Workplace Learning: how we keep track of relevant information

Kerstin Bischoff, Eelco Herder, Wolfgang Nejdl

L3S Research Center / University of Hannover
Appelstraße 9a
30167 Hannover, Germany
{ bischoff, herder, nejdl }@l3s.de

At the workplace, learning is often a by-product of working on complex projects, requiring self-steered, need-driven and goal-oriented retrieval of information just in time from documents or peers. The personal desktop provides one rich source for learning material and for adaptation of learning resources. Data within that personal information space enables learning from previous experience, sharing tacit and explicit knowledge, and allows for establishing context and context-aware delivery of learning material. Results from personal desktop studies and the corresponding technologies have therefore great potential to enhance TEL. Thus, this paper (1) provides a short overview of desktop organization and search studies as well as applications and (2) envisions tighter incorporation of desktop research for innovative TEL infrastructures.

1 Introduction

In our information-based society with its rapidly changing demands, knowledge gets outdated pretty fast. As people often change jobs, the notions of lifelong learning and workplace learning have gained quite some attention. To facilitate these types of learning, our perspective on learning has to be broadened [22]. Knowledge workers, who spend most of their time on retrieving, processing, creating and manipulating knowledge, rely on efficient access to data in different formats [22, 24]. Besides external (corporate) repositories, PC desktops – including possibly connected desktops of colleagues – provide a rich source of valuable material for informal learning.

This paper is structured as follows. In the next section we motivate why personal (semantic) desktop environments are important for learning. In section three, we give a short overview on desktop organization and usage studies. The fourth section describes approaches toward supporting information access and delivery. Finally, we outline how innovative learning scenarios can effectively employ such techniques.

2 Learning in Context

The PC desktop provides a lot of learning resources in various formats. Personal information - including documents, emails, web cache, notes, calendars, links, instant

messaging, all connected to the users and their peers – provides a rich source of prior working experiences [4, 22]. Some may be well-structured learning objects, but most resources are just documents, emails and visited websites. The desktop can provide rich data about a user's activities and interests to enable context-aware delivery of information.

These personal resources encourage *learning from experience*: learning at the workplace that is less based on instruction than on collaboration with peers and learning in action and by reflection [1, 11]. Technological support to raise awareness of relevant knowledge and solutions from the past enables the integration of continuous experiential learning into work processes. It offers a great opportunity to connect new information to prior knowledge and experience in a meaningful way and to make us aware of what we know and of potential gaps we have. These rich, personal repositories are likely to be more effective for disseminating highly context-specific – often *tacit* - knowledge [11]. Moreover, as learning takes place at the workplace itself, no or only minor transfer of *context* is necessary.

3 How do we keep track of relevant information?

Given the increasing amount of information on our PCs, several studies have examined organizational behaviour on the personal desktop. In this section we review research and studies that focus on the strategies that people employ to refind their documents, emails and information encountered on the Web.

Documents typically contain frequently used information closely related to current tasks, which later becomes *archived* [2]. Some users systematically order their documents, others just *pile* them [18]. Users may pile because they can't properly classify the information [16] or anticipate future usage and retrieval. On the other hand, piles may serve as *reminders* [18]. Barreau and Nardi [2] found that electronic documents are usually organized into thematic folders. A proper folder structure provides means for relocating documents and timely reminders. Users tend to place items that need to be paid attention to on the desktop or some other place where they likely will notice them. However, archiving old information is often not considered worth the effort [2].

For refinding documents, users often engage in 'orienteering behavior': instead of providing an exact query, they navigate to the target document in smaller steps [25]. As defining a query is often as hard as efficiently organizing the data [2, 16], location-based orienteering – skimming through a list of folders – is preferred to keyword search. According to [2], users only employ search tools after other unsuccessful trials. A possible explanation is that current text search tools do not support the rich associations that people use as retrieval cues [7, 16].

