Organization and Usage of Learning Objects within Personal Computers

Paul-Alexandru Chirita, Kerstin Bischoff, Martin Wolpers, Vladan Devedžić, Alexandra Cristea, Leo Sauermann, Sergey Chernov, Eelco Herder, Zoran evarac, Jehad Najjar, et al.

Network of Excellence in Professional Learning

PROLEARN

European Commission Sixth Framework Project (IST-507310)

Deliverable	***D7.6 Organization and Usage of Learning Objects within Personal Computers***

Editor *Paul – Alexandru Chirita*

Work Package *WP7000: Knowledge Work Management*

Status *V1.00*

Date *29.11.2006*

The PROLEARN Consortium

1. Universität Hannover, Learning Lab Lower Saxony (L3S), Germany
2. Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany
3. Open University (OU), UK
4. Katholieke Universiteit Leuven (K.U.Leuven) / ARIADNE Foundation, Belgium
5. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (FHG), Germany
6. Wirtschaftsuniversität Wien (WUW), Austria
7. Universität für Bodenkultur, Zentrum für Soziale Innovation (CSI), Austria
8. École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
9. Eigenössische Technische Hochschule Zürich (ETHZ), Switzerland
10. Politecnico di Milano (POLIMI), Italy
11. Jožef Stefan Institute (JSI), Slovenia
12. Universidad Polictécnica de Madrid (UPM), Spain
13. Kungl. Tekniska Högskolan (KTH), Sweden
14. National Centre for Scientific Research “Demokritos” (NCSR), Greece
15. Institut National des Télécommunications (INT), France
16. Hautes Etudes Commerciales (HEC), France
17. Technische Universiteit Eindhoven (TU/e), Netherlands
18. Rheinisch-Westfälische Technische Hochschule Aachen (RWTH), Germany
19. Helsinki University of Technology (HUT), Finland
20. imc information multimedia communication AG (IMC), Germany
21. Open Universiteit Nederland (OU NL), Netherlands

Document Control

Title: Organization and Usage of Learning Objects within Personal Computers

Type: Public

Editor: Paul – Alexandru Chirita

E-mail: chirita@l3s.de

Origin: L3S

AMENDMENT HISTORY

Version	Date	Author	Description/Comments
0.10	19.07.06	Paul - A. Chirita	First general draft outline.
0.20	08.08.06	Alexandra Cristea, Maurice Hendrix	Contribution from TU/E.
0.25	10.09.06	Vladan Devedzic	Draft ideas from the Univ. of Belgrade.
0.30	20.10.06	Kerstin Bischoff, Paul - A. Chirita	Draft ideas from L3S / Desktop usage survey.
0.40	20.11.06	Sergey Chernov	KTH visit report, Desktop organization study.
0.50	26.11.06	Vladan Devedzic, Zoran Ševarac	Complete contrib. from the Univ. of Belgrade.
0.60	26.11.06	Stefania Costache, Leo Sauermann	Semantic Desktop workshop report.
0.75	29.11.06	Martin Wolpers, Jehad Najjar, Erik Duval	CAMA workshop report.
0.85	29.11.06	Kerstin Bischoff	Complete Desktop usage survey.
0.90	04.12.06	Paul - A. Chirita	Introduction, merging contributions, etc.
0.95	08.12.06	Eelco Herder, Paul - A. Chirita	Review and minor changes.
1.00	27.12.06	Paul - A. Chirita	Worked in the comments from the ProLearn reviewers.

Contributors

Name	**Company**
Kerstin Bischoff	Universität Hannover (L3S)
Martin Wolpers	Katholieke Universiteit Leuven (KU Leuven)
Vladan Devedzic	University of Belgrade
Paul – Alexandru Chirita	Universität Hannover (L3S)
Alexandra Cristea	Technische Universiteit Eindhoven (TU/e) / University of Warwick
Leo Sauermann	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
Sergey Chernov	Universität Hannover (L3S)
Eelco Herder	Universität Hannover (L3S)
Zoran Ševarac	University of Belgrade
Jehad Najjar	Katholieke Universiteit Leuven (KU Leuven)
Erik Duval	Katholieke Universiteit Leuven (KU Leuven)
Maurice Hendrix	Technische Universiteit Eindhoven (TU/e) / University of Warwick

Table of Contents

1. Introduction

Given that in our information-based society and its rapidly changing environment knowledge is soon out-dated, people and organizations continuously need to update their skills to remain competent and competitive [European Commission, 2001]. Informal learning processes like Peer-To-Peer teaching and self-steered knowledge delivery have been identified as crucial for learning in an organization setting, especially for SMEs [Rosenberg, 2001, 2006; Cross, 2003]. Particularly knowledge workers, who spend most of their time retrieving, processing, creating and manipulating knowledge, rely on efficient access to data in different formats like documents, Web pages, software code, email messages, etc. in order to fulfill the immediate information need triggered by Personal Information Management activities [Stumpf et al., 2005].

However, due to increased capacities of hard-disk drives, as well as cheaper and faster storage and processing power, the amount of data stored on the PC Desktop is continually increasing. Moreover, personal and organizational resources are spread over multiple, purpose-specific applications not searchable in an integrated manner through search tools offered by mainstream operating systems. Thus, finding documents in one's own information space is sometimes more difficult and takes longer than searching the World Wide Web. The recent arrival of Desktop organization and search applications reveals the rich opportunities for Desktop applications within the e-Learning area and beyond.

Prior deliverables in WP 7 outlined the importance of Knowledge Work Management (KWM) for professional learning arrangements. More specifically, they identified Key Research Areas (D7.1), structured them into an integrated model (D7.3), presented a framework for identifying related research and best practices (D7.2) and reported activities in the KWM field (D7.4). Deliverable 7.5 gave recommendations and guidelines for KWM and emphasized the need to focus on PC Desktop specific research as it strongly influences the learning process. Hence, to promote the integration of Desktop related Knowledge Management and Technology Enhanced Learning this deliverable aims at increasing the awareness of Desktop research within the Professional Learning community and at familiarizing the e-Learning researchers with the state-of-the-art in the relevant areas of Personal Information Management (PIM), as well as with the currently on-going activities and some of the regular PIM publication venues.

The deliverable is organized as follows. After having introduced the reader into the Desktop specific e-Learning topics (Sections 1.1 and 1.2), we describe the major PIM events and activities organized within the context of ProLearn (Section 2). Then, in Section 3 we present an overview of the already available specific Desktop research results. More specifically, we first briefly survey the currently existing publications, and then discuss in more detail those approaches which have been developed in the ProLearn community.

1.1. Desktop Documents and their Organization

Living in the information age with the associated problem of information overload, people permanently encounter, receive or access more information than they actually need and process. The PC Desktop as workplace and tool represents the personalized subset of this information world people create to respond to this challenge [Bruce et al., 2004]. It forms a personal collection of resources that are considered important for current or future personal or work-related activities. As such, the Personal Desktop spans files in various formats, emails, notes, calendars, Web cache, links, instant messaging, etc. This space is only loosely structured into folders and files. Moreover, nowadays information accumulates fast over time. Take for example emails, which may easily produce highly cluttered inboxes [Whittaker & Sidner, 1996], due to the increased amount of necessary content management. Therefore, effectively managing one's information space is crucial in order to keep track of the personal resources, as well as to easily locate them when needed.

In order to design Desktop tools that effectively support users in working with their information, many studies have observed people's organization behavior considering their practices of filing and searching, as well as handling interruptions through notification. Results show that people have many problems with current Desktop systems and invest much effort in organizing and/or finding needed items, and thus enhanced applications for Desktop organization and searching are necessary. Based on these findings, some approaches have been taken to facilitate organization and re-access. Systems like Beagle++ [Chirita et al., 2006a] focus on improving search facilities by integrating different resources and allowing for more flexible and intuitive browsing and querying. Similarly, related issues of task switching and interruption are addressed by systems that reduce the overhead trading off value of immediate notification and disruptive impact. Learning within the PC Desktop environment can directly benefit from the advantages introduced by all these tools, not only for workflow and learning support, but also for the delivery of learning objects.

1.2. Learning Objects on the PC Desktop

Learning objects (LO) to be organized and found on the PC Desktop can have various formats. While there may be well-structured learning objects, units or courses created by authors according to a well-chosen instructional design, most resources especially for informal learning processes are stored documents, emails and visited Websites. These may be used to give instruction, document prior working experiences, etc. [see Rosenberg, 2001; 2006], thus turning the Desktop into an integrated medium for working and learning. In fact, teachers use most often PCs for preparing their learning objects, learning units, and assessment materials, either at their offices or at home, where they are also used to Desktop based working environments. Moreover, testing and reusing the learning objects they make often includes instructional designers and teachers as well, and again it is predominantly an office where this happens, using Desktop technology.

In order to use the learning technology on their Desktop computers successfully, end-users need appropriate software tools. Learners need learning environments that allow them to connect to learning management systems, courses, learning object repositories, assessment systems, and the like. Learning object (Content) authors need all that as well, plus tools that allow them to compose learning objects starting from various resource files, annotate them with standards-based metadata, and upload them to learning object repositories. Teachers also need PC based assessment tools, various learner monitoring tools, grading tools, and administrative tools. All end users often use learning management systems. In addition, there are a number of education-specific software tools that are often very handy to end users and in fact research in e-learning often results in creation of such tools. As an example consider *Neuro-Fuzzy Reasoner* (*NFR*) for learner modeling [Sevarac, 2006], a software component capable of learning a set of predefined fuzzy rules which can be successfully used for addressing learner modeling issues. Its development came from practical needs of teachers who use Desktop PC tools to manage their classes.

The next chapter describes in more detail several events and activities relevant for the Desktop based Professional Learning research.

2. Events and Activities relevant for the Desktop based Professional Learning

Identifying the right places to seek information is a crucial task in any research activity. This section will therefore familiarize the reader with the on-going Desktop Research events organized within the context of ProLearn. We will first present an overview of two recent workshops, CAMA and SemDesk, the former one being especially organized by respected members of the Professional Learning community. In fact, we believe that these two events, together with the Personal Information Management Workshop organized in conjunction with the ACM SIGIR Conference, represent some of the most specifically Desktop oriented research meetings of 2006. The last part of this section will briefly describe the L3S research visit to the KTH group of Human Computer Interaction, a highly important department when bridging over highly user oriented domains such as Personal Information Management and e-Learning.

2.1. CIKM Workshop on Contextualized Attention Metadata (CAMA)

Providing flexible efficient access to relevant information remains a challenge. Approaches based on attention metadata present novel opportunities to deal with this challenge. Attention metadata indicate what end users do with specific information objects. End user applications (authoring tools, search tools, repositories, email applications, etc.) can generate attention metadata. Collecting, managing, merging, analyzing and exchanging these distributed metadata enables smart tools to exploit data about user experiences with heterogeneous tools to provide more flexible user interactions, while respecting user privacy. For example users of one application can be recommended relevant materials by analyzing the attention they gave to information residing within other applications they interacted with.

CAMA 2006 was held in conjunction with the 15th ACM Conference on Information and Knowledge Management (CIKM 2006) in Arlington, USA, on November, 6-11, 2006. Facilitating the usage of contextualized attention metadata is a cross-disciplinary research topic that combines research results and challenges from communities including databases, information retrieval, knowledge representation, knowledge management, user modeling, recommender systems, learning technologies, etc. The main objective of the workshop was to bring together researchers, industrial practitioners, and developers working on relevant topics to discuss recent advances and identify future challenges.

The main topics of the workshop included (1) Attention metadata collection frameworks, (2) Management of large volumes of attention metadata, (3) Real-Time stream based attention metadata processing, (4) Methods and algorithms for analysis of attention metadata, (5) Information extraction in attention metadata streams, (6) Privacy and Security, (7) Personal Information Management, (8) Representation of attention metadata, (9) Applications of attention metadata streams, and others.

The workshop program consisted of six paper presentations and two keynote speakers. The three short and three long papers were organized in two tracks dealing with the two main issues of attention metadata, namely capture and usage of metadata focused on attention.

