Personal Information Management: PIM 2009

ASIS&T 2009 Workshop, November 7-8, 2009, Vancouver, Canada.

Accepted Papers

Paper Presentations

- Robert Capra. *A Survey of Personal Information Management Practices* (short)
- William Jones. *Providing for Paper, Place and People in Personal Projects* (full)
- David Elsweiler, Martin Hacker and Stefan Mandl. *Visualising PIM Behaviour with Markov Chains* (short)
- Manas Tungare and Manuel Perez-Quinones. *Mental Workload at Transitions between Multiple Devices in Personal Information Management* (full)

Posters

- Leslie Thomson. *No Room for White Space: The Personal Information Management of a Printing Company Employee in a Home Office* (full)
- Florian Evequoz and Denis Lalanne. *"I Thought You Would Show Me How to Do It" - A Study on Self-Analysis of PIM Practices* (full)
- Leonardo Galicia, Victor M. Gonzalez and Jesus Favela. *Examining Personal Activity Record by Knowledge Workers* (full)
- Liadh Kelly, Daragh Byrne and Gareth J.F. Jones. *The Roll Of Places And Spaces In Lifelog Retrieval* (short)
- Deborah Barreau, Laura O'Neill and Amanda Stevens. *The Relationship Between Research and Practice: What Are We Learning About Teaching PIM?* (short)
- Christina M. Finneran. *Metaphor Analysis as a Method for Holistic PIM Research* (short)
- Karl Voit, Keith Andrews and Wolfgang Slany. *Why Personal Information Management (PIM) Technologies Are Not Widespread* (short)
- Sarah Henderson. *Guidelines for the Design of Personal Document Management User Interfaces* (full)
- Jakob Eg Larsen and Maciej Luniewski. *Using mobile phone contextual information to facilitate managing image collections* (short)
- Stephen Lackey. *Integrating Place and Time with Tasks: Adding Transportation to PIM* (short)

Demos

- William Jones. *Writing New Stories for the Same Old File System* (demo)

A Survey of Personal Information Management Practices

Robert Capra
School of Information and Library Science
University of North Carolina
Chapel Hill, NC

rcapra3@unc.edu

ABSTRACT

In this paper, we describe results of a two-part written questionnaire completed by 47 people about how they manage personal information and transfer information among electronic devices. We found three main methods used to transfer files among computers: email to self, USB drives, and network storage. For transferring information among electronic devices such as digital cameras and MP3 players, participants reported using the software and cables that came with the devices to transfer files, or simply using a USB connection. Interestingly, participants reported that they did not typically transfer files from their cell phone to their other devices or computers. With regard to saving information found on the web, almost all participants reported using bookmarks, and over half sent email with the information to themselves. We also asked users to estimate their number of bookmarks, number of email folders, and number of email messages in their Inbox. These were all found to follow exponential distributions, indicating a large number of users with small numbers and a few users with large numbers. These results extend our understanding of users' personal information management behaviors and help inform the design of PIM tools.

1. INTRODUCTION

We administered a two part questionnaire that asked about users' practices in managing their personal information and transferring information among computers and electronic devices. The questions were administered as part of a broader study of how users refind information found on the Web that was conducted in the spring of 2007. In this paper, we focus on the administration, analysis, and results of the questionnaire data.

2. METHOD

Forty seven (47) participants were recruited broadly from the University of North Carolina community using posts to listservs. One of the main listservs used for recruitment is for campus-wide informational messages and is distributed to all students, faculty, and staff that opt-in. We recruited a diverse sample of mainly graduate and undergraduate students with a few faculty and university staff: age (M = 26.0 years, $s.d.$ = 8.3), gender (17 male, 30 female), status (undergrad/no degree = 27, graduate student/primary degree = 14, faculty/staff = 6). Over 25 different departmental affiliations were represented (e.g. Biology, Economics, Nursing, Religious Studies, Languages). Participants generally had a large amount of experience conducting searches on the web (M = 9.5 years, $s.d.$ = 2.0) and all participants reported using a computer on a daily basis.

The questions were administered in two parts, one during each of two sessions of the main study. The sessions were scheduled about one week apart. The first questionnaire asked about demographic information and participants' experience using computers to search for information. The second questionnaire was more detailed and asked about the types and locations of computers used, electronic devices used (cell phone, MP3 player, PDA, digital camera) and how information was transferred among the devices. The second questionnaire also inquired about methods used for saving information and asked participants to give estimates of how many email folders and email messages they had. The questions from the second questionnaire are shown at the end of this paper.

Participants were given a paper copy of the questionnaire and asked to write their responses directly in the spaces provided. The questionnaires were administered at the start of each session, before participants did any of the tasks for the main study. Up to eight participants worked at the same time, but each completed their own questionnaire independently.

3. RESULTS

3.1 Computers Used

We asked about what types of computers were used in a typical week. Almost half the participants, 49% (23), used both desktop and notebook computers, while 40% (19) used only a notebook, and 11% (5) used only a desktop. Almost all users, 96% (45), reported using a Windows-based computer in a typical week, with 13% (6) using the Macintosh OS and 4% (2) using Linux/Unix (note that some participants used more than one OS in a week). The locations that they used computers also varied and many participants reported using computers in multiple locations: 68% (32) reported using a computer at home, 45% (21) at work, 62% (29) at school, and 55% (26) in a portable location. With regard to the computers used, 85% (40) of participants considered one to be their "primary" computer. Of these, 75% (35) considered their notebook computer to be their primary computer. The university has a laptop computer requirement for undergraduate students and encourages all students to purchase laptop computers, so this result is not surprising.

3.2 Transferring Data Among Computers

We asked participants to describe: 1) how they "transfer, share, and synchronize files among the computers" that they used, and 2) what problems they encountered when doing so. These were open response questions so the participants could write answers in their own words. A single coder (the author) coded each response into an open list, resulting in a list of coded responses and frequency counts. Table 1 shows the most frequently mentioned methods of transfer and the number of participants who mentioned each method. Table 2 shows commonly mentioned problems with transferring files.

Emailing files to self, or leaving email attachments in the Inbox was the most commonly mentioned method of transferring files among computers, mentioned by almost half our participants

(47%). However, participants noted that using email to transfer files has limitations. They mentioned having problems emailing large files, and some described how they used email for small files and USB drives or CD/DVDs for larger files.

