991) "How Decisions Happen in Organizations," *Human-Computer Interaction*, 6, pp. 95-117.

Nardi, B., S. Whittaker and H. Schwarz (2000) "It's Not What You Know, It's Who You Know: Work in the Information Age," *First Monday*, 5 (5), URL: http://firstmonday.org/issues/issue5_5/nardi/index.html

Nielsen, J. (1993) *Usability Engineering*. Academic Press Inc.

Preece, J., Y. Rogers, H. Sharp, D. Benyon, S. Holland and T. Carey (1994) *Human Computer Interaction*, Addison Wesley Reading.

Orr, J. (1996) *Talking about machines: An ethnography of a modern job*. Ithaca, NY: ILR Press, an imprint of Cornell University Press.

Randall, D., J. O'Brien, M. Rouncefield and J.A. Hughes (1996) "Organisational Memory and CSCW: Supporting the 'Mavis Phenomenon'," In *Proceedings of the 6th Australian Conference on Computer-Human Interaction*. Hamilton, New Zealand: IEEE Computer Society.

Randall, D., M. Rouncefield and J.A. Hughes (1995) "Chalk and Cheese: BPR and ethnomethodologically informed ethnography in CSCW," In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work*. pp. 325-340. Stockholm, Sweden: Kluwer Academic Publishers.

Rouncefield, M., J.A. Hughes, T. Rodden and S. Viller (1994) "Working with 'constant interruption': CSCW and the Small Office," In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*. pp. 275-286. Chapel Hill, NC: ACM Press.

Ruggles, R. (1998) "The state of the notion: knowledge management in practice," *California Management Review*, 40 (3), pp. 80-89.

Satyanarayanan, M. (1996) "Fundamental challenges in mobile computing," In *Proceedings of Fifteenth annual ACM symposium on Principles of distributed computing*, pp. 1-7, Philadelphia, PA: ACM Press.

Scarbrough, H. (1998) "Path(ological) dependency? Core Competencies from an organizational Perspective," *British Journal of Management*, 9, pp. 219-232.

Silverman, D. (1985) *Qualitative Methodology and Sociology: Describing the Social World*, Ashgate Publishing Group.

Terveen, L.G., P.G. Selfridge and M.D. Long (1993) "From 'folklore' to 'living design memory'," In *ACM 1993 Conference proceedings on Human factors in computing systems*, pp. 15-22, Amsterdam, The Netherlands: ACM Press.

Terveen, L.G., P.G. Selfridge and M.D. Long (1995) "Living Design Memory: Framework, Implementation, Lessons Learned," *Human-Computer Interaction*, (10) 1, pp. 1-37.

Weiser, M and J., Morrison (1998) “Project Memory: Information Management for project teams,” *Journal of Management Information Systems*, 14 (4), pp. 149-167.

Wenger, E. (1998) *Communities of Practice: Learning, Meaning, and Identity,* Cambridge University Press.

Whittaker S., D. Frohlich, O. Daly-Jones (1994) “Informal communication: what is it like and how might we support it,” In *Proceedings of ACM 1994 Conference on Human Factors in Computing Systems*, pp. 130-137, Boston, MA: ACM Press.

Voss, A. and T. Kreifelts (1997) “SOAP: Social Agents Providing People With Useful Information In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, pp. 291-298, Phoenix, AZ: ACM Press.

Yakemovic, K.C.B. and E.J. Conklin (1990) “Report on a development project use of an issue-based information system,” In *Proceedings of the ACM 1990 Conference on Computer Supported Cooperative Work*, pp. 105-118, Los Angeles, CA: ACM Press.

First paper[1]

Surveying the World Wide Web

Abstract

The World Wide Web (the Web) is the main driving force behind the rapid diffusion of Internet technology. As a result, we are beginning to live a significant part of our lives in Cyberspace. Measuring and monitoring our surroundings is an essential human activity that helps us both to understand and shape the world we live in. Substantial efforts have in the past years been invested into further understanding the Internet in general and the Web in particular through, for example, surveys of user attitude and behaviour, maps of Internet traffic, and indexing of content. Very little research has, however, investigated how to measure and monitor the contents of Web sites based on a combination of linguistics and data visualisation measures. Many efforts have demonstrated the use of techniques from within a particular discipline such as information retrieval, data mining, or autonomous agents. This paper, however, explores issues related to the monitoring of contents and changes to the Web based on a range of measures. The paper aims to demonstrate the principles behind the application of semi-automatic measurement instruments to forward our understanding of the Web as a body of textual traces of human activity. The paper suggests five basic types of measures for studying the Web: volume, density, vocabulary, structure, and relative measures. A survey of 82 Swedish Web sites was conducted using semi-autonomous Web robots for information retrieval and filtering based on techniques from linguistics and information visualisation. Examples demonstrate how such data can be applied to summarise site contents,

[1] Fagrell, H. and C. Sørensen (2000) "Surveying the World Wide Web," *Scandinavian Journal of Information Systems*, 11 (1), pp. 25-50.

identify site topic, map site structure, and compare Web sites. The results are discussed and related to emergent issues, such as Web navigation, electronic commerce and the management of knowledge.

1. Introduction

The Internet has experienced an explosive rate of growth since it's inception in the late sixties, initially designed as an experimental network for scientists and US Government contractors. During the 1980 it gained widespread popularity among faculty, staff and students at universities and research centres as an information infrastructure. Restrictions for commercial use of the Internet were lifted in 1991, and that year the World Wide Web (the Web) (Berners-Lee et al 1994) was introduced to the public. These two events proved to be a very strong driving force for diffusion of the technology to most aspects of public and private domain around the world at a previously unseen rate (Leiner et al 1997; Guice 1998; Hannemyr 1998; Zakon 1998). As a result, many spend an increasing proportion of both their professional and private lives in this new "world" of bits (Mitchell 1995). We search for information, purchase products and services, and interact with people using this particular information infrastructure. We also experience new problems related to living in and with this new world, such as finding our way, finding specific information, remembering where we have been, and in general getting an overview (Li 1998; Sørensen 1998; Nielsen 1999). In the physical world we have been monitoring an abundance of both cultural and natural attributes over hundreds of years. Monitoring and measuring are fundamental activities for understanding the world we inhabit and shape and substantial efforts have been put into understanding and describing both the Internet and the Web (Dodge 1999). The behaviour and attitudes of people using both the Internet and the Web have been surveyed. Others have studied how to represent, map, visualise and analyse the Internet and the Web using quantitative measures. Most approaches that have been suggested for characterising the contents of the Web, however, tend to focus on the application of a particular technique, and a large proportion of the efforts are directly related to

support search engines indexing the Web or to visualising the hyperlink structure of the Web. There has been relatively little research that has explored the contents of the Web as a body of text undergoing constant change, in an attempt to provide an overview and to compare the contents and structure of Web sites.

This paper asks the question; How can we apply semi-automatic measurement instruments to better understand and present the contents of and changes to sites on the World Wide Web? This article presents results based on a comprehensive survey where 82 Swedish Web sites within six sectors were retrieved a number of times throug seven weeks in 1996. The process was initiated by the specification of the Web measures to be calculated. A series of Web robots were developed and tested. After selecting the most appropriate type of robot for the task at hand, the 82 Web sites were selected and the robots were employed as a means of semi-automatic Web retrieval (Leonard 1997; Schneiderman and Maes 1997; Schubert et al 1998; Wooldridge and Jennings 1999). Techniques from computational linguistics and information visualisation were used in order to filter and represent the Web site contents (Tufte 1983; Tesitelová 1992; Young 1996). Applying techniques from computational linguistics enabled us to provide overviews of not only superficial syntactical changes to the Web sites, but also to gain some insight in more substantial changes to the type of contents of the Web sites as well as the degree of sophistication of the language employed. The task was basically one of collecting, detecting and displaying information based on a stable need, from a dynamic and unstructured source (Oard 1997). The process was semi-automatic, in that the robots were manually configured with a list of sites to be visited. The data collection was automatic, and data aggregation was partly automatic and partly manual. It was based on measures selected from each of the five basic categories; volume, density, vocabulary, structure and relative measures, we demonstrate how to summarise site contents, identify site topic, map site structure, and compare Web sites. The examples are all drawn from the newspaper sector. Given the focus on measuring and understanding changes, this strata appeared to be the most interesting.

Obviously, a concern could be raised regarding the age of the data presented, given the survey was conducted throughout 1996. Given the rapid changes to the Web, the actual data collected may now

be atypical. Given the focus on newspaper *Web* sites, Eriksen et al (2000) illustrates very well in their analysis of how a Swedish, a Norwegian and a Danish newspaper Web site all have changed from 1996 to 1999. However, two issues need to be considered here. Firstly, the delay in publishing the results are caused by factors outside the authors' control. Secondly, the important aspect of this paper is not to provide an actual up-to-date overview of the Swedish Web. The purpose is to demonstrate how the approach selected can add to the way we understand and present the Web. Given our choice of perspective where we look at the Web as individual sites containing hyper-linked text with embedded graphics, relatively little has changed. Although there is an increase in more sophisticated Web sites where a significant proportion is generated dynamically, the end product is still a Web page. Although there is a drive towards digital conversion of various media, involving streamed video and audio as an alternative to textual contents, the emergence of mobile computing with small networked information devices and standards for that platform has actually reinforced the importance and "centre-stage" of textual information.

In the following section we survey related research and outline the problem setting. Section 3 suggests a set of five basic measures for surveying the Web. Section 4 presents the instruments and procedures for collecting and analysing data in the survey. Section 5 presents examples of survey results applying Web measures from each of the five basic types suggested in Section 3. Section 6 discusses the approach and relates it to emergent issues within Web development and use such as Web navigation, electronic commerce, and the management of knowledge. It is argued that there is a need for mechanisms providing an overview of both the structure and the contents of complex Web sites. We conclude that this approach can inform future design of a number of specialised Web services such as navigational support, search engines and advertisement.

2. Related Research

We are increasingly relying on information and communication technology when we interact with others or document our activities. This implies that the traces we leave behind increasingly are electronic as opposed to being paper based. Little research has, however, investigated how we can survey and analyse changes to the Web, when viewed as textual traces of human activity. Other types of traces have been studied previously. In 1992, for example, Hill et al (1992) discussed electronic traces as implications for design of user interfaces. Whittaker et al (1998) have investigated USENET news messages for aspects of people's conversational strategies, demographic belonging and interaction frequency. Greenhalgh (1997) has investigated the traces of movement of avatars in a virtual collaborative environment to inform requirements for look ahead, bandwidth and special treatment of some participants.

A number of approaches for surveying the Web consisting of a dense weave of texts, pictures, interactive components, CGI scripts, have been suggested and applied since its conception in the early 1990. This section relates the approach applied in the survey of the Swedish Web to other research efforts. It is beyond the scope of this paper to compile and discuss all related research. Dodge (1999) however, presents the most comprehensive list of references we have found, and his index on has proved a valuable resource (see www.cybergeography.org).

In this paper we investigate the lessons that can be learnt about changes to the Web from semi-automatic data collection and analysis. The unit of analysis is the individual Web site defined by the registered domain name, and viewed as hyperlinked texts. The project surveyed the publicly accessible part of the Swedish Web. Within that, we chose 82 Web sites within six sectors. A Web robot surveyed the hypertext contents of the chosen Web sites a number of times during a six week period. During each measurement the robot collected the entire public Web site by traversing the hyperlink structure for the site. The data was downloaded and subjected to further analysis. The survey method can be characterised as the application of semi-autonomous robots (Leonard 1997; Schneiderman and Maes 1997; Schubert et al 1998; Wooldridge and Jennings 1999) to the World Wide Web combined with

information filtering (Oard 1997) based on techniques from linguistics (Tesitelová 1992) and the visualisation of information (Tufte 1983; Young 1996). Because of the magnitude of the task of surveying the Web, it was important to stratify the sampling, i.e. select a target population. The survey, therefore, focused on collecting and analysing data from Swedish Web sites within relatively few sectors, e.g., newspapers, companies registered on the stock-exchange and government agencies. This is not significantly different from surveying the physical world. Geographers, sociologists, economists and statisticians are also forced to stratify their areas of inquiry. The annual Swedish statistics report (SCB 1996) for example, only contains a tiny fraction of attributes measured, which in turn only represents an infinite fraction of the attributes measurable. This paper only attempts to investigate the publicly accessible part of the Web. It was not the intent to study aspects of Web sites which were not publicly available, such as, server activity logs and restricted access-areas. In the following we present and discuss research related to our apporach. Section 3 presents the measures applied in the survey and Section 4 presents and discusses the survey method in further detail.

A major area of research with regard to both the Internet and the Web concerns relationships between users and the technology. A number of surveys have investigated demographics, behaviour and attitude of Web users. Pitkow and Kehoe (1997) present an ongoing series of comprehensive demographic surveys of Web-use patterns, conducted by researchers at Georgia Tech. (www.gvu.gatech.edu /user_surveys/). Hoffman et al (1996) also present a study of the use of the Web. Electronic Commerce on the Web is continuously surveyed by the CommerceNet/Nielsen Internet Demographic Survey (CommerceNet 1996) from questionnaire data. Some researchers have applied methods measuring user behaviour when accessing the Web. Garofalakis et al (1999) show how to optimise the structure of Web sites based on surveys of user behaviour. Tauscher and Greenberg (1997) analyse how people revisit web pages with the intention of informing the design of better history mechanisms for browsers. This approach relates to research within user modelling where models of user preferences and behaviour can inform the design of new functionality (Allen 1990; Ambrosini et al 1997; Maglio and Barrett 1997). The approach adopted in this paper does not investigate the

relationships between users and the Web. Instead it analyses the Web as such, and attempts to derive lessons from looking at the hypertexts as language and structure.

Viewing the Web as a "world" of bits naturally raises the issue of space. In geometry, space is defined by two concepts: topology and metric. If we use the geometrical definition of space as a metaphor, the Web's topology can loosely be describes as a graph with nodes and uni-directional links. The nodes represent retrievable documents, i.e., files containing texts, images, links, and several other types of information. However, a plain distance metric does not capture the phenomenon accurately. Increased physical distance between the computers connected in the network does not necessarily lead to higher transaction costs. The metrical aspects can, however, be based on other variables than distance. Other researchers using the geometric metaphor considers the Web's metric to be calculations on how to traverse the graph formed by the link structure (Drew and Hendley 1995; Mukherjea and Foley 1995; Girardin 1996). Chakrabarti et al (1999) apply link clustering algorithms to determine authorities on topics based on the link topology.

Researchers have studied the Web in order to describe emergent properties. For example, Palmer and Eriksen (1999) and Eriksen et al (2000) study newspapers on the Web as new forms of news products. Smithson (1999) has produced a ranking of 100 web sites representing commercial organisations from 8 sectors with respect to the support for electronic commerce. Characteristic for this type of investigation of the Web is a qualitative approach based on manual navigation of the Web sites. This paper, however, explicitly subscribes to a quantitative approach based on semi-automatic techniques. Given the aim of measuring and monitoring changes to a relatively large number of web sites, the software agent approach was deemed more suitable for our purpose compared to the direct manipulation approach offered by the conventional Web browser (Schneiderman and Maes 1997). As argued by Nielsen (1999), the dramatic increase in the size and complexity of the World Wide Web will lead to new challenges for user access methods, and the application of semi-automatic filtering and retrieval techniques thus seems to be a viable approach.

A number of research efforts are concerned with the visualisation and mapping of both the Internet and the World Wide

Web in order to provide support for navigation (Dodge 1999). These address issues such as: maps of the Internet structure, Internet repositories and indices, statistics of Internet traffic and size, and visualisation of Web spaces. For example, Girardin (1996) and Drew and Hendley (1995) present visualisation of hyper-link structures. Barry and Batty (1994) analyse the diffusion of the Internet in order to predict future growth. Dodge (1996) applies a spatial metaphor to analyse the Web using Geographical Information System (GIS) technology. Young (1996) presents an extensive survey of 3D information visualisation research. The approach adopted in this paper uses information visualisation techniques to represent the results of the semi-automatic survey. Visualisation is a means rather than a goal in itself however.

Guan and Won (1999) promote keyword-based data mining of the Web using pre-defined patterns. This approach is suitable for identifying and extracting pre-defined patterns, and not appropriate for characterising the contents of a Web site without prior assumptions about the nature of the contents. Similarly, Kumar et al (1999) applies semi-automatic techniques for identifying communities of common interest based on Web contents. Stenmark (1999) applies a similar approach combined with mutual recommendation functionality (Oard 1997; Resnick and Varian 1997) within a company intranet.

Bray suggests collecting data and performing statistical analyses on volume and density measures of the Web (Bray 1996). The project, furthermore, looks at the relative link topology between Web sites. Bray applies software robots for data collection. This approach has a number of similarities to the approach we have applied. There are, however, also major differences. Bray's survey of the Web is based on the Open Text Index, November 1995, covering 1,5 million pages. The parameters analysed are, however, few and they are mainly volume and density measures, e.g., distribution of page sizes, number of embedded images, and types of file extensions. These are combined with structural measures such as a ranking of sites most often referred to, and other inter-site linking measures. The inter-linking measures are applied to illustrate proximity of sites through a spatial mapping.

Bray's approach and the one adopted in this paper both apply the Web site and page as the two basic sample units. It could be argued that defining the granularity of the survey based on site names is

biased. Surveying sites, results in focusing on institutionalised entities on the Web. One way of taking this into consideration is to calculate "links-to-site" sets. Bray, for example, calculates the rankings of most popular site referenced to in the pages. While Bray focuses on few and relatively simple measures for a large sample, we have chosen to measure more parameters, and to apply a deeper analysis of the contents of the pages.

3. Web Measures

The Web can be viewed as a body of text containing two fundamentally different types of data: the contents and the tags. A tag is in HTML (Hyper-Text Markup Language) meta-data describing the layout and linking structure between the text, graphics, audio and interactive components. In more advanced markup languages such as XML or SGML, the markup types can be defined by the user (Khare and Rifkin 1997; Lassila 1998; Rein 1998). Analyses of HTML hypertexts, therefore, concern both aforementioned types of data. This paper suggests the application of the five basic different types of measures presented below. These are drawn from the research reviewed above and from computational linguistics (Tesitelová 1992). Table 1 summarises the measures and provides examples of actual measures within each type. Most of the measures are based on compilations of frequency lists of words retrieved from the Web sites. A frequency list for a Web site can contain the tokens found, which are all the separated words, or it can be filtered further to only contain the types found, i.e., a list of unique tokens.

Volume measures count total numbers of constituents in the hypertext, such as, bytes, pages, link errors, tokens, types, headings, interactivity, internal and external links. The number of bytes and pages provide measures of the size of a site. The number of link errors reflects how well it is maintained. The total number of tokens and types provide contents-based volume metrics for a site. Interactivity is measured by counting forms, CGI-script and Java-applets. Measuring headings, external links (to other sites) and internal links (within the site) provide quantitative measures for "page-layouts."

Density measures relate volume measures to each other, making it possible to express more general site properties. Examples of density measures are: Bytes pr. page, average number of tokens pr. link error, and number of external links per page.

Vocabulary measures analyse site text vocabulary applying the computational linguistic measures: Stemming, Guiraud and theoretical vocabulary (Tesitelová 1992). Stemming is a technique, which classify common words according to common meaning, e.g., reading, read, reads. Guiraud is a measure reflecting vocabulary richness. It is calculated by dividing the number of types by the square rooted number of tokens. This measure does not incorporate the size of the text, and subsequently fails on both extremely small and large texts. Because of the large variations in the size of Web sites we have used theoretical vocabulary as a complement to Guiraud. Theoretical vocabulary is not sensitive to the text size, but because it is computed based on a frequency list of types, it is computationally more complex than Guiraud. Theoretical vocabulary reflects the expected number of types if the tokens are reduced. The measure is calculated as follows: Assuming that a text containing N number of tokens is reduced to M number of tokens. Let V be the number of word types. The possibility that all occurrences of a particular word type is eliminated in a reduction is $(M/N)^i$. If TN is the original number of types, the theoretical vocabulary will be (TM) (see Figure 1). We reduced the number of types (M) to ten thousand. Both Guiraud and theoretical vocabulary values increases when the vocabulary gets richer.

$$T_M = T_N - \sum_V V_i \left(\frac{M}{N} \right)^i$$

Figure 1: The theoretical vocabulary formula.

Structural measures provide quantitative measures representing the spatial property distance. Two structural measures are applied in this paper, directory structure for the Web site and mean distance of the hypertext. The former is relatively straightforward, and the latter reflects whether the site link-structure is deep or flat by giving the average number of the smallest amount of clicks on links needed to get from the start page to any other page.

Relative measures compile various differences between sites. We have used lexical equality of frequency lists as a relative measure to detect if two sites used the same type of language. This approach does not take into consideration where in the text a particular word appeared. Lexical equality can be calculated from frequency lists of tokens or of types. The types-based method does, compared to the token-based approach, not consider the frequency of a particular word. The token-based method can potentially result in a bias towards non-context carrying words. Context-carrying words often have low frequencies. They are often nouns or verbs and explains more about the texts than highly frequent words such as *and*, *or* and *I* can do. Lexical equality is expressed as a percentage of equality between two frequency lists. Calculating the lexical equality of every combination of a group of sites results in a matrix representing the lexical distance map. The values in the matrix can be visualised using a clustering algorithm (Jain and Dubes 1988). This yields a two-dimensionally representation of the relative distance between the frequency lists according to the clustering metric. Hagman and Ljungberg (1995) have demonstrated how clustering of lexical equality can be applied to compare conventional newspaper articles. Clustering provides a semi-automatic means for identifying patterns across Web sites. Olsen et al (1993) have used clustering of keywords characterising individual documents to obtain an overview of a hypertext document collection.

Measure	Description	Examples
Volume	Count absolute numbers of hypertext atoms (e.g. the tags and the text). This constitutes all raw data collected from which the remaining measures are calculated.	Number of separated words (tokens) or different words (types) within a site
Density	Density calculations based on the volume measures.	Number of errors pr. page. The standard deviation for tokens pr. page.
Vocabulary	Identifies the richness of the used vocabulary.	Guiraud or theoretical vocabulary

Structure	Attempts to measure the site hierarchy, depth and width of the link tree.	The average number of clicks on internal links needed to get from the start page to any other page.
Relative	Compare different data sets.	Lexical equality measure identifies whether two texts deal with the same topic, or have a similar content.

Table 1: The five types of measures from which the actual ones applied in the paper is drawn. For each type, one or more examples are given.

4. Survey Setting

In the survey we applied semi-autonomous Web robots for retrieval of Web site contents (Leonard 1997; Schneiderman and Maes 1997; Schubert et al 1998; Wooldridge and Jennings 1999). The process of bringing about the Web robot for the survey can be characterised in terms of the following nine activities: (1) specification of Web measures to be calculated; (2) design and construction of software robots; (3) small-scale tests of robots; (4) selection of robot for survey; (5) Web site selection; (6) data collection; (7) data aggregation; (8) data analysis; and (9) documentation of results.

A (Web) robot is a program that automatically traverses the Web's hypertext structure by retrieving a document, and recursively retrieving all documents referenced. The term "recursively" does not limit the definition to any specific traversal algorithm. The robot can apply various heuristic algorithms to the selection and ordering of documents to be retrieved. A Web browser is normally not in itself considered a Web robot since it is operated by a human user and does not automatically retrieve referenced documents (Sørensen 1998). If robots do not contain rules stipulating when to terminate the retrieval of documents, they might attempt to retrieve all the public pages on the Web. The termination criterion can be defined relative to a certain link structure depth, or be based on a predefined number of retrieved documents. The criteria applied in our experiments are defined relative to the public pages within a given site or domain. Web robots or Web agents are also frequently referred to as "Web Wanderers," "Web

Crawlers," or "Web Spiders." These names are however, misleading as they give the impression that the software itself moves between Web servers. Since the research reported here did not use the relatively rare mobile agent technology (Huhns and Singh 1997; Schubert et al 1998) the Web robot simply visited sites by requesting documents from them. This technique is similar to the one used by search engines such as Altavista (www.altavista.com) and Hotbot (www.hotbot.com) for indexing the Web.

The survey was conducted on 82 Swedish Web sites. All of the sites surveyed were found in the Swedish University Network (SUNET) link collection. Although all sites were Swedish with server address within the ".se"-domain, this was no guarantee for the Web server physically being located in Sweden. For example, the server *www.ericsson.se* was physically located in the Netherlands. We did not have resources to investigate the entire Swedish Web, and we intended to use our contextual knowledge about the selected sub-strata during the analysis. In order to survey a cross section of Web sites representing both public and private organisations, we chose sites from the six sectors listed in Table 2:

Sector	Count	%
A-List, i.e., companies on Swedish stock-exchange	24	29,3
Municipalities	11	13,4
Newspapers	8	9,76
Political parties and interest groups	13	15,9
Government agencies	19	23,2
TV- and radio stations	7	8,54

Table 2: Frequencies and percentages of the 82 sites analysed.

Three different robot prototypes were constructed. The first one was an extension of the maintenance robot MOMSpider (Fielding 1994), implemented in perl and used for validating links and generating statistics. Due to performance problems with perl, a second robot was developed in C++. During the development of the second robot we came across ht://Dig (available at URL http://htdig.sdsu.edu/) implemented in C++. It is constructed to index local networks, such as Intranets, but with some adjustments it served our purpose perfectly. All the data documented in this paper is collected by this robot. Using an existing agent architecture is highly beneficial since developing a reliable

architecture is a substantial effort (Wooldridge and Jennings 1999). The robot conformed to the de facto ethical standard for Web robots (Eichmann 1994; Koster 1997). This entails that the robot did not squire resources from human users by retrieving pages at high speed. It also ensured that the robot identified itself to the Web server allowing the Webmaster to contact the owner of the robot if problems should occur.

Table 3 shows key sampling data on the total amount of hypertext retrieved. It also provides information on size, download time, number of tokens and types, and the calculation time for the frequency lists for both the largest and the smallest Web site.

All sites	310 mega-bytes uncompressed hypertext. 21 mega-byte URL-lists
Largest site and frequency list	Ericsson, Oct 14. 17,500 kilobytes hypertext downloaded in 10 hours. 2000 kilobytes frequency list. 2,489,999 tokens and 163,636 types calculated in 3500 seconds (with an optimised C-program)
Smallest site and frequency list	Dagens Industri, Oct 14. 178 kilobytes hypertext downloaded in 12 minutes. 17 kilobytes frequency list. 11,014 tokens and 1,948 types calculated in 7 seconds (with an optimised C-program)

Table 3: Key sampling data on the total amount of hypertext sampled: the largest site and frequency list, as well as the smallest site and frequency list.

In order to reach a sufficient depth of analysis, we chose to focus on one of the sectors surveyed, the newspaper Web sites. In general, they changed more frequently compared to sites in other sectors. As an example, there were no changes during the sample period to any of the A-list companies' Web sites. The newspapers sites were collected on five different occasions in 1996, namely, September 23rd, September 30th, October 14th, October 29th and November 4th,. The newspapers are listed in Table 4. The data aggregation was conducted with a variety of small programs implemented in several different languages, such as, C, perl and awk. Part of the data aggregation process was automated by perl scripts "glueing" the various programs together. Standard statistical packages (DataDesk and Microsoft Excel) were used for calculations and hypotheses testing. We have also used data clustering (Jain and Dubes 1988) to visualise relative results in order to establish patterns in the data material.

Newspapers	Description	Launch Date
Aftonbladet	National evening paper	August 1994
Arbetet Nyheterna	Regional morning paper	March 1996
Dagens Industri	National business daily	June 1995
Göteborgs Posten	Regional morning paper	August 1995
Hallandsposten	Regional morning paper	September 1995
Nerikes Allehanda	Regional morning paper	May 1995
Sydsvenska Dagbladet	Regional morning paper	August 1995
Svenska Dagbladet	National morning paper	June 1995

Table 4: The Swedish newspaper Web sites analysed, with a indication of the month when the Web service had been launched.

5. Results

The extent of the survey was such that only parts of the data analysis can be presented. We have chosen to focus on the following five aspects: (1) Summarising the contents of a site; (2) Identifying topics presented at a site; (3) Mapping the changes to a site structure; (4) Comparing the contents of several sites; and (5) Surveying a messy world. To illustrate aspects (1), (2), and (3), we show the results from analysing the two newspaper sites Göteborgs Posten and Sydsvenska Dagbladet. To illustrate the type of analysis performed in (1) and (4), results from comparing the eight newspaper Web sites are shown.

5.1 Summarising Site Contents

How can the results of a survey of a site be summarised? We have chosen to present key-data in the form of tables. Table 5 shows the types of information presented in Table 6 for Göteborgs Posten and Table 7 for Sydsvenska Dagbladet. Göteborgs Posten (GP) is geographically located in Göteborg (Gothenburg) and operates in the western part of Sweden. GP was the second largest morning newspaper in Scandinavia with an average circulation of 273,000 on weekdays and 306,000 on Sundays. These figures have not changed substantially since 1996. Sydsvenska Dagbladet (SD) is also a regional

morning paper. Both GP and SD initiated a Web site in August 1995. Since there are only five observations from the sites, we can only perform a tentative qualitative analysis of the data. Furthermore, some of the variables did not change much during the sample-period. Those variables are bytes, pages, types, links, bytes pr. page, types pr. page, links pr. page and theoretical vocabulary.

Measures	Explanation
Bytes	Number of bytes at the site
Pages	Number of pages at the site
Tokens	Number to tokens, i.e., separated words
Types	Number of types, i.e., number of unique words
Error	Number of link errors
Links	Number of links
Headings	Number of html document headings
Gif imgs	Number of gif images on the site
Jpeg imgs	Number of jpeg images at the site
Bytes/pg	Average number of bytes pr. page
Tokens/pg	Average number of tokens pr. page
Types/pg	Average number of types pr. page
Lnk Err/pg	Average number of link errors pr. page
Links/pg	Average number of links pr. page
Headin/pg	Average number of headings pr. page
Largest pg	The largest page encountered
Guiraud	*Guiraud* calculated for the site
Theo.Voc.	*Theoretical Vocabulary* calculated for site
Mean Dist	The average distance from the root to any other page

Table 5: The types of key data from samples of Swedish Web site.

We can appreciate the rate and extent of change of Sydsvenska Dagbladet's site when comparing the standard deviation between the size of the samples which turns out to be more than the total size of many of the other newspaper sites. Hallandsposten only had a maximum of 601k bytes, Dagens Industri had 176k bytes, and Arbetet 621k bytes.

The theoretical vocabulary gives a quantitative measure of the diversity of a text. In general we would expect a more diverse text to represent a more complex and diverse choice of language. The average

for all sites is 1819 and the top score is 2617 (The Royal Library). This makes the 2437 average for Göteborgs Posten, and 2405 for Sydsvenska Dagbladet quite high. The theoretical vocabulary increased slightly over time on the SD Web site and minor fluctuations could be detected on the GP site. Apart from changes to the vocabulary, no radical changes could be measured on the Sydsvenska Dagbladet's site. This is, of course not a reflection of lack of update of the information on the site. It merely informs us that the contents may have changed, but the type of contents have not. Furthermore, the relatively small changes may not be a significant predictor for substantial changes to the type of texts on the site.

The sample clearly shows that something happened to the GP site between sample II and IV (See Table 6). Firstly, the number of link errors were 5, 4 and 3 in the previous samples, and suddenly increased to 85 and 91 in sample IV and V. The mean distance also changed from 2.3 to 6.6. This indicated a major restructuring of the site transforming it from having a flat links-structure to a more deeply one. The site also had obtained nine interactive forms in sample IV from having no forms in sample III. There had also been a complete change to the sites set of outgoing links. The two most popular external links in sample IV were *www.realaudio.com* and *www.netscape.com*, which occurred twenty times each. These links did not occur at all in any of the previous observations. In sample I-III the most popular links were *www.westnet.com* and *www.sunet.se* and they were used about 25 times each. This clearly indicated an overall change of the site's page and link-structure layout.

Measure	I	II	III	IV	V	Average	Std.dev
Bytes	5113k	5152k	5209k	5629k	5980k	5417k	376665
Pages	912	914	929	902	943	920	16
Tokens	502k	509k	512k	569k	606k	641k	45874
Types	58210	58661	59171	59102	61812	59917	1407
Errors	5	4	3	85	91	38	46
Links	9943	10083	10229	9719	10351	10065	247
Headings	1574	1541	1543	1494	1605	1551	41
GIF imga	144	117	136	177	179	150	27
JPEG imgs	11	14	13	29	31	20	10
Bytes/pg	5606	5637	5608	6241	6342	5887	371.16
Tokens/pg	1108	557	552	631	643	698	282.83
Types/pg	66.71	64.18	63.69	65.52	65.55	65.13	1.20
Lnk Err/pg	0.0055	0.0044	0.0032	0.0942	0.0965	0.0408	0.0498
Links/pg	10.90	11.03	11.01	10.77	10.98	10.94	0.011
Headin/pg	1.73	1.69	1.66	1.66	1.70	1.69	0.03
Largest pg	35kb	43kb	40kb	44kb	51kb	43kb	5918
Guiraud	58.0	58.1	58.5	55.4	56.1	54.2	1.4
Theo.Voc.	2455	2454	2458	2409	2410	2437	25
Mean Dist	2.3	2.3	2.3	6.6	6.7	4.0	2.4

Table 6: Key data from the five samples of the Göteborgs Posten Web site.