In order to have a highly interactive environment for the full day workshop, we started the first session by giving the floor to the workshop participants. They were asked to state their definition of attention metadata and how they envision using it in their contexts. The summary of their answers yields a number of definitions of attention metadata:

Attention metadata describes the behavior with which users deal with information. It can be collected by observing the user's activities. It possibly allows deriving conclusions on the user's interests and goals by relating activity to content. Furthermore, by relating attention metadata of several users, behavioral patterns might be observed.

Steve Gillmor, our first keynote speaker, elaborated on current developments with GestureBank, a currently set-up non-profit organization that strives to open the usage of attention metadata to a wider public. GestureBank provides a solution for secure and reliable storage of private attention metadata. As claimed by Steve, enabling the user to stay in control of his/her data is their main concern.

Our second keynote speaker, David Wiley (Utah State University), discussed an approach using attention metadata to drive Web based reading in working groups. The approach is built on observing group member Web based reading behavior and the usage of the information to facilitate reading recommendations through RSS feeds.

The workshop attracted the participation of 16 researchers arching across various communities, ranging from database and information retrieval to knowledge management and e-Learning. The size was small enough to allow an interactive format and continuous discussions between and during presentations. Participants found the workshop very productive and successful. More details about the workshop can be found online at http://ariadne.cs.kuleuven.be/cama2006/.

Two research issues have been identified to focus the further work on. The usage of contextualized attention metadata has not yet been researched to its full extent. Very preliminary work emerges that shows highly encouraging results. Nevertheless, commercial and industrial uptake is not yet to be seen, as real applications of such data have not been realized. Emerging fields are for example advanced usage statistics, behavioral pattern mining and finally advanced personalization to meet the requests of the single user.

2.2. ISWC Workshop on the Semantic Desktop (SemDesk)

There is a definite need for smarter and more fine-grained computer support for Personal Information Management and handling information in general. We thus held the second edition of the workshop on the Semantic Desktop and Social Semantic Collaboration in conjunction with the fifth International Semantic Web Conference (ISWC2006). The workshop was organized by Stefan Decker, Jack Park, Leo Sauermann, Sören Auer, and Siegfried Handschuh.

The research community needs to blend the boundaries between personal and group data, while simultaneously safeguarding privacy and establishing and deploying trust among collaborators. First, the Semantic Web provides technology to handle any generic data. Second, LOM and other e-learning initiatives provide metadata and standards to empower users. Finally, Peer-To-Peer systems, social software and CSCW allow for collaboration and data exchange. Yet each of these addresses only parts of the picture. The series of SemDesk workshops aim for a merge of these thrusts.

In the 2006 edition, 28 papers were submitted to the workshop, out of which 2 position papers and 11 full papers were accepted for publication. Reviewers came from an international program committee, some of them also members of the ProLearn community. Just like in the previous year, the workshop hosted its own poster and demo session, where practical implementations and projects were shown. These, together with the four paper presentation sessions, completed the one day duration of the workshop. About 40 people attended. The exchange between groups continued and we saw substantial improvement of existing work and also new work presented. The collaboration between attendees is a key factor: For example, last year two collaborations were started at the workshop, one of the results being presented this year, namely the paper by Fernandez-Garcia et al. [2006].

As a result of the workshop we established a moderately large Personal Information Management community, communicating via an online platform at the Website http://www.semanticDesktop.org. Moreover, the papers were published as CEUR-WS workshop proceedings and are all available at the CEUR-WS Website without charge. Most of them are also briefly presented in the Section 3 of this deliverable.

2.3. L3S Visit to the KTH Group of Human Computer Interaction

To identify user patterns for Desktop Data Management, L3S has conducted a set of interviews with 10 different users, representing 10 highly different professional areas. One of the main conclusions of this study was a high diversity of Desktop usage, where there could be no single user model. In order to investigate which user specific feature contributes to which usage pattern, a joint workshop was scheduled for one week within May 2006, in Stockholm, with and hosted by the KTH Group of Human Computer Interaction. The Swedish partner group possesses a strong expertise in user study methodologies, so their input was highly relevant and useful for future interview planning and analysis. Specifically, the feedback was provided by Dr. Bosse Westerlund and Rosa Gudjonsdottir.

The series of meetings started with the presentation of the related user study results. Possible Desktop usage scenarios were modeled by the KTH group with a paper based sketch of a graphical user. Each scenario has been simulated step-by-step, videotaped and supported with the detailed task workflow comments. Later on, the KTH colleagues presented three sample user models, where for each user they specified age, sex, name, occupation, appearance, work-related skills and hobbies, some other preferences, etc.

The second part of the visit concentrated on discussions over several issues related to the first round of experimental interviews conducted within L3S, as well as to the planning of a future, enhanced usage study. Several conclusions drawn include: (1) More attention should be paid to critical cases; there could be no average case as people use the PC Desktop quite differently; (2) the quantity of the 10 to 20 subjects is sufficient for a qualitative user study; (3) the variance in perceived privacy could be insignificant if one compares it with an influence of the personal factors; (4) the interviews should be videotaped in front of user’s workplaces, as this adds to the regular usage tasks the important context information, and therefore makes the interview materials suitable for a post-analysis, in case the researchers come up later with any additional hypotheses.

Taken into account the comments from the KTH colleagues, a report has been written, parts of it being described in Section 3.4.5. Moreover, a second round of interviews is planned to be performed next spring. While the number of recipients will remain roughly the same, they will belong to only one professional area.

3. Research Results

This section gives an overview of the existing research in Personal Information Management, relevant for the Professional Learning researcher. We first look into several approaches to assist and investigate generic Desktop Data Organization. We then present some of the task management support and content specific metadata generation techniques. Subsequently, the second part of the chapter dives into the details of some PIM related research results which have been developed in the context of ProLearn. It first introduces the MOT system for adaptive course authoring, then a malleable schema approach for generating contextual metadata, and in the end it depicts the results of a Desktop oriented usage analysis survey.

3.1. Organizing Personal Documents

We have already seen that Desktop Data Organization is gaining momentum, as no widely accepted solution has been proposed so far. This section gives an overview of the currently existing approaches, first to study current data organizational patterns, and then to design applications that exploit these patterns to produce a better Desktop activity experience.

3.1.1. Patterns in Desktop Data Organization

Since the amount of personal or work related information spread over various applications on our PCs is continuously increasing, managing one's personal information space seems critical to effectively working with and re-finding important information despite overload. In order to identify how to better support people in performing tasks, in organizing and retrieving their information, many studies have examined organizational behavior on the personal Desktop and drawn conclusions for the design of advanced PIM systems. As most of these studies focus only on a specific aspect of PIM, we structured this section accordingly: First, we review the strategies people employ to keep track of their (1) documents, (2) emails and (3) information encountered on the Web; then, the second part of our survey discusses (4) how these resources are integrated into tasks.

Documents have been found to predominantly carry "working" information closely related to current tasks, as well as to frequently used information, which later becomes "archived" for future re-use [Barreau & Nardi, 1995; Cole, 1982]. In an early study on desk organization of print based information, Malone [1983] reported two basic keeping methods for documents: (1) systematically ordered (titled) files and (2) loosely arranged piles of individual items. One reason for the frequently employed piling may be cognitive difficulties in classifying information. Landsdale [1988] examined in detail the difficulties in categorization and setting up folder structures anticipating future usage and retrieval contexts. More importantly, organization in piles serves as a reminding function by enforcing serendipitous encounters [Malone, 1983]. In their studies on filing practices of electronic documents, Barreau and Nardi [1995] found that documents are usually organized into folders by location or category such as projects to ease re-finding and enable being reminded. Users tended to place items requiring actions to do on certain Desktop places or on an upper level of a directory, where they are likely to notice it (thus acting as behavioral triggers). They avoid setting up elaborate filing schemes for archiving old information because it is currently not worth the effort. Ravasio et al. [2004a] observed that people arranged their Desktop screen as a temporary storage facility and active workplace for fast orientation by sight mainly by clustering according to thematic proximity or document type. They showed that filing is a continuous process. Henderson [2005] described that labeling folders follows some stable patterns, e.g., folder names most often include genre, task, topic or time. Thus, it seems that classification depends much on context [Kwasnik, 1991]. Interestingly, Teevan et al. [2004] found the general preference of small steps using local context (orienteering behavior) over direct search, even if reliable cues were known. This is because search always requires remembering some part of the information, and thus defining a query can be as cognitively hard as efficiently organizing the data [Marshall & Jones, 2006]. Moreover, current mainstream text search tools are not even easy to handle and do not support the rich associations people have with their information (as retrieval cues) [Cutrell et al., 2006b; Landsdale, 1988].

Email predominantly carries ephemeral information [Barreau & Nardi, 1995]. Its management is overloaded with functions like task management and personal archiving [Whittaker & Sidner, 1996]. Most users have full inboxes and few folders for archiving mail, mainly because (1) filing information is cognitively difficult requiring time and effort, and more importantly because (2) current items shall be easily accessible and function as reminders. They distinguished three user types or strategies: No filers, Frequent Filers and Spring Cleaners. Only frequent filers who delete and file emails on a daily basis preserve opportunistically being reminded by the few current items in their inbox and are more effective in their use of folders, yet they invest much for this. Bälter [1997] refined this model theoretically to differentiate Spring Cleaners and to account for possible transitions between pro-organizing and anti-organizing strategies encouraged by time pressure and email overload. Duchenaut and Bellotti [2001] pointed out the role of the new habitat email in document exchange, which is promoted by embedding in context and workflow. They found shallow file hierarchies to be common, which is explained by preference for visually scanning an ordered list over clicking with the mouse in order to have immediate access. Folders were commonly organized by criteria like sender (the most salient retrieval cue of emails according to Whittaker et al. [2006]), organization, project or personal interests. Location based search and sorting of mails by sender, date or subject with subsequent browsing were popular strategies employed in looking for a message; global text search tools, as well as automatic filters were found to be less frequently used.

To examine how people assure that important information encountered on **Websites** is re-found at the time it is needed, Bruce et al. [2004] empirically collected a list of common keeping methods: sending an email with the URL to self, printing out a Webpage, bookmarking, saving it to a file, pasting the URL into a document, adding a hyperlink on a personal Website, writing down notes and the URL, copy the link to a toolbar or creating a note in MS Outlook. Other strategies, actually presenting leaving behavior, were entering URL from memory, searching, and accessing from a known access point. Thus, often a conscious decision is made to do nothing to include the to-be relevant information into the personal information collection, being convinced of the ability to easily re-locate it. In a second re-finding study participants were almost always able to quickly re-find the information using the first method they tried [Bruce et al., 2004]. Most often URLs were directly entered, found via bookmarks, Web searching or access through another Website. Thus, three of the most often employed strategies did not present active keeping. Though the authors concluded that people employ appropriate methods for keeping with the information overload, reliance on Web search tools seems high (in contrast to Desktop ones). Although frequently utilized, bookmarks are hardly ever organized even by users extensively filing documents or emails. All in all, organizing strategies seem to be consistent over time [Boardmann & Sasse, 2004].

A major problem users face in current Desktop systems is **fragmentation of information access**; documents, emails or bookmarks are spread over the PC and bound to special purpose applications. Whittaker et al. [2006] discuss two approaches to solving the fragmentation problem: (1) centralization, i.e., the integration of all PIM tools into one application, or (2) information extraction, making information within emails, bookmarks, etc. accessible to all PIM applications. While centralization reduces the need for transferring and entering information into several special tools, it is not well realized yet. Extraction, on the other hand, has the disadvantage of not keeping context. Although both approaches offer distinct improvements over the current situation, a combination seems to be best. At the moment, users may utilize extra monitors to extend screen space and partition between main task and secondary tasks (like email) in order not to have to switch too often between application windows [Grundin, 2001]. Ringel [2003] reported that stable mappings to separate subtasks or tasks were more common on both extra monitors and virtual Desktops. Examining the notions of task switching and interruption, Cutrell et al. [2001] conducted a study on the effects of notifications by instant messaging on memory and hence task performance. They found notifications to be especially harmful when occurring early during a task as this could cause forgetting the primary goal. In general the disruptive impact is greater on fast, stimulus driven tasks, whereas in performing a cognitively demanding task more time to switch to a notification and more reminders may be needed. Prior research showed that disruptions have a serious negative impact if the content is very similar to the current task [Gillie & Broadbent, 1989]. However, training or expertise in handling interruptions [Hess & Detweiler, 1994], as well as recovery techniques [Linde & Goguen, 1987] can visibly reduce this disruptive impact on memory.