The use of USB "flash drives" (also referred to as "pen drives" or "thumb drives") was a commonly reported (34%) means of transferring files among different computers. Some problems mentioned in using USB drives were running out of space on the drive and occasional problems getting different computers to recognize the USB drive.

Method of transfer	**Percentage of participants**
Email to self	47% (23)
USB drive	34% (16)
Don't transfer much – use mostly one computer	30% (14)
Shared Network Drive / Storage	21% (10)
FTP to a server	11% (5)
Burn CD/DVD	9% (4)

Table 1. Methods of Transferring Files Among Computers

Participants also reported transferring files by putting them on a shared network drive or server that is accessible from the computers being transferred among. Two main methods were described: using a shared network drive (21%), and using FTP/SSH to transfer files to a server (11%). We should note that the university provides students with space on a network drive which can be accessed either as a shared network drive or through FTP/SSH. One problem mentioned with this approach is that the network drive was not always accessible due to network and connection problems. About a third of our participants (30%) primarily used a single computer and reported not doing much file transfer.

Problems when transferring files among computers	**Percentage of participants**
File transfer/compatibility problems	20% (9)
File size too big	17% (8)
Hard to find / most recent version	11% (5)

Table 2. Problems Transferring Files

Participants mentioned two common problems that seemed to cut across the method used to transfer the files. The first had to do with file transfer or compatibility problems. Participants described situations in which files did not open properly after transfer, not having the proper software to open files on the computer transferred to, and damaged or missing files. In the second problem class we observed, participants described difficulties in finding a transferred file and in knowing which file was the most recent version.

3.3 Electronic Devices

We asked about how participants transferred, shared, or synchronized files among the different electronic devices and computer that they used. These were also open-answer questions and were coded using a method similar to that described in the previous section.

Many participants described using software and cables that came with device to transfer files, or by connecting the device through a USB port and then copying files. Responses included using iTunes with iPods (9 responses), memory cards with cameras (5), and the use of USB cables with digital cameras (10) and iPods (6). A few responses (4) described using email to transfer files from a cell phone to a computer. However, a number of participants (13) specifically mentioned that they do not typically transfer files between their cell phone and computer.

3.4 Web Information Keeping Methods

We asked participants if they used any of six specific keeping methods to save "interesting or important web sites". The six methods were based on keeping practices identified from previous research studies of PIM practices, especially [3]. The six methods and percentage responses are shown in Table 3.

Keeping Method	**% Response**
Save as a bookmark/favorite	98%
Send myself an email with the URL	57%
Write it down on paper	36%
Copy the URL into a file	23%
Save a copy of the web page to my hard drive	17%
Save it to del.icio.us or other bookmarking site	11%

Table. 3. Web Information Keeping Methods

Aula, et al. [1] conducted a study in which they asked participants to rate how often (never, rarely, sometimes, often. almost always) they used various finding and refinding methods. Comparing their results and ours provide insights and also some points for further exploration. In our study almost all users (98%) reported using bookmarks. Aula et al. found a similar number (92.4%). However, in the Aula et al. study, the median response was that bookmarks were used "sometimes", indicating that many users do use bookmarks, but only sometimes. Aula et al. reported that emailing URLs to self and writing URLs down on paper were not commonly used strategies. However, many of our participants reported using these methods. Further study is needed to understand more about the context and frequency of this use.

3.5 Bookmark Use

Participants were asked to write down an estimate of how many bookmarks they had on their primary computer. User-reported estimates can be inaccurate, but they are easy to obtain, can provide an idea of the general size or trends of the measure, and have been used in other studies [1]. Our participants reported having an average of 49.5 bookmarks (*median* 30, *s.d.* 64.1) with a minimum of 2 and a maximum of 300. One outlier greater than 10 standard deviations from the mean was removed from this

analysis (the outlier indicated 1000 bookmarks and 20000 email messages). These results indicate that while 98% of our participants reported using bookmarks, half reported having 30 or fewer. Our data compares with that from Boardman and Sasse's study published in 2004 [2] and the survey results from Aula et al. published in 2005 [1]. Boardman and Sasse reported an average of 16.8 bookmarks per user, with a minimum of 0 and a maximum of 180. Aula et al. reported a larger average of 220 bookmarks and 30 folders, with 14.4% having less than 50 bookmarks.

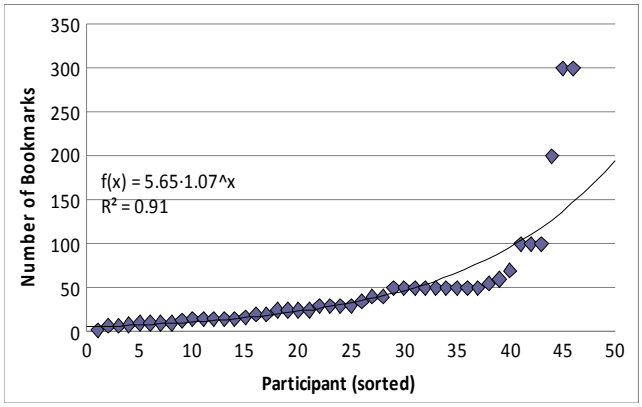


Figure 1. User-reported estimates of number of bookmarks

Figure 1 shows a graph of the estimated number of bookmarks reported by our participants, sorted by size. An exponential trend line has been computed and displayed with an R^2 value of 0.91, indicating a good fit. This suggests that the number of bookmarks that people have follows an exponential distribution.

3.6 Email Folders and Inbox

We also asked participants to give estimates of the number of email folders they had and the number of email messages in their Inbox. Figures 2 and 3 show these numbers, sorted in increasing order. For each, an exponential trend line has been computed and displayed. The R^2 values are 0.95 and 0.94, indicating good fits for both curves. This suggests that the number of email folders that people have and the number of email messages in their Inboxes follows an exponential distribution.