Although much weaker, SD also showed a change between sample III and IV, with increases in both the size of the site and in number of pages. Here the mean distance, however, remained virtually unchanged. In both GP and SD the tokens, types, Guiraud, and theoretical vocabulary showed that the types of texts did not change substantially. As an example, SD had an increase in tokens of around 9.5% over the period. This might not seem much, but the sample period was only 7 weeks, which roughly translates to 70% increase per year. It is not unrealistic to assume a steady growth, since the Web site was started in August 1995. Göteborgs Postens Web site has experienced a growth from an estimated access of 3000 people/day in 1996 to around 27,000 people/day in 1999 and has expanded dramatically in both size and type of material published (Eriksen et al 2000).

Overview information as presented in this section, can, amongst others, provide indications of major redesign of a Web site, and

perhaps also help pointing at what types of changes have been implemented. The overview information can also support cross-site analyses of relative technological advancement, in terms of, for example, comparing the use of interactive components, automatically generated pages, and linking structure. If users accessing Web sites are interested in notification of major changes, and if a Web site is indexed regularly, triggers could be configured, notifying about changes exceeding the defined threshold (Senanayake 1998). In a world of constant change, it could be more valuable to only be informed about significant changes and not only of any change occurring.

Measure	I	II	III	IV	V	Average	Std.dev
Bytes	12887k	13130k	13854k	14316k	14437k	13725k	694378
Pages	1327	1353	1438	1520	1543	1436	96.5
Tokens	1550k	1574k	1642k	1700k	1713k	1636k	73066
Types	124015	125301	128940	131420	132127	128361	3710
Errors	21	18	17	22	21	20	2.17
Links	11052	11271	13095	13852	14042	12662	1417
Headings	1855	1876	1820	1930	1950	1886	53.5
GIF imga	321	305	288	300	307	304	12
JPEG imgs	162	159	178	191	198	178	17
Bytes/pg	9711	9704	9634	9418	9356	9565	166
Tokens/pg	1168	1163	1142	1118	1110	1140	25.89
Types/pg	93.46	92.61	89.67	86.46	85.63	89.56	3.52
Lnk Err/pg	0.0158	0.0133	0.0118	0.0145	0.0136	0.0138	0.0015
Links/pg	8.33	8.33	9.11	9.11	9.10	8.80	0.43
Headin/pg	1.40	1.39	1.27	1.27	1.26	1.32	0.07
Largest pg	426kb	426kb	426kb	426kb	426kb	426kb	0
Guiraud	70.4	70.6	71.1	71.3	71.4	71.0	0.4
Theo.Voc.	2383	2388	2409	2421	2422	2405	18
Mean Dist	4.5	4.4	4.5	4.6	4.6	4.5	0.08

Table 7: All key data from the five samples of the Sydsvenska Dagbladet Web site.

5.2 Identifying Site Topics

Imagine that arriving at a Web site, you could be presented with a pre-prepared list of topics covered by the Web site. As a first step in conducting a more in-depth analysis of the contents of Web sites, we chose to apply a semi-automatic technique for identifying Web site topics. The technique is based on the frequency list of types, i.e. different tokens, for the site. The frequency list is filtered using a stop-list of non-context bearing words such as; *are*, *is*, and *it*. Subsequently a word-stemming algorithm is applied in order to find syntactically related words and dividem into word categories. This, for example, results in the words: *Göteborgs, Göteborg, Göteborgare, Göteborgarna,* and *Göteborgarnas*, from Göteborgs Posten's Web site being classified as one category (See Table 8). The last step in the process is to sort the list of word categories according to occurrence. This technique is also applied manually at newspapers, amongst others the British newspaper The Guardian, to highlight the core contents of political speeches. Here, a list of most common word categories in political speeches of political opponents are compared, to bring forward a deeper understanding of differences in opinion between the political parties. Table 8 presents the most frequent word categories from Göteborgs Posten and Sydsvenska Dagbladet based on the sample September 23rd. Although both Web sites are Swedish regional newspapers, the contents of the Web sites differed significantly. The GP Web site clearly still had a substantial Web coverage of the 1996 Atlanta Olympic Games: 3: Os, 7: USA, 18: Final, and 19: Atlanta. It also devoted substantial amount of text to the topic of films (1227 occurrences of words in this category). Sydsvenska Dagbladet, on the other hand, did not have any of these two topics represented on the list. SD's Web site was mainly devoted to information for a Swedish-Danish investor club, helping people locating appropriate shares. This can, amongst others, be seen from the large proportion of Swedish and Danish company names on the list: 2: Plm, 5: Trelleborg, 7: Skanska, 8: Danisco, 9: Astra, 10: Ericsson, 13: Pharmacia, 14: Carlsberg, 16: Oticon, and 19: Althin.

Words associated with the publishing institution itself or the geographical area in which it is located, were also quite frequent: (A) *Göteborgs Posten*; 1: Article, 2: Gp, 4: Göteborg, 5: Posten, 6: Copyright. (B) *Sydsvenska Dagbladet*; 1: Malmö, 6: Lund, 11: Getinge, and 12:

Sydsvenskan. An initial analysis seems to indicate that Göteborgs Posten was using a language more site-centric than that of Sydsvenska Dagbladet, relative to the rest of the text.

Göteborgs Posten		Sydsvenska Dagbladet	
Freq	Topics and Synonyms	Freq	Topics and Synonyms
2387	Article/Artikel: Artikeln	14035	Malmö: Malmoe, Malmös
2127	Göteborgs Posten/Gp	11744	Plm (Swedish company)
1840	Olympic Games /Os	10177	Novo: Novos (Danish company)
1738	Gothenburg/Göteborg: Göteborgs, Göteborgare, Göteborgarna, Göteborgarnas	10173	Nordic/Nordisk: Nordiska, Nordiskt, Nordiske
1584	Mail/Posten: Post, Postade, Postar, Poste, Postens, Poster, Posterna, Postum	9910	Trelleborg: Trelleborgen, Trelleborgs (Swedish company)
1478	Copyright	9422	Lund: Lunden, Lunds (Swedish City)
1282	USA	8710	Skanska: Skanskas (Swedish company)
1259	Year/År	8528	Danisco: Daniscos (Danish company)
1227	Film: Filma, Filmad, Filmade , Filmar, Filmare, Filmares, Filmarna, Filmas, Filmat, Filmats, Filmen, Filmens, Filmer, Filmerna, Filmernas	7087	Astra (Swedish company)
1196	Page index/Sidindex	6949	Ericsson: Ericson, Erikson, Erickson, Ericssons, Eriksson (Swedish company)
1192	Previous/Föreg	6908	Getinge: Getingar, Getingarna, Getingen (The stemming did not recognise the difference between the name of a geographic place (Getinge) and the Swedish word for wasp (geting))
1157	Copy	6520	Sydsvenskan: Sydsvenska, Sydsvenskans, Sydsvensskans (Synonymous for the newspaper)
982	Sweden/Sverige: Sveriga Sveriges	5874	Pharmacia (Swedish company)

624	Day/Dag	5791	Carlsberg: Carlsbergs (Danish company)
623	Swedish/Svenska: Svensk, Svenskar, Svenskarna, Svenskarnas, Svenskars	5532	Kr (Krona is Swedish and Danish currency)
509	Germany/Tyskland: Tysklands	5102	Oticon: Oticons (Danish company)
506	Russia/Ryssland: Rysslands	5048	Registration/reg
500	Final: Finalen, Finalens, Finaler, Finalerna, Finale	4671	Tele: Talet, Talar, Tala, Talade, Talades, Talan, Talare, Talarna, Talas (The stemming did not recognise the difference between *tele* (as in telephone) and *tala* (the Swedish word for speech))
484	Atlanta: Atlantas	4651	Althin (Swedish company)
		4648	Denmark (abbrev)/Danm: Danma, Danmar

Table 8: Table of the most common words at the Götebortgs Posten and Sydsvenska Dagbladet Web sites on September 23rd, after filtering and stemming of the frequency lists.

This semi-automatic technique of identifying Web site topics is an example of filtering and summarisation, i.e., eliminating unwanted information and representing a large amount of information with a smaller one (Nielsen 1999). Interesting and quick overviews could be provided by the Web site if techniques such as the ones demonstrated here were applied as a service to visitors. In a more automated version, it could also provide important high-level information about trends in the development of a Web site over time by comparing lists of most common topics on a regular basis.

5.3 Mapping Site Structure

How can a measure provide an overview of the structural aspects of a Web site? One way is, of course, to compile a graph representing the inter-document linking. Such an approach would provide a very detailed model, but also one where advanced technology for information visualisation would be necessary in order to represent the complexity of the information space (Young 1996). The purpose of such a model would, further more, be to represent the full complexity of the

hyperspace in order to, for example, support navigation through the use of sophisticated virtual reality technology. The use of this type of information visualisation techniques, where the full complexity of the structure is conserved, can automatically lead to information overload, as can be seen from Figure 2, where we have shown the link structure of the informatics.gu.se Web site at Göteborg University. Projects such as Apple's Project-X have made attempts to alleviate this problem without reducing the amount of detailed information available (Young 1996).

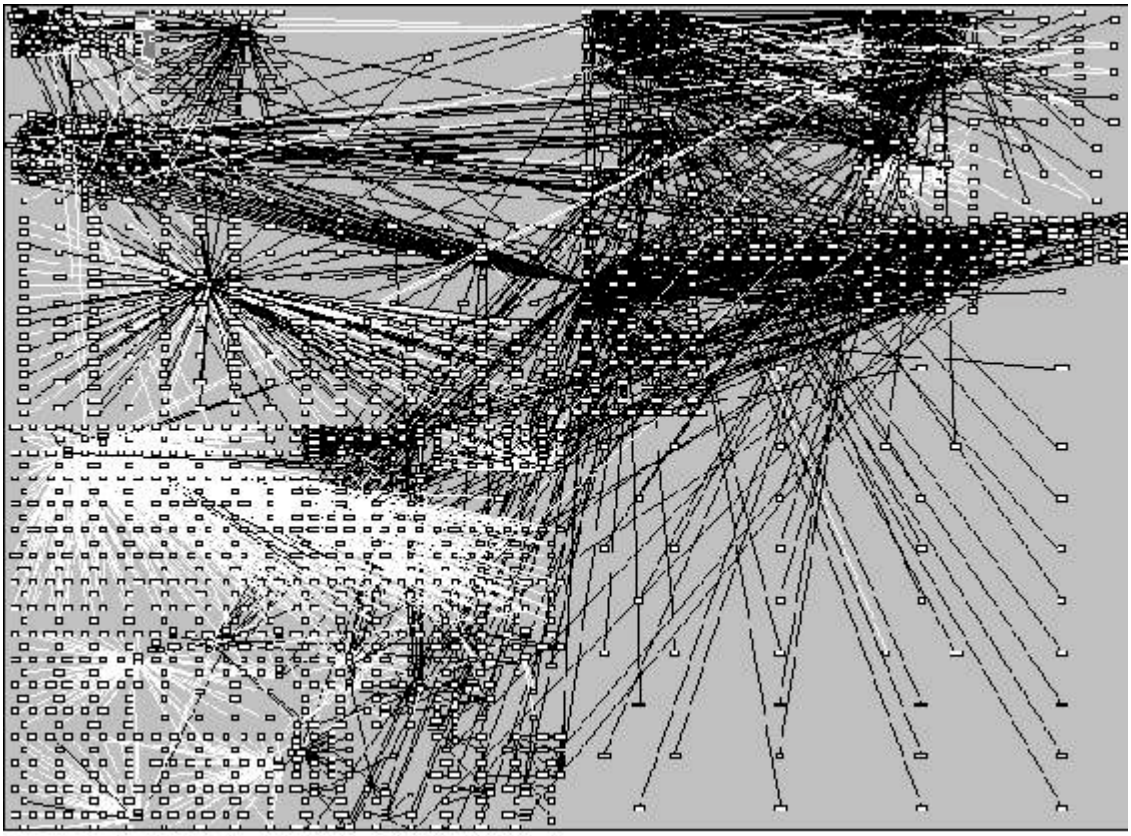

Figure 2: The link structure of the informatics.gu.se site at Department for Informatics, Göteborg University.

How can we present a picture which does not so easily lead to information overload, but instead provides an overview where information on purpose has been discarded in order to provide the overview? Chakrabarti et al (1999) analyse link structures to establish hubs of authority. Li (1998) uses a similar technique as the basis for constructing a search engine index. If, however, the aim is to detect and visualise change to a Web site, then the exact structure of the Web directory is not of the utmost importance. It is also important to be able

to reflect changes. One first reduction of complexity is to look at the directory structure in which the html files are stored. This, of course, implies that the data is analysed as a tree structure, and not as a graph. For large sites, this, however, can also lead to problems of information overload. We have, therefore, adopted the CyberGeo maps concept as a visual mechanism for providing overview of Web site structure (Holmquist et al 1998). CyberGeo maps provide a means for monitoring changes to the names and dates of files and directories by visualising the site's directory and document structure.

Figure 3 shows five consecutive CyberGeo maps for Göteborgs Posten. On a given map, each of the many discs represents a document. Applying a simple hash function on the document file name each disc is placed in a circle around the centre. The size of a document is represented by the size of the disc. First time a document is encountered, it is represented as a white disc. If a particular document has not been changed or deleted since previous measurement, the disc representing the document will assume a slightly darker shade of grey. Documents, which are unchanged will gradually, change from white to black as each measurement is made. The innermost of the concentric circles is the directory root. Documents in the same sub-directory are placed in an outer circle within ten degrees of its parent directory. This approach to presenting a high-level view of the Web site obviously only provides a simplified view. The mapping technique does not take into consideration that a Web site potentially is a complex graph with links across and within documents (see Figure 2), and potentially with documents only serving as aliases or references to other documents somewhere else in the tree structure. It assumes that it makes sense to view the Web site as a tree structure.

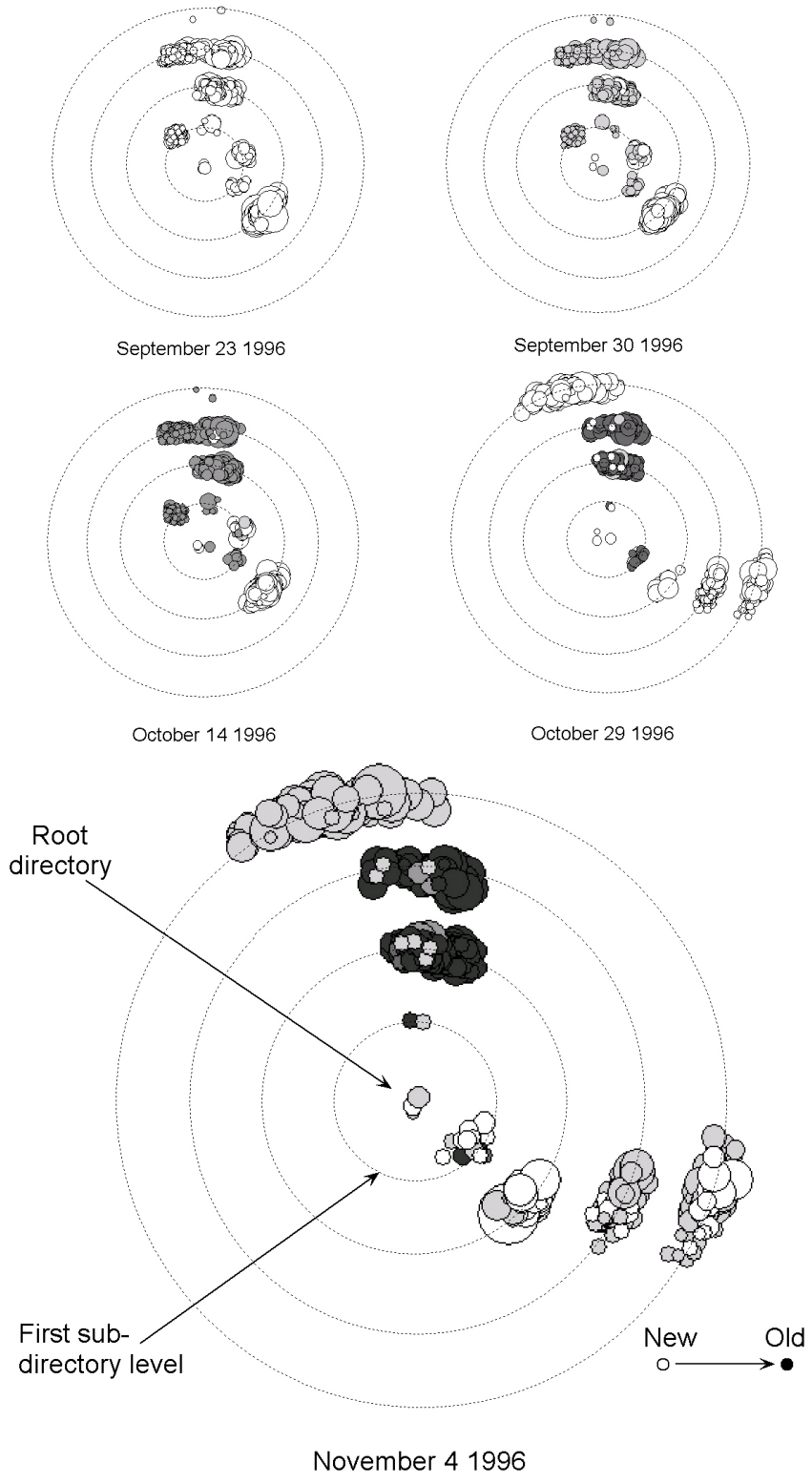

Figure 3: Five CyberGeo maps of Göteborgs Posten showing the changes in directories for the Web site from September 23rd to November 4th 1996. New files and directories, i.e., with changed name and/or time stamp are shown as white circles. Files with name and time stamp which are identical to the previous search will for each time be drawn in a darker shade.

The reorganisation of the Göteborgs Posten site mentioned in Section 5.1 between sample II and IV can very easily be detected from the CyberGeo map sequence in Figure 3. However, Figure 4 illustrates even more clearly how a re-organisation of a Web site will be visualised by the CyberGeo map. Here, two consecutive CyberGeo map pictures for the Omgroup Web site (www.omgroup.com) illustrate a major redesign, where the document representations are almost turned 90° clockwise as a result of almost all documents being renamed.

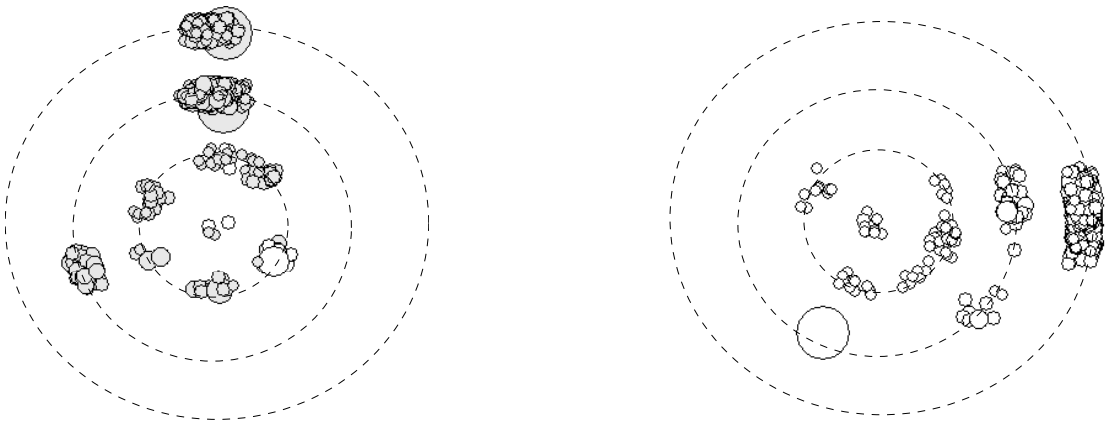

Figure 4: Example of a CyberGeo map shift due to major revamp of Web site. The radical change in the CyberGeo map of this site represents the fact that the Web site was revamped, amongst others resulting in most directories and files being renamed. The large area on the second circle on the right map figure represents the company's annual report.

5.4 Comparing Several Sites

Having analysed how we can summarise the contents and topics of a Web site, amongst others, supporting the comparison of two sites, we now turn to the issue of comparing a number of Web sites to each other with respect to the language used. At this level, all of the eight newspaper sites were analysed in relation to each other by a cluster analysis on lexical equality of both types (see Figure 5), and tokens. We did not find significant differences between clustering tokens and types, so only the type-based approach is shown. Based on the list of types, the clustering algorithm will group together observations based on the degree of commonality between the observation and all other

observations. The clustering algorithm visualises the percentages of lexical equality between sites in two-dimensions. There are no axes in the figure, only relations, e.g. the upper right corner is the least equal to the lower left corner. Each sample is represented as a plot, and each of the sites were sampled five times, except Svenska Dagbladet where the webmaster blocked out our robot using the robots exclusion standard (Koster 1997) during the last two samples.

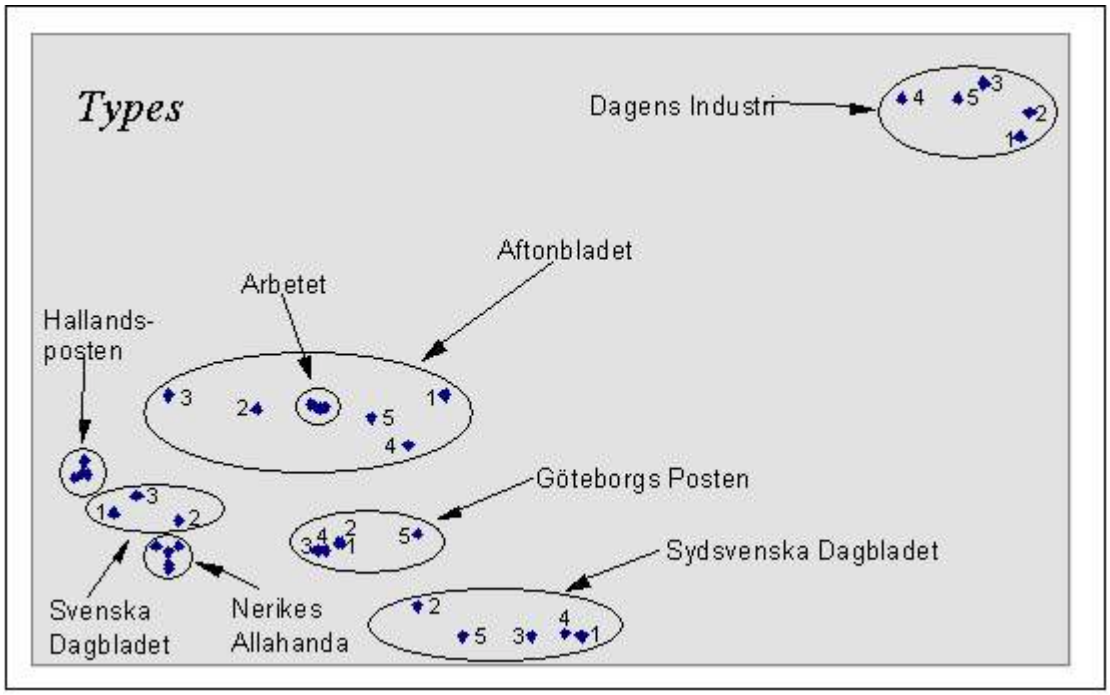

Figure 5: Clustering on lexical quality of types for the newspapers at five occasions.

As seen in Figure 5, none of the sites changed their language significantly during the period, and the samples are therefore within a small region in the clustering. In particular, the language on the Web sites for Arbetet, Hallandsposten, and Nerikes Allehanda changed very little over the period, which is not surprising since none of the other variables collected change significantly either. The variation in tokens and types within one particular site was in most cases less than the variation between sites, making it possible to represent each site as a region (See Figure 5). As clearly demonstrated by Figure 5, Dagens Industri was quite different from the others Web sites. The reason for this is most likely that it is a financial newspaper that, due to the limited scope, used a different language compared to the others. Arbetet and Aftonbladet are both associated with the Swedish Social

Democratic Party. Although only the editorial in Aftonbladet has a distinct political flavour, it seems as though the language of the two newspapers is much the same. These two sites were quite similar in terms of tokens used, but as Figure 5 shows an even greater similarity between the two when clustering types, i.e. comparing the two lists of distinct words. Techniques such as clustering frequency lists of tokens or types can provide a valuable overview of the relative similarities between a number of Web sites.

5.5 Surveying A Messy World

One of the main problems in surveying the World Wide Web is that the HTML “programming” environment does not enforce strong typing. If a particular statement is incorrect, the Web browser will just skip it, making it difficult for the author to ensure correct markup. This, amongst others, results in an extremely “messy” environment to survey. Table 7 shows to the left the set of correct HTML tags at the time of the survey, and to the right examples of markup encountered during the sampling. The table does not indicate any relationship between the proportion of correct and incorrect tags. It is basically the list of tag types divided into correct and incorrect ones. With the emergence of authoring technologies supporting the automatic generation of HTML documents, this problem could be expected to diminish over time.

Correct HTML Tags			Incorrect HTML Tags		
a	Em	option	abcsubmit	gave	ps
address	Font	p	adress	grin	rfpicr
align	Form	param	ahref	hrnosha	shift
applet	Frame	pre	aired	de	silence
area	Frameset	right	aircheck	htm	stations
b	h[1..6]	script	and	http	svd
base	head	select	are	il	tdalign
basefont	hr	small	bgsound	imgsrc	tdcolspan
blink	html	strong	bodybgcol	inputtyp	tdnowrap
blockquot	i	sub	or	e	tdwidth
e	img	table	border	is	textareaw
body	input	td	brt	it	rap
br	left	textarea	by	jberg	thank
caption	li	th	ceneter	jpd	tsfinfo
center	kbd	title	centre	krone	valign
cite	link	tr	clear	m	vk
code	map	tt	color	marquee	w
dd	menu	u	colw	means	wbr
dir	meta	ul	embed	moore	were
div	noframes	width	emp	name	width
dl	ol		front	nobr	wireless
dt			fontsize	of	www
				palign	

Table 7: Listing of the HTML tags found during the survey, distinguishing between the correct and incorrect ones.

6. Discussion

This paper discusses important lessons we can learn about the contents of and changes to the fastest growing technology in the world, the World Wide Web, through employing semi-automatic retrieval and aggregation of Web sites. The paper investigated this question by presenting and discussing results from an experimental survey where Web robots supported the collection and analysis of 82 Swedish Web sites over a seven-week time-span from September 23 to November 4

1996. It should be noted here that during the course of the experiment a couple of the sites banned robots from accessing data.

We have demonstrated how changes to the contents of Web sites can be detected through analysing changes to the basic parameters, such as number of bytes or link errors. CyberGeo maps provided a technique for mapping and monitoring changes to the Web site architecture. Clustering lexical equality of tokens and types across the 8 Swedish newspaper sites and representing the result in a two dimensional plot illustrated linguistic similarities and differences across Web sites. The results did not only demonstrate the sampling instruments' ability to detect changes, but also that a closer calibration of the instrument must be conducted over a longer time-span than seven weeks in order to amplify the sensitivity to significant changes. For example, the samples showed an increase in number of tokens of 9.5% over the seven weeks. If scaled to one year this amounts to around 70%. We recognise that this is a significant change, but compared to what? In general, a sampling period of only seven weeks most likely proved to be too short for obtaining results showing large variations.

In general, compiling frequency lists has provided us with much deeper material about the contents of the Web sites, compared to the analysis conducted by Bray (1996). This, however, is associated with a much smaller sample. Common readability formulas such as Coleman-Liau grade level and Bormuth grade level could be used to further analyse the sampled texts. These indexes determine a readability grade level, based on characters per word and words per sentences and are therefore relatively easy to calculate. Word processors, such as, Microsoft Word uses these types of indexes in their grammar-checking facilities. Also, a further analysis of differences and similarities between the language used in newspaper Web sites compared to the printed papers could further enlighten the discussion of new media (Tesitelová 1992; Hagman and Ljungberg 1995; Palmer and Eriksen 1999; Eriksen et al 2000).

In order to detect statistically significant differences between Web sites within the six sectors, we conducted a number of discrete statistical analyses using contingency tables and Chi-Square tests. This analysis was based on the measures showed in Section 3, and showed some relative trivial results, such as a dependency between

site size and number of link errors. This analysis revealed a number of dependencies between, for example, the size of a site and the number of link errors. It did, however, not show any further results in terms of significant differences between sites from different sectors. The reason for this could be the very short sampling period, and the absence of substantial contextual parameters which we could not establish without a more qualitative approach.

What are the possible implications of this research? All of the techniques employed can be operationalised further and made subject for increased automation, and thus provide support for increasingly automatically generated sophisticated maps of the Web. As the Web increases in size and complexity, there will be an increased need for supporting overview and navigation. The unstructured nature of the Web, which has been identified as one of its primary assets, is however, recognised as beginning to cause severe navigation problems. There is no centrally maintained classification scheme guiding people when they wish to access information. The Web can be navigated using a bottom-up search through one of the Web search engines (Type 1) or top-down through index pages providing taxonomies for Web contents (Type 2) (Gudivada et al 1997). Because of the dynamic nature of the Web, there is considerable overhead associated with locating and maintaining links to desired resources on the Web. The more extensively an individual uses the Web, the greater this overhead will be. Software agents have been promoted as a means of supporting navigation (Maes 1994; Krulwich 1997; Maglio and Barrett 1997; Schneiderman and Maes 1997). Current efforts suggest semantic markups supporting specific markup types characterising Web contents (Khare and Rifkin 1997; Wired 1997; Lassila 1998; Rein 1998). This structuration of the Web would support standardisation of text types within domains and supporting increased automatic analysis of Web page contents, for example within electronic commerce (Glushko et al 1999).

Currently, the issues of electronic commerce and the management of knowledge are at the fore of both research and public debate. The electronic market is increasingly becoming associated with the Internet infrastructure and the Web most often serves as the preferred front end. Companies are finding the Web a productive place to market products, and it is important to understand how the

company Web site can play an important role (Smithson 1999). For potential customers, the ability to easily obtain substantive information about the contents of a Web site is considered as important as being able to locate a shop on the high-street. Similarly, if for example, a car company wishes to know where to market its most recent model, it would be extremely useful to know where cars are intensively discussed and where they are not. Topic spotting on selected Web sites could provide an index to the contents of the site. This is why software agents have been suggested as a means of mediating in an electronic market (Glushko et al 1999; Maes et al 1999)

The management of knowledge has also frequently been linked to Internet and Web based systems supporting people in sharing knowledge and interacting (Scarbrough et al 1999). If people in organisations stores textual traces of their activities in intranets, there will be an inherent need for semi-automatic techniques that provide an overview of and access to this information (Stenmark 1999; Robertson et al 2000).

In both these cases, there will be substantial challenges involved in creating patterns and structure from unstructured bodies of text with the purpose of answering questions or connecting people. Data mining structured databases involves very complex calculations, and data mining the unstructured textual traces on the Web is an even greater challenge (Etzioni 1996; Kawano and Hasegawa 1998; Chakrabarti et al 1999; Guan and Wong 1999). We believe that one viable strategy is to apply relatively simple techniques for abstracting databased on pragmatic heuristics. Nielsen (1999) argues for the need for functionality supporting: (1) Aggregation - showing a single unit that represents a collection of smaller ones; (2) Summarisation - representing a large amount of information with a smaller one; (3) Filtering by eliminating unwanted information; and (4) Elision which is example-based representation. The techniques demonstrated in this paper provides examples of all of these categories, based on techniques from computational linguistics and information visualisation with filtering being the central technique. For example, the CyberGeo maps summarises an entire complex Web site in a relatively simple image and indicates if substantial changes has been made.

The results and approach presented in this article have provided one perspective with regard to the discussion of how to conduct quantitative surveys of the Web. It is because the Web is a highly dynamic and interactive information space, we must apply state-of-the-art computational power to study it. This will increasingly present itself as a challenge, and so far it has mainly been an issue for the professional IT community and not one carried out in a wider context. We are, however, increasingly living in Cyberspace, and will increasingly have to rely on high-level representations in order to make sense of the world and navigate it. Only time will tell to what degree most people will accept an agent-based human-computer interaction paradigm as opposed to a more conventional direct manipulation paradigm (Schneiderman and Maes 1997)

Acknowledgements

This research is partly funded by the Swedish Transport and Communications Research Board (Kommunikationsforsknings-beredningen) grants for the Internet Project (internet.informatik.gu.se). A big thanks to Leif Grönquist and Peter Ljungstrand for assistance on computational linguistics and to Michael Mandahl for bright ideas. Thanks to Maxine Robertson for extensive proof-reading. Thanks to the two anonymous reviewers providing us with valuable critique. Also thanks to all our other colleagues in the Internet Project for constructive comments.

7. References

Allen, R. B. (1990) “User models: theory, method, and practice,” *International Journal of Man-machine Studies*, 32 (5), pp. 511-543.

Ambrosini, L., V. Cirillo, and A. Micarelli (1997) “A Hybrid Architecture for User-Adapted Information Filtering on the World Wide Web,” In *User Modelling: Proceedings of the Sixth International Conference*, pp. 59-61, Vienna, Austria: Springer Verlag.

Barry, B. and M. Batty (1994) “The Electronic Frontier - Exploring and mapping cyberspace,” *Futures*, 26 (24), pp. 699-712.

Berners-Lee, T., R. Calliau, A. Luotonen, H. F. Nielsen and A. Secret (1994) “The World-Wide Web,” *Communications of the ACM*, 37 (8), pp. 76-82.