To summarize, users face many problems in managing their data, especially regarding information fragmentation and overload. While location based browsing is often successfully used to find personal files, classification and structuring are time consuming and cognitively hard. Search tools are not frequently used because they lack important features, are not easy to handle and do not support functions like reminding. For filing email little effort is commonly invested, which may be related to time pressure, immense overload and the reliance on finding them by sorting or targeted search. Automatic assistance in filing (multiple classification), search facilities offering meaningful attributes, reminding, task management and the associated problem of disturbance are critical in efficient PIM support. The next chapter presents some applications aiming at such improvement, concentrating on infrastructures for organizing Desktop data.

3.1.2. Infrastructures for Desktop Data Organization

Due to problems users have in managing their information for future re-finding, various approaches and tools have been proposed to organize and search personal information spaces more naturally and efficiently in an integrated manner [Oren, 2006]. This section will introduce the reader to some of these applications and the research contributions they brought. Such enhanced storing and retrieving of information is important for sharing resources such as learning materials in informal Peer-To-Peer teaching, as well as in authoring formal learning objects, for instance, to contribute to digital libraries like the ARIADNE Knowledge Pool System [Duval et al., 2001].

Beagle++ [Chirita et al., 2006a] is a Linux based Desktop search system which indexes all personal documents of a user and generates additional metadata describing these documents, other resources, as well as their relationships. Triggered by modification events such attention switches, Beagle++ automatically annotates the materials the user has read, used, written or commented upon. Context or task specific metadata can be afterwards extracted to implicitly delimit scenario specific information spaces [Ghita et al., 2005a, b; Brunkhorst et al., 2006]. Figure 1 shows an overview of its underlying ontologies, depending on activity contexts, for instance depicting files annotated with publication metadata, file system attributes, Web history, as well as emails. Each considered context is visually contoured and the inter-context relations between resources depict how various activities interact.

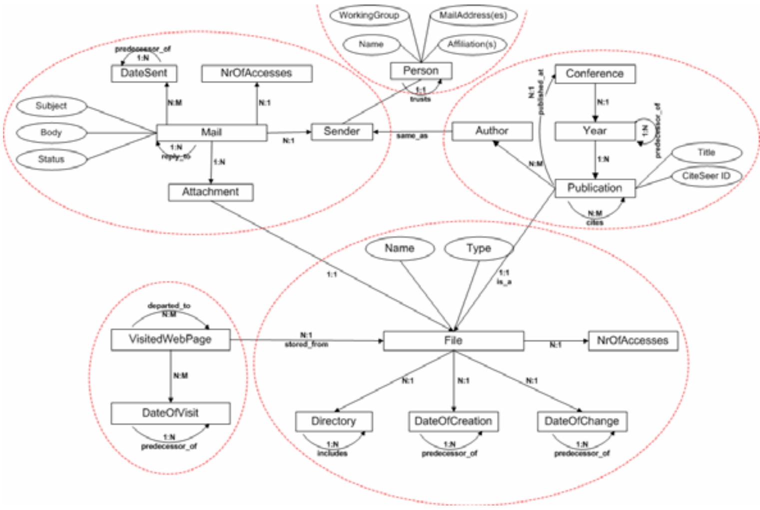

Figure 1: RDF schema of contextual metadata from Beagle++ [Chirita et. al., 2006a].

Semex [Dong & Haley, 2005b] is similar in that it uses an RDF domain model describing personal information items of various sources with classes and relationships among them, which are also exploited for retrieving associated objects. Its interface is designed to support browsing by association. However, Semex's most research efforts focus only on reference reconciliation, that is, extracting correctly all possible references to real world objects (like variant names), and less on establishing ontologies adequate for personal information and various user contexts. Xiao and Cruz [2006] continue the Semex work by presenting a toolkit for semantically managing personal resources and for easily creating applications interacting with an RDF database. The knowledge engineering problems of developing semantic Desktop applications and integrating them with various languages and platforms are further discussed within the CALO project, which develops innovative software to support decision-makers [Chaudhri et al., 2006]. A Case Study in Engineering a Knowledge Base for an Intelligent Personal Assistant shows challenges and sound, pragmatic solutions with respect to used tools and knowledge representation languages.

To facilitate the re-finding of known items stored on one's Desktop ***Stuff I've seen (SIS)*** [Dumais et al., 2003] addresses the problem of information fragmentation by building an unified textual index for information from a range of different sources (documents, email, Web pages, appointments, etc.). Based on the familiarity of the user with his previously seen information, rich contextual cues such as time, people, thumbnails and previews are used to support retrieval and presentation. The interface is designed to enable easy query specification and refinement by iteratively and interactively filtering and sorting the results for example by resource type or the predominantly used time dimension. A user study showed the general acceptance of the interface especially regarding search of files and emails and also user preference for retrieving and displaying results by date over rank (followed by author). Based on these results, as well as on research on episodic memory and time as associative cue, SIS was extended to comprise timeline visualization, where important personal landmarks (digital photographs and calendar events) and important public landmarks (holidays, news events) were displayed together with results to aid in locating the target of a keyword search. In the study, significant time savings were found, users prizing the timeline and especially work related landmarks. [Ringel et al., 2003]

Phlat [Cutrell et al., 2006a] is a SIS follow-up enhanced to allow tagging information with multiple meaningful, personal annotations. Tags can be used like categories as filters for both searching and browsing. Their conclusion that "filing will become less important and will be replaced by more general tagging systems" seems reasonable. Nevertheless, further evaluation must show that task management, especially reminding, is adequately supported without hierarchical structures and location [see Whittaker et al., 2006].

MyLifeBits [Gemell et al., 2002, 2006] aims at fulfilling Vannevar Bush's vision of capturing a user's whole life electronically [Bush, 1945], storing and integrating all sources of personal information like documents, emails, phone calls, images, videos etc. The platform was developed around two fundamental features of Bush's Memex: annotations and links. Manual annotations are especially important for meaningful, intuitive search of pictures by content, as well as for organizing information. Since due to the sheer volume of data some valued items would be hard to find or forgotten, (bidirectional) linking is also crucial for allowing associative browsing and serendipitous encounters. Finally, MyLifeBits supports the user in tagging data and scales such a flat tagging system without loosing overview because of missing hierarchical classification.

Haystack [Karger et al., 2005; Huynh et al., 2002; Haystack] also generates annotations and provides dynamic collection views, but it focuses on agents exploiting user specific and pre-defined ontologies to construct views. It employs semi-structured data modeled in RDF, combining exact Boolean database search with fuzzy text retrieval techniques in order to give the user more control over the form in which information is managed. Users can decide which information items and which relationships to other objects are to be stored, viewed and retrieved. Moreover, together with the categories that can be assigned to an object via a checkbox, attributes can also be utilized to refine the view of collections. Thus, combined with an additional browsing advisor and a text search box, Haystacks' interface provides support for search, as well as for associative browsing or orienteering.

One of the few projects concerned with linking personal resources explicitly to educational content is ***Sidewalk*** [Maier et al., 2006]. Manual marking of different types of resources (e.g., MS Office, PDF, html, audio, video, images, etc.) as superimposed information is used to annotate, link, classify, restructure, and reorganize information. For example concepts or topics in a lesson can be connected to manually marked, relevant regions in a document or a certain group of slides. Thus, elaboration, orientation and navigation are facilitated by creating a concept map of a topic with the corresponding personal learning resources integrated. Selection of different subdocuments important for the learning task at hand is provided in a Sidepad application. Here, URI-like markers to selected subdocuments are shown with labels and comments and may be grouped by the learner for instance to classify related work. Hence, organization as keeping and categorizing personal documents relevant to specific learning contexts is facilitated together with the access to the relevant parts of resources [Murthy et al., 2006].

Addressing common problems users have with the current Desktop metaphor and its strict folder hierarchies and dozens of separate applications, ***Lifestreams*** presents a new time-based metaphor to make managing one's electronic life more fluid, natural and thus easier [Freeman & Fertig, 1995; Freeman & Gelernter, 1996]. To support an informative "glance view", the representation of documents is based on activity streams and consists of descriptive attributes such as time of creation, content type, a summary line and the first lines of text documents. Newly created documents are opened and added to the stream, incoming mails slide in and open files are offset to the side, where editing is done within the corresponding applications. Though LifeStreams supports intuitive strategies for finding having time as retrieval cue, it replaces the super-ordinate aspect "location in the Desktop hierarchy" with "location in the timeline" [Bellotti et al., 2003].

Finally, as Desktop search and organization became such an important task, the **industry** has also manifested a high interest in this area over the recent years. Thus, several tools have been released for free (e.g., Google Desktop Search[1], MSN Desktop Search[2], etc.). Some providers have even integrated their PIM applications into the operating system, such as Apple[3]. Moreover, the open source community has also manifested its interest in this area, the most prominent approaches being Gnome Beagle[4] (now also integrated into SuSE) and KDE KAT[5], developed within the Mandriva community. Many other commercial applications exist (e.g., Copernic, Yahoo! Desktop Search, X1, Scopeware Vision, PC Data Finder, etc.), but as our main focus is on reviewing research innovations, we will not dive into the particularities of each of these tools.

The applications presented so far focused on overcoming information fragmentation and the disadvantages related to the Desktop metaphor by providing means to efficiently organize and search one's personal information space. In the next section we describe approaches addressing the task management support in particular, as well as the related issues of task switching, interruption and context-awareness.

3.2. Task Management Support

Tasks are central in working activities. The value of resources and communication is mainly determined by their relation to the current context. Thus, employing techniques for capturing tasks and user contexts is crucial to support easy access to all related documents and task-specific materials, or to prevent harmful interruptions while working. This section surveys several recent approaches tackling this important organizational goal. It first presents some general PIM context detection and notification techniques, and then brings those context oriented approaches specific to the e-Learning community into the centre of attention.

[1] http://desktop.google.com
[2] http://toolbar.msn.com
[3] http://www.apple.com/macosx/features/spotlight
[4] http://www.gnome.org/projects/beagle
[5] http://kat.mandriva.com

General Desktop Task Analysis. ***TaskMaster*** [Bellotti et al., 2003] addresses missing task support in existing email tools. It considers tasks as the major aspect around which to organize emails, attached documents or sent URLs, and models them through thrasks, communication threads enhanced by equally treated attachments or links. These thrasks are built by grouping related messages together by analyzing message data. They can be combined or broken into sub-thrasks in order to individually optimize the overview on related resources. Inbox and outbox are displayed together and attachments are not hidden inside an email, but directly visible as an equally important task resource. Task specific meta-information like deadline, reminders, and to-dos can also be added and visualized for example as a warning bar changing color, as an alarm bell, etc.

UMEA [Kaptelinin, 2003] supports projects as high-level goals by organizing resources and integrating applications around these goals. Based on activity theory UMEA provides context by tracking interaction histories. It offers an architecture with integrated components for access to various resources and a project specific history of user interaction. Within a defined project, related resources can be easily accessed, sub-tasks may be defined, or dates and deadlines set. All these user interactions are also monitored and tagged to the project, the resulting log entries being displayed to support activities by presenting context in terms of earlier activities and handled resources. Similarly, virtual workspaces for specific tasks or projects is constructed manually in ***ROOMS*** [Henderson & Card, 1986] or ***Task Gallery*** [Robertson et al., 2000], both task management environments employing a spatial metaphor (room or whiteboard) to organize project resources. Related systems for e-Learning like ***Desktop*** [Portugal et al., 2000] employ virtual spaces with rooms, halls and desks offering various communication and coordination tools, both for collaborative and individual working and learning.

In contrast, an approach for automatic task detection is taken within the ***TaskTracer*** project [Stumpf et al., 2005]. Since the costs for task switching may be high in terms of remembering logical chains, execution steps or previous decisions, a Desktop tool effectively monitoring currently performed tasks could yield more efficient interruption recovery, improved personal and workgroup information management, as well as within group workflow detection and analysis. Machine learning is applied to learn and predict tasks from traced interactions with the operating system, as wells as from speech-to-text converted phone calls. A Folder Predictor seems the most cost-saving component in this environment. Learning correlations between (predicted) tasks and folders, the application suggests a folder for saving or opening resources closely to their affiliated files, thus saving costs for interactions. Similar approaches have been proposed, but they vary with respect to the analyzed resources [Segal & Kephart, 1999, 2000] and utilized algorithms [Canny, 2004].