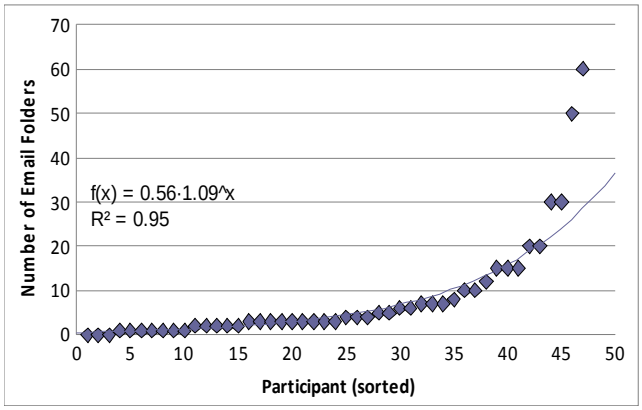


Figure 2. User-reported estimates of number of email folders

We were interested to see if there was any correlation between the number of folders and the number of email messages. A simple hypothesis here is that people with more folders are "filers" and might be more likely to file messages, resulting in fewer messages in their Inbox at any given time. However, a correlation analysis resulted in a Pearson's r = -0.18, indicating only a small negative correlation between these two variables.

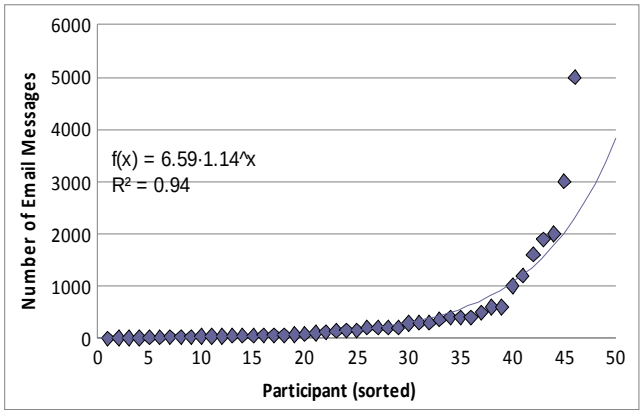


Figure 3. User-reported estimates of Inbox size

4. SUMMARY AND CONCLUSIONS

Participants in our study used three main methods for transferring files among the computers they use: 1) email to self, 2) USB drives, and 3) network-based storage accessible from many different computers. The choice of method may be influenced by file size due to limitations that many email systems place on the size of attachments.

Our participants typically used either simple USB connections or the software and cables that came with devices such as digital cameras and MP3 players to transfer files between these devices and their computer. Files were not typically transferred between their cell phone and computer.

We found that participants made use of bookmarks, email, and writing down notes as primary methods to save information found on the Web. While bookmarks were reported as being used by almost all our participants, half estimated having 30 or fewer bookmarks.

An analysis of user provided estimates of the number of their bookmarks, email folders, and email messages in their Inbox showed that these all follow an exponential distribution. An correlation analysis found little correlation between the number of folders and Inbox size.

Across all the types of transfer that we examined, email played a role. Our participants reported using email to transfer files between computers, a few used it to move files between their computer and cell phone, and many used it to save information found on the Web. The "overloaded" functions of email have been reported on before, e.g. [4][3]. Our study further illustrates the important "common denominator" that email plays in people's need to access data from multiple locations and on multiple devices. These results contrast with the conclusions of Aula et al. [1] that experienced users do not often email themselves as a method for saving information. Our study was conducted approximately two and a half years after Aula's which may account for these differences, or they may be due to differences in

the populations studied. This is an area for which additional research is needed.

These results extend our understanding of how users manage their personal information and help inform the design of PIM tools.

5. ACKNOWLEDGMENTS

This work was supported in part by NSF grant IIS 0812363.

6. REFERENCES

[1] Aula, A., Jhaveri, N., and Käki, M. 2005. Information search and re-access strategies of experienced web users. In Proceedings of the 14th international Conference on World Wide Web (Chiba, Japan, May 10 - 14, 2005). WWW '05. ACM, New York, NY, 583-592.

[2] Boardman, R. and Sasse, M. A. 2004. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM Press, New York, NY, 583-590.

[3] Jones, W., Bruce, H. and Dumais, S. (2001). Keeping Found Things Found on the Web. In Proceedings of the Tenth International Conference on Conference on Information and Knowledge Management. ACM Press, New York, 119-126.

[4] Whittaker, S. and Sidner, C. 1996. Email overload: exploring personal information management of email. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground (Vancouver, British Columbia, Canada, April 13 - 18, 1996). M. J. Tauber, Ed. CHI '96. ACM Press, New York, NY, 276-283.

Questions

1.Describe the computer(s) that you use in a typical week.
- a.Computer type: Desktop or notebook
- b.Operating system: Mac, Windows, Linux, other _____
- c.Location: portable, home, work, school (library, lab, etc.), other _____

2.Do you consider one of the computers listed above as your primary computer? If so, which one?
- a.Yes or no
- b.If yes, which one? _____

3.Briefly describe how you transfer, share, and synchronize files among the computers that you use. For this question, consider only the computers you use – mobile devices will be asked about in a later question. If you use only one computer, just enter “one computer” for this question.

4.List and briefly describe common problems that you have transferring, sharing, and/or synchronizing files among the computers that you use. For this question, consider only the computers you use – mobile devices will be asked about in a later question. If you use only one computer, just enter “one computer” for this question.

5.Which of the following devices do you own and use in a typical week?
- a.Cell phone
- b.MP3 player (iPod, etc.)
- c.Personal digital assistant (Palm, Pocket PC, Blackberry, etc.)
- d.Digital camera

6.Briefly describe how you transfer, share, and synchronize files among the devices and computers that you use.

7.Briefly describe any problems that you commonly have transferring, sharing, and/or synchronizing files among the devices and computers that you use.

8.What methods do you typically use to save interesting or important web sites? (check all that apply)
- a.Save as a bookmark/favorite
- b.Send myself an email with the URL
- c.Copy the URL into a file
- d.Save a copy of the web page to my hard drive
- e.Save it to del.icio.us or other bookmarking site
- f.Write it down on paper
- g.Other (please specify) _____

9.Estimate how many bookmarks you have on your primary computer.

10.How many email folders do you have?

11.How many email messages are in your inbox?