Bray, T. (1996) “Measuring the Web,” In *Proceedings of the Fifth International World Wide Web Conference*, Paris, France, URL: http://www5conf.inria.fr/fich_html/papers/P9/Overview.html

Chakrabarti, S., B. E. Dom, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, and J. Kleinberg (1999) “Mining the Web’s Link Structure,” *IEEE Computer*, 32 (8), pp. 60–67.

CommerceNet (1996) CommerceNet Nielsen Internet Demographic. URL: http://www.commerce.net/nielsen/

Dodge, M. (1996) “Mapping the World Wide Web,” *GIS Europe*, 5 (9), pp. 22–24.

Dodge, M. (1999) The Geography of Cyberspace. URL: www.cybergeography.org

Drew, N. S. and R. J. Hendley (1995) “Visualising Complex Interacting Systems,” In *Proceedings of the Conference on Human Factors in Computing Systems*, Denver, CO: ACM Press.

Eichmann, D. (1994) “Ethical Web Agents,” In *Proceedings of the Second International World Wide Web Conference*, Chicago, USA, URL: http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Agents/eichmann.ethical/eichmann.html

Eriksen, L. B., C. Ihlström, and J. Palmer (2000) “News: From Product to Service,” In *Planet Internet*, K. Braa et al (eds.), pp. 87-112, Lund: Studentlitteratur.

Etzioni, O. (1996) “The World Wide Web: Quagmire or Goldmine?” *Communications of the ACM,* 39 (11), pp. 65-68.

Fielding, R. T. (1994) “Maintaining Distributed Hypertext Infostructures: Welcome to MOMspider’s Web,” In *Proceedings of the First International World Wide Web Conference*, Geneva, Switzerland, URL: http://www1.cern.ch/PapersWWW94/fieldings.ps

Garofalakis, J., P. Kappos, and D. Mourloukos (1999) “Web Site Optimisation Using Page Popularity,” *IEEE Internet Computing*, 3 (4), pp. 22–29.

Girardin, L. (1996) “Mapping the virtual geography of the World-Wide Web,” In *Proceedings of the Fifth International World Wide Web Conference*, Paris, URL: http://www.girardin.org/luc//cgv/www5/index.html

France, Glushko, R. J., J. M. Tenenbaum, and B. Meltzer (1999) “An XML Framework for Agent-Based E-Commerce,” *Communications of the ACM*, 42 (3), pp. 106-114.

Greenhalgh, C. (1997) “Analysing movement and world transitions in virtual reality tele-conferencing,” In *Proceedings of the Fifth European Conference on Computer Supported Cooperative Work*, pp. 313-328. Lancaster, UK: Kluwer Academic Publishers.

Guan, T. and K. F. Wong (1999) “KPS – a Web Information Mining Algorithm,” In *Proceedings of the 8th World Wide Web Conference*, Toronto, Canada. Elsevier, Amsterdam. **W3C**

Gudivada, V. N., V. V. Raghavan, W. I. Grosky, and R. Kasanagottu (1997) “Information Retrieval on the World Wide Web,” *IEEE Internet Computing*, 1 (5), pp. 58-68.

Guice, J. (1998) “Looking Backward and Forward at the Internet,” *The Information Society*, 14 (3), pp. 201-211.

Hagman, J. and J. Ljungberg (1995) “Brute Facts versus Institutional Facts of Language as a Foundation for IR,” In *Proceedings of the 5th European-Japanese Seminar on Knowledge Bases and Information Modelling*, Sapporo, Japan.

Hannemyr, G. (1998) An Even Briefer History of the Internet. Oslo University. Department for Informatics. URL: http://www.ifi.uio.no/~inint/emne01a.htm.

Hill, W. C., J. D. Hollan, D. Wroblewski, and T. McCandless (1992) “Edit wear and read wear,” In *Proceedings of the ACM 1992 Conference on Human Factors in Computing Systems*, pp. 3-9, Monterey, CA: ACM Press.

Hoffman, D. L., W. D. Kalsbeek, and T. P. Novak (1996) “Internet and Web use in the US,” *Communications of the ACM*, 39 (12), pp. 106–108.

Holmquist, L.E., H. Fagrell and R. Busso (1998) “Navigating Cyberspace with CyberGeo Maps,” In *Proceedings of IRIS 21*, pp. 391-399, Sæby, Denmark.

Huhns, M. N. and M. P. Singh (1997) “Mobile Agents,” *IEEE Internet Computing*, 1 (3), pp. 80-82.

Jain, A. K. and R. C. Dubes (1988) *Algorithms for Clustering Data*. NJ: Englewood Cliffs: Prentice-Hall.

Kawano, H. and T. Hasegawa (1998) “Mondou: Interface with Text Data Mining for Web Search Engine,” In *Proceedings of the 31st Hawaii Conference on System Sciences*, Los Alamitos. IEEE Computer Society Press.

Khare, R. and A. Rifkin (1997) “XML: A Door to Automated Web Applications,” *IEEE Internet Computing*, 1 (4), pp. 78-87.

Koster, M. (1997) The Robots Exclusion Protocol. URL: http://info.webcrawler.com/mak/projects/robots/exclusion.html

Krulwich, B. (1997) “Automating the Internet: Agents as User Surrogates,” *IEEE Internet Computing*, 1 (5), pp. 34-38.

Kumar, S. R., P. Raghavan, S. Rajagopalan, and A. Tomkins (1999) “Trawling the web for emerging cyber-communities,” In *Proceedings of the 8th World Wide Web Conference*, Toronto, Canada. Elsevier, Amsterdam. **W3C**

Lassila, O. (1998) “Web Metadata: A Matter of Semantics,” *IEEE Internet Computing*, 2 (4), pp. 30-37.

Leiner, B. M., V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolff (1997) A Brief History of the Internet. URL: http://info.isoc.org/internet-history/brief.htm.

Leonard, A. (1997) *Bots: The Origin of New Species*. San Francisco: HardWired.

Li, Y. (1998) “Toward a Qualitative Search Engine,” *IEEE Internet Computing*, 2 (4), pp. 24–29.

Maes, P. (1994) “Agents that reduce work and information overload,” *Communications of the ACM*, 37 (7), pp. 31-40.

Maes, P., R. H. Guttman, and A. G. Moukas (1999) “Agents that Buy and Sell,” *Communications of the ACM*, 42 (3), pp. 81–91.

Maglio, P. P. and R. Barrett (1997) “How to Build Modeling Agents to Support Web Searchers,” In *User Modelling: Proceedings of the Sixth International Conference*, pp. 5-16, Vienna, Austria: Springer Verlag.

Mitchell, W. J. (1995) *City of Bits: Space, Place and the Infobahn*. MIT Press, USA.

Mukherjea, S. and J. D. Foley (1995) “Visualizing the World-Wide Web with the Navigational View Builder,” *Computer Networks and ISDN System* (Special Issue on the 3rd International Conference on the World-Wide Web.)

Nielsen, J. (1999) “User Interface Directions for the Web,” *Communications of the ACM*, 42 (1), pp. 65–72.

Oard, D. W. (1997) “The State of the Art in Text Filtering,” *User Modeling and User-Adapted Interaction: An International Journal*, 7 (3), pp. 141-178.

Olsen, K. A., R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams (1993) “Visualization of a Document Collection with Implicit and Explicit Links - The Vibe System,” *Scandinavian Journal of Information Systems*, 5, pp. 79-95.

Palmer, J. W. and L. B. Eriksen (1999) “Digital Newspapers Explore Marketing on the Internet,” *Communications of the ACM*, 42 (9), pp. 33–40.

Pitkow, J. (1997) The WWW User Population: Emerging Trends. Slides presented at GlobeCom 1997 GVU Center. Georgia Institute of Technology. www.gvu.gatech.edu/user_surveys

Rein, L. (1998) “XML-Enabled Tools,” *IEEE Internet Computing*, 2 (3), pp. 16-19.

Resnick, P. and H. Varian (1997) “Recommender systems,” *Communications of the ACM*, 40 (3), pp. 56-58.

Robertson, M., C. Sørensen, and J. Swan (2000) “Managing Knowledge With Groupware: A Case Study of a Knowledge-Intensive Firm,” In *Proceedings of the 33rd Hawaii Conference on System Sciences*, Los Alamitos. IEEE Computer Society Press.

Scarbrough, H., J. Swan, and J. Preston (1999): Knowledge Management: A Literature Review. Issues in People Management. London: Institute of Personnel and Development.

SCB (1996) *Statistisk Årsbok 96.* Stockholm: Statiska Centralbyrån. URL:www.scb.se

Schneiderman, B. and P. Maes (1997) "Direct manipulation vs Software Agents: Excerpts from debates at IUI 97 and CHI 97," *Interactions*, 4 (7), pp. 42-61.

Schubert, C., R. Zarnekow, and W. Brenner (1998) "A Methodology for Classifying Intelligent Software Agents," In *Proceedings of the Sixth European Conference on Information Systems*, ed. Walter R.J. Bates, vol. I, pp. 304-316.

Senanayake, C. R. (1998) WebSpy: Towards Supporting Persistence of Interest on the Web. Masters Dissertation. Computer Science, Aston University, UK.

Smithson, S. (1999) The 1999 World Wide Web 100. Department for Information Systems. The London School of Economics and Political Science.

Sørensen, C. (1998) "Where Have You Been Today? Investigating Web Navigation Support," In *Proceedings of IRIS 21*, pp. 875-890, Sæby, Denmark.

Stenmark, D. (1999) "Capturing Tacit Knowledge Using Recommender Systems," In *Proceedings of IRIS 22*, Vol. III, pp. 275-296, Keuruu, Finland.

Tauscher, L. and S. Greenberg (1997) "Revisitation Patterns in World Wide Web Navigation," In ACM SIGCHI '97 Proceedings of the Conference on Human Factors in Computing Systems, Atlanta, Georgia. ACM Press. **XF**

Tesitelová, T. (1992) *Quantitative Linguistics,* Amsterdam: John Benjamins Publishing Company.

Tufte, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.

Whittaker, S., L. Terveen, W. Hill, and L. Cherny (1998) "The dynamics of mass interaction," In *Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work*, pp. 257-264, Seattle, WA. ACM Press.

Wired (1997) Automatable Web. Wired Magazine, August 1997, pp. 41.

Wooldridge, M. J. and N. R. Jennings (1999) "Software Engineering with Agents: Pitfalls and Pratfalls," *IEEE Internet Computing*, 3 (3), pp. 20–27.

Young, P. (1996) Three Dimensional Information Visualisation. Visualisation Research Group. Centre for Software Maintenance. Department of Computer Science. University of Durham. Computer Science Technical Report, 12 (96).

Zakon, R. H. O. (1998) Hobbes' Internet Timeline v4.0. URL: http://info.isoc.org/guests/zakon/Internet/History/HIT.htm

Second paper[2]

IntraNews: A News Recommending Service for Corporate Intranets

Abstract

It is often problematic for employees in large companies to keep up to date with what is going on within the organisation. Lately, corporate intranets has been introduced as a channel for organisational communication. An intranet enables easy access to, e.g., ongoing project information, technical documentation, organisational newsletters and policies. However, even with an intranet it is still a non-trivial task to find out what is going on. The tools that are applied to retrieve information, e.g., search engines (Altavista) and portals (Yahoo) are not easy to use on intranets. For example, a heavy workload prevents people from browsing collections of link and search engines are primarily aimed at supporting users with clear goals. To find out what is going on in a large organisation needs to be treated with another strategy. In this paper existing techniques are first introduced and compared from the viewpoint of their applicability to provide people, in a large company, with information about what is going on. A new application for such circumstances, named "IntraNews," is then introduced. IntraNews was designed based on implications from empirical work at a large European company group. The system is a recommender service specialised for intranets with a lot of users.

---

[2] Fagrell, H. (1999) "IntraNews: A News Recommending Service for Corporate Intranets," In *Proceedings of Computer Supported Cooperative Work in Design*, pp. 323-328, Compiègne, France: Université de Technologie de Compiègne.

1. Introduction

A trend in many large companies is that employees are responsible for keeping themselves up to date with what is going on in the organisation. To accomplish this corporate intranets has been introduced as a new channel for organisational communication. Finding relevant information on an intranet is, however, not an easy task. The exiting technologies that are applied to retrieve information, e.g., search engines (e.g., Altavista), portals (e.g., Yahoo), interface agents (e.g., Lieberman 1997) and retrieval agents (e.g., Voss et al. 1997), can be difficult to use especially when trying to find out what is going on in the company. Also, a heavy workload may prevent people from using the techniques in the first place since they require a quite high investment in time to gain benefit. The technologies have inherited some of the characteristics from traditional information retrieval and information filtering, i.e., to mainly support users with a clear goal or filter out information based on what a user has previously done.

To find out what is going on in society as a whole is rarely considered as a problem. We have our news channels that provide us with a selection of relevant information. We can select the news that we like from a variety of channels, e.g., television, radio, newspapers or even the Web.

News channels can also be found in large organisations. There is often an internal communication department that is responsible for dissemination of information to employees. A problem for such departments is to find out what is interesting for the employees to know so that correct information is disseminated. If we consult an introductory book on news reporting (e.g., Menkler 1997), news is describe as being things that concern a lot of people and are immediate (of current importance). What does this mean for the use of intranets? Well, the activity on an intranet does in a way mirror what is going on in an organisation. For example, if a document is never accessed it is likely that its content is of little value for a majority of the personnel. Interesting documents are on the other hand likely to be found and

accessed, partly because they are often passed along to colleagues (see, Maltz and Ehrlich 1995).

This paper describes a system, named IntraNews, that help people to find out where interesting things are happening on intranets. The system is a news recommending system based on collaborative filtering (for overview see, Resnick and Varian 1997). The design of IntraNews was informed by empirical work at a large European corporate group. The empirical work consisted of interviews about the use of the intranet and the work situation in general and analysis of the intranet server access logs.

This paper is organised as follows: in the next section related solution strategies are surveyed related to the problem. Section 3 describes the research background. In section 4 the system is described in detail. Section 5 is a discussion about the solutions. The paper ends with summary and future directions.

2. Strategies for Information Seeking

The techniques that are offered to retrieve information from intranets are limited when it comes to detect what is going on in the company.

A common techniques is to use *search engines*, i.e., systems that allows the user to submit Boolean queries resulting in a list of links to matching documents (e.g., AltaVista, Lycos). This fairly classical information retrieval model is mainly suited for short-term interests where the user has a clear goal, e.g., what is the telephone number to X (cf., Belkin and Croft 1992)? Search engines also require the users to enter a correct and precise query, while their true interests are often rather vague or at least difficult to express in a precise way. Among other, Furnas et al. (1987) has identified that keyword-based searching is often ambiguous, and formulating a good query is thus not easy. Moreover, it would be impossible for the users to express what is interesting news before they know what is offered.

Another common approach is to develop *portals*, i.e.; collections of links that are categorised manually (cf., Yahoo!). A problem, here, is that the manual method is very slow which results in that many links are out of date and that immediate things are missed. There are also

problems with categories, e.g., a document is often in more than one category and the users of a category system may not have agreed on the use of every category. Another problem is that the meaning of a categories drift over time (cf., Bowkar and Star 1991) making the system even more difficult to use. Even if there was some interesting news in the category system the navigation in such systems are likely to be very time consuming. Time, however, is something very valuable in many organisations today.

A fairly recent approach to the problem is to use *browsing agents*, i.e., programs that assists the user when browsing (see, e.g., Lieberman 1997). The problem here is that the user needs to browse a lot before the agent gives good assistance. This assistance can unfortunately not find new information that the users has not expressed as interesting, i.e., it can only recommend the users information based on past browsing behaviour. News may not be detected with this method.

Another kind of agents are *retrieval agent*, i.e., programs that search the intranet over longer periods based on a profile that the users has declared (see, e.g., Voss et al. 1997; Krulwich 1997). This approach has the same problems as the browsing agents, i.e., there is a great risk in missing news because they are based on a profile of the users past behaviour. Also, the immense flexibility and possibilities makes the agents complicated to use and require much of the users time before they pay off.

A problem, which is true for all techniques descried above, is that use in non-desktop situations is largely overlooked. As Bellotti and Bly (1996) argue, new problems and possibilities occur when people are “walking away from the desktop.”

3. Research Background

The empirical work took place at a large European corporate group, with offices in several European countries and the United States. The research was conducted according to what Dahlbom (1996) describes as “the new informatics.” For a more detailed description (see, Ljungberg et al. 1998). The methods used are both qualitative and quantitative.

The study is based on interviews and observations, accompanied by analysis of statistical data form intranet servers (based on Hyper-Text Transportation Protocol (HTTP)). This was done in order to get a general understanding of how work is actually carried out in the company, and how it relates the intranet use. Eighty hours of observations and 14 qualitative interviews, lasting between 45 and 90 minutes, have been conducted to find background data and collect user suggestions.

Most of the work in the company is organised in projects and the projects participants are often geographically separated in different buildings and floor plans. Email has been the major computer-based tool for information sharing. There has, however, been a trend towards using the intranet for sharing, albeit there is no espoused reward system for publishing on the intranet. Some departments, e.g., internal communication and policies, are today using the intranet as a channel for a majority of their information dissemination.

The intranet was implemented in 1995 as the result of a joint effort between the Group's IT company and the Group Headquarters. Today (May 1999), the intranet has about 30,000 users, 500,000 documents, and 250 servers. Some servers have just over 3000 user sessions per weekday. A user session is a suit of requests (excluding graphics) documents, but it is considered as terminated after 30 minutes of inactivity.

4. IntraNews

4.1 Design philosophy

In this section we will describe the system, IntraNews, in detail, but fist we would like to give an overview of the principles that has guided the design of the system.

The idea, here, is that the accesses to documents mirror what is interesting for an audience came from analysis of the access logs. When the users where asked why some of the documents seemed to be

popular they could always give good explanations. A few examples of answer were “that’s a technical specification that just changed,” “that’s the news project documentation guidelines” and “that’s next weeks lunch menu (for the local restaurant).” One of the systems developers that first introduced the intranet at the company said that the main reason for its success was that managers could download the latest financial figures in Microsoft PowerPoint format. In a general sense this implies that material of current importance are distributed on the intranet.

How are these interesting documents found then? Well, we found that it is common that people recommend documents to each other that they find interesting. As one interviewee puts it “It is common that I get good links (to an intranet document) during coffee breaks, or via e-mail, and when I have found a good link I usually pass it on to colleagues.” Sharing is distributed by “word of mouth,” i.e., the documents are being talked about.

The findings made us think about the problem as a matter of finding and reporting news. In an introductory book on news reporting (Menkler 1997), news is describe as being, for example, things that concern a lot of people, are talked about and are immediate (of current importance). These things inspired us how the analysis of the access logs, which is described in the next section. More general requirements for IntraNews was that it should be seen as news service that provide

- new and relevant information, i.e., news, but the user must determine its importance
- delivered to the users, i.e., sent to the user based on that something interesting has happened
- with transparent access, i.e., in stationary and mobile situations

4.2 Architecture

In figure 1 the architecture of the system is illustrated. The components are looked at up close.

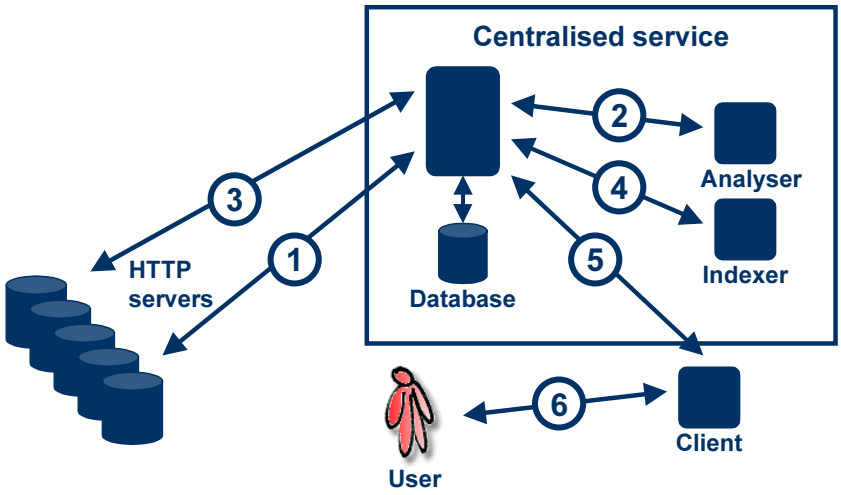

Figure 1: The architecture of IntraNews.

1. Retrieve the access log files from the HTTP servers.
2. Analyse the logs and find candidate documents for indexing
3. Retrieve the candidate documents
4. Create index based on documents
5. A stream of document links is sent to the client based the user's profile
6. The client presents the document links. The user gives feedback and it updates the user's profile in the central database.

4.3 Retrieving Access Log Files

Since the access logs are generally very large (up to 100 megabyte) and growth of some access logs can be up to 1 megabyte per weekday it is not preferable to retrieve the whole files over the network at every update. Therefore, we use a Common Gateway Interface (CGI) script the transmission time could be reduced by just sending the changes. The CGI script is manually installed at each HTTP server. Even if there are many HTTP server brands, the script is very similar for all and it is very easy to install. New HTTP servers that are installed can be found in the internal search engine's robot report log. Figure 2 show a sample of the CGI script for the Apachie HTTP server. The date and

time are sent as a parameter to the CGI script and every entry that is newer than the parameter in the access log are sent back.

```
while (!$done) {
  <LOG>;
  $row = <LOG>;
  ($da, $mo, $y, $time) = ($row =~
/(\d{2})\/(\w{3})\/(\d{4})\:([\d\:]{8})/);
  if ( &dcmp($mo, $l_mo) ==1) {
    $pos -= $day_size*15;
    seek (LOG, $pos, 0);
  } elsif ($da > $l_da) {
    $pos -= $day_size;
    seek (LOG, $pos, 0);
  } elsif ($time > $l_time) {
    $pos -= $day_size/6;
    seek (LOG, $pos, 0);
  } else {
    $done = true;
  }
}
```

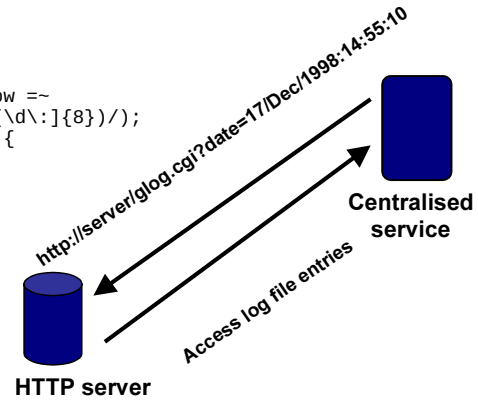

Figure 2: The model for how the CGI script that reduces the amount of data that needs to be transmitted. The example code is in perl and for the Apachie server.

4.4 Analysing the access logs

By analysing the access logs, we can determine candidates for indexing. What we want to find out here is where is (1) documents that many users access and (2) that change frequently or are updated. Every request to an HTTP server is stored as an entry in its log.

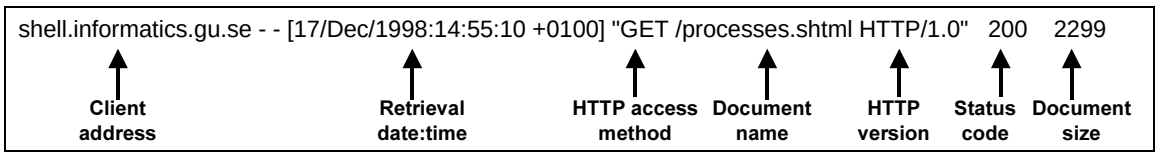

Figure 3: An entry from an access log of an HTTP server.

Figure 3 is illustrating an example of an entry with the address to the accessing client computer, the date and time (Greenwich Mean Time plus one), the access method, the requested file, the HTTP version, the status code (200 = document ok) and the size of the requested document.

Table 1 is an example of what is being stored in the central database for each unique intranet document.

Stored record	Example
Document address (key)	www.hotel.company.se/projects/s80/
Average number of accesses per weekday	730
Average number of unique clients per weekday	400
Document size	12760

Table 1: The stored record for a document.

The server and document name is the key in the database. Documents that are accesses with the method POST are just ignored since they most likely have dynamic content generated by a script. Other dynamic documents, i.e., the ones that use the GET method, are not a problem either, because the different parameters will make our program store the access as a unique document. In either way, if a dynamic document with the same set of parameters is being accessed a lot it should be recommended. A more serious problem is that documents can have many names, i.e., when a directory is accessed a default document is sometimes sent. Often the name of this document is “welcome.html” or “index.html,” but other names can also be found. We know that documents may be missed because of this, but on the other hand, only one of two documents with the same content would be recommended anyhow.

If a document has changed is determined by its size. It is highly unlikely that a changed document will get the exact same size in bytes. Some documents that have a dynamic date included that change the document. However, the date was of equal length and thus does not change the document size. Banners included in many documents are likely to get a lot of accesses, but since banners seldom change and has very little text they will not be recommended.

Some browsers have a cashing mechanism to increase network performance. This decreases the number of accesses for documents that are re-accessed. On the other hand, the popular documents will in general get more traces in the access logs anyhow and from unique clients.

The proper values for when a document should be selected have to be tuned for the local conditions of the organisation. In our case the conditions can be summarised as:

- the document has to be accessed at least 20 time per weekday
- the document has to be at least 1000 bytes

- the document has to change (in size) at least ones a week.
- extra priority is given to document that increase heavily in accesses or unique clients, i.e., 30% on a weekday.
- root pages for many servers were excluded since we know that they never contain new information and are the default start document for many users

4.5 Creating the index

The index is updated once every third hour based on ideas from by traditional Information Retrieval (Salton 1971; van Rijsbergen 1979). The system starts by retrieving the full text from the documents that has been found by the log analyses (see figure 1).

First all documents, e.g., HTML, Microsoft Word and Adobe Acrobat, are converted to plain text. Each document is treated as a "bag of words," where only the number of times a word occurs in a document is regarded, and not the original word order. Each document is modelled as a vector in a high-dimensional semantic space (Salton 1988). Typically, the number of documents (and vectors) is quite large and since most documents consist of HTML files, the average document is quite short. The next step is to filter out high frequent and low content words, e.g., "project", and "the", using a stop-word list. Third, the remaining words are put on their basic form using a dictionary of 300,000 words that has been generated from intranet documents texts. Fourth, each term is weighted with the TF-IDF (Salton 1971).

After this is done the remaining, context-bearing, words that are left in a single document can be very few. But in the whole corpus, the total number of unique words remaining after the filtering is still quite large, about 10,000. This results in a very large, but sparse matrix, a phenomenon known as the sparse data problem with is very computational demanding. This matrix, representing the relationships between all the documents, would be even more sparse, and misleading, if no filtering had been applied. To address this problem, Latent Semantic Indexing (Deerwester et al. 1990) is applied, which reduces compresses the matrix a great deal, while still keeping the most important semantic dimensions, in some implicit sense. The

result is compressed document vectors that contain the “essence” of the modelled semantic relationships. In our case we use 50 elements in each vector. An example of a vector is illustrated in figure 4.

$K = [k_1, k_2, k_3, ..., k_n] = [0.00023177, 0.00006104, -0.00006202, ..., 0.00069215]$

Figure 4: An example of a document represented as a vector.

As we mentioned above, many documents are short, which makes them hard to model adequately. These are just thrown out. These techniques can be applied on a few thousand documents, but using them for the World Wide Web is currently impossible due to practical computational limitations.

4.6 Making sorting possible

To determine equality between document vectors and the profile vector we must be able to compare two vectors. This is done by calculating cosine on the angle between two vectors. The value can vary from 1 to –1 and if the value is 1 there is no angle between the vectors and if the value is –1 the angle is 1800 between the vectors. The cosine for the angle is determined from the scalar product of the vectors and the absolute value for the vectors (see figure 5).

$$COS\theta = \frac{(a_1 b + a_2 b_2 + ... + a_n b_n)}{(\sqrt{a_1^2 + a_2^2 + ... + a_n^2}) \bullet (\sqrt{b_1^2 + b_2^2 + ... + b_n^2})}$$

Figure 5: The formula for computing cosine on the angle between vector a and vector b (for an overview of linear algebra see Lay (1997).

When the user gives feedback that a document is interesting it can be the basis for a profile used to select the documents that are to be of recommend. The profile is stored in a database on the central server. It is represented by a vector, which makes it possible to compare with the document vectors. A genetic algorithm (see, Russell and Norvig 1995, p. 620) is used to manipulate the profile. Our implementation is very simple and based on three operator, i.e., selection, crossover and mutation. The selection operation is just based on the document that the user has selected as interesting and a profile vector is created.

When new documents are selected as interesting crossover operations are executed between the document vectors and the profile vector. The mutation operation is used to ensure that the profile vector has a bit of randomness. This is important when seeing it as a news service, i.e., not specialising on content too much.

4.7 Client side interface

The client side of the system has two interfaces, i.e., a browser based and one for mobile devices.

The first step for every user is, however, to register to the service. This is done by filling in a form that is provided on the intranet. After this a unique URL is sent to the users via email. The users can access the system from every stationary computer on the intranet by requesting the URL with a Javascript compatible browser. When the URL is accessed a new window pops up (see figure 6).

Figure 6: The user interface on the stationary computers.

The windows has a field with rolling text that is the title (or the first heading) of the documents that are recommended. Because of the roller the user can get an overview of what is recommended to read without keeping all attention on the interface. This part of the user interface was inspired by the product PointCast with “pushes” information to its users[3]. If a text string seem interesting the user can pull down the menu labelled “news” and select the document. This will result in that the browser window opens the document for the user to read. If user finds the document interesting the button with the label “interesting?” could be pressed. This will affect the stream document by sorting them on how equal they are to the documents that have been selected as interesting. The user determines the number of documents that should

[3] More information about the product PointCast can be found at their web site (http://www.pointcast.com/). A discussion about push technology can be found in Franklin and Zdonik (1998).

be recommend. In figure 6 the number of documents to recommend is 30. The user can change this at any time by just changing that number.

Today the recommendations are distributed to mobile phones periodically (and event driven) with the Short Message Service[4]. It is unfortunately only possible to read the headlines for a few documents and no feedback to the system is possible. It is, however, possible to access the system from network connected Windows CE.

A prototype client has been developed on an emulated version of the Wireless Application Protocol (WAP)[5]. With this prototype the system can be accessed in its full functionality, but WAP phones are not on the market for the moment unfortunately.

5. Discussion

The system described in this paper is designed informed by empirical work at a large European company group. The group's intranet has servers that logs the users activity. This serves as a basis for collaborative filtering of what is happening within the company. The users can register to the service and get recommendations of news delivered to their desktop computer and mobile phones.

Our system uses the access log and thus no extra load is put on the user for the construction of the collaborative filter. Infoscope (Stevens 1992) uses a similar assumption when the reading time for USENET news articles serve as a basis for collaborative filtering. Moreover, Morita and Shinoda (1994) show in a study, also of USENET news, that the reading time measurement actually produced better recall and precision in a text filtering experiment than using documents explicitly rated by the user as interesting. Other approaches such as matching profiles provide users with new resources through social filtering. Siteseer (Rucker and Polanco 1997) does this by matching the users' bookmark files and Yenta (Foner 1997) extracts

[4] The Short Message Service is a part of Global System for Mobile communication (GSM). More information about GSM can be found at the GSM association (http://www.gsm.org/).

[5] In short WAP makes Web browsing possible on mobile devises. More information about WAP can be found at the WAP forum (http://www.wapforum.org/).

information from many sources, e.g., Web documents and emails. This approach needs a critical mass of users to pay off (see Maltz and Ehrlich's (1995) discussion of the "cold start problem"), in contrast to our system, where even one single user could benefit.

Avery and Zeckhauser (1997) explored the problems that content owners may manipulate recommender systems by recommending them self very strong. This is hardly going to be a problem here, since the system operates within one organisation.

SOAP (Voss et al. 1997) sends queries to Web search engines and uses collaborative filtering for the users in the system to solving users' long-term information interests. To recommend web documents Fab (Balabanovíc and Shoham 1997) use explicit ratings from users, while Phoaks (Hill and Terveen 1996; Terveen et al. 1997) extracts recommendations of web documents from USNET news articles. GroupLens (Konstan et al. 1997) recommends USENET news articles by matching users based on their behaviour in the system and explicit rating of news articles.

6. Summary and Future Directions

This paper adds to the existing body of research by introducing the idea that interesting documents on an intranet could be viewed as news. An application, IntraNews, is described in detail and presented in the setting of a large European company group. IntraNews locates relevant documents by analysing access logs from intranet servers and use it as a basis for collaborative filtering. The filtering is based on where on the intranet

- particularly much people visit (accesses)
- particularly much activity going on (documents changes)

IntraNews could be seen as a recommendation service that delivers news to the users' desktop computer and mobile phones.

The future plans for this research is to improve the system technically and explore the usefulness of organisational news based on intranets. For example, we will apply other techniques from corpus linguistics, e.g., automatic text genre determination and divide long

texts that discuss several topics into smaller chunks, each one treated as a single document (see, e.g., Hearst 1997). Another issue is to get the system to work better for mobile users. In doing this we aim to evaluate news consumption in mobile settings.

Acknowledgement

This research is a part of the Mobile Informatics programme that is funded by the Swedish Research Institute for Information Technology (SITI).

The author would like to thank the colleagues at Viktoria, especially Peter Ljungstrand, Dick Stenmark, and Fredrik Ljungberg.

7. References

Balabanovíc, M and Shoham, Y. "Fab: content-based, collaborative recommendation," *Communications of the ACM*, 40 (3), pp. 66-72.

Belkin, N. J. and W.B. Croft (1992) "Information Filtering and Information Retrieval: Two sides of the same coin?" *Communications of the ACM*, 35 (12), pp. 29-38.