Modeling tasks in terms of attentional states, [Horvitz et al., 2003] explore ways to make computer systems context aware to better suit interaction and prevent harmful disruption. In order to infer user's current attention focus and workload, various contextual clues as gaze-tracking, tracking of gestures, device interaction, calendar entries, time and user preferences are used in the cross-device ***Notification Platform*** [Horvitz et al., 2002; Oliver et al., 2002; Horvitz & Apacible, 2003] to decide when to interrupt and notify about an incoming message. Priorities for incoming notifications are learned from both implicit and explicit user feedback. Decisions about user's attention status are made using Bayesian networks, influence diagrams and Hidden Markov Models, and the expected value of alternate actions like delay vs. immediate interruption are calculated. ***The Scope*** [van Dantzich et al., 2002] is a related notification summarization tool designed to balance between keeping track of new incoming messages from various sources and distracting attention from the primary task in order to enable the user to decide at a glance which notifications to look at in more detail and when.

User interactions have also been successfully exploited in various systems, especially for building proactive recommender systems and with Just-In-Time retrieval agents [see e.g., Budzik & Hammond, 1999; Rhodes & Maes, 2000]. Their main underlying idea is that observable behaviors like reading time, Web navigation or saving, bookmarking, printing, forwarding, copy-and-paste, etc. provide useful information about a user's interests [Oard & Kim, 2001]. This way, based upon the idea that two files accessed within a small window of time should be related to each other, Chirita et al. [2006b] employed such interactions and file accesses to identify working contexts for clustering search results. Similarly, a method for re-retrieving information obtained through past Web browsing by recording low-level actions (click links, opening windows, entering keywords, etc.) is presented by Morita et al. [2006]. They analyzed these actions to obtain "semantic periods" and used automatic logs to remind the user about such periods. Finally, implicit relevance feedback techniques can also be employed to better track user's actions within their learning environments in order to feed them back into social filtering systems like in [Najjar et al., 2005].

Desktop Task Analysis for e-Learning. In informal Peer-To-Peer teaching, elaborate presence management is central. ***BuddySpace*** [Eisenstadt & Dzbor, 2002] enhances the concept of presence in group communication tools by exploiting current locations, tasks, projects or topics of interest. Similarly, Schmidt [2005] and Schmidt and Braun [2006] also emphasize the importance of task and context-awareness in online-learning at the workplace. In selecting a teacher for informal teaching, context is important to prevent overload on teacher's side. Decision on immediate delivery of a request should follow negotiation between learner and teacher contexts regarding urgency, task, subject and relationship. Organizational status could be taken into account as well, because some learners may feel uncomfortable to contact experts particularly of higher status. Hence, the authors propose a third type of learning process, a context-steered one, which bridges the gap between the typical course-steered learning by offering adaptivity to the user (e.g., learning style, preferences, prior knowledge) and the self-steered learning supported by knowledge management tools. Useful information for context-modeling may thus be personal goals, recently acquired knowledge or competencies, preferences regarding interactivity level or semantic density (taken from IEEE Learning Object Metadata standard[6]), and so on.

In the ***LIP*** [Schmidt, 2005] system, such context-steered learning is constructed by monitoring learner's activities while interacting with different applications. Small learning units are compiled from the learning materials and recommended by deducing from the existing domain knowledge. During compilation, learning objects that meet the knowledge gap are identified and prerequisite learning objects are recursively added. Finally, the resulting learning programs are pruned by features related to user's context and ranked according to organizational preferences such as estimated learning time.

Also intending personalized, situated, adaptive retrieval on demand, **Raft** [Specht & Kravcik, 2006] integrates collaborative learning activities in more complex task contexts. In a biological field trip, physical context, user sensors and messages are used in inferring distinct tasks, for which predefined learning material is delivered and to which new manually annotated resources can be added and shared. Such context-capturing techniques may also be

[6] LOM: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

introduced in e-Learning scenarios to enhance mobile or ubiquitous learning as in [Bull et al., 2003] or [Ogata and Yano, 2004].

3.3. Metadata for e-Learning and Beyond

Both data organization and task analysis employ metadata as a foundation for implementing their approaches. In fact, most systems depicted in Sections 3.1 and 3.2 do utilize metadata within their infrastructures. This section will move our survey further: It will present those research activities specifically targeted at building qualitative metadata for other applications (including generic PIM, e-Learning, etc.), yet without concentrating on any of them.

Accounting for the importance of context-awareness, Ochoa and Duval [2006] discusses the Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects. Four types of recommendation metrics are detailed: Link Analysis Ranking, Similarity Recommendation, Personalized Ranking and Contextual Recommendation. Though designed for Learning Objects, the authors show that these metrics could also be applied to rank and recommend other types of reusable components as software libraries.

The capture and usage of attention metadata to efficiently support personal ways of Desktop usage is also dealt with by Belizki et al. [2006]. To achieve this, contextual information across heterogeneous media types, file formats, and applications should be annotated and linked. The paper presents a light weight system which monitors the file structure and automatically generates semantic metadata based on user activities.

Archner and Delcambre [2006] discuss about Capturing and Re-Using Human attention in Corporate Decision Making. Decision making in the corporate environment involves gathering, organizing, de-conflicting, and using fine-grained information from diverse sources and formats to evaluate scenarios and choose courses of action. Though the pertinent information is readily available, and is of human-manageable scale, it is typically not integrated and ready for use. Tools for organizing this information are scarce, in part because the desired structure is often not pre-defined or available in an existing ontology. The authors introduce a construct called a manifestation, along with the notion of collections of manifestations, to support entity-centric decisions based on the requirements and constraints of a corporate decision making environment.

A new approach for modeling semantically rich metadata is to use Wikipedia pages as ontological concepts to annotate documents [Fernandez-Garcia et al., 2006]. Similarly, keeping metadata already available on Websites has also been proposed by Reif et al. [2006]. There, a Semantic Clipboard is suggested as an extension to existing clipboards in order to allow end users to exploit existing RDF data published on Websites or pasted from a Desktop application. Finally, Schandl and King [2006] emphasized the need for efficient metadata management for unstructured content and pointed out how extensions to the well-known WebDAV protocol could be designed to allow for retrieval, storage and management of metadata in current operating systems.

3.4. Specific Results within the ProLearn Context

In this section we move towards more detailed presentations of three ProLearn specific Desktop investigations. We thus first discuss about adaptive course authoring with MOT and about generating malleable schemas for e-Learning. Then, in the second part of this section, we depict the results of a study about usage patterns for personal resources.

3.4.1. Adaptive Course Authoring with My Online Teacher (MOT)

One of the reasons for which adaptivity in e-Learning is more accepted than actually realized is the cumbersomeness of the authoring process: creating rich personalized courses for a great number of users means creating both a large number of relevant course alternatives, at different granularity, as well as the adaptive mechanism to switch between these alternatives. Hendrix et al. [2006] present an approach to add publications stored on an author's Desktop via automatically generated metadata to effectively enrich an e-Learning course differentiating between advanced and beginner students.

They use the metadata automatically generated by Beagle++ [Chirita et al., 2006a, see also Section 3.1.2.] to allow description and retrieval of the appropriate articles. Since these are not enough to generate a complete course, a manual annotation step allows addition of additional content, as well as of attributes like pedagogical weights and labels, which are necessary to build the final adaptive course product. Specifically, the authors add information about the hierarchical structure and order of the material in the context of a lesson, as well as additional pedagogical annotations describing for which students the material is best suited (e.g., beginner versus advanced).

Two models are employed: (1) a domain map and (2) a lesson model. A domain map is composed of a hierarchy of concepts with concept attributes, containing or linking to e-Learning content alternatives describing the same concept. Additionally, domain concepts can also connect to related concepts. A lesson is composed of a hierarchy of sub-lessons. The lesson represents a filtered, pedagogically labeled, weighted and ordered version of the concept attributes.

When enriching the domain and lesson models, one course should get the right information in the right place in the associated hierarchy. To achieve this, the program first queries the Sesame [Sesame] database, in which Beagle++ stores all metadata of articles on a user's Desktop, using as search terms title and keywords of each domain concept found in the current lesson. To ensure that every resource is only added once, the place with highest 'relevance' is sought. Relevance is computed by ranking based on the set of keywords belonging to the current domain concept and the set of keywords found in the current article. If a resource is ranked equally for two domain concepts, it is added to the topmost concept. The number of articles to be added to any concept is limited to 3, as adding too many articles probably confuses the learner rather than help her. Furthermore, the data and metadata imported (and transformed) from Sesame can be reordered and edited using MOT (see Hendrix et al. [2006]). Figure 2 shows an extract of editing in the lesson environment.

Figure 2: Adding manual metadata in a MOT lesson [Hendrix et al., 2006].

In the example, the paper called ‘Adaptive Educational Hypermedia: From generation to generation’ and all its relevant metadata have been added automatically during extraction to the ‘Adaptive Hypermedia’ domain concept, at the highest level in the hierarchy. The paper ‘Adaptive Authoring of Adaptive Educational Hypermedia’, a more specific one, has been added to the domain concept ‘Authoring of Adaptive Hypermedia’, in a lower position in the hierarchy. The author can now manually add pedagogical labels to this new material, labeling the material ‘adv’ for advanced learners. Finally, export of the enriched course from MOT to RDF is also possible in order to let it be exploited within external applications as well [Hendrix et al., 2006].

To conclude, Hendrix et al. [2006] described how the Semantic Desktop can be interfaced with a sophisticated adaptive hypermedia authoring environment to provide rich, automatically generated metadata about resources, as well as relationships between these resources. Thus, a course can be enriched with both automatically generated metadata and other materials relevant for an advanced course added manually using the MOT authoring system.

3.4.2. Malleable Schemas for Generating Contextual Metadata

For the case in which such semi-automatic course enrichment is realized while metadata schemas change or evolve (i.e., new tasks make new metadata attributes necessary, additional data sources with additional schemas are added over time, etc.), conventional schemas are not suited. One solution, proposed by Cristea and Nejdl [2006] relies on malleable schemas (introduced by Dong & Halevy [2005a]) to flexibly describe metadata, and then on fuzzy querying and matching based on these schemas to flexibly retrieve metadata. Malleable Schemas offer a mechanism by which the modeler can capture the imprecise aspects of the domain during the modeling phase in a well principled fashion. Thus, queries can be imprecise in their references to schema elements and closely related schema elements will be considered as well in query processing. Such imprecise-ness is useful for example in the case of different tools for file storing, evolving tools, if only part of a schema is integrated or in general for handling less structured data. Hence, malleable schemas help in capturing the important aspects of the domain at modeling time, without having to commit to a strict schema. They provide a simplified, unified view on a collection of related and overlapping base schemas using regular objects and properties. An object can represent a file, a publication, or any other data concept on the Desktop. A property can be either an attribute or a relationship (to another object). This model can be easily converted into RDF.

An extended example schema, on which queries on several partially known (or evolving) schemas can be based, e.g., to extract information for enriching a course with articles, allows for three types of impreciseness: flexible objects, flexible relations or property names and flexible paths. Thus, if for instance one is looking for an article written by the author "Cristea", which was sent to the current user by "Nejdl" via email, but one does not know the exact structure of the RDFS base schema, articles can be returned even if they do not match the exact terminology or the relations stated in the query. First, the precise naming of the object class 'Article' (or 'Publication', 'File', 'Attachment') could be unknown. Second, the name of the authors searched for could correspond to a property called 'Name', but it could also be 'firstName' and 'lastName'. In addition, the connection between the object, e.g., sender, and an (imprecise) attribute can be indirect and imprecise (of undetermined length). Consequently, a query becomes compact and almost like natural language. It resembles the way people ask questions: they only remember partially terminology and relations, and some information about the structure. Moreover, individual representations remembered by people differ. Still, individuals are capable of communicating with each other despite different models (schemas).