Providing for Paper, Place and People in Personal Projects

William Jones, Harry Bruce, Elisabeth Jones, Julia Vinson

The Information School, University of Washington

ABSTRACT

How do people go about planning and completing personal projects? What can be done to help? These questions are important in their own right. Also, in a digital age of information, managing a project often means managing many forms of information over extended periods of time including paper documents, electronic documents, email messages, and several forms of web information (conventional web pages, blogs, wikis, etc.). Questions, therefore, have relevance to the study of personal information management (PIM). This article describes qualitative results gleaned from an in-depth study of people completing personal projects. Participants from a range of professions and backgrounds each selected a personal project meeting certain criteria (the project could be freely discussed, involved several forms of information and was expected to last for several more weeks). For each participant, progress on the selected project was then tracked through a series of situated interviews over a period four to twelve weeks. Results point to the enduring importance of paper and "place" in any system of supporting tools. Participants used paper in several ways – to brainstorm, to remind, to motivate and to track their efforts. Participants often needed to give information a place – whether in physical or digital space. Other people were also an important factor in a project's timely completion. People sometimes complicated or impeded a participant's efforts to complete a project. But, more commonly, other people were a source of motivation and assistance. Study results suggest that the factors of paper, place and people should each be considered in efforts to support personal information management.

Keywords: Personal information management, human information behavior, ethnography, problem-solving, project planning

Introduction

At any point in time, most of us are working on several different projects. Some of projects are work-related (e.g., "complete annual report"); some projects are not (e.g., "buy a new car"). Some of our projects are part of a larger project involving other people (e.g., "submit a plan for re-structuring my group as part of the larger company re-organization" or "get legal advice as part of my work on the board of directors for our condominium"). These projects are "personal" to us because they're important to us and because we're responsible. The projects won't get done without us. Sometimes we have help and are mostly doing the planning and supervision. On other occasions, we may be doing things mostly on our own from start to finish.

The study of how people manage projects in their lives has relevance to several fields of inquiry. For example, personal projects involve planning. As such, results have relevance to the basic study human cognition. Mumford, Shultz & Van Doorn (2001) note that the study of planning in psychology has proceeded in "fits and starts" over the past 50 years and remains under-developed.

The study of personal projects is also relevant to the study of personal information management or PIM. A key challenge in the study of PIM is to understand how people manage information of

several forms over extended periods of time (Jones, 2007). Personal projects in our lives often involve information in several forms – paper documents, electronic documents, email messages, and several forms of web information (conventional web pages, blogs, wikis, etc.). And personal projects often last for weeks or months from start to finish. Personal projects are, therefore, potentially a manageable unit of analysis for the larger study of PIM.

A project such as "Trip to Boston" is composed of tasks such as "make plane reservations" or "get travel authorization". A number of studies in recent years have looked at how people manage tasks in their lives (for example, Bellotti et al., 2003, 2004; Czerwinski et al, 2004; Gwizdka, 2002; Wolverton, 1999). These studies point to the challenges people face as they are interrupted throughout a normal working day and must switch between several tasks.

More recently, studies have begun to provide insight concerning how people go about managing personal projects and the information needed to complete these projects. In a study looking at how people organize different forms of information (files, email messages and web references) Boardman and Sasse (2004) found that projects were a common basis for creating and naming file folders. Projects, sometimes the same projects, were also frequently reflected in the choice of email folders. But, perhaps more commonly, as Bergman et al. (2006) discovered in another study, information relating to a project was "fragmented" across very different organizations – one for documents, another for email messages and another for web references. A study by Jones et al. (2005) revealed that the structure of subfolders under a project folder often served multiple, albeit ad hoc, purposes in a person's efforts to manage a project. Subfolders were, for example, a reminder of tasks to be done as well as a means of grouping the information needed to complete these tasks.

These studies are especially useful for their informational scope. The studies do not narrowly focus on the use of just one form of information (e.g., email messages or web references). Instead, the studies look at how people organize across different forms of information. This larger scope of inquiry reveals important patterns and problems in PIM. From all three studies, for example, we learn that folder structure is generally more elaborate for a person's electronic documents and other files than for email messages and web references. And all three studies point to a problem of information fragmentation: A person's informational challenges are often multiplied by a proliferation of informational forms each with its own organization and its own constellation of supporting tools.

The study described in this paper takes an additional step in the study of PIM by looking at how people manage different forms of project-related information as projects unfold over a period of time.

The Study

The study involved 27 participants (14 female), ages 19 to 49. Current job or professional endeavor for participants ranged widely. Included in the sample were students (two undergraduates, three masters level students, three doctoral candidates), software engineers (including one video game designer and one system administrator), teachers, librarians and administrators (including two workers in non-profit organizations).

During a preliminary interview, participants were asked to list several projects that they were currently working on and then to select from this list a project that: 1. could be discussed freely, 2. involved several forms of information, and 3. was expected to last for several more weeks. Selected projects ranged widely (see Table 1). Some were for work; some were not. Some involved other people; some involved only the participant.

Table 1. A sampling of projects selected by participants

Image print suite, depicting past, present, and future images of women
Arranging a group visit for children and their mentors to see the children's incarcerated parents
Converting paper files to electronic files
Coordinating a number of local charitable organizations to facilitate cooperation
Curriculum Map for Language Courses
Design a video game
Learning the procedures for new job as a librarian
Writing a guide to fly fishing
Making an interactive (electronic-enhanced) stuffed animal
Masters of information management capstone project
Organizing a summer institute course, in collaboration with 3 faculty members
Planning Star Wars game campaign
Preparing the reading lists for doctoral general exam
Testing new advertising targeting software for his job
Training for a triathlon

For the selected project, participants then completed, depending upon their availability, from two to five follow-on sessions each lasting from 60 to 90 minutes and occurring over a period of four to twelve weeks. Participants were paid $15 per hour for their time. The primary focus of follow-on sessions was the selected project[1].

Participants were asked to show and describe their organizations of project-related paper documents, electronic files, email, and web references. The researcher posed questions to better understand the "how" and "why" behind the various uses and organizations of project-related information. More holistic questions were also asked about the participants' satisfaction with their organization systems, and how those systems might be improved. Follow-up interviews explored any changes that had occurred in the participants' organization strategies in the intervening weeks between sessions, as well as the participants' evolving attitudes towards their chosen strategies. All interviews were audio recorded and transcribed. In the analysis phase, the interview transcripts were coded using a constant comparative technique to elicit overarching themes.