Bellotti, V. and S. Bly (1996) "Walking away from the desktop computer distributed collaboration and mobility in a product design team," In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*, pp. 209-218, Cambridge, MA: ACM Press.

Bowker G. and S.L. Star (1991) "Situations vs. Standards in Long-Term, Wide-Scale Decision-Making: The Case of the International Classification of Diseases," In *Proceedings of the 24th Annual Hawaii International Conference on System Sciences*, Kauai, Hawaii, vol. IV, IEEE Computer Society Press.

Dahlbom, B. (1996) "The New Informatics," *Scandinavian Journal of Information Systems*, 8 (2), pp. 29–48.

Deerwester, S., S.T. Dumais, T., Landauer, G.W., Furnas and R.A. Harshman (1990) "Indexing by latent semantic analysis," *Journal of the Society for Information Science*, 41 (6), pp. 391-407.

Foner, L. (1997) "Yenta: A Multi-Agent, Referral Based Matchmaking System," In *Proceedings of the first international conference on Autonomous agents,* pp. 301-307, Marina del Rey, CA: ACM Press.

Franklin, M. and S. Zdonik (1998) “Data In Your Face: Push Technology in Perspective,” In *Proceedings of ACM 1998 international conference on Management of data,* pp. 516-519, Seattle, WA: ACM Press.

Furnas, G., T. Landauer, L. Gomez and S. Dumais (1987) “The Vocabulary Problem in Human-Systems Communication,” *Communications of the ACM*, 30 (11). pp. 964 – 971.

Hearst, M. (1997) “TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages,” *Computational Linguistics*, 23 (1), pp. 33-64.

Hill, W. and L. Terveen (1996) “Using frequency-of-mention in public conversations for social filtering,” In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*, pp. 113-121, Cambridge, MA: ACM Press.

Konstan, J.A., B.N. Miller, D. Maltz, and J.L. Herlocker, L.R. Gordon and J. Riedl (1997) “GroupLens: applying collaborative filtering to Usenet news,” *Communications of the ACM,* 40 (3), pp. 77-87.

Krulwich, B. (1997) “Automating the Internet: Agents as User Surrogates,” *IEEE Internet Computing*, 1 (5), pp. 34-38.

Lay, D.C. (1997) *Linear Algebra and its Applications*. 2nd edition. Addison Wesley Longman, Inc.

Lieberman, H. (1997) “Autonomous Interface Agents,” In *Proceedings of ACM 1997 Conference on Human Factors in Computing Systems*, pp. 67-74, Atlanta, GA: ACM Press.

Ljungberg, F., B. Dahlbom, H. Fagrell, M. Bergquist and P. Ljungstrand (1998) “Innovation of new IT use: Combining approaches and perspectives in R&D projects,” In *Proceedings of the Fifth Biennial Participatory Design Conference*, pp. 203-209, Seattle, WA: CPSR and ACM Press.

Maltz, D. and K. Ehrlich (1995) “Pointing the way: Active collaborative filtering,” In *Proceedings of ACM 1995 Conference on Human Factors in Computing Systems*, pp. 202-209, Denver, CO: ACM Press.

Mencher, M. (1997) *News Reporting and Writing*. Seventh Edition. McGraw-Hill.

Morita, M. and Y. Shinoda (1994) “Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval,” In *Proceedings of the 17th International ACM Conference on Research and Development in Information Retrieval,* pp. 272–281, Dublin, Ireland: ACM Press.

Resnick, P. and H.R. Varian (eds.) (1997) “Recommender systems,” *Communications of the ACM,* 40, (3), (Entire issue).

van Rijsbergen, C. J. (1979) *Information Retrieval*, 2nd Edition, Butterworths. London.

Rucker, J. and M.J. Polanco (1997) “Siteseer: personalized navigation for the Web,” *Communications of the ACM*, 40 (3), pp. 73-76.

Salton, G. (1971) *The SMART Retrieval System*, Englewood Cliffs, Prentice-Hall, Inc., New Jersey.

Salton, G. (1989) *Automatic Text Processing*, Addison-Wesley, Inc.

Stevens, C. (1992) "Automating the Creation of Information Filters," *Communications of the ACM*, 35 (12), p. 48.

Terveen, L., W. Hill, B. Amento, D. McDonald and J. Creter (1997) "PHOAKS: a system for sharing recommendations," *Communications of the ACM*, 40 (3), pp. 59-62.

Voss, A., T. Kreifelts and V. Paulsen (1997) "SOAP: Recommendations from Social Agents," In *Proceedings of Computer Supported Cooperative Work in Design,* pp. 168-173, Bangkok, Thailand: International Academic Publishers.

Third paper[6]

Exploring Support for Knowledge Management in Mobile Work

Abstract

This paper reports fieldwork from the electrical utilities industry, examining the suitability of current knowledge management perspectives to the day-to-day work of mobile staff. Reporting the results of the empirical study, we make a distinction between four aspects of local and mobile "knowledge management" as it took place in the mobile work setting: *sharing*, i.e., several parties exchange knowledge; *indexing*, i.e., one party explains to another what knowledge to retrieve; *diagnosing*, i.e., two parties make sense of how to interpret a situation, and; *foreseeing*, i.e., one party (or more) uses knowledge to project the future. We compare and contrast the empirical findings with current knowledge management perspectives, and outline an initial sketch of a framework for "practical knowledge management."

---

[6] Fagrell, H., S. Kristoffersen and F. Ljungberg (1999) "Exploring Support for Knowledge Management in Mobile Work," In *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work*, pp. 259-275, Copenhagen, Denmark: Kluwer Academic Publishers.

1. Introduction

Organisational memory and knowledge management have inspired important debates within the field of CSCW (e.g., Randall et al 1996; Bannon and Kuutti 1996). This paper reports from an empirical study where these two issues were examined in the day to day work of mobile service engineers. The study also explored the concept of local knowledge since its importance has been much recognised in the field (see, e.g., Rouncefield et al 1994; Randall et al 1996). Local knowledge has mainly been conceived as specific to place and procedures and the ways in which it appears in mobile work settings have not been investigated. Moreover, there has been a tendency, within CSCW, to localise activities to an individual user, and thereby missing to enhance the flexible and unexpected collaboration that often emerges out of a mobile work situation (cf., Luff and Heath 1998).

To collect empirical data we observed (Hammersley and Atkinson 1993) and interviewed the mobile staff (Patton 1990). The study lasted for about five months with an average intensity of two days per week. Large parts of the empirical material was transcribed and analysed according to the principles of grounded theory (Glaser and Strauss 1967).

Our analysis provides a critique that challenges some of the current conceptions of knowledge management, but at the same time gives opportunity for new mobile applications. We suggest a conceptual framework based on how knowledge sharing mechanisms map into action. Our point is that knowledge management in mobile settings is *social* and *dynamic*. It is this *social* mapping process that deserves our attention, although it may not in itself be suitable for computerisation.

The structure of this paper is as follows: Section 2 outlines the related work and is followed by the research background and setting. Section 4 presents an analysis of the fieldwork. Section 5 discusses the results, the related research and derives design implications. Section 6 concludes the paper.

2. Related Work

Organisational memory is an established theme in CSCW. Research contributions on the topic involve empirical studies, systems design, as well as conceptual debates. Organisational memory systems are typically based on hypermedia that links different sources of information, e.g., records. Recently, the notion knowledge management has been introduced into the CSCW field (cf., Greif 1998). It seems to have much in common with organisational memory. However, an emphasis on *managing large repositories of information* with information retrieval and artificial intelligence techniques can be noted (see, Abecker et al 1998; O'Leary 1998).

One of the first organisational memory systems described in the literature was gIBIS (Conklin and Begeman 1988). The aim of gIBIS was to make decision processes explicit by capturing the argumentation.[7] Another system is "TeamBuilder," of which the objectives is to support team members in identifying expertise and collaborate more effectively (Karduck 1994). "Answer Garden" aims to help organisations to capture and enable retrieval of experiences made by their employees (Ackerman 1994). In "Answer Garden 2," there are also features for finding and interacting with experts directly (Ackerman and McDonald 1996).

Under the knowledge management label, Terveen et al (1993) report from a project of creating a "living design" for software developers. The system provides access to, and updating features of design-relevant information. More recently Edwards and Mahling (1997) explored design implications for knowledge management in the legal domain. In addition to their advice about putting knowledge "chunks" into organisational context, they suggest that roles should be assigned for assuring relevance, accuracy, and periodically review the material (similarly to library work).

Empirical studies on knowledge management have aimed at eliciting design implications, e.g., by analysing the work of a telephone hotline group (e.g., Ackerman and Halverson 1998) and expertise location in a software development company (McDonald and Ackerman

[7] The gIBIS is now a commercial product called QuestMap (http://www.gdss.com/OM.htm).

1998). Studies have also evaluated organisational memory systems in use (e.g., Ackerman and McDonald 1996).[8]

The conception of an organisational memory has been criticised for diverging from remembering, as it actually takes place in organisations (Randall et al 1996; Bannon and Kuutti 1996). A similar position is taken by Kovalainen et al (1998) who sees organisational memory as an "artefact mediated process." Many ideas on the topic derive from the decision-making perspective on organisations. This perspective has been heavily criticised by, among others; March (1991) who argues that this is simply not the way things happen in organisations. Another issue is the uneasy relation between learning and memory. As Weick (1979) puts it: "to organise[9] is to reduce variety and remember but to learn is to keep variety and to forget." This type of learning, however, is of a different character then the less ambitious learning-by-doing that is mainly considered in this paper. Orr (1996) describes an example of this sort of learning where the staff of mobile photocopier repairmen uses stories to assistant the communities learning. To further explore this, we turned towards the concept *local knowledge*.

Geertz (1983, p. 167) views local knowledge from the perspective of "to-know-a-city-is-to-know-its-streets." Here, local knowledge could be similar in many different geographically dispersed workplaces, but *it is created through local experiences and used locally* by a limited group of people that are working more or less together. In CSCW, local knowledge has been described as: knowledge of the particularities (Rouncefield et al 1994), what makes work run smoothly (Randall et al. 1995), to know who knows what, who is busy, who is worth asking about "x" (Bowers et al 1996).

The focus of this paper is the mobile work setting. Recently, it has been argued that there is a need to explore how mobility takes place in collaboration (Luff and Heath 1998). "Walking away from the desktop" means new problems and possibilities (Bellotti and Bly 1996). Luff and Heath (1998) identify three types of mobility: *micro mobility,*

---

[8] Systems that link users with similar interests are also known under the label 'recommender systems' (see Foner 1997, Kautz et al 1997).

[9] Weick (1979) make a difference between organisation and organising arguing that the noun form encourages us to see the phenomenon as a static thing, whereas the verb encourages us to see the phenomenon as a dynamic process. This is similar to the suggestion by Randall *et al* (1996) to shift from using noun memory to the verb remembering.

local mobility and *remote mobility*. Micro mobility is the way in which the artefact may by mobilised and manipulated for various purposes around a relatively circumscribed, or "at hand" domain. Local mobility is more scattered, for instance, walking between rooms, floors and buildings at a local site, e.g., that of product designers at a consulting firm (Bellotti and Bly 1996), personnel at London underground (Luff and Heath 1998), bank officers at a customer service centre (Kristoffersen and Rodden 1996). Remote mobility is when remote users interact with each other using technology, e.g., construction foremen visiting work teams. The case study of dispersed and mobile IT support staff, described in Kristoffersen and Ljungberg (1998), involves both remote and local mobility.

3. Research background

Göteborg Energi (GE) is a publicly owned Swedish energy provider with approximately 900 employees and 3,000 million SEK (approx. 333 million *euro*) in yearly revenue. Every year GE provides just over 4,700 GWh electricity, 3,600 GWh district heating and 700 GWh of natural gas. Their 350,000 customers can be anything from a single flat to heavy industrial production facilities or sports stadiums.

A couple of years ago the Swedish electricity monopoly was disbanded and the market de-regulated. GE suddenly had to cope with more than a hundred suppliers instead of one and more competitors in the provider layer. As a consequence of this, the company had to undergo substantial organisational changes and become more customer-oriented than before, in order to answer to the competition and keep their advantage and market share.

3.1 Large scale introduction of GIS

GE is currently engaged in significant IT investment as a way of transforming the company. The old mainframe-based systems are being replaced with a client-server solution and much paperwork is computerised. One of the investments has been to computerise maps of

the electricity network. This activity is finished and many kinds of information have been integrated with a Geographical Information System (GIS). One example of this is that 15,000 cable boxes, i.e., the connecting nodes for houses and streetlights, are represented in the graphical view of the map. Additionally, information is provided about the cable box's components and to which customer's it is serving.

The GIS is accessible on the companies 850 networked PCs. The next step is to make it accessible for the 300 employees that do a majority of their work away from a networked stationary PC. The idea is to equip GE's 200 vehicles so that the GIS can also be accessed from the vehicles. The expected benefit of the mobile system is that less driving back and forth would be required to get information, e.g., maps, from stationary computers.

We followed a pilot project to evaluate the types of techniques under actual work conditions. A car was equipped with a laptop with network access via the Global System for Mobile communication (GSM)[10]. The laptop was connected to a colour printed making it possible to download maps and print them in the car.

3.2 The local service station and roles

GE has several local service stations and they are responsible for a limited geographical area in which they are located. The responsibilities are for *installations, maintenance* and *error repairs*. The work assignments are generally distributed from the central administration office, but lately the stations' personnel have been allowed to initiate customer relations, which is a big change. They are also to a great extent autonomous at an operational level. There are three major roles at the station, namely electricians, planners and administrators.

Electricians work in pairs, mainly for security reasons, but also so that skills in both high and low voltage are represented. The work pairs get their assignments from the planner, but they often do the final co-ordination with the customers themselves. The work in the field generally involves planned work, i.e., installations and maintenance, but also emergencies, e.g., power failure, which has to be

[10] More information about GSM can be found at the GSM association (http://www.gsm.org/).

taken care of immediately. Almost all of the electricians at GE are on a rolling timetable for 24-hour on-call duty. In these situations they may have to operate beyond their ordinary geographical areas. The electricians have cellular phones and pagers as standard equipment.

The planner schedules assignments by allocating resources, such as personnel, raw material and vehicles. An important issue for the planner is to manage co-ordination with contract entrepreneurs that could be large construction companies or a one-man firm with a bulldozer. It is also the planner's responsibility to make security certificates of finished jobs and inform customers and neighbours of potential disturbances. The planner has an office in the main building with a phone and a networked computer. However, the planner uses his cellular phone very much since most of his working time is spent in the field. Constant communication is vital because of the need continually to change priorities. As a planner put it:

> We think we have control of the situation. We have projects and deadlines. But we never do one job at a time. We have five or six assignments simultaneously. And then new customers call and want the electricity connected to their houses. No matter how much work we have, we must do it within two days.

Planners administrate complexity, and try to control resources that are constantly changing.

The administrators work in the main building. One of them does accounting, communication with the central office and takes care of the customer telephone calls. The other one mainly creates prospects for future installations in the area, e.g., new housing and industries.

The first thing that happens (about 7 am) on a workday is that most employees at the station meet, having an informal coffee break. They talk about the work and how it should progress during the week. Afterwards, assignments are divided among the work pairs (of electricians). The planner and the work pairs negotiate the division of work and decide how to proceed. When this is done, the cars are loaded with appropriate equipment. After this the pairs go to the customers that are at several geographical locations. If the schedule is not rearranged, they will return to the service station during the day. They will normally spend the afternoon doing some more work in the field,

followed by paper work back at the office.[11] More time is spent in the field on bigger assignment on projects and maintenance. Figure 1 gives an overview of the staffs geographic distribution and roles.

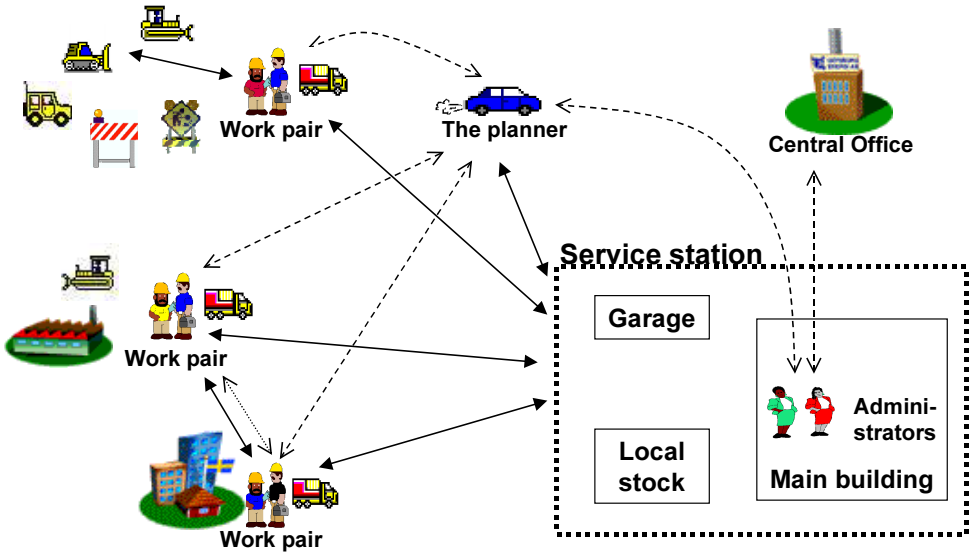

Figure 1: The communication flow (dotted arrows) and the mobility of the workforce.

Our observations consisted of following the electricians and the planner at the station, in the vehicles and in the field (see figure 1).

4. Results

Reporting the results of the empirical study, we make a distinction between four aspects of local and mobile "knowledge management" as it took place in the mobile work setting investigated. These are:

- *Sharing*: Several parties exchange knowledge.
- *Indexing*: One party explains to another which knowledge to retrieve.
- *Diagnosing*: Two parties make sense of a situation, i.e., how it should be interpreted.

[11] Other systems, e.g. billing and order, where accessible, but not used in the mobile environment during the time of our study.

- *Foreseeing*: One party (or more) uses knowledge to project the future.

Let us now consider these types in more detail.

4.1 Sharing

We found three types of sharing, namely ephemeral, persistent and stories. The one categorised as *ephemeral* took place at the daily coffee meeting. When people meet they discuss their work. This is where they get updated and informed. The following paragraph is an example of an event that was repeated every morning with only small variances.

> The planner comes in and pulls out his agenda and starts to discuss some reorganisations of today's work. He throws out a couple of questions of what the work pairs are doing this day and how long they think that it will take.

The planner tries to see if there is anyone that can reschedule today's program if something urgent comes up. Since this is a collective effort, the electricians gain knowledge of what their colleagues are doing. In many other studies the "talking out loud" and "overhearing" has been identified to give colleagues peripheral awareness what is going on and inviting the to give feedback (e.g., Heath and Luff 1991). This mainly occurs during the morning meeting, but it was also observed at other places around the station. It makes it possible for the planner to call the correct work pair if a reorganisation is necessary. It is also useful for the electricians since they become aware of their colleagues' whereabouts in case they need to ask someone. If the planner calls them with a question or assignment they cannot answer, they may know whom else to recommend.

If someone knows something that concerns everyone, it is often brought up on the morning meeting, as the following example shows.

> George says so everyone can hear: "There may be a need for more personnel at Tuve. I was there yesterday … it seemed like there was a lot of things that they would like us to do." [The others were listening and some nodded in agreement]. The planner takes a note and says that he is going to call them [the customer].

This helps the planner to plan ahead and it notifies the other electricians to prepare for what they might be doing next.

The formalised organisational sharing functions are to a great extent embedded in documents (forms and maps) that electricians fill in during and after assignments. We call this *persistent* sharing. These would then be sent to other organisational units, e.g., map surveyors and customer billing.

One example of how people share persistent knowledge was a *filer* the electricians used when they were on duty (during nights and weekends). When on duty, the electricians may have to visit sites where they rarely go, or have not been before. This could be problematic. For example, it could be difficult to remember details from time to time. It could also be difficult to get hold of people for assistance. To cope with the problem, the electricians have developed a (paper) *filer*. The *filer* is indexed by street address and contains information about local conditions of specific places. Some illustrative examples of entries from the *filer* are:

- *Karl Gustav's street 22-30: "The cable box is behind the basement door in the backyard of Engelbrekt's street 45."*
- *Erik Dahlberg's stairs: "Call the janitor on 111 22 33 to get the key to entrance 5. In case of no answer call the FM housing centre on 111 44 33."*
- *Viktor Rydberg's street 42: "The box is in the heating room. The key is in the cupboard and its got a wood piece on the ring."*

The content of the *filer* is designed to provide information that is relevant for on-call duty work. It bridges some of the representational shortcomings of the maps, e.g., location of keys and doors. The electricians who are on-call duty update the *filer* when they encounter something they find worthwhile noting. Several electricians have referred to the *filer* as being useful.

A way of sharing how to deal with problems that occur on a regular, but unplanned basis is to tell *stories* of previous problems and their solutions. These stories are told in order to transfer knowledge in the work group. They were told in any place when the work pace was low. Telling these stories was a way of producing, maintaining and communicating knowledge. They also form "traces" of the community's

common experience and knowledge that had been built up over time. The stories often have a specific topic, e.g., the following one that is about finding a place.

> A little chat on how much the landscape had changed during over the years leads into a story about "the old days." Roy started telling a story about how he had problems in finding a certain work site in the sixties. This was when a big suburb was being built. The story circles around problems of finding a site and that the map was wrong. When driving around in the woods without being able to find the construction site, Roy finds that his car is running out of petrol. Just as the car starts to cough, he sees a petrol station downhill and manages to get the car to roll down. When refuelling the car he looks around. And there… he sees the small gap between the trees that lead to the construction site where he is supposed to do his job.

We heard several version of this story. The story is not exactly about diagnosis and specific problems solving, as discussed by Orr (1991, p. 160). It is an allusion for how to find places. The literal places or situation that story refers to is not of current interest; it is from the sixties! However, some hints may still be useful, e.g., a small road often leads to a construction site and there may be settlements nearby a petrol station.

Another topic for the story telling is how customers may be taken care of. It can for example be frustrating when the customers are not really prepared. Here is an illustrative example:

> We got the place were there was this guy who was stressed up beyond belief, he just could not talk… [laugh] So we helped him out. I mean we could not just leave and there is always something we could do... I gave Dennis [planner] a call, but he was not answering – I knew he shouldn't but anyway…

It is the planner's job to take care of these situations. However, this is sometimes prohibited by a high workload, and the electricians may have to solve it themselves. They have to negotiate with the customer and come up with a solution. Moreover, the attitude of the stories is to relieve some pressure from the planner (and for the customer too).

4.2 Indexing

One way in which knowledge management takes place in the mobile setting is what could be called “indexing.” For example, when people explain to someone else “how to find a particular place,” they would typically relate to recognisable physical objects, e.g., “do you remember where we did this and that work two years ago.” Whether or not the other party (one or several) knows a particular object becomes obvious in the conversation. For example, “You mean where…?! Yes, I know where it is…,” would be a typical example of someone recognising a “sign,” while simply being quiet would indicate the opposite.

Let us now consider an example of how people use this kind of knowledge.

> The planner receives a call to his cellular phone (while driving). It’s from a construction site where they “want service now.” The planner says that he is going to “fix it.” Then he hangs up and makes a phone call. “Jonsson?… could you go to ‘project Hjalmar’? They need ‘construction electricity’ [preliminary connection used by construction firms]… Can you fix it immediately? They know that you’re going there.. the same place as the last time.” He hangs up.

The task in question was quite small, and the planner just wanted to get it out of the way. The planner knows that the electricians are away on maintenance work that is not time-critical. Therefore, he could tell the client that the two electricians will show up. The planner also knows that the electricians he calls have been doing work there before, and they know exactly where it is; thus, they are the best-fit work pair at this point.

4.3 Diagnosing

“Diagnosing” is related to “indexing.” In “diagnosing” one or more people collaboratively seek to make sense of a situation. One example of is when electricians encounter a cable that is not on the map. If it is not determinable whether or not the voltage is up, the cable has to be “shot,” i.e., putting it out of operation forever. This could mean for example that a housing area looses electricity, but it is, of course, better then that someone gets injured. This is, however, a decision that

the electricians would collectively take with the planner as a last-way-out-solution.

A more interesting thing is that maps cannot be strictly followed since minor inaccuracies are very common. The following paragraph is an illustrative example of an error in the map is "diagnosed."

> A work pair was going to connect power to new houses. At the site the map did not seem to fit the landscape. Chris and Bernie started turning and shuffling the map back and forward while discussing intensively. After a few minutes Chris says that the construction firm may have built an extra lane of houses that we do not to have on the map. This turned out to be correct and the work could continue, now with additionally six houses to connect power to.

In this case Chris suggested "there is an extra lane of houses," and since Bernie seemed to agree, this was the way in which they made sense of the situation.

One of the reasons maps do not match the landscape is that landscapes change. Moreover, the surveyors and the electricians do not always agree on the updates of the maps, which results in compromises. Confusion may occur, but the electricians can still make sense of the map since they know about these circumstances. Another difficulty is that other companies have equipment in the ground (telecom cables and water pipes). Even if this information is shared among the companies there is always be the risk that the maps differ.

4.4 Foreseeing

"Foreseeing" is using knowledge to make suggestions about the future. Let us consider one example. When the electricians are at a specific place they would typically project beyond the task at hand. For example:

> John says to the civil engineer at when they have finished a job: I can see that you are going to call us later this week to get those poles moved over there… I can call my boss and see if we could do it right away? The civil engineer gladly accepts the offer.

Here, John knows that there will be a need to “get those poles moved over there” in the near future, and for that reason, he could provide the client with a good offer. He uses his knowledge to “foresee” how things will proceed. It is tightly coupled to the circumstances that are unveiled for the electrician at a specific place and the local knowledge of what possible action to take.

4.5 Practical knowledge management: an initial sketch

Based on the empirical observations, we have started to develop a framework of “practical knowledge management.” The objective is to join the debate of knowledge management in a constructive way, that is to say make suggestions that may be helpful in practice.

We start by examining the relations between the sharing mechanisms and their utilisation. We are fully aware of the fact that these relationships may not be clear-cut. Nevertheless, there seems to be connections between the mapping

- from ephemeral into indexing,
- from persistent into diagnosing, and
- from stories into foreseeing.

One example of how ephemeral sharing maps into indexing is when the re-organisation of personnel relies on the plans of the day as stated at the daily morning meeting. For example, a morning meeting we observed involved a discussion about “how much personal a particular assignment will need.” Later during the day, we observed how Chris called Ronnie (the planner), requesting more people.

> Chris: “We need more people here. Roy said that they had it slow today. Should I call him and ask them to come?” He hands up and calls Roy.

In this situation, Chris explains for Ronnie that Roy had told him previously that he and his peers will not have a very high workload this particular day. When Chris finds out he needs more people, he (seemingly) therefore makes the suggestion (for Ronnie) to call Roy,

and in doing so, he explicitly refers to what Roy has told him. Thus, ephemeral sharing was mapped into indexing.

Persistent sharing mechanisms, such as maps and documents, were used for diagnosing. Most often, diagnosing took place in the field where the local conditions of the place could be matched with the records. The *filer* for instance, is an obvious support in some diagnosing processes.

The administrators have got a lot of persistent record at the office. When they are in contact with the electricians it is most often some information at the office that is needed, e.g., customers address or telephone numbers.

Stories contain a more abstract type of knowledge. The stories map into actions of foreseeing and diagnoses. Not only situations, problems and solutions are communicated in the stories, but also opportunities for giving better service.

A system based on these categories could support improvisation and help the electricians to gain an overview of the customer's needs and available resources. It could also be useful in some cases of minor emergencies for on-call duty. Bigger assignments, on the other hand, could still be centrally co-ordinated.

5. Discussion

Knowledge management and *organisational memory* have inspired important debates within CSCW. Often these are contrasted with reference to *local knowledge*. Bringing these perspectives together, we have proposed new categories of "practical knowledge management" that cover the layered abstraction and creative mapping of local knowledge to new places and procedures. Exploring knowledge management in the day to day work of electricians, we found four important ways in which it took place. To summarise:

- Sharing, i.e., the exchange of knowledge between parties.
- Indexing, i.e., one party explains to another what knowledge to retrieve.

- Diagnosing, i.e., two parties make sense of a situation, i.e., how it should be interpreted.

- Foreseeing, i.e., one party (or more) uses knowledge to project the future.

For each of these we would like to compare and contrast some central views of organisational memory and knowledge management literature:

- Making the argumentation explicit

Central to the argumentation of organisational memory is the notion of *rationale*, i.e., that supplying the motivation and structure of the "memory creation process" (Conklin and Begeman 1988; Yakemovic and Conklin 1990). It follows from this focus that an organisational memory system should try to make the *decision processes explicit* by capturing all essential aspects. In a sense, everything relevant should be saved and made available to anyone who might need it.

- Finding and interacting with experts

Another important objective of organisational memory is to support team members in finding and interacting with experts, and to accomplish such work more effectively (Ackerman and McDonald 1996; Karduck 1994).

- Putting knowledge in chunks

In knowledge management, the central objective is often to capture, organise and share the knowledge of employees for the achievement of a common, strategic goal. This objective necessitates breaking up knowledge in manageable chunks of various types, for instance, administrative data, facts and figures about the operations, declarative knowledge, procedural knowledge, and analytic knowledge (Edwards and Mahling 1997).

- The roles of knowledge

In knowledge management it is also suggested that roles should be assigned for assuring relevance, accuracy, and periodically review of material. It is also suggested that various techniques from automated retrieval reasoning are central to this kind of learning organisation.

We will not, in the paper, argue against the central tenets of these two related disciplines. We think, however, that the picture

emerging from the following comments may be one on the basis of which some design implications may be drawn.

5.1 Sharing

In our case, mechanisms were implemented in attempts to persistently share knowledge through artefacts. On the other hand, we also found ephemeral sharing that was generated for a purpose and for dedicated receivers (explicitly seeking it). In this respects our findings are much closely aligned with those of Bannon and Kuutti (1996), Randall et al (1996), and Anderson and Sharrock (1993), who claim that knowledge is actively constructed for the individual in a particular context. We think that the shear volume of relevant information that conceivably could be captured in this case is prohibitive of the former approach.

Organisational memory has to assume that people are willing to be experts for "unknown others," thus taking on additional work by explicating and storing experiences in a common information space. A related assumption is that users of these systems will actually trust that they contain reliable, updated information. We did not observe knowledge explicating and storing activities that were external to work itself, i.e., at least not in a formal way. Thus, the practical use of knowledge as the manipulation of "knowledge chunks" does not agree with our observations. One central design implication would be to avoid redundant work imposed by explicating knowledge creating chunks as well as allowing users to partake in designating experts. The role of experts is one of several roles that were not explicit in the organisation. It seem hard to see what could be gained by forcing the organisation to allocate and manage roles such as "gatekeepers" and "knowledge librarians" (Edwards and Mahling 1997).

5.2 Indexing

In traditional organisational memory, "knowledge" is usually structured by a theory of argumentation. For some aspects of knowledge, in our case, such as rules and formal procedures, this is certainly an appropriate approach. Also, when the solution space may

be anticipated, i.e., when and if the problem relate to a specific geographical location only. We found in other instances, that people organise information in an individual and situated fashion. Thus, one design implication could be to allow personal, rather than general mark-up, indexing and organisation of the data captured.

It is often rightly assumed that people could be helped by experts (or rather get assistance) in solving problems. However, knowing who the expert is for the particular problem at hand, does not seem to be a matter of formal roles. We found that people often rely on informal, individual networks in combination with open requests. Knowing who should have been asked may not be known until after the problem is solved. Put simply, *knowing who* is sometimes a major part *knowing what*.

The ways in which we observed people collaboratively indexing knowledge, imply that it is a complex phenomenon, which is related to action rather than representation, and thus, cannot easily be broken down into general chunks. This also shows that all members of an organisation could serve as "experts," "gatekeepers," "knowledge librarians," etc., depending on the situation.

5.3 Diagnosing

Many organisational memory systems, such as Answer Garden, require that the requesters of knowledge know, in advance, how to formulate a question. This is often possible when the problem space is limited, for instance when it is (yet) an instance of a problem with new parameters (call the janitor first – what is the telephone number?). We found, however, that in many instances this is undertaken as a collaborative task, during which the problem-solving process is already ongoing. It is a known phenomenon in information retrieval that to express an interest in a formulised query is difficult (see, Marchionini 1995). This could mean that the organisational memory system should not be conceived separately from the task-support for these users. As Anderson and Sharrock (1993, p. 149) puts it: "patterns of knowledge and patterns of action define each other." These ideas are not so far from the work on "living design" (Terveen et al 1993) and social activity indicators (Ackerman and Starr 1995).

Diagnosing was a truly collaborative effort that seldom was preceded by one person explicitly trying to find and interact with an expert. It would be necessary to engage assistance in implementing and evaluating solutions as well as giving advice on a more abstract level.

5.4 Foreseeing

Traditionally, organisational memory has been concerned with the past. The majority of requests documented in our case similarly tries to resolve a current problem using previous experiences. We also found, however, some instances of future projections and business planning on the basis of situated problem solving. This could means that knowledge management should be seen in connection with a relaxed workflow design. Foreseeing is about creating possibilities for action and thus making a business opportunity for the people present in the situation. Thus, it is not always necessary to find experts that fit the "problem," but rather to define a problem that may be solved by the experts that are already engaged.