Due to its interpersonal nature, **email**, and the ways users handle it, differs much from other information items. Email predominantly carries *ephemeral* information [2] that is only needed for a short time, such as memos, to-do-lists, and mail messages. Email plays an important role in everyday life, supporting activities as contact management, personal archiving and document exchange [9, 26, 27].

Many users keep almost all of their emails in the Inbox, as archiving costs time and effort. Besides, the Inbox serves as a list of reminders [27]. As far as archiving is concerned, users may be classified as *no filers*, *frequent filers* and *spring cleaners* [27]. Only frequent filing provides effectiveness and the chance of being reminded, but it does not always compensate for the archiving time. [9] found shallow file hierarchies – organized by sender, organization, project – to be common to have immediate access. Location-based search and sorting of mails by sender or date with subsequent browsing were popular strategies employed in looking for a message. Search tools and automatic filters were less frequently used.

Bruce et al. [4] empirically collected a list of common keeping methods for important **information encountered on Web sites**. Among them: sending an email with the URL to self, printing out a Web page, bookmarking, saving it to a file, pasting the URL into a document, writing the URL down. These methods are not heavily used; though. And although browsers provide several means for relocating information found earlier on the Web – including the back button, bookmarks, URL auto-completion and the history toolbar – these tools do not provide the functionality needed for refinding information. In a study on revisitation [19], we found that users particularly had issues in relocating a page visited weeks or months before, as the Web address mostly did not reside in the browser's memory – and not in the user's memory either. This left users with little more choice than a *repeated search*, which often turned out to be unsuccessful, due to the user's inability to replicate the original query, or due to the fact that the original query did not directly led to the desired page. The results showed the need for better support for *orienteering* behaviour [25].

Concluding, users face many problems in managing their data. While location-based browsing is often successfully used to find personal files, classification and structuring is time-consuming and cognitively hard. Even less effort is commonly spent on archiving emails. Search tools are not frequently used, because they lack important features. **Fragmentation of information access** exacerbates these problems: resources are spread over the PC and bound to specific applications [26]. Thus, assistance in (multiple, flexible) filing, search facilities offering enhanced attributes, and reminding, integrated desktop infrastructures as well as task management are critical.

4 Systems supporting information access on the desktop

Various approaches and tools have been proposed to organize and search personal information spaces more naturally and efficiently. The next sections will introduce some of these innovative applications and the contributions they brought.

4.1 Integrated search infrastructures

Beagle++ [6] is a desktop search system that indexes all personal documents and generates additional metadata that describes these documents, other resources, as well as their relationships. Triggered by modification events, Beagle++ annotates the

material that the user has read, used, written or commented upon. *Haystack* [15] also generates annotations and provides dynamic collection views, but it focuses on agents exploiting user specific and predefined ontologies. It supports search, as well as associative browsing or 'orienteering'. *Stuff I've seen* (SIS) [10] uses rich contextual cues such as time, people, thumbnails and previews to support retrieval and presentation. SIS was extended to comprise timeline visualization, where important personal and public landmarks (photographs, calendar or news events, holidays) were displayed together with results of a keyword search [21]. *Phlat* [8] is a follow-up, enhanced to allow tagging information with multiple meaningful, personal annotations. Similarly, the integrated platform *MyLifeBits* [12] was developed around *annotations* and *links*. As an alternative to filing, manual annotations (or tags) serve for organizing information, and for meaningful, intuitive search or browsing by content. Linking enables associative browsing and serendipitous encounters.

4.2 Recommending personal information

While the former approaches enhance the refinding of information, users may often be *unaware* of information related to their current work already existing on their desktops – or they do not have the time to search for it. Just-in-time Information Retrieval (JITIR) Agents [20] *proactively recommend* relevant resources, by modelling the user's preferences and tasks from the user's current activities and interactions like Web navigation, saving, or printing.