Results

The study produced an enormous amount of data. The focus of this paper is on results gleaned from a qualitative analysis of transcripts.[2] As Malone (1983) notes, the value of such a qualitative analysis is often in the insights and compelling examples that result. In some cases, examples and insights deserve special focus in follow-on studies. In other cases, examples and insights may point directly to implications for tool design.

---

[1] Eighteen of 27 participants were able to complete all five sessions of the study. These participants, during sessions two and four, completed a series of performance tasks whose results are not described in this paper.

[2] An article including quantitative results of the study is forthcoming.

Participant comments point to the recurring importance of three factors in the planning and completion of personal projects:

- **Paper**. All participants reported using paper in one way or another during the completion of their selected project. In particular, paper was used in initial brainstorming and in to-do list management.
- **Place**. The concept of place, space and location figured in various ways in the statements of participants. Several participants expressed a desire that information relating to the project should be in the "same place" – in view or easily accessible. **People**. Participant comments pointed to the problems that can arise when other people are resistant to a project-related procedure or an organizational scheme. But more often participant comments revealed the ways that other people provide h**elp th**rough direct assistance or provision of useful information. Potentially more important, participant comments suggest that other people can be a crucial source of motivation and emotional support.

Each of these factors is discussed in turn.

Paper

Fourteen of 27 participants indicated that they used paper in some form for the initial planning of a project. Fifteen of 27 participants indicated that paper figured into their procedures for to-do list management.

The utility of paper to-do lists went beyond this, however. One participant noted that the act of writing the to-do list was more valuable to her than referring back to it later on: *"But I find that once I write it once in the planner I very rarely need to refer to the planner to remember to do things."* – TF147.

This participant also made lists of tasks completed as a way to assess and reward: *"I also make a lot of lists after the fact, to show that I accomplished things in a given day. Making the lists after I have completed the tasks gives me great satisfaction and will also remind me if I have forgotten anything. "*– TF147.

Beyond task lists, for some participants, the tangibility of paper was tied to a sense of accomplishment. As one participant, a doctoral student in communications, noted, *"Umm, paper things, I actually, paper filing is, is fairly satisfying to me. It, uh, because it gets accomplished. Umm, you can see something from it."* NB187.

Another reason for using paper laid in its ability to attract attention when posted:

"For personal items, such as birthday reminders or doctor's appointments, I use a wall calendar. ...I utilize whatever organization tool seems most appropriate for the project -- sometimes that is a flip chart or binder." GH130.

"*For personal tasks that need completing, I keep paper lists (for instance, person errands, shopping lists, etc). I usually keep these on my desk at home, or taped to my front door. For tasks that need completing at work, I keep a paper list on my desk, and also add reminders in my Outlook calendar."* FX191.

Place

The comments above suggest the importance of "place" in participants' decisions to use paper. Paper can easily be placed (on a wall, mirror, front door, etc.) to attract attention. The importance of place goes well beyond paper, however. The impact of digital place, of "knowing where to go" has also been repeatedly affirmed in studies looking at how people access digital information (Barreau and Nardi, 1995; Teevan et al. 2004). Indeed, Bergman et al. (2008) observe a preference for browsing as a means of accessing electronic documents notwithstanding recent dramatic improvements in and widespread availability of desktop search facilities. Consistent with the studies cited in the introduction, participants in the current study used folders as a way to "place" digital information. Use of folders was especially apparent for the organization of digital

documents and other files: All but two of the participants had at least one layer of subfolders under a project file folder.

Participant comments also point to the importance of place and the related concepts of visual space, location and control. For example, a participant described a problem that arose because an item was "placeless" (i.e., it was automatically placed in a temporary folder by an application).

"I'll tell a story and hopefully it'll make some sense. Umm, I, I had a, a document as an email attachment. I, I opened that document, and then I left the network and I started to travel with the document open, making changes on it, saving regularly as I was working on it. Umm, I got to my destination, I was required to hand that document in, I got back on the network, somehow I had closed the document in the mean time, and I knew it was there. I had saved it a bunch of times. It was nowhere that I could easily find. It took me about forty-five minutes of diligent searching on the computer to find it in some hidden Windows temporary folder where they had stashed it." NQ149.

Problems of control can also arise when a document is in a shared space. An administrator (FG130) had all her electronic documents on a shared file server. Midway through the interviews, her most important document was either moved or deleted without her knowledge. "Control" figured prominently into her comments in subsequent interviews.

A graduate student expressed a desire for a tool that would gather all project-related information into the same place:

"You know, something that puts all the stuff in once place instead of having all these different places for, you know, all the electronic stuff, you know, I use del.icio.us for web references, I use my Mail app for email, I use, um, Things for the project, you know, some of the project information but sort of task-coordination, you know, sort of the organization of the project. And so everything's in its own little place and it might be nice if there was some way to have that all in one place." KT199

Similarly, another participant expressed a desire for a tool that *"would help me to organize and consolidate all of the information sources that I use for the project, uh, it would help me save time on finding, uh, information that I need, it would also...help me to organize, uh, new information, uh, emails, Bookmarks, documents and whatnot, uh, incorporate into the project organization." TE200.*

Finally, a third participant described an ideal tool that*: "would allow me to link everything together for every accession. Um. … so it just would be something that unified all of the separate tools and databases that I use." KT182.*