5.5 Conclusion

This paper adds to the existing body of empirical work in CSCW by dealing with a novel class of work, namely mobile service work. Our empirical results provide a sound grounding for further design and evaluation of distributed support for practical knowledge management in a mobile setting. The analysis challenges some conceptual underpinnings of work in both sides, but, at the same time, this critique opens windows of opportunities for implementers of CSCW systems in what seems to be, for researchers and consultants alike, a very compelling domain.

Acknowledgement

This research is a part of the Mobile Informatics programme that is funded by the Swedish Research Institute for Information Technology (SITI).

We would like to thank our colleagues at the Viktoria Institute, especially Magnus Bergquist and Peter Ljungstrand. Finally, thanks to Wayne Lutters for a thoughtful review.

6. References

Abecker, A., A. Bernardi, K. Hinkelmann, O. Kühn, and M. Sintek (1998) "Towards a Technology for Organizational Memories," *IEEE Intelligent Systems*, 1998, 13 (3) pp. 40-48.

Ackerman M.S. and B. Starr (1995) "Social Activity Indicators for Groupware," In *Proceedings of the 8th ACM symposium on User interface and software technology*. pp. 159-168. Pittsburgh, PA: ACM Press.

Ackerman M.S. and C. Halverson (1998) "Considering an organization's memory," In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work*. pp. 39-48. Seattle, WA: ACM Press.

Ackerman, M.S. (1994) "Augmenting the organizational memory: a field study of answer garden," In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*. pp. 243-252. Chapel Hill, NC: ACM Press.

Ackerman, M.S. and D.W. McDonald (1996) "Answer Garden 2: Merging organizational memory with collaborative help," In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*. pp. 97-105, Cambridge, MA. ACM Press.

Anderson, R. and W. Sharock. (1993) "Can Organisations Afford Knowledge?" *Computer Supported Cooperative Work*, 1993, 1 (3) pp. 123-142.

Bannon, L.J. and K. Kuutti (1996) "Shifting Perspectives on Organizational Memory: From Storage to Active Remembering," In *Proceedings of the 29th Hawaii Conference on System Sciences. Vol. IV*. pp. 156-167. Los Alamitos. IEEE Computer Society Press.

Bellotti, V. and S. Bly (1996), "Walking away from the desktop computer distributed collaboration and mobility in a product design team," In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*. pp. 209-218, Cambridge, MA: ACM Press.

Bowers, J., J. O'Brien and J. Pycock (1996) "Practically accomplishing immersion: cooperation in and for virtual environments," In *Proceedings of ACM 1996 Conference on Computer-Supported Cooperative Work*. pp. 380-389, Cambridge, MA: ACM Press.

Conklin, J. and M.L. Begeman (1988) "gIBIS: a hypertext tool for exploratory policy discussion," In *Proceedings of ACM 1988 Conference on Computer Supported Cooperative Work*. pp. 140-152. Portland, OR: ACM Press.

Edwards, D.L. and D.E. Mahling (1997) "Toward knowledge management systems in the legal domain," In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*. pp. 158-166, Phoenix, AZ: ACM Press.

Foner, L. (1997) "Yenta: A Multi-Agent, Referral Based Matchmaking System," In *Proceedings of the first international conference on Autonomous agent*s. pp 301-307, Marina del Rey, CA: ACM Press.

Geertz, C. (1983) *Local Knowledge: Further Essays in Interpretive Anthropology*, New York: Basic Books.

Glaser, B. and A. Strauss (1967) *The Discovery of Grounded Theory*. Chicago: Aldine.

Greif, I. (1998) "Everyone is talking about knowledge management, " In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work*. pp. 405-406. Seattle, WA: ACM Press.

Hammersley, M., and P. Atkinson (1993) *Ethnography. Principles and practice*, Second edition, London: Routledge.

Heath C. and P. Luff (1991) "Collaborative Activity and Technological Design: Task Coordination in London Underground Control Rooms," In *Proceedings of the Second European Conference on Computer-Supported Co-operative Work*. pp. 15-22. Amsterdam, The Netherlands: Kluwer Academic Publishers.

Karduck, A. (1994) "TeamBuilder: a CSCW tool for identifying experts and team formation," *Computer Communications*, 1994, 17 (11) pp. 77-87.

Kautz, H., B. Selman and M. Shah (1997) "Referral Web: combining social networks and collaborative filtering," *Communications of the ACM*, 1997, 40 (3) pp. 63-65.

Kovalainen, M., M. Robinson and E. Auramäki (1998) "Diaries at work," In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work*. pp. 49-58. Seattle, WA: ACM Press.

Kristoffersen, S. and F. Ljungberg (1998) "MobiCom: Networking dispersed groups," *Interacting with Computers,* 1998, 10 (1) pp. 45-65.

Kristoffersen, S. and T. Rodden (1996) "Working by Walking Around. Requirements of flexible interaction management in video-supported collaborative work," In *Proceedings of Human Computer Interaction*. pp. 315-329. London, UK: Springer-Verlag.

Luff, P. and C. Heath (1998) “Mobility in Collaboration,” In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work*. pp. 305-314. Seattle, WA: ACM Press.

March, J.G. (1991) “How Decisions Happen in Organizations,” *Human-Computer Interaction*, 1991, 6, pp. 95-117.

Marchionini, G. (1995) *Information Seeking in Electronic Environments*. Cambridge Series on Human-Computer Interaction. Cambridge University Press.

McDonald, D.W. and M.S. Ackerman (1998) “Just talk to me: a field study of expertise location,” In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work*. pp. 39-48. Seattle, WA: ACM Press.

O'Leary, D.E. (1998) “Enterprise knowledge management,” *IEEE Computer*, 1998, 31 (3) pp. 54–61.

Orr, J. (1996). *Talking about machines: An ethnography of a modern job*. Ithaca, NY: ILR Press, an imprint of Cornell University Press.

Patton, M.Q. (1990) *Qualitative Evaluation and Research Methods.* New York: Sage.

Randall, D., J. O'Brien, M. Rouncefield and J.A. Hughes (1996) “Organisational Memory and CSCW: Supporting the ‘Mavis Phenomenon’,” In *Proceedings of the 6th Australian Conference on Computer-Human Interaction*. Hamilton, New Zealand: IEEE Computer Society.

Randall, D., M. Rouncefield and J.A. Hughes (1995) “Chalk and Cheese: BPR and ethnomethodologically informed ethnography in CSCW,” In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work*. pp. 325-340. Stockholm, Sweden: Kluwer Academic Publishers.

Rouncefield, M., J.A. Hughes, T. Rodden and S. Viller (1994) “Working with ‘constant interruption’: CSCW and the Small Office,” In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*. pp. 275-286. Chapel Hill, NC: ACM Press.

Terveen, L.G., P.G. Selfridge and M.D. Long (1993) “From ‘folklore’ to ‘living design memory’,” In *ACM 1993 Conference proceedings on Human factors in computing systems.* pp. 15-22. Amsterdam, The Netherlands: ACM Press.

Weick, K.E. (1979) *The social psychology of organising*. Second Edition. McGraw-Hill: New York.

Yakemovic, K.C.B., and E.J. Conklin (1990) “Report on a development project use of an issue-based information system,” In *Proceedings of ACM 1990 Conference on Computer Supported Cooperative Work*. pp. 105-118. Los Angeles, CA: ACM Press.

Forth paper[12]

Empirically informed knowledge management systems in mobile domains

Abstract

This paper reports from an empirical study of how knowledge management takes place in the work of news journalist at a radio channel. Based on the field data, we have in cooperation with the work force studied derived implications for the design of knowledge management systems for mobile workers. The implications focus on the *evolving, dependent tasks of news journalists*. Having reported the implications from the fieldwork, we relate our findings to the so called "repository view" of knowledge management, which dominates the field.

[12] Fagrell, H. and F. Ljungberg (2000) "A Field Study of News Journalism: Implications for Knowledge Management Systems," Accepted for publication in *Proceedings of the Sixth Biennial Participatory Design Conference*, New York, NY: CPSR and ACM Press.

1. Introduction

Knowledge management has received much attention in research recently. So far, much research on the topic has been theoretical and rather abstract. Topics of concern involve the epistemology of organisational knowledge (Cook and Brown 1999), the sharing of knowledge in communities of practice (Wenger 1998), the categorisation of organisational knowledge (Blackler 1995), comparative reviews of organisational learning and knowledge management concepts (Swan et al 1999). This research may be valuable in many situations, but it does not offer much guidance for designers and users of knowledge management systems. Another strand of research on knowledge management conducts empirical work on micro-level, analysing knowledge in organisations (Fagrell et al 1999; Randall et al 1996; Anderson and Sharock 1993; Ackerman and Halverson 1998). Clearly, the level of detail in this research is more suitable for designers. However, the research in this category has not been so concerned with design of knowledge management systems. In most cases, the empirical results have been used to criticise current systems and theoretical frameworks. Yet a third strand of knowledge management research has been technologically oriented. The main focus has been on developing knowledge management systems. This research has rarely been based on empirical investigations of how knowledge management really takes place in practice. Accordingly, the assumptions on which the systems are based on are not informed by empirical work.

Against this background we call for the need to conduct empirical studies of knowledge management with the explicit objective of informing the design of knowledge management systems. This call is answered by the research reported in this paper.

The fieldwork we report was conducted at Radio Sweden. Our study investigated the work of news reporters and the ways in which they use and manage knowledge during the working day. Previous PD and CSCW research on the news domain is represented by Kensing et al (1997) and Bellotti and Rogers (1997), but their focus has not been

on *mobile* reporters. We focus on mobile personnel for several reasons. First, many jobs in western society seem to become increasingly mobile. Accordingly, the need to understand mobile work and design applications that really match the mobile setting is emergent. Second, mobility has been largely overlooked by designers of CSCW systems (Luff and Heath 1998). Most research has been concerned with stationary settings; thus there is a need to broadening the scope of the field. Third, in the case of knowledge management, the main focus has been on the PC and stationary settings. However, the work of key personnel in many organisations is highly mobile.

Knowledge management has been used to describe many different things, ranging from "organisational learning efforts to database management tools" (Ruggles 1998). Needless to say, there is no general agreement on a definition of the concept. However, there has been a tendency in the literature to treat knowledge management in a Tayloristic way, which may not only result in unsuitable IT systems but also deskilled workers. Furthermore, knowledge management and related concepts, like organisational memory and organisational learning, are sometimes used inter-changeably. We view knowledge management as a social, dynamic process that has to be considered carefully, and in cooperation with workers to be understood correctly.

Clearly, in a work environment the ability to manage knowledge, i.e., find, create, use, is important to accomplish work. The objective of the study we present in this paper is to investigate this in practice with the ambition to provide designers of mobile knowledge management systems with design implications.

2. Research background

An introductory textbook on news journalism suggests, in a top-down fashion, that events which fall into the following categories qualifies as news items (Mencher 1997):

- Events that are likely to affect many people, e.g., a hurricane.
- Events that are immediate, e.g., election results and traffic jams.

- Events involving well-known people or institutions.
- Events in the circulation or broadcast area.
- Events that reflect clashes between people or institutions.
- Events that deviate sharply from the expected and the experiences of daily life
- Events that are being talked about

In order to help designers of interactive systems, these types of lists need to be complemented with detailed studies of how people actually go about to do their work. This has been much recognised in CSCW (e.g., Button and Dourish 1998) and PD (e.g., Ljungberg et al 1998; Kensing et al 1997; Crabtree 1998).

The empirical study we conducted used observational techniques for data collection. All together we observed the work for approximately 300 hours. We made copies of many of the documents produced and discussed by the journalists.

Having transcribed the field notes, we started the coding of the empirical data. We went through the data, made notes, and labelled data of interest. The analysis of the empirical data was aimed to identify where and how knowledge management occurred in the day-to-day work of the journalists. In the analysis, we used HTML to tag the field data and make links. We involved the workforce studied in the analysis of the empirical data. We presented our preliminary results for the journalists and they had a chance to make suggestions to assure a correct and effective analysis. The suggestions we outline in this paper should be viewed as the result of a joint effort between the authors and journalists.

The empirical research was conducted at a public service radio station in Gothenburg, Sweden, during the period of the Swedish election in September 1998. The station employs about 100 people and broadcasts nationally and in the local region. The study mainly concerns the channel responsible for the local region. A recent survey shows that 64% of the population in the region (approx. 650,000 people) listen to the channel every week. [13]

[13] The source is RUAB (http://www.ruab.se/Engltxt.htm) a company that continuously carries out radio audience measurements.

The study was focused on three radio programs:

- News update. The news broadcast every hour.
- Gothenburg Direct. A program that covers what happens in the region. It involves interviews, music, and discussions.
- Election Extra. A program dedicated to the upcoming election. It covers news, interviews with politicians, etc.

The three programs run separately, with dedicated staff, but the news collected is often shared. The idea to investigate and report a particular news item can originate from a telegram from a news agencies like Reuters, an article in a news paper, a tip-off from the public, press releases, information found on the Web, etc.

The station is an open plan office (see figure 1). In the middle is a large table where the journalists join together in daily meetings. The studios and the work desks have PCs that share disks and printers via local area network. All PCs have Internet access, and there is a stationary telephone on each desk. The field journalists (figure 1) have mobile phones, but also a microphone that is connected to a transceiver on a rucksack. This equipment broadcasts to a van where the signal is amplified and sent to the studio.

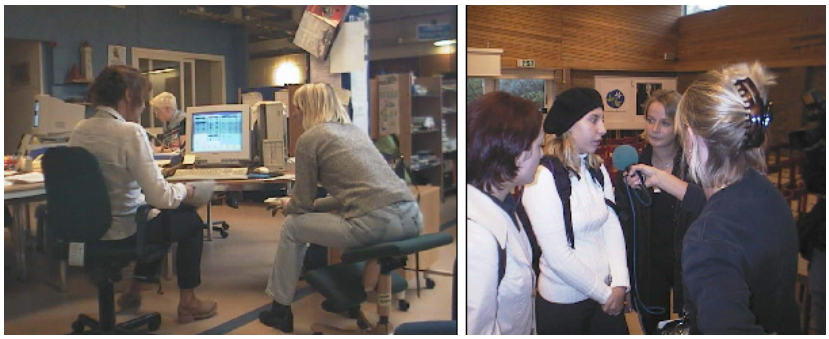

Figure 1. A discussion in the office landscape (left) and a field interview (right).

To keep track of upcoming events the journalists use “diaries.” A dairy is a paper-based shared calendar. An entry of a diary can be the date of a press conference, a court verdict, etc. This information can also be found in an old, but widely used document management system called MANUS. The MANUS system also contains other types of documents, e.g., telegrams, broadcast reports and program manuscripts.

3. Empirical results

In this section we report the main results of the empirical study. Let us first give a brief overview of the work process of the radio station.

3.1 The Work Process

Every morning at 8.30 there is a scheduled meeting that is lead by the editor of the News update. The meeting takes place at a large table in the middle of the office. Journalists from all local programs participate, usually about 10-15 people. Journalists join and leave as the meeting progresses. The editor has a prepared list of potential news items that is passed to the participants. The list, which has been assembled using dairies, MANUS, local newspapers, etc., serves as a starting point for discussions about "what should be done today."

When the meeting ends at about nine o'clock everyone has an overview of the preliminary schedule of the day. At this point the three programs usually have short meetings, where the discussion from the morning meeting continues. The journalists also decide who should do what and how activities should be co-ordinated.

It is important to assure originality of news items. This is often initiated at the morning meeting, but needs often to be continued by the journalists individually throughout the day.

A program can contain all kinds of formats, even for a single news item. The formats are:

- Recorded interviews: Recordings from telephone or field interviews are either transmitted from the field to the studio or recorded on tape. The recordings, called "clips," are always edited before broadcast.

- Live interviews: A live interview can be done in the studio, on the telephone, or in the field.

- Talkers: Talkers can always be found as part of a program, e.g., introducing other material or just reading important messages. A talker is either recording or read live by a journalist.

During a live interview the possibilities to change the focus is limited and it is up to the individual journalist to frame the conversation and ask the appropriate questions. When the material is recorded, however, there are some possibilities to conform the material. The purpose of the editing is to make the report more understandable in the time frame of the programs. Some final touches are also done to adjust the report so that it fits in the overall agenda of the program.

The material is sometimes insufficient, i.e., the right questions were not asked. This is not good since the collection of material is very time-consuming and the editing are often done in the last minute. To have a gap in the program where a clip was supposed to be is very frustrating for the journalists, but they usually manage it by rearranging the parts of the program and use some backup material, e.g., play extra music.

Let us now describe how the journalists in general and the field reporters in particular manage knowledge to get the work done.

3.2 The Preparation of News items

Our primary focus describing the management of knowledge, in the work of the journalists, is on the process of preparing news items. This process involves two main tasks, which we call *exploring* and *elaborating* on news items. Exploring is the process to investigate potential news items and initiate them. Elaborating is the process of researching and framing the initiated news items to produce reportable news items.

Exploring and elaborating are knowledge intensive processes, where the journalists continuously access information resources and colleagues to identify, assess, validate, frame, research, news items on which they are working. Furthermore, the processes are interrelated in that you while exploring a news item also partly elaborate on it, and vice versa. The processes are also co-operative in many ways. Most work tasks are collective and you would not be able to do so much work on an individual basis. The explicit distinction between exploring and elaborating is, of course, a simplification. There are cases when it is not so easy to know when the journalists explore and elaborate on a news item.

The management of knowledge in the preparation of news items relies on both people and information systems. In reporting the ways in which this happens, we highlight some important questions that the journalists need to consider. By understanding these questions - how they are managed - and the knowledge involved, we may be well prepared to inform the design of knowledge management systems.

3.2.1 Exploring news items

Exploring is the highly collaborative process to identify, assess and validate news items. The morning meeting serves as an important forum for doing this. At the morning meeting, the journalists bring up, discuss, criticise, assess, validate, news items in a highly collaborative fashion. However, the exploration of news items continues through out the day. The process of exploring can be understood as a combination of *what* news items to report and how to report it. These two issues are thought of in a dualistic way: the journalists do not think "what" without thinking of "how," and vice versa. Below let us describe important considerations in the process of exploring potential news items.

What have others done?

One important source for news items is what other media and competitors have done. In many cases, the journalists argue for a particular news item with reference to another other media or competitor. Consider the following excerpt from a morning meeting.

> Pedro says: "I would like to make a case of VF ice hockey and build on what they said in Metro" [Metro is a free newspaper that is available on the tram]. Mary answers "Great … to move further on I mean … since probably a lot of people have read it."

In this case, the ice hockey news was brand new, but it had already been reported by Metro. This is not a problem, however, but an advantage. The reason is that many people have read about the story in the newspaper this morning. The topic is new and people are familiar with it, and for that reason they want additional information, another analysis, and so on. Accordingly, it is a very good topic to move further for the radio journalists.

What have we done?

It is important to follow up on stories that have been reported internally. Many stories are not one shot, e.g., a court trial can run over a long period and are reported as it progresses. There are also issues that are always important, e.g., health care quality and economic politics. Most journalists have the responsibility of being especially well-informed on one such area. The following example illustrates a follow up discussion at the morning meeting.

> Erica says: "Anything new about that 'abused couple from last week' ... Alice?" Alice responds: "I don't know, the police did not want to comment on the accusations and that was it." "Tell us what happened," Thomas asks. Alice answers: "Well, it was a father and a son who were walking home from the church and got physically abused by someone... it happened to be police officers in civilian clothing doing some narcotics raid. Perhaps they [the couple] were mistaken for drug dealers?" Erica answers: "Can you ask them if they want to talk about it as guests in the studio?" "Well… sure, I need to check if there is going to be a trial... and I will talk to Roy about it," Alice responds.

In this case, Erica starts the discussion and gives the word to Alice who had been working on the topic. Thomas asks Alice to inform everyone about the case. This happens often and results in that most journalists know about what their colleagues are doing. Erica also suggests that Alice should talk to Roy, who knows a lot about police issues. Alice accepts Erica's suggestion on *how* to take the case further.

Is it unique in the repertoire?

It is important that the channel gives a fair proportion of attention for current issues. For example, some political parties are very focused on getting media attention, but it is preferable that all political parties are getting equally much attention. This is of course difficult to accomplish when the workload is high and some parties constantly keep in touch, while others are very difficult to reach. Consider the following example.

> At a morning meeting Lena says: "I can interview that fellow from the left wing party?" Annie [producer] answers: "Perhaps the left wing party have had bit to much attention this week? Adam had one [left wing politician] in the studio yesterday and Thomas reported from their press conference." Annie continues "I know that

> they are 'pushy' but there are other parties as well." "Yeah, you are probably right," Lena agrees.

Since Annie is the producer she had a better overview of the repertoire of the channel. Lena becomes convinced that she should leave the issue to rest that day. The producers are responsible for the repertoire, but a part for the morning meeting it is practically accomplished through ongoing conversation among journalists.

Is it valid?

All ideas for news items are validated collectively. The following example is an illustration of this.

> Erik says: "An upset emigrant called me and said that there has been too little information about the election in foreign languages… perhaps we should do something with it?" Adam [Election Extra] answers that "It's not true … there has been a lot…" Some other people agree and Harry says "all political parties have had gatherings for different groups… with interpreters… but only like three people showed up."

Here, Erik has been misinformed, but it is taken care of because of the others' knowledge about the issue. This minimises the risk that valuable resources are put on news items that are not accurate. It also minimises the journalists' ability to misuse their position, e.g., to help friends that would benefit from media coverage.

In this case a discussion was started on why emigrants does not seem to care about the local election. Some notes were than taken and stored in MANUS system under the category "issues-to-watch." Under the right circumstances it may become a news item at some other point.

Is it possible to do?

How to conduct a report is often a matter of what is possible to. An event can be considered too difficult to report in relation to its news value. In other cases the possibilities for a report must to be assessed in a little timeframe. This is illustrated in following example.

> Mary [the producer of the News update] gets a tip-off on the telephone that there are very long waiting time in the switchboard system at Sahlgrenska [The largest hospital in the region]. Mark and Carrie [News update journalists] say that they have the same

> experiences. Karin is on her way to Sahlgrenska right now to do another report, but Mary stops her. Mary asks Karin: “Could you please go to the switchboard as well and try to get an explanation and an interview?” Mary and Karin start a discussion. Mary recalls that she heard something about the switchboard system. Mary queries the MANUS-system and finds a telegram from couple of months earlier about it. The telegram contains the text: “progress without any disturbance” and the name of the responsible manager. Karin takes a note and leaves.

In this case, Mary considered the tip extra carefully since she knew that the problem would affect a lot of people. That the two colleagues, Mark and Carrie, had similar experiences imply that it is true. Furthermore, it is easy to report because Karin is going there anyway.

Mary's suspicion that she heard something about the switchboard system was confirmed by the telegram and the promise to “progress without any disturbance” could be used in an interview. The name of the manager may be useful for Karin. The story was also, predicted to fit the genre of the News update.

3.2.2 Elaborating on news items

This category of observations concerns the ways in which initiated news items are researched and framed. As more information about a news item gets known it may drift in focus, relevance and validity. Nevertheless, when a new item enters the elaboration phase it almost always results in a report, even though the initial plan and expectations may differs from the final result. Below let us describe important considerations in the process of elaborating on news items.

What have others done?

It is important to have an overview of what has been done previously by others when framing a news item. In the following example Adam is going to report about an airport, Säve, which has been threatened to be closed down.

> Peter says to Adam while pointing in a newspaper: “It says here that Säve is expanding... more helicopters traffic.” Peter goes to Adam's desk and they start to search in the external archives for “Säve” and a name that was mentioned in the newspaper article. Adam and Peter browse the results and Adam says: “Oh yeah, I remember... I guess it's best to just call them.” “I agree, a live

> interview at 14.35 would be great," Peter says and hurries away. Adam makes some notes before he calls.

Peter knew about Adam's task, thus he recommended the newspaper article. Adam's search gave him an overview about the topic that was enough to continue. Adam could rediscover the details without having to go through complete and exhaustive records. The information came from several sources, i.e., the newspaper and external archive system, which would be difficult to include in a single system.

What have we done?

It is also important to know what has been done internally on topic to facilitate the research and framing. In the following example Emma wants to elaborate on a new framing.

> Emma walks in a hurry to Annie and says: "... about the angle on the Nazi issue. How come the extreme right wing has grown stronger in Europe?" Karl overhears and says: "... extreme right wing has always been strong in Europe." "Yea right, but have we done anything on it," Emma continues facing Annie. "I don't know," Annie says and queries the MANUS-system. Annie browses the results and says: "No, seems as if we haven't done anything about that... Peter has been on to the Nazis..., Steve as well." "Ok, great I will go and talk to Peter then," Emma says and walks to Peter's desk where she saw him recently.

In this case Emma was in a hurry with little time for preparations. The new angle was collaboratively validated with Annie and Karl who seemed to approve. The search in the internal archives revealed that Peter had been working on the topic earlier. It was likely that he knew more about the topic than Emma did. Thus, talking to Peter could increase quality of the report. Steve was also mentioned, but since Emma had seen Peter recently, he was the obvious choice.

Are there any new issues that should be considered?

Things that are planned in the morning may turn out to be irrelevant later during the day as more things arise. The following example illustrates when new facts arise, which affects the task.

> Mark's phone rings. It is a police officer who comments on traffic accident report that was just in the broadcast. Mark exchanges a few word and hangs up. Mark queries the archives. Then, he says to Adam and Jessica: "It was a police officer who said that the

> traffic report was wrong. They have sent a new telegram about it. I can fix it before next broadcast." "Great, I missed that telegram... peculiar title," Adam replies and continues "Erica is at lunch now, but is going to continue on that track. Send her an email."

Here, some facts were broadcasted that was not correct. This happens quite often for accident reports, where the uncertainties are usually high. Mark informed Adam and Jessica about the change. Adam told Mark to inform Erica as well, but since she was at lunch, Mark sent her an email. The event rapidly changed Erica's possibilities of doing her task.

How does it relate to the repertoire of the channel?

To care about the repertoire of the channel is especially important throughout the day when the programs run concurrently. Two programs should not cross-report news items without different foci. The following example illustrates when it has broken down, but is repaired:

> Alice who works on the program "Election extra" arranges so that a political scientist could be interviewed about the possible outcomes of the election. A bit later on, Alice becomes aware that "Gothenburg direct" is going to interview a political scientist too. This makes her concerned that there may be "too much" political scientists. A bit later Alice approaches, Erica, the producer of Gothenburg direct asking about the political scientist in a gentle way. Erica says that they have dropped that part of the program. This makes Alice happy…

The journalists at a program needs to know what is being done in other programs, and visa versa, so that the overall picture can be cared for. This is normally done at the morning meeting, but in this case it had broken down. The reason for this was that Alice had not mentioned that there was going to be a political scientist in the program, since it was not settled at that point. It was not settled at the morning meeting for the other program either.

What is possible to do?

It is especially difficult to research and frame news items for journalists in the field. It is also more difficult to plan what may happen in the field. If the plan for a news item is distorted the journalist must reframe it and to be able to create a quality report. If

there were to be no report it would be very expensive, since the resource could be used for another news item instead. The following example illustrates this.

> John is in the field and has just finished a task when Pedro calls on his mobile phone. Pedro tells him that the board of an ice hockey club, Frölunda Indians, has just announced that they are going to give a press conference. John remembers an article from a morning newspaper "The Metro," where the club was accused of tax avoidance when paying salaries. Pedro asks John to go there and do a report for the 12:30 news. At the site John waits outside. He chats with some people from newspapers, television, etc. When the board of the club arrives, John asks the chair of the club, "Could we do an interview for the 12.30 news?" The chair replies immediately: "No, I won't give any interviews today." The press conference starts and John puts a microphone on the table in front of the board of the club (see figure 2). He also makes a quick call on the mobile phone to the studio, informing them about the chair seems unwilling to give interviews. The press conference starts. John asks the first question and a cascade of questions from other journalist's follow. The microphone records the conversation and transmits it to the studio.

Figure 2: John at the ice hockey clubs press conference (left) and reporting after it from the field (right).

> John calls the studio (again) and says, "I do not think Hasse is will give an interview… Perhaps we should use a recording … I will try again to get Hasse" This happens ones again during the press conference. The press conference ends. John waits for the chair outside the room. He tries to persuade him, but it doesn't work. He

> calls the studio and suggests that Peter Anderson, who knows everything about ice hockey, should be contacted. He could assist the editing, he knows how to frame the story, etc. During the last call John assists the editing by suggesting where important things seemed to have happened. John sits on a bench outside the arena (see figure 2) and composes how to report the event. After a while, the studio records John reading the report. This is finished at 12.20. The report is broadcasted 12.30.

What happens here is that the chair refuses to be interviewed and puts John in a difficult position. He must complement the Metro article with some additional analysis or information. Almost everything that was said at the press conference was in the article and John did not know very much about sports politics. However, by noting what the chair said and by giving an introduction, John manages to produce material that - after some editing – is of reasonable quality. John said after the press conference, that a talker would have been “too passive” adding little value compared to the newspaper article. He also said that he knew that Peter Anderson who knows everything about ice hockey and could assist in the framing of the news item. Unfortunately, Peter was not accessible, because he was not on duty.

Later that day the topic was debated in the program Gothenburg direct. At that point, Thomas, a journalist with special knowledge on the issue, was present which resulted in a good discussion.

4. Implications for Design

In this section we discuss what the empirical results imply for the design of knowledge management systems. We seek to summarise the results from the study that we think are most important for designers of interactive knowledge management systems, discuss what implications these observations may have for design, and relate them to the literature.

Observation #1: *The main purpose of making searches in archives is to get an overview of information, not to find "the right answer"*

In the process of exploring and elaborating on news items, the journalists frequently used archives of different types, e.g., the MANUS system. However, most often the use of archives aimed at getting an overview of the issues of concern, not finding specific information about something. To illustrate this, let us return to the airport case described previously. The journalist starts to work with the news item, and makes therefore a search in the archive (MANUS), he would be more interested in the list of headings returned from the system than the individual instances as such. In a sense, the list of headings is the answer. This is often enough for the journalist to continue to explore and elaborate the news item. The overview helps the journalists to put the news item into context. The implication of this observation is as follows:

Implication #1: *Provide easy access to surveys of archives based on the work item in focus*

So, how does the implication match the literature? In fact, we can trace similar thoughts in a wide range of research contributions. For example, according to the empirical results reported by Bowker (1997), the ways in which nurses use medical records sits well with our implication. The main use of the record is overview and classification to facilitate remembering. March (1991) provides a similar line of reasoning in an essay on the problems of rational choice. However, the focus of his essay is not exactly knowledge management.

In literature under the label of knowledge management, the implication above has rarely been considered. In most cases, the focus has been on providing detailed information exclusively (e.g., Edwards and Mahling 1997). In well-known knowledge management systems, such as Answer Garden (Ackerman 1994; Ackerman and McDonald 1996), Project memory (Weiser and Morrison 1998), and Designer assistant (Terveen et al 1995), there are search features. However, the ways in which users can survey search results are very limited. Clearly, these systems are much more oriented towards finding "the matching record" (which sometimes is very important) than providing the user with an overview. There are some exceptions, however, as represented by gIBIS (Conklin and Begeman 1988) and RepTool

(Jordan et al 1997). These systems offer features for overview information. However, the information to overview is very different from our case: design rationale in long-term projects. Accordingly, the possibilities to overview the types of information that the news journalists are interested in have not been provided by any system up to now.

Observation #2*: Information in archives is often used to locate expertise*

The journalists often have difficulties to remember who knows something about a special issue. Therefore, in order to locate expertise, they often use the internal archives to investigate who have published on a particular topic, i.e., who can assist them in exploring and elaborating on a news item. Quite often the names of the author found in the system reminds the journalist of whom to ask. Usually, the colleague to ask is someone known, but for one reason or another he or she could not be remembered before seeing the name in the archive. This implies the following implication:

Implication #2*: Attach authors to all information. This would help users to find the expert*

A similar finding was reported by McDonald and Ackerman (1998). They found that programmers often used the change histories of source code files to identify expertise. The importance of "knowing who knows" has been highlighted by, e.g., Randall et al (1996). In our case, the main reason why the journalists want to discuss the news items with others does not seem to be a matter of incompetence. Rather, it is often a question of getting a "second opinion" and another context of the work.

In Answer Garden 2, there are features that route users to expertise if the system does not contain solutions to their problem (Ackerman and McDonald 1996). Such features are also provided in the system described by Streeter and Lochbaum (1988). The user queries the system with a keyword, and receives a list of people who can assist. Similar features are provided by the Referral Web system, described by Kautz et al (1997). The importance of expertise location has also been acknowledged by the designers of the Designer assistant (Terveen et al 1995) and the Project memory (Weiser and Morrison 1998), but no explicit support is included in the systems.

***Observation #3**: Expertise location is determined by who is accessible*

Expertise location is dependent on who is available. Accordingly, knowing the author of certain information is not enough. One also needs to know if they are available for interaction. In the field, in particular, the journalists have problems to know that (if they even know whom to contact!). They often used the telephone to find out, which is probably not the most efficient way. Furthermore, when the journalist is in the field, it is not always easy to articulate the local conditions when talking to the expert. These observations imply:

***Implication #3**: Provide easy access to available authors of information, especially for field reporters*

The problem of locating people in a distributed and mobile work context was found in a field study by Bellotti and Bly (1996). McDonald and Ackerman (1998) also found the importance of geographical distance in expert location. The importance of distance in collaboration has also been reported by, among others, Kraut et al (1993) and Randall et al (1996). The political problems of expertise location, e.g., status, have been discussed by, among others, Allan (1977) and Randall et al (1996). We did not observe such problems in our study.