Rhodes and Maes [20] describe the implementation of three different agents. The *Remembrance Agent* monitors the user and continuously searches the desktop or databases for related items matching the current task. Suggestions are displayed in a side window. *Margin Notes* links Web pages to personal files by rewriting the source code on-the-fly. *Jimminy* bases his suggestions on various environment-aware sensors contained in a shoulder-mounted wearable computer. The results are shown on a head-mounted display. *Watson* [5] additionally uses a simple and explicit task model to interpret user actions in order to anticipate a user's information need.

4.3 Supporting tasks and processes

Tasks are central in working. The value of resources is mainly determined by their relation to the current context. *TaskMaster* [3] organizes emails, attached documents or sent URLs around tasks. These communication threads – or 'thrasks' – are built by analyzing message data. Task specific meta-information – deadlines, appointments, to-dos – can also be added and visualized. A similar approach is followed in *UMEA* [14], which uses *projects* as organizational units and provides interaction history as context. By contrast, the *TaskTracer* project [24] employs machine learning to learn and predict user tasks from traced interactions with the operating system/applications. Based on learned correlations between tasks and folders, a Folder Predictor suggests a folder for saving or opening resources, thus saving interaction costs.

We will now illustrate how TEL scenarios can benefit from such techniques.

5 Information finding in innovative learning scenarios

Authors of e-Learning content may be supported in creating learning material by their own desktop resources, for instance by semi-automatically *enriching* a course with available publications, in order to adapt to different knowledge levels [13]. Systems in the Sidewalk Project [17] allow for manually marking and linking one's own resources to concepts of a lesson; these links could be created automatically to provide a personalized, enriched concept map that promotes elaboration and motivation.

The benefit of establishing context and providing easy access to already existing resources seems even more important in workplace settings. In these situations, advanced features – like recommendation of experts and reuse of previous experience – as well as techniques for context sharing seem promising. In addition to providing elaborate, semantically enriched and flexible browsing and search facilities, JITIR agents can continuously recommend relevant resources from the repository, in which similar prior problems and solutions are described. Delivery can be personalized and contextualized by compiling a profile that is built from keywords from the desktop or the current task. As an example, the LIP system supports the situation-aware retrieval of resources adapted to the current context [23]. The created context information can also be reused as metadata for learning resources and information fragments.

Conclusion and Outlook

Workplace learning requires advanced information finding functionalities to retrieve relevant knowledge. This paper provides a short overview over relevant research, motivating learning in context, and giving a discussion of information finding and organizing strategies, approaches and systems. We also describe state-of-the-art systems supporting information access on the desktop, which provide advanced search, recommendation or task support functionalities. Finally, we sketch some first ideas on how formal and informal learning can be enhanced with such techniques, as an outlook to future research and towards innovative learning solutions at the workplace.

Acknowledgements: the research reported in this paper has taken place in the contexts of the PROLEARN Network of Excellence and the European IST TenCompetence. Both projects are member of the Professional Learning Cluster.

References

[1] Agyris C. & Schön, D. A.: *Organizational Learning*: *A Theory of Action Perspective*. Reading, MA, Addison-Wesley (1978).
[2] Barreau, D. & Nardi, B. A.: Finding and Reminding. File Organization from the Desktop. *SIGCHI Bulletin*, 27(3) (1995) 39-43.
[3] Bellotti, V., Ducheneaut, N., Howard, M. & Smith, M. I.: Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. *Proc. CHI 2003* (2003) 345-352.