The above three types of impreciseness can be resolved in various ways: term similarity in objects, relations or properties can be resolved via Wordnet (synonyms, hyponyms, hypernyms, and antonyms) or by other similarity measures. In [Zhou & Nejdl, 2006], for instance, an information retrieval approach based on vector space representations and Wordnet is used. In Cristea and Nejdl [2006] this model is extended based on the malleability definition given above: First, the malleable schema terms and relations are expanded by binding imprecise classes, attributes, relations and structures to their respective correspondents in the base schema set. For this step, a set of rules for term expansion are used. Then, the expanded queries must be verified for the validity of relations, and ranked according to verification results, to similarity to malleable schema or according to tuple content. Finally, the output is pruned by cutting off items which are invalid or have the rank below a certain threshold.

This method is useful for static courses, and is even more important and efficient in the context of adaptive, personalized courses, where the richness of alternatives is mandatory in order to care for learners with different needs and different levels of knowledge. The method can further be extended to the automatic enrichment of courses used for other types of adaptation. For example images can be automatically retrieved from the Desktop to enrich a course with information for learners with visual preference. The method also addresses one other distinctive feature of the semantic Desktop (i.e., to evolve in time), supporting this evolution through malleable schemas. Further work will focus on evaluation as well as on the extension to support adaptive navigation rules for runtime adaptation. In general, automatic metadata generation and mapping of different schemas are important research topics with respect to enhancing both the exchange of learning material in Peer-To-Peer networks or Open Learning Repositories [Allert et al., 2002; Nejdl et al., 2002] and in Digital Libraries on educational resources [Duval et al., 2001, Ochoa et al., 2005; Cardinaelis et al., 2005].

3.4.3. Patterns in Using Personal Documents

Motivation. The Desktop becomes one of the richest sources of personal information and learning materials. Occasionally, users can get an overview of a new subject entirely from the materials stored on their hard disk. While the Desktop is an important learning source, incomplete attempts were made to understand how effectively it is used and what can be done to enhance it. Nevertheless, several works were devoted to particular aspects of the Desktop organization (see also Section 3.1). For example Gwizdka [2002] explicitly evaluated the efficiency and effectiveness of the information finding tasks in the MS Outlook email client and compared it with their own prototype. The reported results prove that email finding time can be improved significantly if several specific user requirements are taken into account. Here, this idea is extended onto other information needs on a Desktop, considering not only the emails but also other document access types, such as blogs, forums, Web pages, instant messages (IM), databases, separate files and folders, address books and calendars.

For the analysis of the Desktop activity scenarios, first a number of prior publications have been selected in order to extract the possibly useful metadata fields for the different Desktop usage scenarios. The results of this literature review are presented in the Table 1, where the Source column corresponds to the publication with a particular scenario, Metadata / Data column specifies a set of possibly useful documents and metadata fields, and Privacy reflects our estimation of data privacy level. We included a privacy level estimation for each metadata type, since the privacy issue is crucial for PIM.

Source	Metadata / Data	Privacy
[Xiao & Cruz, 2005]	author, book	L
[Osterfeld et al., 2005] [Franz & Staab, 2005]	chat message	H
[Osterfeld et al., 2005] [Franz & Staab, 2005]	instant message: date/time, sender, receivers, previous/next message	L
[Gemmell et al., 2004]	music metadata	L
[Cheyer et al., 2005] [Gemmell et al., 2004]	event, task, person, photo	M
[Gemmell et al., 2004]	document, phone call, chat transcript, email message, saved search results	L, M, H
[Dumais et al., 2003]	document title, date, rank, author, mailto, file type, mail-cc, email has attachment, message type, message read, path, size, title.	L
[Cheyer et al., 2005]	“normalized” contacts, project.	L
[Cheyer et al., 2005]	person’s expertise	H
[Ghita et al., 2005b]	title, citations, conference, year, path, number of accesses, co-citations, movie genre, director, actor, rating.	L

Table 1: The data and metadata fields used in the Desktop activity scenarios from recent publications. The privacy levels are: H – high (cannot be shared), M – medium (can be shared with some users), L – low (can be shared). The values can be combined, i.e., there might be different privacy levels for different parts of information within the same document access type.

The selected scenarios share a common limitation, as most of them are strongly related to the research activities. We believe that the scientific community alone does not represent the whole range of professions. Therefore, we need an overview of a broader example of a real-world Desktop data usage.

Selected Tasks. In order to collect the non-scientific examples of Desktop use, we conducted a set of interviews, about 3 hours each. We intentionally selected respondents which do not belong to the academia. Similar to the study described in [Ravasio et al., 2004a, b], our subjects cover 12 professions (one person per category). They had both technical and non-technical backgrounds, with experience level ranging from extensive daily use of computer for both work and entertainment to irregular use of computer only in leisure time. The survey was intended to learn the following features of desktop data:

1. A description of an average working day and usual Desktop tasks.
2. The data management habits, the main file storages / folders and file placement patterns, file and application usage statistics, most used file formats, and user estimation of the data privacy levels.
3. Suggestions for metadata enhancements and new search capabilities.

The main questions were about what people store on their Desktops, how they organize information, what and how they search, the interviews being transcribed on the fly and analyzed afterwards.

Data Usage. The set of document access types and corresponding data usage levels are presented in Table 2. Users answered how often they read or send emails and instant messages, access files and folders, browse the Web, and so on. It is interesting that some of them consider only a couple of applications for everyday use. For example a designer uses the image editor and viewer 90% of her working time. We also noticed that people have various habits in cleaning their Desktops. While some keep only documents from the past few months, others have archives from several previous years. Therefore, the difficulty of finding and re-using documents depends on the history and richness of the personal data. The make-up of the Desktop materials is also obviously influenced by the amount of time spent in front of the computer, as well as by the level of a user expertise. From the data usage patterns we drew several conclusions:

- Several access types like browsing and email exchange are used by all respondents.
- Interestingly, the Web access is not used by every user, especially since some companies restrict Web usage during working days.
- Blogs and forums are used only by two users, so they are not very common yet.
- The usage patterns are very diverse and user-specific.

The last conclusion suggests that the data usage can hardly be averaged among many users. This is an important observation, which limits any Desktop analysis to a very narrow user target group. It indicates the clear need both for personalizing the Desktop tools and for a specific user model for each data management solution.

Profession	Emails	IM	Files/ Folders	Web	Blog/ Forum	Calendar/ Address book
Newspaper chief editor	A	A	A	A	A	A (paper-based)
Consultant	A	N	U	A	N	U
Technical lead manager	U	A	A	A	N	N
Designer	U	N	A	U	N	U (paper-based)
Sales analyst	A	A	A	N	U	U
Bookkeeper	U	N	A	N	N	U (paper-based)
Engineering manager	U	U	A	U	N	U

Software developer	A	A	A	A	N	A
Lawyer	U	U	U	A	N	N
Architect	U	U	A	U	N	U
Dental mechanic	U	N	A	U	N	U

Table 2: Data types and applications usage levels. The values are: A – actively used, U – used, N – not used. The difference between "actively used" and "used" is loosely defined, based on a common sense. For example, 5-7 emails per week are considered as "used", while 5-7 emails per day are reported as "actively used". Note that some users still use paper-based calendars and address books.

Privacy Perception. We asked users if they can share their data, share with restriction to a particular group of people, or cannot share it at all. The answers were translated into privacy levels, which we defined as "low", "medium" and "high" respectively. Often users specified different privacy levels within the same document access type. For example they are ready to share some part of their emails, but not all of them. The distribution of privacy level estimates among the user and document access types are shown in Table 3.

All users strictly dislike revealing their IM conversations, blogs and address books, while the majority of them would freely share local copies of Web pages. The email messages have various privacy levels, though they often include very private data. As expected, files and folders have a full range of privacy estimates. The privacy levels distribution also supports the hypothesis that a specific user model is crucial for any Desktop analysis.

Profession	Emails	IM	Files/ folders	Web	Blog/ Forum	Calendar/ Address book
Newspaper chief editor	M,H	H	L, M, H	L, M	H	L, M, H
Consultant	M, H	-	H	L	-	-
Technical lead manager	H	H	L, M, H	L	-	-
Designer	L, M	-	L, H	L	-	H
Sales analyst	M, H	H	L, M, H	L	H	-
Bookkeeper	L, H	-	L, H	-	-	-
Engineering manager	H	H	L, M, H	-	-	-
Software developer	L, H	H	L, M, H	L, H	-	H
Lawyer	H	H	L, M, H	L	-	-
Architect	M, H	H	L, H	L, M	-	-
Dental mechanic	H	-	L, M, H	L	-	-

Table 3: Information privacy levels. The privacy levels are: H – high (cannot be shared), M – medium (can be shared with some users), L – low (can be shared). The values can be combined in the sense that different privacy levels are possible for different parts of information in the same document type.

Metadata Elements. We collected the metadata fields, which are either used now or can be of use in future, either for personal use or for information exchanges. They seem to be mainly borrowed from the original data structure. For example each email has metadata fields like Sender, Recipient, Title, etc. Some fields correspond to a specific DB structure, like Version, Company, etc. In the overview of the metadata fields presented in Table 4, we do not distinguish between existent and suggested fields. The metadata fields include both content and activity based metadata. The general purpose documents like photos and music are common for the most users. Other documents are very domain specific, being characterized by highly specialized content and metadata elements.

Profession	Suggested metadata	Information type
Newspaper chief editor	Author, company, region, event description from a column (rubric) classification, author, journal name and issue number, genre.	Articles, Web-pages and photos
Technical lead manager	Phone model, vendor, and date of the last modification.	Manuals, specifications, internal documents
Designer	Vendor, picture title, topic (after a hierarchical classification), concurrently opened files, price, materials used	Stitch pictures, drawings from completed projects
Sales analyst	Person (associated with a picture), project, file size, URL, partner, sender, regular activity association (planning, implementation, etc.), concurrently opened files.	Photos, music, documents
Engineering manager	Creator, contributor, version, product, unit, related documentation, development state, manufacturer, product, related drawings.	Drawings and sketches; specifications, project documentation
Software developer	Creator, contributor, version, language, purpose, development state, documentation, manufacturer, version, language.	Source code files, documentation
Lawyer	Source of publication (it can be useful to divide publications into “claim is supported / not supported” for claims like “X action is allowed under Y conditions”).	Laws and comments, cases
Architect	Topic, subtopic (urbanism, landscape, etc.), material type, picture creation date, author, magazine title.	Pictures – design examples
Dental mechanic	Picture creation date, title, photo creation location, daytime.	Photos

Table 4: Existing and suggested metadata fields of interest.

Conclusion. Generic Desktops have a mixture of the general-purpose and domain-specific materials. First, the quantity of personal objects is highly correlated with the level of user expertise. The study displayed a large diversity in possible usage activity levels. Second, users vary in their privacy preferences, but many types of documents are considered as highly private, which makes it difficult to spread and re-use materials. Finally, while there are several generally demanded metadata fields, the majority of users would still prefer domain specific ones.

It thus seems unrealistic to consider general Desktop data management solutions. Rather a separate personalization is required for each particular user model. Future research will thus focus on the user modeling part. Also, the possible next steps include a new set of interviews among a targeted set of the users with the same occupation, possibly within the same company. This will give an insight into which user model parameters can or cannot be exploited by professional domains only.

3.5. Discussion

We have seen quite a large number of research investigations addressing various aspects of supporting Desktop specific activities. Nevertheless, Personal Information Management is still in its childhood, and is only now gaining momentum. For example some usage pattern studies exist, yet they clearly still miss several important perspectives, such as analyzing the resource access characteristics. Similarly, several document organizational systems have been developed, some of them for our specific area of interest (i.e., e-Learning), yet fast finding of personal information still remains an open issue. And so on.

The same is valid when analyzing information exchanges between users: There exist a relatively large amount of approaches to generate metadata (see Table 1, as well as Sections 3.1.2, 3.2 and 3.3), yet minimal progress has been made in the direction of exploiting these metadata across Peer-To-Peer social networks or communities of users. Clearly once a user has already annotated a shared resource (document, web page, etc.), others could benefit from this information as well. Several scenarios can be envisioned for such resource exchanges between different users:

- If some user has a new resource on her Desktop, an application could first search the Peer-To-Peer network for other peers storing the same item, and then automatically annotate it with the respective distant metadata, if any. Similarly, if some metadata exists already, the information exchange tool could be exploited to enrich the current annotations with data from other peers.
- If a user is searching for other resources similar to a personal document, it might be more efficient (at least in terms of quality of results) to search its neighbors within a social network, using either a summary of the resource contents, or better, only the metadata associated to them. The latter approach is both faster, as it employs only small pieces of text for communication across the network, and more reliable, as two resources sharing very similar annotations are very likely to be related as well.