There are PC based applications like ICQ (www.icq.com) that help people to find out if colleagues are available (or busy, etc.). However, these systems are not very sophisticated when it comes to expert location. Furthermore, their support for mobile workers is very limited. The Babble system (Erickson et al 1998) can help people to get an overview of ongoing conversations between people, which could be used to know who is available. However, using Babble in our case would not be straightforward in this context. For example, there would be very many conversations going on simultaneously. Furthermore, the mobile work context would be difficult to deal with (e.g., how to represent and change the setting of a journalist?). As far as we are concerned, there is no system that provide sophisticated mobile support for expertise location.

***Observation #4**: It is important to handle overlapping news items*

The work at the radio station is parallel, but also distributed and mobile. As a consequence, there is sometimes a risk of cross reporting news items, i.e., that two programs reports the same issue. Cross

reporting matters for the repertoire of the radio channel. For example, a political party should not receive too much attention, there should not be too many interviews with ice hockey players, and so on. Therefore, it is very important for the journalists to know the work of the colleagues. Due to the mobile and distributed organisation of work, but also because the work items evolve (and new are initiated) through out the day this is not always easy to handle. This implies:

***Implication #4**: Help the journalists to keep each other informed about the (evolving and new) work tasks*

This implication relates to issues of co-ordination Schmidt and Bannon (1992), notification (Patterson et al 1996), and session management (Edwards 1994), which have received much attention in CSCW. Co-ordination is about coping with interrelated work tasks. Notification deals with acknowledging events in collaborative systems. Session management is the ways sessions are set up and handled in collaborative systems.

Despite the research on these issues in CSCW, the work on how to handle this in a mobile context has received little attention. Nevertheless, in a knowledge intensive work environment with mobile staff such support seems vital.

***Observation #5**: New information on the task of concern may be very important*

In time dependent work, such as news journalism it is very important to be aware of new information of the topic of concern. For example, if a journalist starts to work on a news item that is planned to be reported three hours later, it may be very important to know what the news agencies, television, other radio channels, etc., report on the topic meanwhile. In other words, new information on the topic needs to be considered in order to prepare a good news item.

It is important to notice that the news items often evolve during the working day, and that the journalist quite often work on the field. From this it follows that:

Implication #5*: Provide staff with new information based on the (evolving) items on which they work. Continuous support is important, and it needs to be mobile*

Modern editorial software, e.g., ENPS (www.enps.com), allows people to define filters that redirect news telegram to their desktop PC. These filter support the long-term interest of users, assuming that the defined profile actually match the interest of users. In our case, the interest in new information is not mainly long-term, even though the journalists obviously have long-term interest (cf., Belkin and Croft 1992) The interest is not short-term either, but rather "task-based." The interest the journalists have in new information is based on the task (news item) on which they work, which is temporary (until the item is reported) and evolving. This type of interest is not supported by current systems. The only system that is a bit related to this is Darwin (Kristoffersen and Ljungberg 1998), of which, the design was informed by a study of a dispersed and mobile IT-support group.

5. Discussion

Systems aiming to support the management of knowledge in organisations increasingly called "knowledge management systems" have traditionally assumed a "repository view" of knowledge (Bannon and Kuutti 1996). From such a perspective a knowledge management system is basically a collection of solutions to problems. The central idea is that the system provides solutions to problems experienced by the users. In cases where the system cannot assist, the user would typically search for the solution elsewhere and update the system with the new piece of knowledge. Some systems also provide features for expertise location. The idea is to assist users to find people who could help solve their problems in cases where the solution cannot be found in the system.

Many researchers have criticised this perspective of knowledge, stressing the importance of context, and the emergent collective nature of knowledge. In many cases, the critique against the assumptions of which current knowledge management systems builds has been rough. Few of these contributions come up with suggestions for how to

improve current systems. They do not make the life much easier for designers and users of knowledge management systems.

The objective of the research presented in this paper is to be a little more specific about how to improve the design of knowledge management systems. We believe that it is difficult not to consider repositories in this setting. Even though no repository is perfect, they do provide valuable help in some situations, and are definitively worth developing further. As Weick and Westley (1996, p. 444) note in a discussion of the relation between organisations and repositories:

> "If organisations are repositories, they are flawed sources of guidance, both because storage is imperfect and because retrieval is an act of reconstruction. Memory is imperfect twice over, which is bad enough. Even worse, organisations face chronically 'novel present.' To rely on repository of built-up knowledge is to rely on approximations rather than certainties. To underscore the approximate character of prior learning, self-designing systems apply lessons of the past while simultaneously questioning their relevance."

So, how could we complement the repository model to provide a better platform for knowledge management?

From problem to evolving task

The point of departure for how the journalists managed knowledge in their work was not a problem, which often is assumed in knowledge management, but the news item, or task, of which they were concerned. Furthermore, the task evolves over time; it is not a fixed entity.

From independent to dependent input

The repository model of knowledge management does not assume any dependencies between "inputs" to the system. Our study implies the opposite: users are very interested in the input of others, simply because they need to make sure that different types of double work does not happen. Therefore, the repository model needs to be put in a co- ordination context, where the management of dependent tasks and inputs is taken seriously.

From information as such to information linked to authors and their accessibility

We could interpret the repository model as promoting a perspective where "the information speaks for itself," the user has a problem, the system contains the solution. Why then would the author of the information be important? As we have seen in our study, it is often very important to know who are the authors. It is also important to know whether or not they are accessible. Therefore, we should add author and accessibility to all information in the repository.

From short and long-term interest to task-based interest

Tasks typically evolve through out the day, and the journalist would be interested in new information related to a particular task as long as the journalist works on it. We can call this *task-based interest* in information. Task-based interest is different from both short term and long term interests, which most often are the only types of profiles discussed.

From "hits" to hits and overview

The main focus of the repository model is the "hits" that a user receives from a query. As we have seen, the overview of hits could be equally important in many cases. Therefore, we need to expand the model from focusing on hits (more or less exclusively), to also consider the overview.

From stationary access to flexible access

The work environment in many organisations of today (including the radio station) is flexible in that people work both at the desktop and while on the move. This requires flexible access to the knowledge management system, i.e., both mobile and stationary clients.

Based on these ways of extending the repository model of knowledge management, we are currently developing a flexible knowledge management platform that:

- starts out from evolving, dependent tasks
- with all information linked to authors and their accessibility.
- with filters that are task-based

- that shows hits and overview, and
- which provides flexible access (stationary and mobile)

6. Summing up

The objective of the study we present in this paper was to investigate knowledge intensive work in practice in order to provide designers of knowledge management systems with suggestions for design. Using qualitative methods, we found that the preparation of news items involved the process of exploring and elaborating. Knowledge management in these processes involved information in archives and colleagues. Based on these observations we elicited five design implications, which is related to the literature. We also used the data to suggest how to enhance the repository model of knowledge management. This model has been criticised in the literature. However, it is our claim that it has important aspects that are worth keeping, and therefore, we ask how the model could be enhanced and improved. Based on the empirical analyses we made the following suggestions: First, the starting point should be *evolving tasks*, not static problems only. Second, the inputs to the system should be treated as *dependent* of each other, not independent. Third, the information in the repositories need to be *linked to authors* and their *accessibility* (the information as such is not enough). Fourth, the user interest in the repository is *task-based* (interesting as long as they work on the task), and not only short term or long term. Fifth, the overview of "hits" in the repository is important, not the individual hits only. Sixth, users need flexible access to the system, not from desktop only.

Our future work involves two main tasks. First, we will complete the development of the knowledge management platform we are developing based on suggestions above. The system, which is called the NewsMate, will be evaluated and tested in practise. Initial user feedback is positive. Second, we will develop a general architecture for flexible knowledge management. The objective is to set up an

architecture that makes it easy to develop knowledge management systems for new domains.

We also suggest more empirically oriented design projects on knowledge management, which develop novel systems based on observations of work and participatory design principles.

Acknowledgements

This research is a part of the Mobile Informatics programme that is funded by SITI.

7. References

Ackerman, M.S. (1994) "Augmenting the organizational memory: a field study of answer garden," In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work,* pp. 243-252. Chapel Hill, NC: ACM Press.

Ackerman M.S. and C. Halverson (1998) "Considering an organization's memory," In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 39-48. Seattle, WA: ACM Press.

Ackerman, M.S. and D.W. McDonald (1996) "Answer Garden 2: Merging organizational memory with collaborative help," In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work,* pp. 97-105, Cambridge, MA. ACM Press.

Allen, T.J. (1977) *Managing the Flow of Technology*. Cambridge, MA: MIT Press.

Anderson, R. and W. Sharock. (1993) "Can Organisations Afford Knowledge?" *Computer Supported Cooperative Work: The Journal of Collaborative Computing,* 1 (3), pp. 123-142.

Bannon, L.J. and K. Kuutti (1996) "Shifting Perspectives on Organizational Memory: From Storage to Active Remembering," In *Proceedings of the 29th Hawaii Conference on System Sciences. Vol. IV*. pp. 156-167. Los Alamitos. IEEE Computer Society Press.

Belkin, N.J. and W.B. Croft (1992) "Information Filtering an Information Retrieval: Two sides of the same coin?" *Communication of the ACM*, 37 (7), pp. 29-38.

Bellotti, V. and S. Bly (1996) “Walking away from the desktop computer distributed collaboration and mobility in a product design team,” In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*, pp. 209-218, Cambridge, MA: ACM Press.

Bellotti, V. and Y. Rogers (1997) “From Web Press to Web Pressure: Multimedia Representations and Multimedia Publishing,” In *Proceedings of ACM 1997 Conference on Human Factors in Computing Systems*, pp. 279-286. Atlanta, GA: ACM Press.

Blackler, F. (1995) “Knowledge, knowledge work and organizations: An overview and interpretation,” *Organization Studies*, 16 (6), pp. 16-36.

Bowker, G.C. (1997) “Lest we remember; organizational forgetting and the production of knowledge,” *Accounting, Management and Information Technologies,* 7 (3), pp. 113-138.

Button, G. and P. Dourish (1998) “On Technomethodology: Foundational Relationships between Ethnomethodology and System Design,” *Human-Computer Interaction*, 13 (4), pp 395-432.

Conklin, J. and M.L. Begeman (1988) “gIBIS: a hypertext tool for exploratory policy discussion,” *ACM Transactions on Office Information Systems,* 6 (4), pp. 303-331.

Cook S.D.N. and J.S. Brown (1999) “Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organizational Knowing,” *Organization Science,* 10 (4), pp. 381-400.

Crabtree, A. (1998) “Ethnography in Participatory design,” *Proceedings of the Fifth Biennial Participatory Design Conference*, pp. 93-105, Seattle, WA: CPSR and ACM Press.

Edwards, D.L. and D.E. Mahling (1997) “Toward knowledge management systems in the legal domain,” In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge,* pp. 158-166. Phoenix, AZ: ACM Press.

Edwards, W.K. (1994) “Session Management for Collaborative Applications,” In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work,* pp. 323-330. Chapel Hill, NC: ACM Press.

Erickson, T., D.N. Smith, W.A. Kellogg, M.R. Laff, J.T. Richards, and E. Bradner (1999) “Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of ‘Babble’,” In *Proceedings of ACM 1999 Conference on Human Factors in Computing Systems*, pp. 72-79. Pittsburgh, PA: ACM Press.

Fagrell, H., S. Kristoffersen and F. Ljungberg (1999) “Exploring Support for Knowledge Management in Mobile Work,” In *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work,* pp. 259-275, Copenhagen, Denmark: Kluwer Academic Publishers.

Jordan, B., R. Goldman and A. Eichler (1998) “A Technology for Supporting Knowledge Work: The RepTool,” In *Information Technology for Knowledge management,* U.M. Borghoff and R. Pareschi (eds.), pp. 79-96. Berlin: Springer-Verlag.

Kautz, H., B. Selman and M. Shah (1997) “Referral Web: combining social networks and collaborative filtering,” *Communications of the ACM*, 40 (3), pp. 63-65.

Kensing, F., J. Simonsen and K. Bødkar (1997) “Participatory Design at a Radio Station,” *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, 7 (3-4), pp. 243-271.

Kraut, R., R. Fish, R. Root and B. Chalfonte (1993) “Informal communication in organizations: form, function, and technology,” In *Groupware and Computer-Supported Co-operative Work,* R. Baecker, (ed.), pp. 387-314. Morgan Kaufmann.

Kristoffersen, S. and F. Ljungberg (1998) “MobiCom: Networking dispersed groups,” *Interacting with Computers,* 10 (1), pp. 45-65.

Luff, P. and C. Heath (1998) “Mobility in Collaboration,” In *Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 305-314. Seattle, WA: ACM Press.

Ljungberg, F., B. Dahlbom, H. Fagrell, M. Bergquist and P. Ljungstrand (1998) “Innovation of new IT use: Combining approaches and perspectives in R&D projects,” In *Proceedings of the Fifth Biennial Participatory Design Conference*, pp. 203-209, Seattle, WA: CPSR and ACM Press.

March, J.G. (1991) “How Decisions Happen in Organizations,” *Human-Computer Interaction,* 6 (2), pp. 95-117.

McDonald, D.W. and M.S. Ackerman (1998) “Just talk to me: a field study of expertise location,” In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 39-48. Seattle, WA: ACM Press.

Mencher, M. (1997) *News Reporting and Writing*. Seventh Edition. McGraw-Hill.

Randall, D., J. O'Brien, M. Rouncefield and J.A. Hughes (1996) “Organisational Memory and CSCW: Supporting the ‘Mavis Phenomenon’,” In *Proceedings of the sixth Australian Conference on Computer-Human Interaction*. Hamilton, New Zealand: IEEE Computer Society.

Patterson, J.F., M. Day and J. Kucan (1996) “Notification servers for synchronous groupware,” In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 122-129. Seattle, WA: ACM Press.

Ruggles, R. (1998) “The state of the notion: knowledge management in practice,” *California Management Review*, 40 (3), pp. 80-89.

Schmidt, K. and L. Bannon (1992) “Taking CSCW seriously. Supporting articulation work,” *Computer Supported Cooperative Work. An International Journal*, 1 (1-2), pp. 7-40.

Streeter, L.A. and K.E. Lochbaum (1988) “An Expert/Expert-locating System Based on Automatic Representations of Semantic Structure,” In *Proceedings of the fourth Conference on Artificial Intelligence Applications*. San Diego, CA: IEEE Computer Society.

Swan, J., H. Scarbrough and J. Preston (1999) “Knowledge Management - The Next Fad to Forget People?” In *Proceedings of the seventh European Conference on Information Systems*, pp. 668-678. Copenhagen, Denmark.

Terveen, L.G., P.G. Selfridge and M.D. Long (1995) “Living Design Memory: Framework, Implementation, Lessons Learned,” *Human-Computer Interaction*, (10) 1, pp. 1-37.

Weick, K.E. and F. Westley (1996) “Organizational Learning: Affirming an Oxymoron,” In *Handbook of Organizational Studies*, S. Clegg, et al (eds.), pp. 440-458. London: Sage.

Weiser, M and J., Morrison (1998) “Project Memory: Information Management for project teams,” *Journal of Management Information Systems*, 14 (4), pp. 149-167.

Wenger, E. (1998) *Communities of Practice: Learning, Meaning, and Identity,* Cambridge University Press.

Fifth paper[14]

NewsMate: Providing Timely Knowledge to Mobile and Distributed News Journalists

Abstract

NewsMate is a knowledge management system supporting distributed and mobile teams of journalists. The system addresses well-known problems of knowledge management systems by: (1) shifting focus from problems to task; (2) incorporating pointers to several sources, even external to the organisation; (3) supporting direct communication between people by articulating local conditions. This paper reports the system and the preliminary results of use experiences.

[14] Fagrell, H. (2000) "NewsMate: Providing Timely Knowledge to Mobile and Distributed News Journalists," Revised version accepted for publication in *Beyond Knowledge Management: Managing Expertise*, M. Ackerman et al., forthcoming.

1. Introduction

Knowledge management has been recognised as a key component to sustain competitive advantage (Ruggles 1998). The chapter focuses on information technology support for Knowledge Management, usually called knowledge management systems. This category of systems have been criticised extensively in the literature, in particular for:

- Offering a passive "repository" rather than actively supporting remembering (Randall et al 1996; Bannon and Kuutti 1996; Hughes et al 1996);
- Not taking into account how work is actually conducted, relying instead on idealised models of human activity (Brown and Duguid 1991);
- Assuming that there is a well-defined problem to be solved through rational choice, which is not always valid (March 1991);
- Not mediating knowledge that is distributed between many sources, of which only a few are explicitly known in advance (Ackerman and Halverson 1998);
- Promoting an objective view of "knowledge," which excludes important aspects like originator and context (Fagrell et al 1999).

This chapter argues for a complementary approach to IT-support for Knowledge Management, in which:

- Knowledge is actively acquired by the user;
- Empirical studies of mobile work informs the design (Fagrell and Ljungberg 2000; Fagrell et al 1999);
- The user can be aided in defining as well as solving the problems;
- Knowledge from many sources is integrated in a way that is tailored to the local use situation;

- Communication links between users and originators of knowledge can be established.

This approach is illustrated with the NewsMate system, a mobile knowledge management system for radio journalists, which aims to provide timely information in a distributed and mobile work setting. The client platform of the NewsMate is a Personal Digital Assistant (PDA) with network access using a mobile phone.

The work domain focused on, in this chapter, is radio news journalism. Journalists "repackage" information to *make news* in understandable and interesting ways. The pace of such work is usually high. Little time is available for preparation. Reporting is often conducted away from editorial staff at the radio station. For this reason they cannot easily adopt traditional desktop-based systems (cf., Bellotti and Rogers 1997).

Journalists often have to report events on topics that they are not fully profound with (Fagrell and Ljungberg 2000). For this reason, the knowledge management support is concentrated on the issue of "timely knowledge," i.e., knowledge that is relevant, and pertinent for the task at hand.

The historical rationale of many traditional knowledge management systems is to help people to find solutions to problems (e.g., Ackerman 1994). It is assumed that the user has a problem, which is not entirely unique. The system is then used to identify an already solved, similar problem. A solution for the old problem is then applied to the new problem.

Clearly, the alternative setting of distributed and mobile radio journalist can provide new insight into Knowledge Management. For instance, there are several conceptual dimensions that describe the novel requirements of such work, by going from:

- *Problem to task* – there is little focus on problems that are external to the task. Accordingly, we depart from tasks rather than problems. For example, the note taking and authoring that journalists conduct prior to reporting are used as input.

- *Closed to open* – traditional system rely on passively storing information internally in the system. We support an open architecture that uses sources internal as well as external to the organisation.

- *Indirect to direct communication* – knowledge can not always be repackaged for distribution without problems (Brown and Duguid 1991). For this reason, direct communication between users and experts is supported.

The rest of this chapter is structured as follows. The next section presents the NewsMate in a use scenario. After this we describe technical details and the user interface. A section with use experiences follow. Next, we discuss our contribution and related work.

2. Use scenario

Let us now go into a scenario where a user is interacting with the NewsMate and colleagues in a typical work situation (figure 1). The results that the system displays are genuine, except for the translation to English, and are taken from a prototype installation at Radio Sweden in Gothenburg.

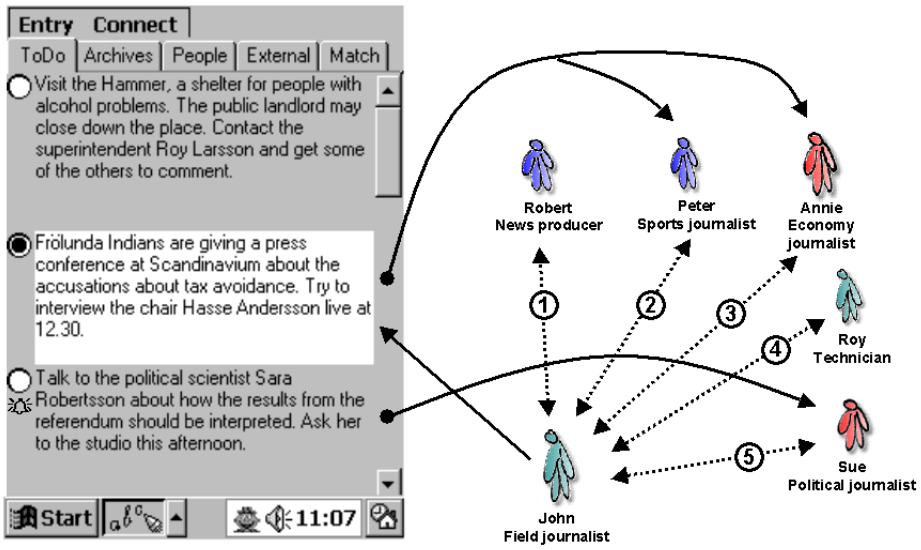

Figure 1: The collaboration between the editorial staff in the use scenario.

The time is 11:00 and John, who is working as a field journalist at a radio station, has just finished a report. As he walks to the car he receives a call on his mobile phone. It is his colleague, Robert, telling

him that the board of an ice hockey club, Frölunda Indians, has just announced that they are going to give a press conference.

Figure 2: Robert is talking to John while taking some notes on the NewsMate. Note the earphone.

John remembers an article from a morning newspaper "The Metro," where the club was accused of tax avoidance when paying salaries. Robert asks John to go there and do a report for the 12:30 news. He points out that an interview with the chair of the board would be great. John and Robert discuss the details of the task and while talking, John uses his NewsMate to make some notes (figure 2).

John's task is now to report from the press conference and do an interview with an appropriate framing. John is neither an expert in taxes nor in ice hockey, but will be able to do an satisfying report if the background and framing comes in place.

John takes a minute to structure his notes into an entry on the To Do tab. Then he connects his NewsMate to the server via the mobile phone, activates the To Do entry and chooses Send to do in the Connect menu (figure 3a).

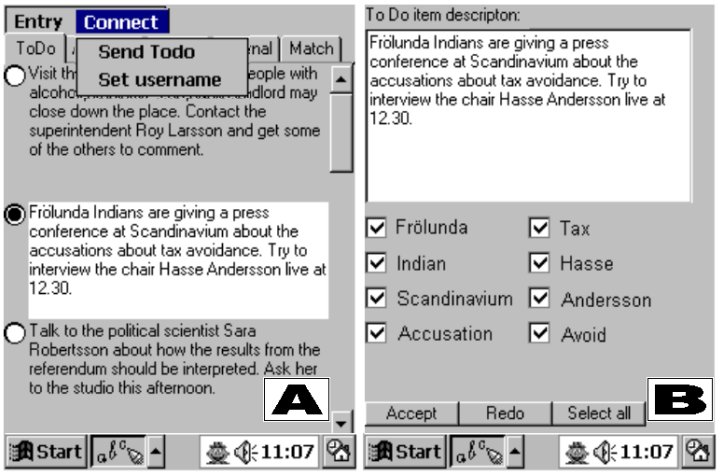

Figure 3: A Send To Do (a) and the selection of keywords (b).

A list of keywords is returned and John chooses the ones that fits and clicks on the Accept button (figure 3b). A few seconds later the results arrive and the mobile phone disconnects. John takes a look at the Archives tab where a list of internal documents is displayed (figure 4a).

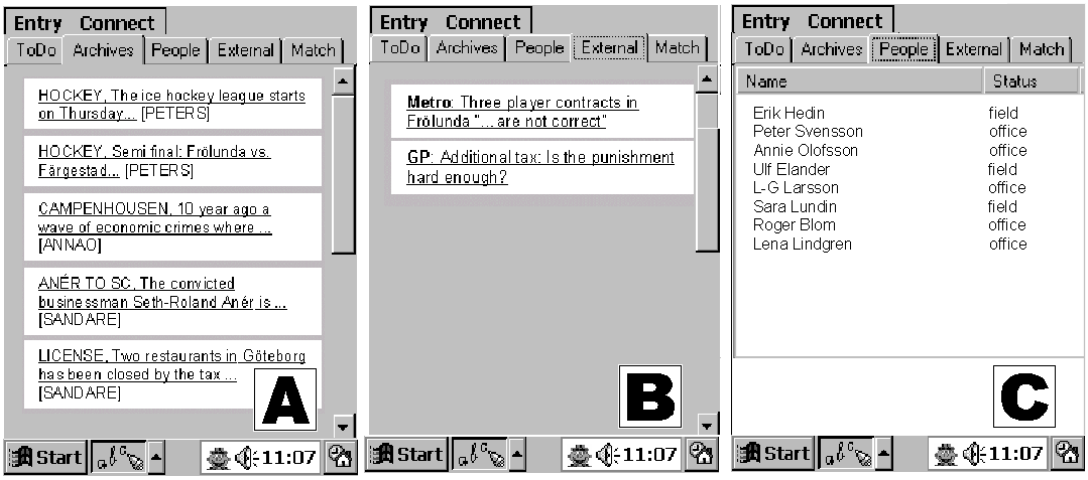

Figure 4: Three tabs on the NewsMate: Archives (a), External (b) and People (c).

The titles of the documents gives John an overview of what have been done internally on the topic. It seems like economic crimes in the restaurant sectors are common and John also notes an article about well-known economic criminal. John realises that perhaps this kind of crime has spread to a new sector. This may be an interesting introduction to the report John thinks.

John continues through the tabs and looks at the External tab (figure 4b) to see what the newspapers and other competitors have

done on the topic. “Just the Metro article of value I suppose,” John says to himself when he realises that he already knew about that.

He also checks the People tab (figure 4c). Here, colleagues who are on duty right now and have been working on the topic previously are displayed. The quality of the report is likely to increase if he discusses the topic with someone more experienced.

When John arrives to the press conference he plans to contact some colleagues who may help him. Since Erik is on the top of the People tab he is the most appropriate colleague to talk to. Erik is, however, out of the office and there is no answer on his mobile phone.

Instead, John calls Annie, who answers (figure 5). John remembers that she has been working on white-collar crime. Annie and John discuss questions like, “are there any similar cases?” and “does any board member risk jail?” Annie thinks that the case is not clear enough to be talked about in matters of punishment yet. They agree, however, that he could ask about “bad accounting” practise. The communication between John and Annie results in that John gets aware of some appropriate angles of the report from an accounting perspective.

Figure 5: Annie and John discuss the case from an economic crime perspective.

John hangs up and enters the room where the press conference has just begun. There are a lot of other journalists and people from the ice hockey club there. The chair immediately states that “I will not give any interviews. We’re giving the press conference. That’s all.” The board’s main message is that they are not guilty, but that they have started an internal investigation. They will not give any further comments until the investigation is finished. John thinks: “OK, what to

do now? I need to have something interesting to report. Let's consult the NewsMate."

Figure 6: John asks Peter about the framing of the report.

John sneaks outside and takes a look at the People tab (figure 4c). The second entry is the sports journalist Peter Svensson. "Oh Peter, of course," John says to himself and gives Peter a call (figure 6). They start a discussion about what has happened. John asks whom else to interview since the chair refused. Peter mentions that the accountant of the club, Thomas Zetterlund, may accept to be interviewed, "He is the next in line to be the chair of the club." John agrees that Thomas Zetterlund is suitable, because they really need an interview to complement the article in the newspaper. John asks some more questions, e.g., "is it common with tax avoidance in ice hockey or sports in general?" and "is this the first case nationally?" Though the use of Peter's expertise and John's knowledge of local conditions the framing is collaboratively established. This kind of situated knowledge is rarely accessible in a knowledge management system. Furthermore, the time constraint makes direct communication with Peter the best way for John to get the knowledge.

The press conference ends and John asks the accountant Thomas Zetterlund if he minds being interviewed. Thomas accepts and John contacts Roy the technician at the studio to co-ordinate the broadcast. Roy tells him to be prepared to go on the air in a minute. John turns to Thomas and waits for Roy to tell him when to start (figure 7).

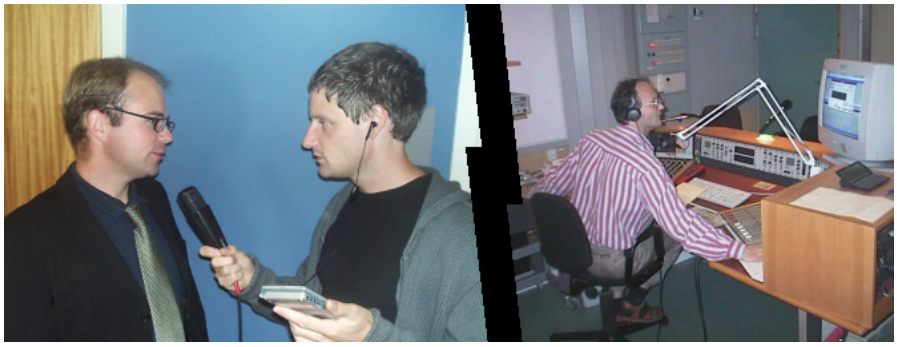

Figure 7: Thomas interviewed live by John (left) and Roy manages the broadcast in the studio (right).

Roy gives a sign and John opens by saying “What we are used to hear from the real estate and restaurant sectors has now entered the professional sports.” This framing of the news was derived from the timely knowledge provided by the NewsMate. John continues by saying, “Here with us is Thomas Zetterlund, the accountant of Frölunda Indians. Thomas, what do you think about these accusations?” Thomas replies, “We use a lot of agents internationally to contract players, and I cannot say for sure whether they have done anything illegal.” He continues with “We are conducting an internal investigation, and I do not want to comment this further until the investigation is done.” After this John asks about the effect on the sport and they elaborate a bit on the question. Then he hears Roy in the earphone saying ten seconds left. John thanks Thomas and ends the interview.

The NewsMate helped John in his efforts to make good quality journalism. He had not enough knowledge to be able to ask the right questions and frame the report. When the chair refused to be interviewed John used the NewsMate to locate someone who could help him to find the second best interviewee. Furthermore, Annie and Peter helped John to find an appropriate angle for the report. A traditional knowledge management system that focus on problems would not have given John any of the support he needed to manage his task since he had no actual “problems.”

Figure 8: John reads an SMS (left) and the Match tab on the NewsMate (right).

John walks back to the car. On his way the mobile phone beeps to signal that he has receives a text message (figure 8). Apparently something that may affect John's work has happened. John connects his NewsMate and a bell is shown on the third To Do entry. The entry concerns an interview with a political scientist about a upcoming referendum (figure 1). He becomes aware that his work might not be unique in relation to some other activities at the channel. The channel's repertoire must be considered as a whole as similar reports should not be broadcasted without different foci.

Figure 9: Sue negotiates with John about the "match."

John activates the entry and checks the Match tab. It appears as if Sue is planning to do something involving a political scientist (figure 8). John gives Sue a call and they both agree that two political scientists on the same day is "too much" (figure 9). They agree that the best thing

is to drop the political scientist from John's program since Sue's program is focused on politics. In this case the NewsMate assisted to prevent a potential cross report. A negotiation had to take place to settle if there was a need to change foci or content of the reports.

3. The Newsmate System

3.1 Platforms

All communication in the NewsMate system is done with TCP/IP sockets via a server. The scalability and maintainability relies on the server, which is implemented in Perl with a Berkley database on the Microsoft Windows NT 4.0 platform. The client-side hardware is a pen-based personal digital assistant, i.e., Casio Cassiopeia E-105 running Microsoft Windows CE 2.11. The client is implemented in Microsoft Visual Basic for Windows CE. The GSM phone (Nokia 6110) connected to a CompactFlash card (MobiFlash) has Internet access via the Point-to-Point Protocol (PPP). The bandwidth is 9600 bits per second and the data transmissions are rarely larger than four kilobytes. The time required to be online is about one minute.

Users that are not online can be actively contacted using the Short Message Service (SMS). SMS is a part of GSM and makes it possible to send and receive text messages with the mobile phone.

3.2 Architecture and prototype installation

The architecture is generic, but requires integration with editorial systems. The prototype installation described here is from Radio Sweden in Gothenburg.

The server keeps track of all events in the system, i.e., very little computation is done on the client. All of the communication between the server and the clients are in plain text that is tagged with XML (eXstensible Markup Language). The server also manages all the

database calls. The database queries are formatted using a stop-list and a stemming dictionary. The stop-list purpose is to filter out high frequent and low content words, e.g., “interview” and “the.” The stemming dictionary puts the words on their basic form, e.g., “accusations” and even the misspelled “accusasion” will all be replaced by “accusation.” The stop-list and stemming dictionary are generated from two years of content from the internal archives (about 15,000 text documents of a total size of 10 megabytes). The stemming list is about 300,000 words (4 megabytes) and placed on the server because of its size. The stop list is 1,500 words (10 kilobytes) and is on the client.

The database of the internal archives is constructed from program reports, proposals for programs and not yet reported surveys and can be searched with good performance. The database can also provide a list of authors of documents that are related to an entry (cf., Streeter and Lochbaum 1988). The list of authors will only contain those that are working at this point in time, i.e., logged on to the local area network. The whereabouts of people is also taken into account and based on if they have checked out a car.

By periodically and automatically gathering the content from the web edition of local newspapers and a web service with local news another database have been constructed. It can be searched in the same way as the internal one, but the original articles are not stored locally. This also includes incoming telegrams from news agencies.

To find out if users are working on similar topics the terms representing the To Do entries are compared. Furthermore, the overlap in search results from the internal and external archives are also compared to find similarities.

The following use case demonstrate the technical actions of each component (figure 10).