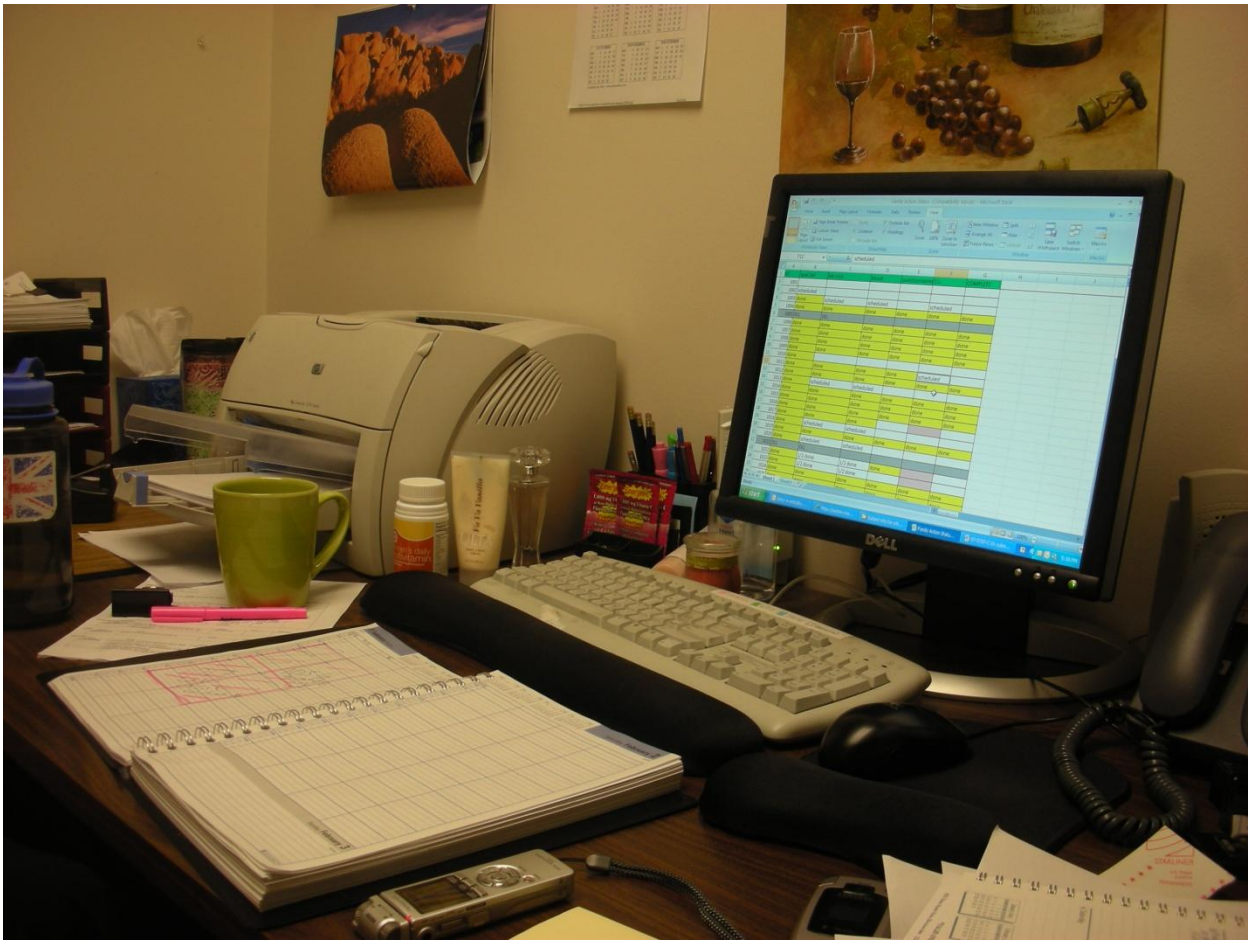


Figure 1. One participant had a scheduling project requiring the coordination of several different forms of information – paper-based and digital.

Another participant, faced with the need to work with several different forms of information, paper-based and digital (see Figure 1), expressed a desire for better use of screen space so that all the information she needed could be visible at the same time: "*I would still do the split screen in quadrants probably so I could see all the information." FL126.*

People

Other people can be a help or a hindrance in a project's completion. E. Jones et al. (2008) describe the importance of a co-adoption factor in the success or failure of a person's efforts to adopt a new system of information management. System success is more likely if other people are also using the system or are at least supporting and appreciating a person's efforts to use the system. Conversely, participants made comments like "why bother?" to suggest that a system is more likely to be abandoned if no one else knows or cares about its use.

People can also support a project through direct assistance or by providing information of direct relevance to a project. Of potentially equal importance, participant comments suggest that other people can be an important source of motivation and emotional support.

For example, people may organize their information for reasons similar to those that motivate us to straighten up our houses when guests are coming. We do so as not to look bad in the eyes of others. But we also benefit from the greater order that results.

One participant said that he spent time organizing information (for a video game he and his team are building) even though he was not sure he would really use the organization that much. When asked, why, he replies, *"I don't want to live like a goober. Cuz I get paid pretty well as a senior*

LD, and they have a couple of juniors and I just can't be, you know, perceived to be less competent than they are."

A doctoral student described her positive experiences working in the presence of other students: *"Umm, the somebody to poke at me, mostly, I mean it's that, you know, that sort of motivational tool. And, and, and some of that I get in, with a writing partner, you know, that we really hold each other to, to producing. Umm, uhh, so sort of a, and I'm getting there just because I'll have to this summer, of a, of a calendar that you have to have this done on these days or you're just not gonna get done." NB187*..

Another participant described his use of a blog both as a way of describing his project (an effort to animate a stuffed animal as a hobby) to others and also as a way of keeping track of project-related information:

"I've been using blogs to collaborate for years now, but so this was just a quick one that I originally actually set up for my girlfriend when she was doing – began to do crafty stuff, but she stopped – she never used it so I just co-opted and began just to throw stuff up there as I saw it when I was sitting bored at work or at home. So umm this is pretty much the documentation of the project so far. I, you know, use web links, I ordered the stuff…" -- FU156.

Several participants also referred to the beneficial effects of the study's interviewer on the project's progress:

"I would like to say that because I am held accountable to tell you something each week, I'm probably moving forward on this at a greater rate than I would have otherwise" – FG130. This participant even asked (only partly in jest) if the interviewer could come back and visit her from time to time as she continued to work on the project they had been discussing.

"It's just, I mean it's actually really helped me to be talking about it because it's just made me, it's made me process how I organize, so it's probably made me, it's made me more organized. And it, and it points out some of the things that are necessary for me" NB187.

"I think I've made writing a curriculum map more interesting than it is and um, [laughs] most teachers would tell you it's boring and it's kind of annoying because you don't think about." SS207 – a teacher creating a teach creating a curriculum map.

Discussion

Results point to the enduring importance of both paper and "place" in any system of supporting tools. We may never go completely paperless (Sellen & Harper, 2002) and perhaps we shouldn't try. Paper for certain uses is tough to beat. Paper and the means to write on paper with pencil or pen are nearly always at hand. Paper can be folded, torn and thrown away. Paper requires no power supply and its information won't be lost with a disk crash (though fires and floods are a different matter). Writing, sketching or doodling on paper is easy and satisfying. There is a "feel" to paper that we may never achieve with digital forms of information.