We refer the reader to [Ghita et al., 2005a] for a discussion on how such resources could be best integrated into the local Personal Information Repository of a user, once they have been collected from other persons.

All in all, we conclude that it is now the right time to focus our research towards the PC Desktop: (1) Several basic results exist, and thus we have where to start from; (2) The infrastructure capabilities are growing, in terms of both storage size and of computing power; (3) Last, but not least, there is more and more data available, which requires enhanced methods for information search and access within the local repository, as well as across social networks.

4. Conclusions

E-Learning has changed the way we learn considerably. One of the important changes is that we rely more on our skills to find some information than actually knowing it ourselves – and to relocate information we have seen before rather than memorizing it. From various studies it has become clear that the contents on the PC Desktop is becoming more and more integrated in the working environment, and consequently, part of the e-Learning experience.

There exists a plethora of research on the storage and management of documents and other information containers, such as emails and Web sites. An interesting observation that applied to most studies is the division between people who spent much time on organizing their information, and those who trust on their ability to find a needle in a haystack. Further, the fragmentation of information access – the lack of an integrated tool for searching various types of documents – is considered an issue and is therefore addressed by various research projects and industrial efforts. This research goes hand in hand with research on task management support, which aims at modeling user tasks and the interaction with user interfaces.

The use of attention metadata for providing personalized access to Desktop based document repositories is gaining attention from both academics and industry. While advanced personalization, based on behavioral pattern mining, is a hot topic, we are still far from industrial uptake and commercial exploitation. One of the preventing issues is the fact that users vary wildly in the usage of their Desktop contents, a variety that is caused by differences in user tasks and personal factors. Another preventing issue is the concern about control for attention metadata. In particular for the use of Desktop data for e-Learning, the boundaries between personal and group data need to be blended.

From a different perspective, research within ProLearn has shown that e-Learning courses can be enriched with publications from authors' Desktops. However, as is often the case, the metadata schemas for publications are subject to change. This may be solved by a technique called malleable schemas, which allows for imprecise queries and flexible query processing.

Summarizing, we conclude that Desktop documents have the potential of enriching e-Learning significantly. First, the need for actively memorizing information is reduced. Second, a large collection of documents with sophisticated tools for searching and browsing provides the flexibility and broadness needed in this information era. Third, the automatic detection of relevant documents provides e-Learning authors with the opportunity of enriching their material without too much extra effort.

Having said this, there are several issues that need to be addressed before Desktop data becomes an integrated part of e-Learning:

- There is the need for tools that provide integrated access to various types of information sources: documents, email history, Web pages visited;
- For the reuse of documents that cover several topics, it may be needed to also model the internal structure of these documents;
- More insight is needed in how users actually interact with the contents of their personal Desktops; rather than aiming for the description of ‘the average user’, it seems to be more worthwhile to describe the whole range of identified usage patterns, one at a time.

References

[Allert et al., 2002] Allert, H., Dhraief, H., Kunze, T., Nejdl, W., Richter, C. (2002): Instructional Models and Scenarios for an Open Learning Repository - Instructional Design and Metadata. In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education, E-Learn 2002. Oct. 15-19, 2002, Montreal, Canada.

[Archner & Delcambre, 2006] Archner, D. W., Delcambre, L.M.L. (2006): Capturing and Re-Using Human attention in Corporate Decision Making. In: Proceedings of the 1st International ACM Workshop on Contextualized Attention Metadata: collecting, managing and exploiting of rich usage information, CAMA 2006. Nov. 10 - 11, 2006, Arlington, Virginia, USA, pp. 17-20.

[Bakshi & Karger, 2005] Bakshi, K., Karger, D. R. (2005): End-User Application Development for the Semantic Web. In: Proceedings of the 1st Workshop on the Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, located at the International Semantic Web Conference. Nov. 6, 2005, Galway, Ireland.

[Bälter, 1997] Bälter, O. (1997): Strategies for organizing email messages. In: Proceedings of HCI'97, Springer: London, UK, pp. 21–38.

[Barreau, 2006] Barreau, D. (2006): Personal Information Management in Context: In: Proceedings of the SIGIR 2006 Workshop on Personal Information Management, PIM 2006. Aug 10-11, 2006, Seattle, Washington, USA, pp. 6-7.

[Barreau & Nardi, 1995] Barreau, D., Nardi, B. A. (1995): Finding and Reminding. File Organization from the Desktop. In: SIGCHI Bulletin. Vol. 27, No. 3, July 1995, pp. 39-43.

[Belizki et al., 2006] Belizki, J., Costache, S., Nejdl, W. (2006): Application Independent Metadata Generation. In: Proceedings of the 1st International ACM Workshop on Contextualized Attention Metadata: collecting, managing and exploiting of rich usage information, CAMA 2006. Nov. 10 - 11, 2006, Arlington, Virginia, USA, pp. 33-36.

[Bellotti et al., 2003] Bellotti, V., Ducheneaut, N., Howard, M., Smith, M. I. (2003): Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. In: Proceedings of the ACM Conference on Computer Human Interaction, CHI 2003. April 5-10, 2003, Fort Lauderdale, Florida, USA, pp. 345-352.

[Boardmann & Sasse, 2004] Boardmann, R., Sasse, M. A. (2004): Stuff Goes into the Computer and Doesn't Come Out. A Cross-tool Study of Personal Information Management. In: Proceedings CHI 2004. April 24-29, 2004, Vienna, Austria, pp. 583-590.

[Bruce et al., 2004] Bruce, H., Jones, W., Dumais, S. (2004): Information behavior that keeps found things found. In: Information Research. Vol. 10, No. 1, Oct. 2004.

[Brunkhorst et al., 2006] Brunkhorst, I., Chirita, P.-A., Costache, S., Gaugaz, J., Ioannou, E., Iofciu, T., Minack, E., Nejdl, W., Paiu, R. (2006): The Beagle++ Toolbox: Towards an Extendable Desktop Search Architecture. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference ISWC 2005, Nov. 6, 2006, Athens, Georgia, USA.

[Brusilovsky, 2004] Brusilovsky, P. (2004): Adaptive Educational Hypermedia: From generation to generation. In: Proceedings of 4th Hellenic Conference on Information and Communication Technologies in Education. Sep 29 -Oct 3, 2004, Athens, Greece, pp.19-33.

[Brusilovsky, 2001] Brusilovsky, P. (2001): Adaptive hypermedia. In: User Modeling and User Adapted Interaction, Ten Year Anniversary Issue. Vol. 11, No. 1/2, pp. 87-110.

[Budzik & Hammond, 1999] Budznik, J., Hammond, K. (1999): Anticipating and Contextualizing Information Needs. In: Proceedings of the 62nd Annual Meeting of the American Society for Information Science.

[Bull et al., 2003] Bull, S., McEvoy, A.T., Reid, E. (2003): Learner Models to Promote Reflection in Combined Desktop PC/Mobile Intelligent Learning Environments. In: Proceedings of the

Workshop on Learner Modeling for Reflection, located at the International Conference on Artificial Intelligence in Education 2003, University of Sydney, Australia, pp. 199-208.

[Bush, 1945] Bush, V. (1945): As we may think. In: The Atlantic Monthly. Vol. 176, no. 1, July 1945, pp. 101-108. online: http://www.theatlantic.com/doc/194507/bush

[Canny, 2004] Canny, J. F. (2004): Gap: A Factor Model for Discrete Data. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2004. July 25-29, 2004, Sheffield, UK, pp. 122-129.

[Cardinaelis et al., 2005] Cardinaelis, K., Meire, M, Duval, E. (2005): Automating metadata generation: the simple indexing interface. In: Proceedings of the 14th international conference on World Wide Web, WWW 2005, May 10-14, Chiba, Japan, pp. 548-556.

[Chaudhri et al., 2006] Chaudhri, V., Cheyer, A., Guilii, R., Jarrold, B., Myers, K., Niekarsz, J. (2006): A Case Study in Engineering a Knowledge Base for an Intelligent Personal Assistant. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference, ISWC 2005, Nov. 6, 2006, Athens, Georgia, USA.

[Cheyer et al., 2005] Cheyer, A., Park, J., Giuli., R. (2005): IRIS: Integrate. Relate. Infer. Share. In: Proceedings of the 1st Workshop on the Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, located at the International Semantic Web Conference, Nov. 6, 2005, Galway, Ireland.

[Chirita et al., 2006a] Chirita, P.-A., Gavriloaie, R., Costache, S., Nejdl, W. and Paiu, R. (2006): Beagle++: Semantically Enhanced Searching and Ranking on the Desktop. In: Proceedings of the 3rd European Semantic Web Conference: The Semantic Web: Research and Applications, ESWC 2006, June 11-14, 2006, Budva, Montenegro, pp. 348-362.

[Chirita et al., 2006b] Chirita, P.-A., Costache, S., Gaugaz, J., Nejdl, W. (2006): Desktop Context Detection Using Implicit Feedback. In: Proceedings of the SIGIR 2006 Workshop on Personal Information Management, PIM 2006. Aug 10-11, 2006, Seattle, Washington, pp. 24-27.

[Chirita et al., 2005a] Chirita, P.-A., Gavriloaie, R., Ghita, S., Nejdl, W., Paiu, R. (2005). Activity-Based Metadata for Semantic Desktop Search. In: Proceedings of the 2nd European Semantic Web Conference: The Semantic Web: Research and Applications, ESWC 2005, May 29 – June 1, 2005, Heraklion, Crete. LNCS 3532, Springer, pp. 439-454.

[Chirita et al., 2005b] Chirita, P.-A., Costache, S., Nejdl, W., Paiu, R. (2005): Semantically Enhanced Searching and Ranking on the Desktop. In: Proceedings of the 1st Workshop on the Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, located at the International Semantic Web Conference, Nov. 6, 2005, Galway, Ireland.

[Cole, 1982] Cole, I. (1982): Human aspects of office filing: Implications for the electronic office. In: Proceedings of the Human Factors Society 26th Annual Meeting.

[Cristea & Aroyo, 2002] Cristea, A. I., Aroyo, L. (2002): Adaptive Authoring of Adaptive Educational Hypermedia. In: Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH 2002. May 29-31, 2002, Malaga, Spain, LNCS 2347, Springer, pp. 122-132.

[Cristea & Calvi, 2003] Cristea, A. I., Calvi, L. (2003): The Three Layers of Adaptation Granularity. In: Proceedings of the 9th International Conference on User Modeling 2003, UM 2003, June 22-26, 2003, Johnstown, Pennsylvania, USA, LNCS 2702, Springer, pp. 4-14.

[Cristea & de Mooij, 2003a] Cristea, A. I., De Mooij, A. (2003): Adaptive Course Authoring: My Online Teacher. In: Proceedings of ICT'03, Papeete, French Polynesia.

[Cristea & de Mooij, 2003b] Cristea, A., De Mooij, A. (2003): LAOS: Layered WWW AHS Authoring Model and its corresponding Algebraic Operators. In: Proceedings of WWW’03, Alternate Education track. May 20-24, 2003, Budapest, Hungary, ACM.

[Cristea & Nejdl, 2006] Cristea, A., Nejdl, W. (2006): Adaptive Authoring of Adaptive Courses on the Semantic Desktop using Malleable Schemas. In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higer Education, E-Learn 2006. Oct. 13-17, 2006,Waikiki Beach, Honolulu, Hawaii.

[Cristea et al., 2005] Cristea, A. I., Smits, D., De Bra, P. (2005): Writing MOT, Reading AHA! - converting between an authoring and a delivery system for adaptive educational hypermedia. In: A3EH Workshop, AIED'05, July, 2005, Amsterdam, The Netherlands.

[Cristea & Stewart, 2005] Cristea, A. I., Stewart, C. D.: Automatic Authoring of Adaptive Educational Hypermedia. (Book chapter) In: Ma, Z. (2005): Web-Based Intelligent e-Learning Systems: Technologies and Applications. IDEA publishing group: Northeastern University, China.