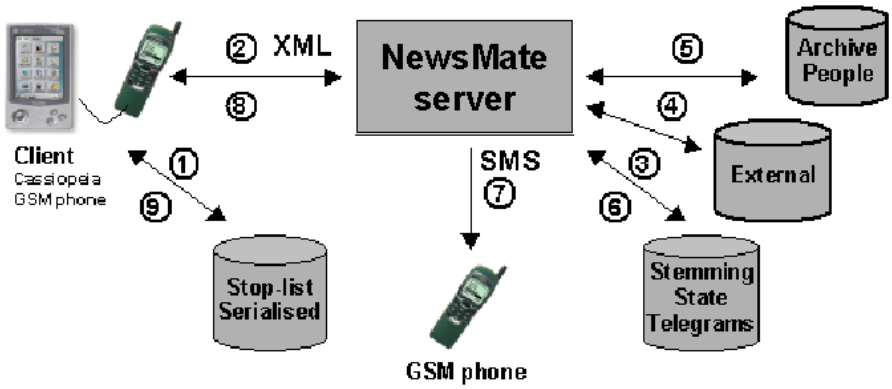

Figure 10: The architecture of NewsMate.

1. The client filters the To Do entry through the stop-list.
2. The client calls the server with the formatted query.
3. The server filters the call through the stemming dictionary to put the terms on their basic form.
4. The server queries the external database and receives the results.
5. The server queries the achieves/people database and receives the results.
6. The server stores the users states and checks for overlaps and if any new and matching telegram has arrived.
7. If any overlaps or matching telegrams are found text messages (SMS) are sent to the concerned users.
8. The results are returned to the client.
9. The client stores the results as serialised objects.

3.3 User Interface

To do

On the To Do tab the currently active entry determines the content of all the other tabs. It lets the user add, edit and remove entries. The user updates the text with the pen. The user can choose Send To Do in

the Connect menu making a list of keywords appear (figure 3ab). The relevant keywords can be selected. If Redo button is pressed a new list of keywords is generated. When the user is satisfied the Accept button is pressed and the results of the other tabs will be returned.

Archives

For each entry that has been sent to the server a list of ten articles (title and ingress) from internal archives is displayed at the archive tab (figure 4a). The user can click on a title and get the full text of the article. The intended use is, however, to remind the user of how the topic had been reported by looking at the list.

External

The External tab lists matches from predefined external sources, e.g., newspapers and telegrams (figure 4b). By clicking on the entries the user can browse the full text articles. The intended use here is once again to remind the user of how the topic had been reported externally by looking at the list.

People

The People tab displays the names of journalists who have been involved in similar tasks and are currently on duty (figure 4c). The Status column indicates if a person is in the office or in the field.

Match

If there is an overlap between To Do entries a notification is issued. The users who are concerned receive an SMS their mobile phones. The user can then connect and a bell icon will be displayed beside the entry in question (figure 1). When the To Do entry is made active, the matching users name and To Do entry text is displayed on the Match tab (figure 8). The match can, of course, also happen when a user has just done a Send To Do. In these cases no SMS will be sent.

In some cases the message will only confirm what the users already know. In other cases it may be preferable to get in direct contact to resolve the matter.

If any new telegrams have arrived, which are similar to a registered To Do item on an offline client, they will be sent in this way as well.

4. Use Experiences

To gain use experiences we have conducted focus groups (Nielsen 1993, pp. 214-216) and evaluative ethnography (Hughes et al 1994, pp. 434-435).

4.1 Focus Groups

The “focus group” technique means that a group of about five to ten users are brought together to discuss issues and concerns about the features of a system.

In our case, the discussions lasted for about 30 minutes with each group and a moderator maintained the groups focus.

Group	Number of Journalists	Organisation	Medium and news type
1	11	Radio Sweden, Gothenburg	Radio, local
2	8	Radio Sweden, Gothenburg	Radio, local and national
3	6	Radio Sweden, Stockholm	Radio, IT group
4	8	Göteborgs-Posten	Newspaper, local
5	4	Göteborgs-Posten	Newspaper, IT group
6	3	Freelance	Radio, television, and newspaper

Table 1: The six focus groups.

The groups were assembled to represent both local and national news journalists operating in different mediums (see table 1). Group 3 and 5 are groups of journalists that partly work with information technology related issues.

Before the discussion started we presented

- the purpose of the NewsMate
- use scenarios
- a tour of the user interface
- a test drive of the system

Everyone had a positive opinion of the system and seemed to think that what the system aims to support is important. The IT-groups

believed that the features of the NewsMate were original. The ability to improving the quality of reports was considered to be the central contribution of the system. This depended of course on individual journalist's motivations and commitments to use the NewsMate.

Slow network

The total time required to get the result after sending a To Do item was about one minute. This was considered acceptable, mainly because no interaction is required during the transmission. A person put it as follows: "Perhaps, one minute is ok, but if the system had instant access the use could be different… especially how often one would use it."

Pen-based authoring

Many of the journalists use paper-based to do lists when preparing reports (cf., Kidd 1994). This seems to coincide with the use of the NewsMate. The "extra" amount of authoring with the pen-based system seemed to be acceptable for them. However, some concern was expressed that the pen-based input is slower than note taking with pen and paper. Some people had already pen-based PDAs for personal use and they did not think that it would be any problem.

Level of detail

The genuine examples that the NewsMate presented in the use scenarios seemed to be on an appropriate level of detail. For example, we did not know why the term "Campenhousen" appeared in figure 4a. The journalists, however, told us that he is a prosecuting attorney in Gothenburg. This kind of local knowledge seems important to accomplish remembering (cf., Ehrlich and Cash 1999; Fagrell et al 1999)

Mobile access to background information

The access to background information in the field would be especially interesting for freelance journalist. The reason, they claimed is that the quality of their reports is judged commercially and increased quality means more contracts.

Some journalists are not so much in the field. There is however a lot discussions going on in places with no computers. In these situations the NewsMate could be useful.

Matching tasks

The feature to incorporate the task support with periodical searches in the telegram system was considered very interesting. Overlapping reports was also recognised as problematic, especially for journalists that work in the field a lot.

Finding expertise

The People feature was liked but some journalists said that there could be some status problems to approach journalists from other units. They did, however, believe that this might change due of the system.

4.2 Evaluative ethnography

The evaluative ethnography was conducted at Radio Sweden in Gothenburg for a period of 50 hours. The researcher followed 10 journalists closely and mapped their current task on the NewsMate. When the journalists had a few moments free they were confronted with the results of the NewsMate and asked to comment. Most comments were tape-recorded and then transcribed. Fourteen cases of news making and reporting were observed of which four representative examples are presented in the following sections. The match feature was only tested with telegram, i.e., overlapping work tasks could not be identified.

4.2.1 The art performance

In a very short notice, Erik had to report from an art performance. The theme of the performance was “Time in sound and vision in the streets of Gothenburg.”

On the way to the performance, the results from the NewsMate were displayed for Erik. Most of the article from the internal archives were about the millennium celebration. The tourist association had received a grant for the millennium celebration and there were some debate on how the money should be spent. Erik said “Oh I remember

this, I think that this event is sponsored by this grant… interesting. Hopefully, someone from the tourist association is there."

The person who had been working on the topic was listed in the People tab, but Erik felt no need to call anyone at that point.

One of articles from the External tab was about different religions way of counting time. Erik looked at it and said: "well good, I need this information for another thing I'm doing just before the New Year 2000."

At the event, there were some people from the tourist association and the artist. Erik exchanged a few words with the people from the tourist association. Then he started to chat with the artist. After calling the studio he started to interview the artist live and some of the sound effects could be heard in the background.

While driving back to the studio Erik said that he found it more relevant to interview the artist about the art performance rather than the person from the tourist association about money spending. Erik claimed that the results of the NewsMate were interesting and got him thinking about possibilities on how to report the event. More specifically, he had asked the artist about a sponsorship and how different religions counted time.

4.2.2 The award winner

Emma was assigned to report from the "Pencil of freedom" prize, which is awarded every year to someone who has been struggled against anti-democratic forces. At the station Emma searched the internal archives and telegrams. She found some background information, e.g., previous winners, but it was not possible to gather information about all potential winners.

At the ceremony a chairman gave a short talk and introduced the winner: "and the winner is Anna Christensen, professor of law, Lund University." Anna Christensen entered the stage and the chairman read the motivation followed by the audience's applauds. The motivation was also provided on paper.

Emma stepped aside and called Erik on her mobile phone. She told Erik who the winner was and read the motivation. Erik was going to use the material in the next news-update broadcast that was only 5 minutes away.

Emma continued to listen to Anna Christensen's short talk. The talk ended and the audience, who was mainly journalists, started to ask questions. After this, Emma did a recorded interview with the winner for about 15 minutes.

After the interview Emma said that she was very satisfied with the recorded material, "It will be a good report this afternoon both locally and nationally."

The NewsMate had returned a lot of results on the External tab and one internal article. The external articles included that the student committee at Gothenburg University like to see Anna Christensen as new principle and legal aspects of a fire catastrophe in Gothenburg.

Emma reaction to the results of the NewsMate was that "she [Anna Christensen] is fairly unknown, but has been seen here and there in the press... I should have used that thing with the student... It would not be appropriate to talk about fire accidents today." She continued "This report worked out fine, but in situations like this, when it is impossible to plan, the system would be helpful."

4.2.3 Health care re-organisation

On a Monday morning Adam had just finished a report at a convention centre when the producer calls on his mobile phone. The producer said that the president of a large hospital was available for an interview about the re-organisation problems of the hospital at the convention centre.

Adam had not read the newspaper articles on the topic from last weekend, but had to do the interview anyway, because he was the only one available at short notice.

After the interview Adam said, "I am uneasy with the report. If I knew the background better it would have been different… The listeners were in a way 'cheated' of a better report." When looking at the results of the NewsMate Adam said "It is hard to say if archives and external would have made a difference, but if I would have been in contact with someone on the People list it would have made a great difference."

Later in the afternoon Adam said that he had talked to Maria and Tom, two of the names on the People tab, to learn more about the topic. “To bad I did not talk to Maria before the interview,” he said.

4.2.4 Problems with airbags

Madeline, the producer, asked Annie to report something about a telegram that arrived a few hours earlier. “Volvo car had to withdraw a lot of cars due to problems with the airbags.” “Perhaps you can talk someone at Volvo… it’s just a short report and if it is interesting someone else can do a ‘follow up’ this afternoon.” “Ok, I can do a short report, but I have to be away in an hour,” Annie answered.

Annie went to her PC and looked at incoming telegrams. She found a telegram saying that Volvo withdraws 135,000 cars because the airbags could go of if exposed to static electricity. Annie wrote down some questions on a paper. Annie looked in the telephone dictionary and found the number to Volvo cars information manager. She called the number and got redirect to the switchboard. “Ok, I understand” she said and hung up. “These information managers change all the time and not even the switch board knows who it is.” Annie called some other number and said: “It is to a person that I have been in contact with before. He probably knows who I could call.” After a while Annie found the right telephone number and called it from a studio telephone that could record the interview digitally. She started to ask a few questions. “How much will this cost Volvo? A quick calculation that the repair of each car will be a $100 and 135,000 car; that would make 13,5 millions,” Annie asked. The manager answers “I will not comment on that, only that Volvo will pay. And by the way it is 235,000 cars. We have sent a new telegram about half an hour ago.” Annie answered: “Sorry about that sir, I have not read that message. Can someone call you later today and record an interview?” The information manager accepted and Annie went back to her PC and looked at the new telegram.

Annie said: “I wanted to record him right now, but I could not since I had not read the latest telegram. Instead, someone else have to record the interview.” Annie wrote a message and recorded it. Then, she told Madeline where to find it and hurried to her next assignment.

The NewsMate had received a message about the new telegram before Annie got in contact with the information manager. When Annie was confronted with the result she said: “If I had known about the new telegram I would have read it first and then called the manager to do a recorded interview.”

5. Discussion

In this chapter we have introduced the NewsMate, a knowledge management system that supports mobile and distributed news journalists. Our contribution is an approach to knowledge management emphasising

- That there is a need to step away from the focus on problem solving and instead incorporate knowledge management support in the task. This is supported through a To Do feature of the NewsMate.
- That the information architecture should consider records from sources internal as well as external to the organisation. NewsMate features illustrating this are Archives and External.
- That direct communication should be actively favoured and based on present conditions. This is illustrated through the People and Match features of the NewsMate.

The use experiences have shown that the NewsMate system is useful. We believe that our approach is well aligned with the critique against traditional knowledge management systems. For example:

- Randall et al (1996), Bannon and Kuutti (1996) and Hughes et al (1996) have criticised knowledge management for offering a passive “repository” view of knowledge. The task and direct communication orientation of the NewsMate provides support for active remembering, sensitive to the human conduct that actually takes place in organisations.
- Ackerman and Halverson (1998) show that knowledge is often distributed between many different sources, only a few, of which, are explicitly known in advance. This is addressed with the open

information architecture of the NewsMate in conjunction with the encouragement of direct communication.

5.1 Related systems

One of the first systems in the knowledge management category was gIBIS (Conklin and Begeman 1988; Yakemovic and Conklin 1990). The aim was to make the rationale for a design decision explicit by capturing the argumentation in a hypermedia system. The system is partially about getting an overview of complex design problems. There is little evidence of knowledge being re-used with this system and the users' tasks except for problem solving. It incorporates other sources of information, e.g., design sketches and code, by linking to them. There are no connections to present activities and communication between people is made indirectly through the system. Similar ideas can also be found in a recent system, the RepTool, which have included graphic views and maps to facilitate remembering (Jordan et al 1998).

A similar approach was offered by Terveen et al (1993), reporting from a project creating a "living design memory" for software developers. This was accomplished by introducing links to associate information units with the pertaining parts of the products. It can be argued that in a sense facts are placed in the context where they are to be used. The system was integrated with the software development process, which is very different from journalism. It is claimed that relevant knowledge exists in the form of "folklore" and that it should be incorporated in the system. To repackage knowledge is, however, problematic [cf., 8]. Instead of getting the "folklore" into the system our approach is to support people communicating the "folklore."

Most of these features can be found in the Project memory system (Weiser and Morrison 1998). The system has also To Do lists, but in this case they are in the form of milestones. This is also the focus of Kreifeldt et al (1993). The To Do feature of the NewsMate is more like an open resource that is used to reduce the complexity of the task.

Knowledge Pump (1998) integrates and refines ideas from intelligent information sharing (Malone et al 1998) and collaborative recommender systems (Resnick and Varian 1997). The system address

the problem of long-term knowledge interests and are based on user profiles. The NewsMate is mainly used to address short-term knowledge interests linked to the task. The journalists' user profiles are only based on published records.

Ackerman and Starr (1995) argue for the importance of social activity indicators, i.e., it is important to be aware of other people's activity in a collaborative system. In the NewsMate this is addressed with the Match feature that informs involved parties in the case of a potential cross reporting.

Answer Garden (1994) aim to help organisations capture and retrieve experiences made by their employees. In short, the Answer Garden lets the users browse a network of diagnostic questions to find the answers they want. If the answer cannot be found the question is routed to experts and later inserted (along with the answers) into the network. The experts can also modify the network, thus the knowledge grows. Answer Garden 2 introduces features that route the user directly to the expert if the solution is not found in the network (Ackerman and McDonald 1996). A collaborative help feature is also added to make the interpretations of de-contextulized information easier. The idea is to remove unnecessary context (Ackerman and McDonald 1996, p. 103), which can be difficult (Fagrell et al 1999). None of the versions of Answer Garden support the user with the task-at-hand. The system requires additional authoring and support for problem solving.

In a field study McDonald and Ackerman (1998) investigates expertise location. Once again the focus was on problem solving in software engineering. The identification of potential expertise with the NewsMate is accomplished through the People feature. Furthermore, no matter how experienced a journalist is, a "second opinion" is always appreciated (cf., Ehrlich and Cash 1999, p. 162).

Co-ordination of work and sharing of experiences were the focus of Kristoffersen and Ljungberg (1998) when designing support for a dispersed IT-support group. They focused on tasks, but they were strictly problem oriented and the communication between people was mainly through the system. The NewsMate is not designed for this, thus co-ordination and task assignment are done elsewhere.

Finally, Kristoffersen and Ljungberg's system also supports the mobility of the staff. It has recently been argued that CSCW designers

have not taken mobility seriously (Luff and Heath 1998). The main requirement for the NewsMate was that it should move beyond the desktop and into the field.

6. Future Directions

The approach to knowledge management that is illustrated in this chapter may also be useful in other categories of work, e.g., sales personnel. To test this a more general version of the NewsMate is under development.

Technical improvements will of course affect that development. For example, in the year 2000 the GPRS (General Packet Radio Services) infrastructure is implemented on the GSM network.[15] The bandwidth is quite high theoretical, but in practice it will be about 50 kilobits per second. The most interesting feature is the instant connection where information can be sent or received immediately, i.e., no dial-up modem connection is necessary. These technical possibilities might open up for new mobile applications similar to the NewsMate.

Acknowledgements

This research is a part of the Mobile Informatics programme that is funded by the Swedish Information Technology research Institute (SITI).

The author would like to thank the colleagues in the Mobile Informatics group at the Viktoria Institute, especially Fredrik Ljungberg, Kerstin Forsberg, Erik Johannesson and Steinar Kristoffersen. The author would also like to thank Sofia Eklund and Jens Bergqvist.

[15] Source: Mobile Lifestreams Limited, www.mobilegprs.com

7. References

Ackerman, M.S. (1994) “Augmenting the organizational memory: a field study of answer garden,” In *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work,* pp. 243-252. Chapel Hill, NC: ACM Press.

Ackerman M.S. and C. Halverson (1998) “Considering an organization’s memory,” In *Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 39-48. Seattle, WA: ACM Press.

Ackerman, M.S. and D.W. McDonald (1996) “Answer Garden 2: Merging organizational memory with collaborative help,” In *Proceedings of the ACM 1996 Conference on Computer Supported Cooperative Work,* pp. 97-105, Cambridge, MA: ACM Press.

Ackerman M.S. and B. Starr (1995) “Social Activity Indicators for Groupware,” In *Proceedings of the eight ACM symposium on User interface and software technology,* pp. 159-168. Pittsburgh, PA: ACM Press.

Anderson, R. and W. Sharock. (1993) “Can Organisations Afford Knowledge?” *Computer Supported Cooperative Work: The Journal of Collaborative Computing,* 1 (3) pp. 123-142.

Bannon, L.J. and K. Kuutti (1996) “Shifting Perspectives on Organizational Memory: From Storage to Active Remembering,” In *Proceedings of the twenty-ninth Hawaii Conference on System Sciences,* Vol. IV, pp. 156-167. Los Alamitos: IEEE Computer Society Press.

Bellotti, V. and Y. Rogers (1997) “From Web Press to Web Pressure: Multimedia Representations and Multimedia Publishing,” In *Proceedings of the ACM 1997 Conference on Human Factors in Computing Systems,* pp. 279-286. Atlanta, GA: ACM Press.

Brown, J.S. and P. Duguid (1991) “Organizational Learning and Communities of Practise: Towards a Unified View of Working, Learning and Innovation,” *Organization Science,* 2 (1) pp. 40-57.

Conklin, J. and M.L. Begeman (1988) “gIBIS: a hypertext tool for exploratory policy discussion,” In *Proceedings of the ACM 1988 Conference on Computer Supported Cooperative Work,* pp. 140-152. Portland, OR: ACM Press.

Ehrlich, K. and D. Cash (1999) “The invisible world of intermediaries: a cautionary tale,” *Computer Supported Cooperative Work: The Journal of Collaborative Computing,* 8 (1-2) pp. 147-167.

Fagrell, H., S. Kristoffersen and F. Ljungberg (1999) “Exploring Support for Knowledge Management in Mobile Work,” In *Proceedings of the sixth European Conference on Computer-Supported Cooperative Work,* pp. 259-275 Copenhagen, Denmark: Kluwer Academic Publishers.

Fagrell, H., S. Kristofferssen and F. Ljungberg (1999) “How Journalists decide which Questions to Ask: Implications for Mobile Knowledge Management,” Technical report VRR-99-16, The Viktoria institute, Gothenburg, Sweden.

Glance, N., A. Damián and M. Dardenne (1998) “Knowledge Pump: Supporting the Flow and Use of Knowledge,” In *Information Technology for Knowledge management*, U.M. Borghoff and R. Pareschi (eds.). Berlin: Springer-Verlag.

Hughes, J.A., V. King, T. Rodden and H. Andersen (1994) “Moving out from the control room: ethnography in system design,” In *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work,* pp. 429-439. Chapel Hill, NC: ACM Press.

Hughes, J.A., S. Kristoffersen, J. O'Brien and M. Rouncefield (1996) “When Mavis met IRIS: Ending the love affair with Organisational Memory,” In *Proceedings* of IRIS'19, pp. 767-788. Gothenburg, Sweden: Gothenburg Studies in Informatics.

Jordan, B., R. Goldman and A. Eichler (1998) “A Technology for Supporting Knowledge Work: The RepTool,” In *Information Technology for Knowledge management*, U.M. Borghoff and R. Pareschi (eds.). Berlin: Springer-Verlag.

Kidd, A. (1994) “The Marks are on the Knowledge Worker,” In *Proceedings of the ACM 1993 Conference on Human Factors in Computing Systems,* pp. 186-191. Boston, MA: ACM Press.

Kreifeldt, T., E. Hinrichs and G. Woetzel (1993) “Sharing To-Do Lists with a Distributed Task Manager,” In *Proceedings of the third European Conference on Computer-Supported Cooperative Work,* pp. 31-45. Milan, Italy: Kluwer Academic Publishers.

Kristoffersen, S. and F. Ljungberg (1998) “MobiCom: Networking dispersed groups,” *Interacting with Computers,* 10 (1) pp. 45-65.

Luff, P. and C. Heath (1998) “Mobility in Collaboration,” In *Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 305-314. Seattle, WA: ACM Press.

Malone, T.W., K.R. Grant, F.A Turback, S.A. Brobst, and M.D. Cohan (1987) “Intelligent information sharing systems,” *Communications of the ACM,* 30 (5) pp. 390-402.

March, J.G. (1991) “How Decisions Happen in Organizations,” *Human-Computer Interaction,* 6 (2) pp. 95-117.

McDonald, D.W. and M.S. Ackerman (1998) “Just talk to me: a field study of expertise location,” In *Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 39-48. Seattle, WA: ACM Press.

Nielsen, J. (1993) *Usability Engineering*. Academic Press.

Randall, D., J. O'Brien, M. Rouncefield and J.A. Hughes (1996) "Organisational Memory and CSCW: Supporting the 'Mavis Phenomenon'," In *Proceedings of the 6th Australian Conference on Computer-Human Interaction*. Hamilton, New Zealand: IEEE Computer Society.

Resnick, P. and H.R. Varian (eds.) (1997) "Recommender systems," *Communications of the ACM*, 40 (3) Entire issue.

Ruggles, R. (1998) "The state of the notion: knowledge management in practice," *California Management Review*, 40 (3) pp. 80-89.

Streeter, L.A. and K.E. Lochbaum (1988) "An Expert/Expert-locating System Based on Automatic Representations of Semantic Structure," In *Proceedings of the fourth Conference on Artificial Intelligence Applications*. San Diego, CA: IEEE Computer Society.

Terveen, L.G., P.G. Selfridge and M.D. Long (1993) "From 'folklore' to 'living design memory'," *Proceedings of the ACM 1993 Conference on Human Factors in Computing Systems,* pp. 15-22. Amsterdam, The Netherlands: ACM Press.

Weiser, M and J., Morrison (1998) "Project Memory: Information Management for project teams," *Journal of Management Information Systems*, 14 (4) pp. 149-167.

Yakemovic, K.C.B. and E.J. Conklin (1990) "Report on a development project use of an issue-based information system," In *Proceedings of the ACM 1990 Conference on Computer Supported Cooperative Work,* pp. 105-118. Los Angeles, CA: ACM Press.

Sixth paper[16]

FieldWise: A Mobile Knowledge Management Architecture

Abstract

The paper presents results of a research project that has aimed at developing a knowledge management architecture for mobile work domains. The architecture developed, called FieldWise, was based on fieldwork in two organisations and feedback from users of prototype systems. This paper describes the empirically grounded requirements of FieldWise, how these have been realised in the architecture, and how the architecture has been implemented in the news journalism domain. FieldWise adds to the field of CSCW by offering an empirically grounded architecture with a set of novel features that have not been previously reported in the literature.

[16] Fagrell, H., K. Forsberg and J. Sanneblad (2000) "Mobile Knowledge Management: Model, Architecture and Applications," Accepted for publication in *Proceedings of ACM 2000 Conference on Computer Supported Cooperative Work*, Philadelphia, PA: ACM Press.

1. Introduction

Recently, the issues of “knowledge management” (Fagrell et al 1999; Ackerman and Halverson 1998) and “mobility” (Luff and Heath 1998; Bellotti and Bly 1996; Bellotti and Rogers 1997) have received much attention in the CSCW literature. The interest in these issues is often motivated by the fact that work in many organisations is “knowledge intensive” and “mobile.” These issues have been explored separately. However, so far few researchers have explored the topic of knowledge management in mobile work domains.

This paper reports the final step of a research project with the objective to design novel and commercially interesting IT support for knowledge management in mobile work. The design has been informed by:

- an empirical study of mobile and distributed service electricians, i.e., observation of work and workshops (Fagrell et al 1999).
- a field study of mobile news journalists at a radio station, i.e., observation of work (Fagrell and Ljungberg 1999).
- experiences with a prototype system for mobile news journalists, i.e., observation of work under real working conditions with the prototype and design workshops (Fagrell 2000).

Based on the results a novel mobile knowledge management architecture, called FieldWise, was constructed.

The kind of typical work organisation that can benefit from a FieldWise implementation is one where:

- people’s tasks are time critical and driven by deadlines
- the result relies on the creativity of autonomous, but interrelated people
- there is a culture of co-operation and sharing of knowledge amongst people
- people are mobile and distributed

Typical work of this sort is, for example, news journalism, sales and real estate brokering. The first implementation of FieldWise is adapted for news journalism. Each new work domain and installation of FieldWise requires a definition of the informational context (cf., Forsberg and Dannstedt 2000) of the organisation, e.g., integration with existing personnel- and record-keeping systems.

FieldWise provides mobile access for five client platforms, i.e., Pocket PC, Windows CE 2.11 (Palm-sized and Handheld), EPOC, and PalmOS. But just offering access to stationary systems cannot solve the knowledge management problems, in mobile work. Several novel requirements must be met.

2. Requirements

This section describes the requirements derived from the empirical work and experiences with prototypes in mobile settings.

2.1 Evolving and interdependent tasks

When people produce material with the objective to accomplish a coherent result there is a need for co-ordination. On the other hand, the autonomous nature of mobile work makes central control unsuitable. The local and unique circumstances that people confront give them knowledge that is situated and local. Also, the actions of people give rise to new local knowledge and situated actions, thus the tasks evolve (cf., Tsoukas 1996; Fagrell and Ljungberg 1999). In situations when the tasks of co-workers are overlapping, it is useful to interrelate the knowledge that each person possesses (Tsoukas 1996). In a mobile and distributed organisation it is never possible for anyone to have complete overview at any point. On the other hand, planning of potentially interesting tasks can be done by a common resource, e.g., an editor at a newspaper.

Empirical research implies that people take notes to reduce the complexity of their tasks (cf., Kidd 1994). Our empirical results suggest that task-related notes are useful in mobile situations to facilitate

remembering (Fagrell et al 1999; Fagrell 2000). In this respect, the notes are a representation of a “prospective memory” (cf., Brown et al 2000).

This suggests that a mobile knowledge management architecture should support the user's tasks, as they evolve. It should also notify the users of interdependencies between tasks, and provide access to tasks that are potentially interesting.

2.2 Overview of records

Having accurate records is not always desirable in an organisational sense. It is often more expensive to store and retrieve than to re-discover (Bowker 1997). Finding out what to do in a situation is a matter of recognising similarities with previous situations, i.e., to remember matching cases and creatively find a category of what needs to be done (Randall et al 2000). Our empirical observations suggest that in mobile situations an overview of what has been done previously can facilitate remembering and ease the rediscovering of relevant details (Fagrell and Ljungberg 2000). Also, feedback annotations that are made in the context of a persistent record can point out ways of taking a task further (cf., Fagrell et al 1999; Fagrell and Ljungberg 2000). For example, it is common that the editorial staff evaluate news reports and discusses improvements at daily meetings. Such annotations can contain suggestions on how to follow-up on a task or how to improve quality in a more general sense.

This suggests that a mobile knowledge management architecture should give an overview of records rather than an exhaustive collection. Annotations to records should also be displayed.

2.3 Location of available expertise

Mobile users often confront situations that they are not totally familiar with. The most efficient way to get help is often to talk to co-workers who have relevant expertise. Knowing who knows has been highlighted as important (Randall et al 1996; McDonald and Ackerman 1998). It is

often a question of getting a “second opinion” on how to solve the task (cf., Ehrlich and Cash 1999).

It is, however, especially problematic for mobile workers to know whom to contact, since they are often away from their co-workers. Accordingly, knowing whom to contact is not enough, they need to be available for interaction as well (Fagrell and Ljungberg 2000).

The expertise does not have to fit exactly, i.e., an expert does not need to be a specialist in the specific topic, but rather knowledgeable on the general genre of the task.

This suggests that a mobile knowledge management architecture should suggest experts and present their accessibility.

2.4 Filtering based on task and long-term interests

Mobile workers are often interested in the latest information that is related to the current task (Fagrell and Ljungberg 2000; Fagrell 2000). For example, a person who is travelling to visit a customer may want to be notified about the customer's latest press releases. After the visit, however, the person is less interested in getting this kind of information. We call this task-based interests.

Task-based interests are different from both short-term and long-term interests. A short-term interest is, for example, to retrieve a phone number based on a query, and long-term interests is to filter email from a specific sender based on a profile (Belkin and Croft 1992). Task-based interests are similar to long-term interests, but are only active for the duration of a task.

This suggests that a mobile knowledge management architecture should have filters based on task-based interests. A task-based interest should easily convert into a long-term interest.

2.5 Dynamic configuration of mobile services

The capabilities of mobile devices change rapidly. What once were simple calendar replacements are now expected to have both Internet and Wireless LAN features. Due to limitations in hardware design

most of these devices are not upgradeable, which means that an organisation must invest in new devices to get all the latest services. This has led to a wide range of handheld devices from different manufacturers with different revisions of the same operating system. Therefore, implementing a mobile system in an organisation requires the support for several mobile devices and their specific capabilities.

Handheld devices are also expensive, which in some organisations has led to device sharing among the employees. A shared device is not usable as standard personal calendar or phone register since the applications can only store data from one user at a time. Custom built applications may work, however, if they are designed with this in mind.

Organisations may want to provide unique content and services specifically tailored for each co-worker based on their personal preferences and work environment, for instance restricting access for freelancers.

This suggests that a mobile knowledge management architecture should allow dynamic configuration of the mobile knowledge support according to current user preferences and mobile device capabilities.

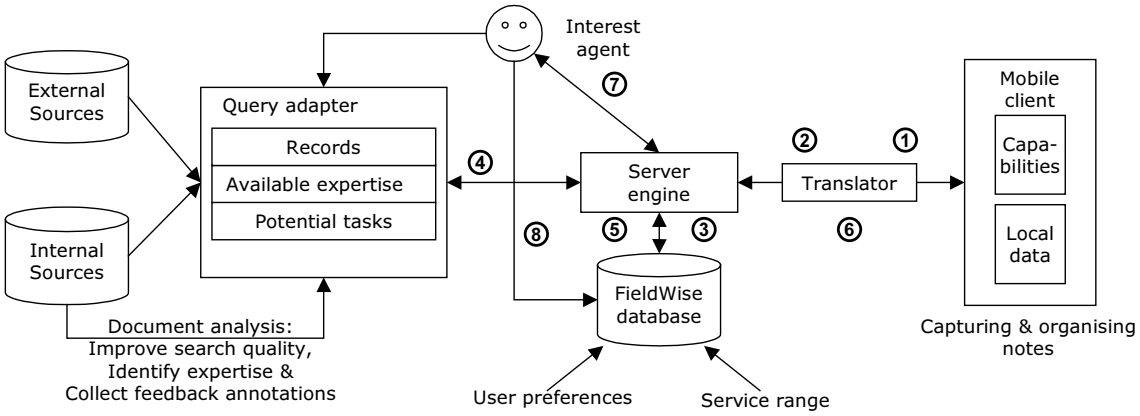

Figure 1: The FieldWise architecture.

2.6 Summing up the requirements

Based on our empirical work and experiences with prototypes, we propose the following main requirements for a mobile knowledge management architecture. It should:

1. Support evolving tasks and notify users of interdependencies.
2. Offer an overview of records, including annotations.
3. Suggest available expertise.
4. Filter information based on task and long-term interest.
5. Enable adaptation to user preferences and mobile device capabilities.

There are of course more general requirements for a CSCW architecture, but our focus is on the specifics of mobile knowledge management.

3. The Fieldwise Architecture

Based on the requirements above, this section describes the general architecture of FieldWise. The following listing describes a use case that illustrates how large parts of the architecture works (figure 1).

1. The mobile client connects and the user's notes are sent to the server engine.
2. The server engine analyses the capabilities of the client and selects the appropriate translator to adapt its response.
3. The server engine creates an interest profile based on the notes and stores it in the FieldWise database.
4. The interest profile is sent to the query adapter, which performs a search in internal and external sources. The server engine receives the results.

5. The server engine looks in the FieldWise database for overlaps with the interest profiles of other co-workers. If there is an overlap, the server engine notifies the co-workers. The response to the user is also adapted according to the user's preferences stored in the FieldWise database.