[4] Bruce, H., Jones, W. & Dumais, S.: Information behavior that keeps found things found. *Information Research* 10(1) (2004).
[5] Budznik, J. & Hammond, K.: Anticipating and Contextualizing Information Needs. *Proc. of the 62nd Annual Meeting of the American Society for Information Science* (1999).
[6] Chirita, P.-A., Gavriloaie, R., Costache, S., Nejdl, W. & Paiu, R.: Beagle++: Semantically Enhanced Searching and Ranking on the Desktop. *Proc. ESWC 2006* (2006) 348-362.
[7] Cutrell, E., Dumais, S. T. & Teevan, J.: Searching to Eliminate Personal Information Management. *Communications of the ACM* 49(1) (2006) 58-64.
[8] Cutrell, E., Robbins, D. C., Dumais, S. T. & Sarin, R.: Fast, flexible filtering with Phlat - Personal search and organization made easy. *Proc. CHI 2006* (2006) 261-270.
[9] Duchenaut, N., Bellotti, V.: E-mail as Habitat. An Exploration of Embedded Personal Information Management. *Interactions* 8(5) (2001).
[10] Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R. & Robbins, D.C.: Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. *Proc. SIGIR'03* (2003) 72-79.
[11] Eraut, M.: Informal learning in the workplace. *Studies in Continuing Education* 26(2) (2004) 247-273.
[12] Gemell, J., Bell, G. & Lueder, R.: MyLifeBits: A Personal Database for Everything. *Communications of the ACM* 49(1) (2006) 88-95.
[13] Hendrix, M., Cristea, A. I. & Nejdl, W.: Authoring Adaptive Learning Material on the Semantic Desktop. *Proc. A3H@AH 2006* (2006).
[14] Kaptelinin, V.: UMEA: Translating Interaction Histories into Project Contexts. *Proc. CHI 2003* (2003) 353-360.
[15] Karger, D. R., Bakshi, K., Huynh, D., Quan, D. & Sinha, V.: Haystack: A Customizable General-Purpose Information Management Tool for End Users of Semistructured Data. *Proc. CIDR 2005* (2005) 13-26.
[16] Landsdale, M. W.: The psychology of personal information management. *Applied Ergonomics* 19(1) (1988) 55-66.
[17] Maier, D., Archner, D., Delcambre, L., Murthy, S., Annareddy, T. J., Cassel, L. N., Gangula, D., Teng, G. B., Fox, E. A., & Murthy, U.: Personal Information Enhancement for Education. *Proc. PIM 2006* (2006) 86-89.
[18] Malone, T. W.: How Do People Organize Their Desks? Implications for the Design of Office Information Systems. *ACM Transactions on Office Information Systems* 1(1) (1983) 99-112.
[19] Obendorf, H., Weinreich. H., Herder, E. & Mayer, M.: Web Page Revisitation Revisited: Implications of a Long-term Click-stream Study of Browser Usage. *Proc. CHI 2007*(2007)
[20] Rhodes, B. J. & Maes, P.: Just in time information retrieval agents. *IBM Systems Journal* 39(3/4) (2000) 685–704.
[21] Ringel M., Cutrell, E., Dumais, S. & Horvitz, E.: Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores. *Proc. Interact'03*. (2003). 184-191.
[22] Rosenberg, M. J.: *E-Learning. Strategies for Delivering Knowledge in the Digital Age*. New York et al., McGraw-Hill (2001).
[23] Schmidt, A. & Braun, S.: Context-Aware Workplace Learning Support: Concept, Experiences, and Remaining Challenges. *Proc. EC-TEL 2006* (2006) 518-524.
[24] Stumpf, S., Bao, X., Dragunov, A., Dietterich, T. G., Herlocker, J., Johnsrude, K., Li, L. & Shen, J.: Predicting User Tasks: I Know What You're Doing! *Proc. AAAI-05* (2005).
[25] Teevan, J., Alvarado, C., Ackerman, M. S. & Karger, D. R.: The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. *Proc. CHI 2004* (2004) 415-422.
[26] Whittaker, S., Bellotti, V. & Gwizda, J.: Email in Personal Information Management. *Communications of the ACM* 49(1) (2006) 68-73.
[27] Whittaker, S. & Sidner, C.: Email Overload: exploring personal information management of email. *Proc. CHI 1996* (1996) 276-283.