[Cross, 2003] Cross, J. (2003): Informal Learning – the other 80%. http://www.internettime.com/Learning/The%20Other%2080%25.htm, verified Nov. 30, 2006.

[Cutrell et al., 2006a] Cutrell, E., Robbins, D. C., Dumais, S. T., Sarin, R. (2006): Fast, flexible filtering with Phlat - Personal search and organization made easy. In: Proceedings of Human Factors in Computing Systems, CHI 2006. April 22-27, 2006, Montréal, Kanada, ACM Press, pp. 261-270.

[Cutrell et al., 2006b] Cutrell, E., Dumais, S. T., Teevan, J. (2006): Searching to Eliminate Personal Information Management. In: Communications of the ACM. Vol. 49, No. 1, Jan. 2006, pp. 58-64.

[Cutrell et al., 2001] Cutrell, E., Czerwinski, M., Horvitz, E. (2001): Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. In: Proceedings of Interact 2001. July 9-13, Tokyo, Japan.

[van Dantzich et al., 2002] van Dantzich, M., Robbins, D., Horvitz, E., Czerwinski, M. (2002): Scope: Providing Awareness of Multiple Notifications at a Glance. In: Proceedings of Advanced Visual Interfaces, AVI 2002. May 22-24, Trento, Italy, ACM.

[Dong & Halevy, 2005a] Dong, X. L., Halevy, A. (2005): Malleable Schemas: A Preliminary Report. In: Proceedings of the 8th International Workshop on the Web and Databases, WebDB 2005. June 16-17, 2005, Baltimore, Maryland, ACM.

[Dong & Halevy, 2005b] Dong, X. L., Halevy, A. (2005): A Platform for Personal Information Management and Integration. In: Proceedings of the 2nd Biennial Conference on Innovative Data Systems Research, CIDR 2005. Jan. 4-7, 2005, Asilomar, California, USA, pp. 119-130.

[Dourish et al., 1999] Dourish, P., Edwards, W. K., LaMarca, A., Salisbury, M. (1999): Presto: An Experimental Architecture for Fluid Interactive Document Spaces. In: ACM Transactions on Computer-Human Interaction. Vol. 6, No. 2, 1999, pp. 133-161.

[Duchenaut & Bellotti, 2001] Duchenaut, N., Bellotti, V. (2001): E-mail as Habitat. An Exploration of Embedded Personal Information Management. In: Interactions. Vol. 8, No. 5, Sept./Oct. 2001, pp. 30-38.

[Dumais et al., 2003] Dumais, S., Cutrell, E., Cadiz, J J , Jancke, G. , Sarin, R., Robbins D.C. (2003): Stuff I've Seen : A System for Personal Information Retrieval and Re-Use. In: Proceedings of the International Conference on Research and Development in Information Retrieval, SIGIR'03. July 28-Aug. 1, 2003, Toronto, Canada, pp. 72-79.

[Duval et al., 2001] Duval, E., Forte, E., Cardinaels, K. , Verhoeven, B., Van Durm, R., Hendriy, K., Wentland Forte, M., Ebel, N., Macowicz, M., Warkentyne, K., Haenni, F. (2001): The Aridane Knowledge Pool System. In: Communications of the ACM, Vol. 44, No. 5, May 2001, pp. 73-78.

[Eisenstadt & Dzbor, 2002] Eisenstadt, M., Dzbor, M. (2002): BuddySpace: Enhanced Presence Management for Collaborative Learning, Working, Gaming and Beyond. In: Proceedings of the JabberConf Europe. May 12-14, Munich, Germany.

[European Commission, 2001] European Commission (2001): Lifelong learning practice and indicators. Commission Staff Working Document.

[Fernandez-Garcia et al., 2006] Fernandez-Garcia, N., Sauermann, L., Sanchez, L., Bernardi, A. (2006): PIMO Population and Semantic Annotation for the Gnowsis Semantic Desktop. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference ISWC 2005. Nov. 6, 2006, Athens, Georgia, USA.

[Franz & Staab, 2005] Franz, T., Staab, S. (2005): SAM: Semantics Aware Instant Messaging for the Networked Semantic Desktop. In: Proceedings of the 1st Workshop on the Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, located at the International Semantic Web Conference, Nov. 6, 2005, Galway, Ireland.

[Freeman & Fertig, 1995]: Freeman, E. T., Fertig, S. J. (1995): Lifestreams: Organizing your electronic life. In: AAAI Fall Symposium, AI Applications in Knowledge Navigation and Retrieval, Nov. 1995. Cambridge, Massachusetts, USA.

[Freeman & Gelernter, 1996] Freeman, E., Gelernter, D. (1996): Lifestreams: A Storage Model for Personal Data. In: SIGMOD Record. Vol. 25, No. 1, pp. 80-86.

[Gemell et al., 2006] Gemell, J., Bell, G., Lueder, R. (2006): MyLifeBits: A Personal Database for Everything. In: Communications of the ACM. Vol. 49, No. 1, Jan. 2006. pp. 89-95.

[Gemell et al., 2002] Gemell, J., Bell, G., Lueder, R., Drucker, S., Wong, C. (2002): MyLifeBits: Fulfilling the Memex Vision. In: Proceedings of the ACM Conference on Multimedia. Dec. 1-6, 2002, Juan Les Pins, France,

[Gemell et al., 2004] Gemmell, J., Williams, L., Wood, K., Lueder, R., Bell, G. (2004): Passive Capture and Ensuing Issues for a Personal Lifetime Store. In: Proceedings of the 1st ACM workshop on Continuous archival and retrieval of personal experiences, CARPE'04. Oct. 14, 2004, New York, USA, pp. 48-55.

[Ghita et al., 2005a] Ghita, S., Nejdl, W., Paiu, R. (2005): Semantically Rich Recommendations in Social Networks for Sharing, Exchanging and Ranking Semantic Context. In: Proceedings of the International Semantic Web Conference, ISWC 2005. Nov 6, 2005, Galway, Ireland, pp. 6-10.

[Ghita et al., 2005b] Ghita, S., Henze, N., Nejdl, W. (2005): Task Specific Semantic Views: Extracting and Integrating Contextual Metadata from the Web. In: Proceedings of the International Semantic Web Conference Workshop on The Semantic Desktop - Next Generation Personal Information Management and Collaboration Infrastructure, ISWC 2005. Nov. 6, 2005, Galway, Ireland.

[Gillie & Broadbent, 1989] Gillie, T., Broadbent, D. (1989): What makes interruptions disruptive? A study of length, similarity and complexity. In: Psychological Research, Vol. 50, pp. 243-250.

[Grundin, 2001] Grundin, J. (2001): Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In: Proceedings of the 2001 Conference on Human Factors in Computing Systems, CHI 2001. March 31-April 5, 2001, Seattle, Washington, USA, pp. 458-465.

[Gwizda, 2002] Gwizdka, J. (2002): TaskView – Design and Evaluation of a Task-based Email Interface. In: Proceedings of CASCON 2002.

[Haystack] Haystack Project, http://haystack.csail.mit.edu

[Henderson, 2005] Henderson, S. (2005): Genre, Task, Topic and Time: Facets of Personal Digital Document Management. In: Proceedings of CHINZ '05, July 6-8, 2005, Auckland, New Zealand, pp. 75-82.

[Henderson & Card, 1986] Henderson, A., Card, S.K. (1986): Rooms: The use of virtual workspaces to reduce space contention in a window-based graphical user interface. In: ACM Transactions on Graphics. Vol. 5, No. 3, pp. 211-243.

[Hendrix et al., 2006] Hendrix, M., Cristea, A. I., Nejdl, W. (2006): Authoring Adaptive Learning Material on the Semantic Desktop. In: Proceedings of the 1st International Workshop on Authoring of Adaptive and Adaptable Hypermedia in conjunction with the 4th International Workshop on Authoring of Adaptive and Adaptable (Educational) Hypermedia (A3H), @AH 2006, June 20, 2006, Dublin, Ireland.

[Hess & Detweiler, 1994] Hess, S. M., Detweiler, M. C. (1994): Training to Reduce the Disruptive Effects of Interruptions. In: Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting.

[Horvitz & Apacible, 2003] Horvitz, E., Apacible, J. (2003): Learning and Reasoning about Interruption. In: Proceedings of ICMI'03, Nov. 5-7, 2003, Vancouver, British Columbia, Canada.

[Horvitz et al., 2003] Horvitz, E., Kadie, C., Paek, T., Hovel, D. (2003): Models of Attention in Computing and Communication: From Principles to Applications. In: Communications of the ACM. Vol. 46, March 2003, pp. 52-59.

[Horvitz et al., 2002] Horvitz, E., Koch, P., Kadie, C. M., Jacobs, A. (2002): Coordinate: Probabilistic Forecasting of Presence and Availability. In: Proceedings of the Eighteenth Conference on Uncertainty and Artificial Intelligence, UAI '02. Aug. 1-4 2002, Edmonton, Alberta, Canada, pp. 224-233.

[Huynh et al., 2002] Huynh, D., Karger, D., Quan, D. (2002): Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF. In: Proceedings of Semantic Web Workshop 2002. May 7, 2002, Hawaii, USA.

[Jones et al., 2005] Jones, W., Phuwanartnurak, A. J., Gill, R., Bruce, H. (2005): Don't Take My Folders Away! Organizing Personal Information to Get Things Done. In: Proceedings of the Conference on Human Factors in Computing Systems, CHI 2005, extended abstracts. April 2-5, 2005, Portland, Oregon, USA. ACM, New York, pp. 1505-1508.

[Kaptelinin, 2003] Kaptelinin, V. (2003): UMEA: Translating Interaction Histories into Project Contexts. In: Proceedings of the 2003 Conference on Human Factors in Computing Systems, CHI 2003. April 5-10, 2003, Fort Lauderdale, Florida, USA, pp. 353-360.

[Karger et al., 2005] Karger, D. R., Bakshi, K., Huynh, D., Quan, D., Sinha, V. (2005): Haystack: A Customizable General-Purpose Information Management Tool for End Users of Semistructured Data. In: Proceedings of the 2005 CIDR Conference. Jan. 4-7, Asimolar, California, USA, pp. 13-26.

[Kelly & Teevan, 2003] Kelly, D., Teevan, J. (2003): Implicit feedback for inferring user preference: a bibliography. In: SIGIR Forum. Vol. 37, No. 2, 2003, pp.18–28,

[Kwasnik, 1991] Kwasnik, B. (1991): The importance of factors that are not document attributes in the organisation of personal documents. In: Journal on Documentation, Vol. 47, No. 4, 1991, pp. 389-398.

[Landsdale, 1988]: Landsdale, M. W. (1988): The psychology of personal information management. In: Applied Ergonomics. Vol. 19, No. 1, 1988, pp. 55 -66.

[Ley et al., 2005] Ley, T., Lindstaedt, S. N., Albert, D. (2005): Supporting Competency Development in Informal Workplace Learning. In: Althoff, K., Dengel, A., Bergmann, R., Nick, M., Roth-Berghofer, T. (eds.): Professional Knowledge Management, Lecture Notes in Artificial Intelligence (Vol. 3782), Berlin et al.: Springer, pp. 189-202.

[Linde & Goguen, 1987] Linde, C., Goguen, J. (1987): Checklist interruption and resumption: A linguistic study. NASA Technical Report NASA-CR-177460, Ames Research Center, Moffett Field, California, USA.

[Maier et al., 2006] Maier, D., Archner, D., Delcambre, L., Murthy, S., Annareddy, T.J., Cassel, L. N., Gangula, D., Teng, G. B., Fox, E. A., U. Murthy (2006): Personal Information Enhancement for Education. In: Proceedings of the SIGIR 2006 Workshop on Personal Information Management, PIM 2006. Aug 10-11, 2006, Seattle, Washington, pp. 86-89.

[Malone, 1983] Malone, T. W. (1983): How Do People Organize Their Desks? Implications for the Design of Office Information Systems. In: ACM Transactions on Office Information Systems. Vol. 1, No. 1, Jan. 1983, pp. 99-112.

[Marshall & Jones, 2006] Catherine M. Marshall, Jones, W. (2006): Keeping Encountered Information. In: Communications of the ACM. Vol. 49, Jan. 2006, pp. 66-67.