6. The results are translated and returned to the mobile client, which presents and stores the results locally.

7. The server engine configures the interest agent to continuously monitor for updates that match the user's interest profile.

8. If the interest agent identifies an update, the FieldWise database is modified and the user is notified.

There are, of course, other use cases that are important, e.g., accessing the common resource of potential tasks and feedback annotations, but they are described in of the next section.

3.1 How the architecture meets the requirements

Evolving and interdependent tasks

The mobile client captures and organises the users' task related notes. Based on the notes, as well as the forthcoming query results, a profile is created, which makes it possible to compare users' tasks to identify interdependencies. As the task evolves, due to situated and local circumstances, the user updates the notes and the profile is updated. If two tasks are significantly similar the concerned users are notified, letting them decide if it is necessary to take action.

A commonly shared list of potentially interesting tasks is accessible to all mobile clients. The user can copy a task from the list and develop it further, i.e., adapt it to local circumstances. This also reduces the extra work that is required for note taking.

Overview of records

The query adapter uses the interest profile to search sources that are within the informational context for the organisation, e.g., internal and external archives. The search result is then processed by the server

engine. The purpose is to provide the mobile clients with extracts of records. The extracts are displayed as summarised lists to provide the user with an overview of what has previously been done on the topic. The complete record can also be accessed, but the idea is that the overview, in itself, should remind the user of the context for the task.

Furthermore, if feedback annotations are available they are presented as a part of the result.

Location of available expertise

Based on the profile, people with related expertise, are displayed on the mobile client. The availability of the people is also presented.

Expertise is automatically identified through analysis of records, e.g., authorship and mentioning of internal documents, using the query adapter. This approach is more suitable than to ask people to explicitly state their area of expertise and knowledge level (cf., Kautz et al 1997).

To find out if someone is available is an adaptation for each installation, but typically organisations have back-end security systems for entering the office. Another alternative is to use work schedules or shared calendars. The server engine keeps track of this information by using the query adapter.

Filtering based on task and long-term interests

As tasks evolve throughout the day, new information becomes available. Also, new or evolved tasks may overlap. The interest agent filters out this information for each task, based on the interest profile stored in the FieldWise database.

As long as a task is active the interest agent notifies its user of relevant updates. When a task is deleted or de-activated the interest agent stops the monitoring. Another alternative is that a task-based interest grows into a long-term interest that does not necessarily represent a current work task, but rather an area of expertise. In this case the interest agents remain active.

Dynamic configuration of mobile services

Each mobile client has unique capabilities. The server engine investigates the capabilities of the mobile client, as well as the current network capacity, and translates its response. For example, less information is sent if the mobile client has a slow network connection.

The organisation can adapt the range of services for specific users by updating the user's preferences, which are stored in the FieldWise database. This also affects the responses of the server engine. The services are distributed as components, making it easy for the organisation to add new services.

Maintaining data (user preferences and tasks) on a mobile device is always a problem. The mobile devices may crash, shared between people, or data may be corrupted over time. For this reason, all client data is replicated in the FieldWise database. A user can roam between different FieldWise installations and thus, the user authentication must be handled by another system. The user can rewind to the latest state of the mobile client. The mobile client can also be the host for transporting the user's preferences to another FieldWise installation.

4. A Fieldwise Implementation

In this section we describe an implementation of FieldWise adapted for the news journalism domain. In this case, FieldWise is integrated with an editorial system, called IMpress, which is running at a Swedish newspaper. The overall implementation is illustrated in figure 2. We use the latest mobile devices and (wireless) network technologies that are available. Let us go into the technical details and then see how the implementation meets the requirements.

4.1 Technical details

4.1.1 Network access

The communication is done with TCP/IP sockets through a server on an internal structure. The clients use GSM (Global System for Mobile communication) and Point-to-Point Protocol (PPP) to communicate in wireless mode. GSM is the largest digital wireless communications standard in the world with its 284 million subscribers (EMC 2000). The

clients are connected to a GSM phone with cable or infrared (IrDA 1.0). Today, the bandwidth for GSM is 9600 bits per second and the time to get online is about 20 seconds. Users that are not online are notified using the Short Message Service (SMS). SMS is a part of GSM and makes it possible to send and receive text messages with the mobile phone.

The clients with PCMCIA slot can use wireless LAN (IEEE 802.11b). The client is then connected instantly with a bandwidth of about 11 megabits per second.

The clients can be connected in a stationary mode with a network card or via a serial cable to a networked PC.

4.1.2 Mark-up languages

The internal structure is encoded in XML (eXstensible Mark-up Language) and we use XSLT (eXtensible Stylesheet Language Transformations) for transformations (Clark 1999). In the implementation the query adapter uses XMLNews. In the news industry there are two emerging XML-compliant mark-up languages: XMLNews (Megginson 1999) and News Information Text Format (NITF). The reason we choose XMLNews, is that it is more extensible as the data is separated from the meta-data, allowing multiple meta-data schemas. To describe user preferences and device capabilities we use an XML application called Composite Capability/Preference Profiles (CC/PP) (Reynolds et al 1999).

4.1.3 Operating systems and development platforms

The server is implemented in Java on the Microsoft Windows 2000 platform using an Oracle database manager, which is also the database manager for the editorial system IMpress. To find out the availability of people we use the shared electronic calendar in Microsoft Exchange.

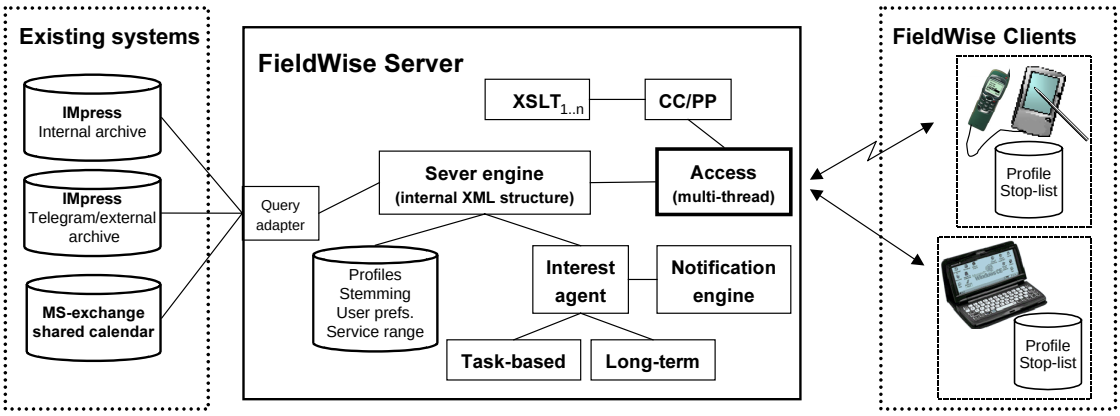

Figure 2: The detail architecture of a FieldWise implementation in the news journalism domain.

4.1.4 Algorithm of the interest profiles

The interest profile is represented with keywords and a blueprint of its search results.

The keywords are filtered out of the task notes through a stop-list, which takes away common words, e.g., he, she and it. For each work domain the stop-list is adapted to filter out high frequency words with low content value. In the news domain the word “interview” is filtered out because it is to common to be a part of interest profile. The stop-list is about 1,500 words (10 kilobytes) and stored on the client. The stop-list is also refined for the informational context for the organisation.

A stemming dictionary is also used to improve the interest profile. It is used to put the keywords on their basic form and manage mis-spellings. For example, the word “Ericsson” will match on different spelling, e.g., “Ericson,” “Erikson,” “Eriksson,” and plural version. The stemming dictionary is generated from two years of text from the internal archive. The dictionary is about 300,000 words (4 megabytes) and is placed on the server because of its size.

Each keyword is weighted with TF-IDF (Salton 1971). Also, a blueprint of the search result is added to the interest profile, giving the algorithm a collaborative filtering aspect, i.e., overlaps between interest profiles can be detected, even if they do not contain the same keywords.

The interest profile is used in conjunction with the database manager's free text functionality. This enables a suitable quality of the search results.

4.2 How the requirements were implemented

Our focus in this section is on the user interface and the specifics of the adaptation for the news journalism domain.

Evolving and interdependent tasks

The application lets the user add, edit and remove tasks and organise the tasks in folders.

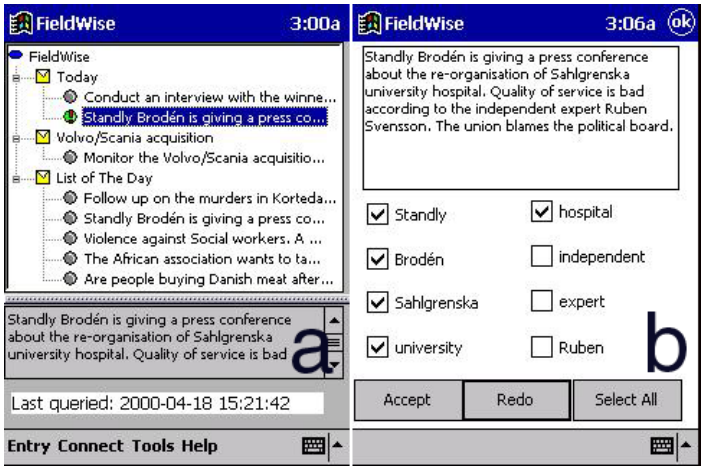

Figure 3: The tree view of tasks and folder (a) and (b) generate keywords. The illustration is on Pocket PC.

All of the folders are personal to the user except the "list of the day," which contains potentially interesting tasks that are shared (figure 3a). The "list of the day" is timely tasks stored in the editorial system, which are suggested by the news editor or by other colleagues. A task can be copied from the "list of the day" folder into another folder of the tree view. The sharing also reduces the amount of input required to construct a new task.

Keywords are automatically generated from the text of the task (figure 3b). The user can choose the keywords that represent what the task is about. Pressing the Redo button generates new keywords. The user does not have to write complete sentences; adding keywords at the end of the text is enough.

If the user is satisfied the Accept button is pressed. After this, the user chooses if the task should be activated. The reason for this is that the users may be interrupted while editing and wants to continue the note taking later on. When the user is online the tasks can be registered by choosing Send task in the Connect menu.

If the task of another co-worker overlaps, the co-workers name and task text is displayed on the Match tab (figure 4d). The co-worker is also be notified (via e.g., SMS).

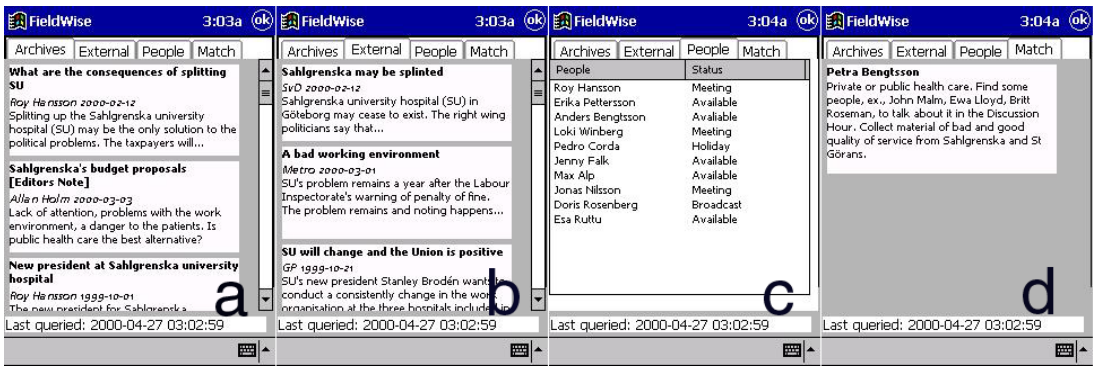

Figure 4: The Archive (a), External (b), People (c), and Match (d) tabs. The illustration is on Pocket PC.

As the task evolves during the day the user updates the notes, re-sends it and receives new information.

Overview of records

All the information on the tabs in figure 4 depends on the active and sent task. The tabs are displayed when a task is registered or Result is chosen in the Entry menu.

The Archive tab displays a list of stories from the internal archives (figure 4a) where each item contains: title, author, publication date and ingress. The External tab displays a list of stories from predefined external sources, e.g., newspaper articles and telegrams (figure 4b). The items on the external list contain title, source, publication date and ingress. Approximately ten items are displayed on each tab.

The intended use of the listed stories is to give the user an overview, i.e., a reminder of what has been done previously. The user can click on a title and get the full text of the story. This is, however, accomplished through a separate system, e.g., a web browser, when the mobile client is online.

Editors evaluate, comment and annotate published stories. This coaching is captured by the editorial system IMpress, and we display this with the article entry on the Archive tab (figure 4a [Editors Note]). By clicking on the Editors Note the user gets the comment (like a Post-it note). Usually these are suggestions of how to follow-up on a story or, on a more general level, how the genre of reporting can be improved.

Location of available expertise

People tab displays a list of co-workers and their availability who have previously been working on a topic related to the current task (figure 4c).

The expertise is identified through the authorship of published articles from the internal archive. We also include documents of the category "issues to watch," that are common in the news domain.

The availability of people is found in the shared electronic calendars system (Microsoft Exchange).

Filtering based on task and long-term interests

As long as a task is active the user is notified about the latest information, e.g., press releases and overlapping tasks of other users.

Today, the notifications are done via SMS (figure 5b), but we also support email. The implementation sends the SMS through a gateway that converts emails into SMS.

A task can be converted into a long-term interest by selecting it and choosing Item options in the Tools menu (figure 5a).

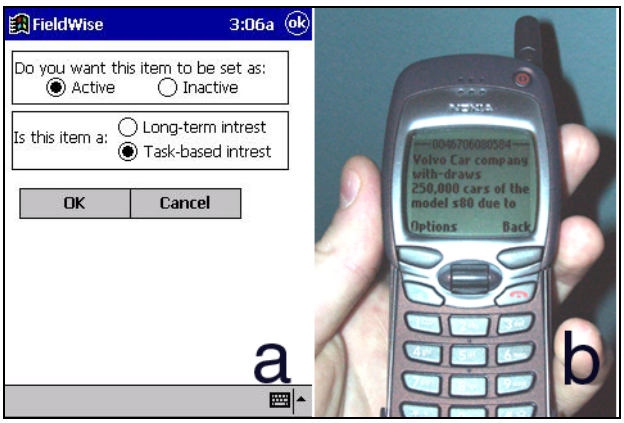

Figure 5: The item options (a) and a GSM phone showing an SMS message (b).

In the case of an overlap between two task-based profiles, both parties are notified.

The functionality of the filter for task-based and long-term interests differs a bit in this implementation. A long-term interest only notifies new information. It does not notify overlaps with a task-based interest. A user with a long-term interest should not take the initiative to approach a colleague that has not asked for help. Instead, if the colleague wants help, the list of expertise can be consulted.

Eventually the colleague's task results in an article that becomes available.

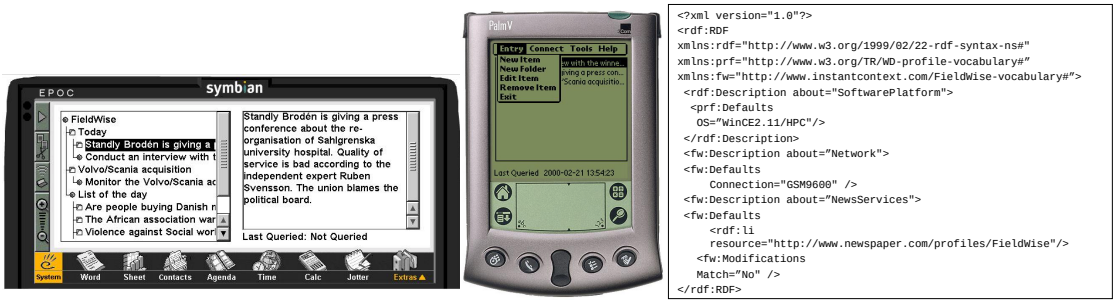

Figure 6: The implementation on EPOC ER5 and Palm OS. The textbox is an extract of a CC/PP identification of a client.

Dynamic configuration of mobile services

The mobile client platforms supported are Pocket PC, Windows CE 2.11 (Palm-sized and Handheld), EPOC ER5, and PalmOS 3.5 (see figure 6).

The mobile devices have different capabilities, e.g., software platform versions, size of display and storage capacity. Also, each device can be connected to the network in a number of ways, e.g., GSM, Wireless LAN and serial cable. Accordingly, the server engine adapts its responses to the capabilities of the device and its network connection. This is done through the selection of a device specific translator (encoded in XSLT), that adapt the responses.

An organisation can define its range of mobile services. For a specific role, or for a unique user, a set of services can be provided. For example, a freelance journalist can be offered a different set of services then the editor. We choose to see each tab on the mobile client as a service, which is dynamically configured.

In order for the server engine to adapt its responses and dynamically configure its services, the clients must be identified. Figure 6 shows an extract of the information needed by the server engine to adapt its responses and dynamically configure its services for the mobile client. The identification is encoded using CC/PP, extended with a FieldWise-specific namespace. It also exemplifies how a default profile, for the range of services, is modified.

The appropriate translator and service configuration is selected depending on:

- **Platform**. Different platforms have different capabilities. For example, the display capacity of the device affects the length of the title and ingress for each article.
- **Connection speed**. If there is a slow connection, less information is transferred to the client. Figure 6 illustrate an identification of a connection with 9,600 bit per second over GSM. For a client with instant access, the notification method is affected as well, i.e., there is no need for SMS.
- **Software version**. Different versions of the client need information in different formats.
- **Services**. The server needs to know the services of the client. Each tab (figure 3) is considered to be a service. Figure 6 illustrate an identification where the Match tab is cancelled.

5. Discussion

The FieldWise architecture as it is described here is especially useful for work organisation where:

- people's tasks are time critical and driven by deadlines
- the result relies on the creativity of autonomous, but interrelated people
- there is a culture of co-operation and sharing of knowledge amongst people

- people are mobile and distributed

The domain of work may be, for example, news journalism, sales and real estate brokering. However, every installation requires integration and adaptation to the local work practice of the organisation. For example, the implementation for the news journalism described here has a special filtering policies for task and long-term interests. The informational context for organisation must be considered, e.g., what are the relevant external sources and how is the information structured?

Since we have not been able to identify any similar architectures for mobile knowledge management the related research is discussed on a feature level.

The features include (1) support for evolving tasks with notifications of interdependencies; (2) overview of records and annotations; (3) suggestions of available expertise; (4) filtering based on task and long-term interest; (5) adaptation to user preferences and mobile devices.

Evolving and interdependent tasks

A foundation for the architecture is that the users evolving task is captured. The representation of the task makes it possible to continuously provide notifications of interdependencies.

Shared task lists has been used by, e.g., Kreifeldt et al (1993), but they display the lists in a common information space, whereas our approach is notify if there is an overlap. Similar to FieldWise, the Yenta system notifies its user of others' with similar interests (Foner 1997). The difference is however that FieldWise match the users based on their current tasks. Also, Cadiz et al (2000) show the importance of matching the awareness tool with the task support in a recent study.

Overview of records

FieldWise use text analysis algorithms to improve search quality. The principle for the visualisation mechanism is to provide overview of records rather than an exhaustive collection.

The quality of the result is based on how well the index works. It needs to be tuned in for the local circumstances. In some cases there is a need to incorporate a separate index to get good quality of the search result. The index does not need to be very advanced to solve the

problem (see, e.g., the NewsMate prototype (Fagrell 2000)). Another app-roach is to use a commercially available product for information filtering, e.g., Autonomy (www.autonomy.com).

In knowledge management systems, e.g., Answer Garden 1-2 (Ackerman 1994; Ackerman and McDonald 1996), Bubble-up (El Sawy and Bowles 1997), and Project memory (Weiser and Morrison 1998) there are search features. The ways in which users can survey search results are very limited. Clearly, these systems are much more oriented towards finding "the matching record" rather than providing the user with an associative overview. There are some exceptions, however, as represented by gIBIS (Conklin and Begeman 1988), Designer assistant (Terveen et al 1995) and RepTool (Jordan et al 1998). These systems offer features to get overview of complex design problems using graphics and hypertext. However, the stationary nature of these systems makes them less useful in the mobile work supported by FieldWise.

Location of available expertise

FieldWise determines if someone is knowledgeable on a topic through the analysis of internal documents. If a person is the author of a document we assume that they have expertise on the topic. The availability of the expertise is an adaptation for each specific FieldWise installation. Emerging technologies that can identify the position of a mobile device can be used. For example, the Ericsson Mobile Positioning System can locate a GSM phone with the accuracy of about 300 meters.

There are PC based applications like ICQ (www.icq.com) that help people to find out if colleagues are available (or busy, etc.). The system is not very sophisticated when it comes to expert location. The Answer Garden 2 (Ackerman and McDonald 1996) and TeamBuilder (Karduck 1994) supports the location of predefined experts. Referral Web helps people to find experts based their relationship in a social network, assuming that topical exercise among co-authors, i.e., an expert can be identified by their participation in co-author relationships or Web page listing (Kautz et al 1997). The Expert Finder (Lieberman and Vivacqua 2000) is tailored for Java programmers and supports the location of expertise by automatically analysing the

programmers source code. None of the systems considers if the expertise is available for interaction.

Filtering based on task and long-term interests

FieldWise use agents to monitor the users task and long-term interests. Notifications are issued about the latest information and task overlaps as long as the interest is active.

Support for long-term interests can be found in several systems, e.g., Bubble-up (El Sawy and Bowles 1997), Fab (Balabanovic and Shoham 1997), IntraNews (Fagrell 1999), Knowledge Pump (Glance et al 1998) and Soap (Voss and Kreifelts 1997) and Yenta (Foner 1997), but neither of them supports task-based interests in combination with notifications.

Dynamic configuration of mobile services

The FieldWise architecture adapt to user preferences and several mobile devices. The users preferences can also be transported between different installations of FieldWise. The architecture is developed to easily deploy future mobile devices and wireless communication technologies.

The only system that we have found that supports knowledge management on mobile devices is Darwin (Kristoffersen and Ljungberg 1998). Darwin supports the distribution and exchange of lessons learned within a dispersed IT-support group.

6. Concluding remarks

In this paper we describe an architecture for mobile knowledge management that can easily be adapted for new mobile work domains, mobile devices and wireless communication technologies. As opposed to similar design efforts the requirements are derived from empirical studies of mobile work.

The research adds to the field of CSCW by offering a set of novel features that have not been reported in the literature previously.

The architecture has been applied in a commercial implementation, which may assure its relevance and usefulness (www.instantcontext.com).

The next implementation of FieldWise focuses on the work domain of mobile sales personnel.

Acknowledgement

This research is a part of the Mobile Informatics programme that is funded by the Swedish Information Technology research Institute (SITI).

The authors like to thank the colleagues in the Mobile Informatics group at the Viktoria Institute, especially Fredrik Ljungberg and Jens Bergqvist. Also, thanks to Adera Business Solutions and ICTech.

A special thanks to Kerry Northrup at IFRA Center for Advanced News Operations.

7. References

Ackerman M.S. and C. Halverson (1998) "Considering an organization's memory," In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 39-48. Seattle, WA: ACM Press.

Ackerman, M.S. (1994) "Augmenting the organizational memory: a field study of answer garden," In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work,* pp. 243-252. Chapel Hill, NC: ACM Press.

Ackerman, M.S. and D.W. McDonald (1996) "Answer Garden 2: Merging organizational memory with collaborative help," In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work,* pp. 97-105, Cambridge, MA. ACM Press.

Balabanovic, M and Y. Shoham (1997) "Fab: content-based, collaborative recommendation," *Communications of the ACM*, 40 (3), pp. 66-72.

Belkin, N.J. and W.B. Croft (1992) "Information Filtering an Information Retrieval: Two sides of the same coin?" *Communication of the ACM*, 37 (7), pp. 29-38.

Bellotti, V. and S. Bly (1996) "Walking away from the desktop computer distributed collaboration and mobility in a product design team," In *Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work*, pp. 209-218, Cambridge, MA: ACM Press.

Bellotti, V. and Y. Rogers (1997) “From Web Press to Web Pressure: Multimedia Representations and Multimedia Publishing,” In *Proceedings of ACM 1997 Conference on Human Factors in Computing Systems*, pp. 279-286. Atlanta, GA: ACM Press.

Bowker, G.C. (1997) “Lest we remember; organizational forgetting and the production of knowledge,” *Accounting, Management and Information Technologies,* 7 (3), pp. 113-138.

Brown, B.A., A. Sellen and K. O’Hara (2000) “A Dairy Study of Information Capture in Working Life,” In *Proceedings of the ACM 2000 Conference on Human Factors in Computing Systems,* pp. 438-445, The Hauge, The Netherlands: ACM Press.

Cadiz, J., A. Espinosa, L. Rico-Gutierrez, R. Kraut, W. Scherlis and G. Lautenbacher (2000) “Coming to the Wrong Decision Quickly: Why Awareness Tools Must be Matched with Appropriate Tasks,” In *Proceedings of the ACM 2000 Conference on Human Factors in Computing Systems,* pp. 392-399, The Hauge, The Netherlands: ACM Press.

Clark, J. (1999) XSL Transformations (XSLT) Version 1.0. W3C Recommendation 16 November 1999. http://www.w3.org/TR/xslt

Conklin, J. and M.L. Begeman (1988) “gIBIS: a hypertext tool for exploratory policy discussion,” *ACM Transactions on Office Information Systems,* 6 (4), pp. 303-331.

Ehrlich, K. and D. Cash (1999) “The invisible world of intermediaries: a cautionary tale,” *Computer Supported Cooperative Work: The Journal of Collaborative Computing,* 8 (1-2), pp. 147-167.

El Sawy, O.A. and G.E. Bowles (1997) “Redesigning the customer support process for the electronic economy: Insights from storage dimensions,” *MIS Quarterly*, 21 (4), pp. 457-484.

EMC World Cellular Database. http://www.gsmworld.com/

Fagrell, H. (1999) “IntraNews: A News Recom-mending Service for Corporate Intranets,” In *Proceedings of Computer Supported Cooperative Work in Design*, pp. 323-328, Compiègne, France.

Fagrell, H. (2000) “NewsMate: Providing Timely Knowledge to Mobile and Distributed News Journalists,” Revised version accepted for publication in *Beyond Knowledge Management: Managing Expertise, M. Ackerman et al. (eds.).*

Fagrell, H. and F. Ljungberg (2000) “A Field Study of News Journalism: Implications for Knowledge Management Systems,” Accepted *for publication in Proceedings of the Sixth Biennial Participatory Design Conference*.

Fagrell, H., S. Kristoffersen and F. Ljungberg (1999) “Exploring Support for Knowledge Management in Mobile Work,” In *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work,* pp. 259-275, Copenhagen, Denmark: Kluwer Academic Publishers.

Foner, L. (1997) "Yenta: A Multi-Agent, Referral Based Matchmaking System," In *Proceedings of the first international conference on Autonomous agents,* pp 301-307, Marina del Rey, CA: ACM Press.

Forsberg, K. and L. Dannstedt (2000) "Extensible use of RDF in a Business context," In *Proceedings of the Ninth International World Wide Web Conference,* Amsterdam, The Netherlands: Elsevier Science.

Glance, N., A. Damián and M. Dardenne (1998) "Knowledge Pump: Supporting the Flow and Use of Knowledge," In *Information Technology for Knowledge management,* U.M. Borghoff and R. Pareschi (eds.). Berlin: Springer-Verlag.

Jordan, B., R. Goldman and A. Eichler (1998) "A Technology for Supporting Knowledge Work: The RepTool," In *Information Technology for Knowledge management,* U.M. Borghoff and R. Pareschi (eds.), pp. 79-96. Berlin: Springer-Verlag.

Karduck, A. (1994) "TeamBuilder: a CSCW tool for identifying experts and team formation," *Computer Communications*, 17 (11), pp. 777-887.

Kautz, H., B. Selman and M. Shah (1997) "Referral Web: combining social networks and collaborative filtering," *Communications of the ACM*, 40 (3), pp. 63-65.

Kidd, A. (1994) "The Marks are on the Knowledge Worker," In *Proceedings of the ACM 1993 Conference on Human Factors in Computing Systems,* pp. 186-191. Boston, MA: ACM Press.

Kreifeldt, T., E. Hinrichs and G. Woetzel (1993) "Sharing To-Do Lists with a Distributed Task Manager," In *Proceedings of the third European Conference on Computer-Supported Cooperative Work,* pp. 31-45. Milan, Italy: Kluwer Academic Publishers.

Kristoffersen, S. and F. Ljungberg (1998) "MobiCom: Networking dispersed groups," *Interacting with Computers,* 10 (1), pp. 45-65.

Lieberman, H. and A. Vivacqua (2000) "Agents to Assist in Finding Help," In *Proceedings of the ACM 2000 Conference on Human Factors in Computing Systems,* pp. 65-72, The Hauge, The Netherlands: ACM Press.

Ljungberg, F., B. Dahlbom, H. Fagrell, M. Bergquist and P. Ljungstrand (1998) "Innovation of new IT use: Combining approaches and perspectives in R&D projects," In *Proceedings of the Fifth Biennial Participatory Design Conference*, pp. 203-209, Seattle, GA: CPSR and ACM.

Luff, P. and C. Heath (1998) "Mobility in Collaboration," In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work*, pp. 305-314, Seattle, WA: ACM Press.

McDonald, D.W. and M.S. Ackerman (1998) "Just talk to me: a field study of expertise location," In *Proceedings of ACM 1998 Conference on Computer Supported Cooperative Work,* pp. 39-48. Seattle, WA: ACM Press.

Megginson, D. (1999) XMLNews-Story (1999-04-05) Technical Specification. http://www.xmlnews.org/

Randall, D., J. O'Brien, M. Rouncefield and J.A. Hughes (1996) "Organisational Memory and CSCW: Supporting the 'Mavis Phenomenon'," In *Proceedings of the sixth Australian Conference on Computer-Human Interaction*. Hamilton, New Zealand: IEEE Computer Society.

Reynolds, F., J. Hjelm, S. Dawkins and S. Singhal (1999) Composite Capability/Preference Profiles (CC/PP): A user side framework for content negotiation. W3C Note 27 July 1999. http://www.w3.org/TR/NOTE-CCPP/

Salton, G. (1971) *The SMART Retrieval System*, New Jersey: Englewood Cliffs, Prentice-Hall, Inc.

Terveen, L.G., P.G. Selfridge and M.D. Long (1995) "Living Design Memory: Framework, Implementation, Lessons Learned," *Human-Computer Interaction*, (10) 1, pp. 1-37.

Tsoukas, H. (1996) "The firm as a distributed knowledge system: a contsructionist approach," *Strategic Management Journal*, 17, Winter Special Issue, pp. 11-25.

Weiser, M. and J., Morrison (1998) "Project Memory: Information Management for project teams," *Journal of Management Information Systems*, 14 (4), pp. 149-167.

Voss, A. and T. Kreifelts (1997) "SOAP: Social Agents Providing People With Useful Information In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, pp. 291-298, Phoenix, AZ: ACM Press.

Gothenburg Studies in Informatics (ISSN 1400-741X)

1. Ulf Sundin. *A Logic Programming Approach to Information Modelling and Database Design*, May 1990. (Licentiate Thesis).
2. Thanos Magoulas and Kalevi Pessi. *En studie om informations-systemsarkitekturer* (in Swedish), February 1991. (Licentiate Thesis).
3. Peter Nordenstam. *Individbaserade relativt öppna informationssystem* (in Swedish). February, 1990. (Licentiate Thesis).
4. Bo Dahlbom and Lars Mathiassen. *Struggling with quality*. The Philosophy of Developing Computer Systems, August 1991. (Revised edition, Computers in Context. The Philosophy and Practice of Systems Design. Blackwell Publishers Oxford: Blackwell, 1993.)
5. Börje Langefors. *Essays on infology. Summing up and Planning for the Future*, Edited by Bo Dahlbom, August, 1993.
6. Bo Dahlbom (eds.). *The Infological Equation. Essays in honor of Börje Langefors*, March 1995.
7. Bo Dahlbom, Frederik Kämmerer, Fredrik Ljungberg, Jan Stage and Carsten Sørensen (eds.). *Design in Context*. Proceedings of the 18th Information Systems Research Seminar in Scandinavia, June, 1995.
8. Bo Dahlbom, Fredrik Ljungberg, Urban Nuldén, Kai Simon, Jan Stage and Carsten Sørensen (eds.). *The Future*. Proceedings of the 19th Information Systems Research Seminar in Scandinavia, June 1996.
9. Agneta Ranerup. *Användarmedverkan med representanter* (in Swedish), August 1996. (Doctoral Thesis).
10. Ole Hanseth. *Information Technology as Infrastructure*, November 1996. (Doctoral Thesis).
11. Fredrik Ljungberg. *Networking*, September 1997. (Doctoral Thesis).
12. Jan Ljungberg. *From Workflow to Conversation*, October 1997. (Doctoral Thesis).
13. Thanos Magoulas and Kalevi Pessi. *Strategisk IT-management* (in Swedish), March 1998. (Doctoral Thesis).
14. Fredrik Ljungberg (eds.). *Informatics in the Next Millennium*. Essays in honor of Bo Dahlbom, June 1999.
15. Urban Nuldén. *e-ducation*, May 1999. (Doctoral Thesis).
16. Lars Erik Holmquist. *Breaking the Screen Barrier*, May 2000. (Doctoral Thesis).

17. Nina Lundberg. *IT in Healthcare. Artefacts, Infrastructures and Work Practices*, June 2000. (Doctoral Thesis).

18. Henrik Fagrell. *Mobile Knowledge*, October 2000. (Doctoral Thesis).