[Miller, 1968] Miller, G. A. (1968): Psychology and Information. In: American Documentation. July 1968. pp. 286-289.

[Morita et al., 2006] Morita, T., Kura, T., Hidaka, T., Tanaka, A., Kato, Y. (2006): Method of Retrieving a Web Browsing Experience Using Semantic Periods. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference, ISWC 2005, Nov. 6, 2006, Athens, Georgia, USA.

[MOT] MOT homepage, http://wwwis.win.tue.nl/~acristea/mot.html

[Murthy et al., 2006] Murthy, S., Murthy, U., Fox, E. A. (2006): Using Superimposed and Context Information to Find and Re-find Sub-documents. In: Proceedings of the SIGIR 2006 Workshop on Personal Information Management, PIM 2006. Aug 10-11, 2006, Seattle, Washington, pp. 90-93.

[Najjar et al., 2005] Najjar, J., Meire, M. , Duval, E. (2005): Attention Metadata Management: Tracking the use of Learning Objects through Annotation.XML. In: World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005, EDMEDIA05., Montreal, Canada, AACE, pp. 1157-1161.

[Nejdl et al., 2002] Nejdl, W., Wolf, B. , Qu, C. , Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmér, M., Risch, T. (2002): EDUTELLA: A P2P Networking Infrastructure Based on RDF. In: Proceedings of WWW'02. May 7-11, 2002, Honululu, Hawaii, USA, pp. 604-615.

[Oard & Kim, 2001] Oard, D. W., Kim, J. (2001): Modeling information content using observable behavior. In: Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology. Pp. 38-45.

[Ochoa et al., 2005] Ochoa, X. , Cardinaels, K. , Meire, M., Duval, E. (2005): Frameworks for the Automatic Indexation of Learning Management Systems Content into Learning Object Repositories. In: Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications, EDMEDIA 2005. June 27, 2005, Montreal, Canada, pp. 1407-1414.

[Ochoa & Duval, 2006] Ochoa, X., Duval, E. (2006): Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects. In: Proceedings of the 1st International ACM Workshop on Contextualized Attention Metadata: collecting, managing and exploiting of rich usage information, CAMA 2006. Nov. 10 - 11, 2006, Arlington, Virginia, USA, pp. 9 -16.

[Ogata & Yano, 2004] Ogata, H., Yano, Y. (2004): Context-Aware Support for Computer-Supported Ubiquitous Learning. In: Proceedings of the 2nd IEEE Workshop on Wireless and Mobile Technologies in Education, WMTE'04. March 23-25, Taoyuan, Taiwan, pp- 27-34.

[Oliver et al., 2002] Oliver, N., Horvitz, E., Garg, A. (2002): Layered Representations for Recognizing Office Activity, In: Proceedings of the 4th IEEE International Conference on Multimodal Interaction, ICMI 2002. Oct. 14-16, 2002, Pittsburgh, Pennsylvania, USA, pp. 3-8.

[Oren, 2006] Oren, E. (2006): An Overview of Information Management and Knowledge Work Studies: Lessons for the Semantic Desktop. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference, ISWC 2005, Nov. 6, 2006, Athens, Georgia, USA.

[Osterfeld et al., 2005] Osterfeld, F., Kiesel, M., Schwarz, S. (2005): Nabu – A Semantic Archive for XMPP Instant Messaging. In: Proceedings of the 1st Workshop on the Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, located at the International Semantic Web Conference, ISWC 2005. Nov. 6, 2005, Galway, Ireland.

[Pettigrew et al., 2001] Pettigrew, K., Fidel, R., Bruce, H. (2001): Conceptual frameworks in information behavior. In: Annual Review of Information Science and Technology. Vol. 35, pp. 43-78.

[Portugal et al., 2000] Portugal, R.C., Guerrero, L.A., Fuller, D.A. (2000): Desktop, a System based on Virtual Spaces to Support and to Promote Collaborative Learning. In: Proceedings of the 3rd International Conference on Collaborative Virtual Environments, CVE 2000. Sep. 10-12, San Francisco, California, USA, pp. 199-200.

[Ras et al., 2005] Ras, E., Memmel, M., Weibezahl, S. (2005): Integration of E-Learning and Knowledge Management – Barriers, Solutions and Future Issues. In: Proceedings of the 3rd Conference on Professional Knowledge Management – Experiences and Visions. Berlin et al.: Springer.

[Ravasio et al., 2004a] Ravasio, P., Guttormsen Schär, S., Krueger, H. (2004): In Pursuit of Desktop Evolution: User Problems and Practices With Modern Desktop Systems. In: ACM Transactions on Computer-Human Interaction. Vol. 11, No. 2, June 2004, pp. 156-180.

[Ravasio et al., 2004b] Ravasio, P., Guttormsen Schär, S., Krueger, H. (2004): The virtual desktop: Selected results of a user study. In: Arbeit + Gesundheit in effizienten Arbeitssystemen. Proceedings of the 50th Spring Conference of the German Association for Ergonomy (Gesellschaft für Arbeitswissenschaften GfA). March 24-26, 2004. Zürich, Switzerland.

[Reder & Schwab, 1990] Reder, S., Schwab, R. (1990): The Temporal Structure of Cooperative Activities. In: Proceedings of the Conference on Computer Supported Cooperative Work, CSCW'90. Oct. 7-10, Los Angeles, California, USA, ACM, pp. 303-316.

[Reif et al., 2006] Reif, G., Gall, H., Morger, M. (2006): Semantic Clipboard - Semantically Enriched Data Exchange Between Desktop Applications. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference, ISWC 2005. Nov. 6, 2006, Athens, Georgia, USA.

[Rhodes & Maes, 2000] Rhodes, B.J., Maes, P. (2000): Just in time information retrieval agents. In: IBM Systems Journal. Vol. 39, No. 3/4, pp. 685-704.

[Ringel, 2003] Ringel, M. (2003): When One Isn't Enough: An Analysis of Virtual Desktop Usage Strategies and Their Implications for Design. In: Extended abstracts of the 2003 Conference on Human Factors in Computing Systems, CHI 2003. April 5-10, 2003, Fort Lauderdale, Florida, USA, pp. 762-763.

[Ringel et al., 2003] Ringel, M., Cutrell, E., Dumais, S., Horvitz, E. (2003): Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores. In: Proceedings of Interact 2003, the 9th IFIP TC13 International Conference on HCI. Sep. 1-5, 2003, Zurich, Switzerland, IOS Press, Amsterdam, The Netherlands, pp. 184-191.

[Robertson et al., 2000] Roberston, G., van Dantzich, M., Robbins, D., Czerwinski, M., Hinckly K., Risden, K., Thiel, D., Gorokhovsky, V. (2000): The Task Gallery: A 3D Window Manager. In: Proceedings of the 2000 Conference on Human Factors in Computing Systems, CHI 2000. April 1-6, 2000, The Hague, The Netherlands, pp. 494-501.

[Rosenberg, 2006] Rosenberg, M. J. (2006): What Lies Beyond E-Learning? http://www.learningcircuits.org/2006/March/rosenberg.htm, verified Nov. 30, 2006.

[Rosenberg, 2001] Rosenberg, M. J. (2001): E-Learning. Strategies for Delivering Knowledge in the Digital Age. McGraw-Hill: New York et al.

[Sauermann, 2005] Sauermann, L. (2005): The Semantic Desktop – a Basis for Personal Knowledge Management. In: Proceedings of I-KNOW'05, June 29-July 1, 2005, Graz, Austria.

[Schandl & King, 2006] Schandl, B., King, R. (2006): The SemDAV Project: Metadata Management for Unstructured Content. In: Proceedings of the 1st International ACM Workshop on Contextualized Attention Metadata: collecting, managing and exploiting of rich usage information, CAMA 2006. Nov. 10 - 11, 2006, Arlington, Virginia, USA, pp- 27-31.

[Schmidt, 2005] Schmidt, A. (2005): Bridging the Gap between Knowledge Management and E-Learning with Context-Aware Corporate Learning. In: Proceedings of Wissensmanagement. (Professional Knowledge Management), 3rd Biennial Conference, WM 2005. April 10-13, 2005, Kaiserslautern, Germany, Revised Selected Papers, LNCS, 3782 Springer 2005, pp. 203-213

[Schmidt & Braun, 2006] Schmidt, A., Braun, S. (2006): Context-Aware Workplace Learning Support: Concept, Experiences, and Remaining Challenges. In: Innovative Approaches for Learning and Knowledge Sharing. In: Proceedings of the 1st European Conference on Technology-Enhanced Learning, EC-TEL 06. Oct. 1-4, 2006, Heraklion, Crete, Greece, Springer, pp. 518-524.

[Segal & Kephart, 2000] Segal, R., Kephart, J. O. (2000): Incremental Learning in Swiftfile. In: Proceedings of the 7th International Conference on Machine Learning, ICML 2000. June 29 - July 2, 2000, Stanford University, Standord, California, USA, pp. 863-870.

[Segal & Kephart, 1999] Segal, R., Kephart, J. O. (2000): Mailcat: An Intelligent Assistant for Organizing E-Mail. In: Proceedings of the 3rd International Conference on Autonomous Agents. May 1-5, 1999, Seattle, WA, USA, pp. 276-282

[Semantic Desktop] Semantic Desktop, http://www.semanticDesktop.org/

[Sesame] Sesame, http://www.openrdf.org/

[Sevarac, 2006] Sevarac, Z. (2006): Neuro Fuzzy Reasoner for Student Modeling. In: Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies, ICALT 2006. July 5 - 7, 2006, Kerkrade, The Netherlands, pp. 740-744.

[Specht & Kravcik, 2006] Specht, M., Kravcik, M. (2006). Authoring of Learning Objects in Context. In: International Journal on E-Learning. Vol. 5, No. 1, pp. 25-33.

[Stumpf et al., 2005] Stumpf, S., Bao, X., Dragunov, A., Dietterich, T. G., Herlocker, J., Johnsrude, K., Li, L., Shen, J. (2005): Predicting User Tasks: I Know What You're Doing! In: Proceedings of the 20th National Conference on Artificial Intelligence, AAAI-05, Workshop on Human Comprehensible Machine Learning. July 9-13, 2005, Pittsburgh, Pennsylvania, USA.

[Stumpf & Herlocker, 2006] Stumpf, S., Herlocker, J. (2006): TaskTracer: Enhancing Personal Information Management Through Machine Learning. In: Proceedings of the SIGIR 2006 Workshop on Personal Information Management, PIM 2006. Aug 10-11, 2006, Seattle, Washington, pp. 105-107.

[Teevan et al., 2004] Teevan, J., Alvarado, C., Ackerman, M. S., Karger, D. R. (2004) : The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In: Proceedings of the 2004 Conference on Human Factors in Computing Systems, CHI 2004. April 24–29, 2004, Vienna, Austria, pp. 415-422.

[Whittaker et al., 2006]: Whittaker, S., Bellotti, V., Gwizda, J. (2006): Email in Personal Information Management. In: Communications of the ACM. Vol. 49, No.1, Jan. 2006, pp. 68-73.

[Whittaker & Sidner, 1996] Whittaker, S., Sidner, C. (1996): Email Overload: exploring personal information management of email. In: Proceedings of the 1996 Conference on Human Factors in Computing Systems, CHI 1996. April 13-18, 1996, Vancouver, Canada, pp. 276-283.

[Xiao & Cruz, 2006] Xiao, H., Cruz, I. F. (2006): Application Design and Interoperability for Managing Personal Information in the Semantic Desktop. In: Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop, SemDesk 2006, located at the 5th International Semantic Web Conference, ISWC 2005. Nov. 6, 2006, Athens, Georgia, USA.

[Xiao & Cruz, 2005] Xiao, H, Cruz, I. F. (2005): A Multi-Ontology Approach for Personal Information Management. In: Proceedings of the 1st Workshop on the Semantic Desktop: Next Generation Personal Information Management and Collaboration Infrastructure, located at the International Semantic Web Conference. Nov. 6, 2005, Galway, Ireland.

[Zhou & Nejdl, 2006] Zhou, X., Nejdl, W. (2006). Desktop Search with Malleable Schemas. L3S, Hannover. Technical Report, April 2006.