On the other hand, there is still much we can learn from paper's use with application to digital tools of information management. Participant comments provide the following takeaways for tool design:

- Support the digital equivalent of paper scraps that make it easy to record thoughts that may have nothing to do with the active application or the information currently in view (see Bernstein et al., 2008).
- Look for situations in which the greater benefit of writing thoughts down may be in the writing itself and not in the subsequent retrieval of the information. In these situations, make writing fast and easy and don't burden users with lots decisions concerning how the information should be organized for later use.
- On the other hand, there are times when information should remain in view even when the need for it has apparently passed. For example, users may want tasks to remain visible in a

list even after they have been checked as complete, as a way of assessing progress and affirming their own achievements.

Similarly, for place, the challenge is not to attempt a faithful virtualization of physical place. Rather, we need to understand which aspects of place most matter in a digital space of information. Participant comments point to the value of the following features in tool design:

- Control. Placing information in folders may give users a sense of control that tagging does not (see Civan et al., 2008). On the other hand, as the case of the missing document from the shared file space attests, this sense of control is sometimes "misplaced".
- Browsing. Users may continue to invest effort in organizing information into folders, notwithstanding the increasing availability and sophistication of tagging and search systems. In part, this may reflect an enduring preference for browsing as a stepwise, contextualized method of information access. On the other hand, we can think of many instances when we're quite happy to "jump" to the desired information. The challenge, then, may be to understand better the circumstances in which people prefer "orienteering" to "teleporting" (see Teevan et al., 2004).
- Integration. Sometimes users may literally want all relevant information to be in a single view. In other cases, though, they may simply want project-related information "nearby". Perhaps the desire is that items of information that are needed in the same context are somehow connected to each other, so that the retrieval of one item flows easily into retrieval of the remaining items. Our experience is too often the opposite. The information needed to complete a task is often scattered across email messages, web pages and documents, paper and digital, with no connection among these diverse formats and information spaces.

Of potentially greater importance than "paper" or "place" may be the factor of "people". Notwithstanding the "personal" in personal information management, participant comments make it clear that social considerations figure large in their efforts to manage their information. People may organize information for the same reasons that they tidy a messy house -- not because information organized or a house tidied is more functional (though they usually are) but rather so as not to look like a "goober" (in the eyes of teammates or guests). Similarly, notes may be written or re-written if there is intent to share the notes with others (Erickson, 1996; Marshall & Brush, 2004).

We want to avoid the bad opinions of other people. On a positive side, we seek out the company of other people. We may, for example, find ourselves monitoring our email, or a message board, or Twitter or our Facebook account even at the expense of the projects we need to complete and other things we need to do.

What if our need for social interaction could be leveraged in our efforts to manage our information? It is rare to find people who truly want to listen to us as we talk about our personal projects and our efforts to manage our personal information. But information tools – including web services and handheld devices – enable new modes of communication and a more conversational style of expression. Can these tools also support good PIM?

We have the example of the participant who blogged about his project. Whether or not anyone actually reads his blog posts, the blogging style is chatty and conversational. Are people more likely to express themselves when such a style of expression is the norm (vs. the more formal style of conference papers, to take a contrasting example)? If so, the expression can provide a context or, more literally, a text within which to weave references to project-related information – and indeed, the participant's blog posts included references to numerous project-relevant web sites.

Or consider the constructed example of someone, call her Jill, who posts a series of photos to Flickr taken from a summer vacation to Italy. She writes captions. The sequence of pictures and their captions tell a story of her summer vacation. Her travel companions comment. Other friends comment. Jill comments on these comments. As this happens the story is told in greater detail. The pictures on Jill's camera or on her hard drive are a source of guilt and foreboding ("I really

should do something with these pictures before I forget… “ “What if I lose them or delete them???!!”). The pictures on the Web set the stage, instead, for an enjoyable interaction between Jill and her friends. Jill’s motivation for this time and trouble is social. But as a by-product, the pictures are organized and annotated not just for the present but, potentially, for a future 20 or 40 years from now when Jill’s memories of the trip have faded.

Conclusion

Paper, place and people. Each is a consideration in the design of tools to support in the management of personal projects and in the management of the information needed to complete these projects. Affordances for paper and place intermingle. We write things on paper, for example, because it is so readily at hand – “where” we happen to be. There is no need to start up a digital device and click to an accepting application. “Place” as a verb gives us a sense of control and a remembrance of actions completed. We can place paper-based information so that it is in view or close at hand.

A challenge is to realize a digital facsimile of these physical world affordances and in ways that don’t also copy the many obvious disadvantages of the physical world (Russell et al., 2006). A paper document cannot, after all, appear at the same time in several places according to our need. And paper documents stay “in place” long after our need for them has passed. We call it “clutter”.

The factor of (other) people is in a class by itself. It may be tempting to place concepts such as “group” and “personal” in opposition to one another. Indeed, in many cases considerations of one trade against the other. The transactions we make to function in a group, for example, must frequently be done with some compromise to personal privacy (Karat, Brodie, & Karat, 2007). But examples described in this article point to another circumstance wherein one supports the other.

We might call it the “toothbrushing” effect. Our motivations may be immediate and social. We take extra steps to document and organize in order to make contact with our friends and colleagues or for the sake of appearances (i.e., so as not to have bad breath or to appear like “a goober”). But the benefits we realize through our efforts can also be lasting and deeply personal.

Acknowledgements

This material is based on work supported by the National Science Foundation (#0534386).

References

Barreau, D. K. and Nardi, B. (1995). "Finding and reminding: file organization from the desktop." SIGCHI Bulletin 27(3): 7.

Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D. G., & Ducheneaut, N. (2004, April 24-29). What a to-do: Studies of task management towards the design of a personal task list manager. Paper presented at the Conference on Human Factors in Computing Systems (CHI 2004), Vienna, Austria.

Bellotti, V., Ducheneaut, N., Howard, M., & Smith, I. (2003). Taking email to task: The design and evaluation of a task management centered email tool. Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2003), Ft. Lauderdale, FL.

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2006). The project fragmentation problem in personal information management, Proceedings of the SIGCHI conference on Human Factors in computing systems. Montreal, Quebec, Canada: ACM Press.

Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., & Whittaker, S. (2008). "Improved search engines and navigation preference in personal information management." ACM Trans. Inf. Syst. 26(4): 1-24.

Bernstein, M., Van Kleek, M., Karger, D., & Schraefel, M. C.. (2008). "Information scraps: How and why information eludes our personal information management tools. ." ACM Trans. Inf. Syst. 26(4): 1-46.

Boardman, R., & Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out" A cross-tool study of personal information management. Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004), Vienna, Austria.

Civan, A., Jones, W., Klasnja, P. & Bruce, H. (2008). Better to Organize Personal Information by Folders Or by Tags?: The Devil Is in the Details. 68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2008), Columbus, OH.

Czerwinski, M., Horvitz, E. & Wihite, S. (2004). A diary study of task switching and interruptions. ACM SIGCHI Conference on Human factors in Computing Systems (CHI 2004), Vienna, Austria, ACM Press.

Erickson, T. (1996). The design and long-term use of a personal electronic notebook: a reflective analysis. In Proceedings of the SIGCHI conference on Human factors in computing systems: common ground (pp. 11-18). Vancouver, British Columbia, Canada: ACM.

Gwizdka, J. (2002). Reinventing the inbox : supporting the management of pending tasks in email. Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems, Doctoral Consortium (CHI 2002), Minneapolis, MN.

Jones, E., Bruce, H., Klasnja, P. & Jones, W. (2008). “I Give Up!" Five Factors that Contribute to the Abandonment of Information Management Strategies. 68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2008). Columbus, OH.

Jones, W. (2007). Keeping Found Things Found: The Study and Practice of Personal Information Management. San Francisco, CA, Morgan Kaufmann Publishers.

Jones, W., Phuwanartnurak, A. J., Gill, R., & Bruce, H. (2005, April 2-7). Don't take my folders away! Organizing personal information to get things done. Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2005), Portland, OR.

Karat, C., Brodie, C., & Karat, J. (2007). Management of Personal Information Disclosure: The Interdependence of Privacy, Security and Trust. In Personal Information Management. Edited by W. Jones & J. Teevan. Seattle: University of Washington Press.

Malone, T. W. (1983). How do people organize their desks: implications for the design of office information-systems. ACM Transactions on Office Information Systems, 1(1), 99-112.

Marshall, C. C., & Brush, A. J. B. (2004). Exploring the relationship between personal and public annotations. In Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries (pp. 349-357). Tuscon, AZ, USA: ACM.

Mumford, M. D., Schultz, R. A., & Van Doorn, J. R. (2001). Performance in planning: Processes, requirements and errors. Review of General Psychology, 5(3), 213-240.

Russell, D. M., Slaney, M., Qu, Y., & Houston, M. (2006). Being Literate with Large Document Collections: Observational Studies and Cost Structure Tradeoffs. In HICSS .

Sellen, A. J. and R. H. R. Harper (2002). The myth of the paperless office. Cambridge, MA, MIT Press.

Teevan, J., Alvarado, C., Ackerman, M., & Karger, D. (2004). The perfect search engine Is not enough: A study of orienteering behavior in directed search. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004), Vienna, Austria.

Wolverton, M. (1999). Task-based information management. ACM Computing Surveys (CSUR) archive, 31(2es).

Visualising PIM Behaviour with Markov Chains

David Elsweiler
Department Computer Science 8 (AI), University of Erlangen
david.elsweiler@i8.uni-erlangen.de

Martin Hacker
Department Computer Science 8 (AI), University of Erlangen
martin.hacker@i8.uni-erlangen.de

Stefan Mandl
Department Computer Science 8 (AI), University of Erlangen
stefan.mandl@i8.uni-erlangen.de

ABSTRACT

This paper presents our initial efforts at visualising personal information behaviour using Markov Chains. We describe a laboratory-based study of email re-finding and use Markov Chains, created from captured user interactions, as a means of understanding the behaviour exhibited. The models we generate not only provide an excellent overview of how the participants interacted with the experimental interface, but, by forcing the experimenters to ask questions they would not normally ask in order to comprehend the models, they also offer a starting point from which a fuller understanding of the exhibited behaviour can be attained. We illustrate this through examples, discuss the advantages and limitations of the approach and outline how we will expand on the work in future research.

1. INTRODUCTION

A key challenge for PIM researchers lies with evaluation. Few techniques exist to help understand how people use PIM tools and, consequently, very few of the many prototypes that have been designed have actually been evaluated. This lack of tool evaluation has been repeatedly identified as a factor restricting progress in the field, e.g. [2, 7]. By studying how people use PIM tools we can understand what interactive support people need when re-finding, evaluate the effectiveness of existing tools and inform the design of more useful tools for managing and re-finding information.

In this paper we present a novel evaluation method, based on the visual analysis of statistical models derived from interaction logs. Although we are still at an early stage with this work and have only used the method on a small dataset, our findings so far have been very positive and suggest that this approach may be helpful in gaining an understanding of PIM behaviour not possible with existing techniques alone. Here, we explain the approach, illustrate the advantages and limitations using examples collected from a laboratory-based study of email re-finding, and continue to outline our plans to develop the approach in future work.

2. BACKGROUND LITERATURE

The main method for PIM evaluations is to use laboratory-based user studies where users are observed in controlled environments. Such studies can provide an understanding of participants' re-finding strategies, such as the teleporting and orienteering strategies observed in [10] and the discovery that people generally prefer spatial browsing over keyword search [1]. Lab-based user studies can also be used to verify the benefits of particular tools, e.g. [9, 8]. However, there are also limitations to such studies. First, they are performed in artificially created environments with the presence of an experimenter, both of which are likely to impact on the participants' behaviour. Second, when any more than a handful of participants are observed over long time periods it becomes difficult to establish fine grain patterns in behaviour. Further, user studies rely heavily on experimenter observations so the findings are often criticised for being anecdotal and open to subjective bias.

An alternative method to laboratory approaches is to use log-file analysis techniques (LFA) to learn about user behaviour in naturalistic conditions out-with the control of the experimenter [5, 3]. LFA examines the quantitative aspects of user behaviour, including the nature of submitted queries and the properties of items accessed. This is an important technique as it allows the capture of a large quantity of data relating to how users behave with systems without the expense and distracting influence of an observer. The data are also less susceptible to subjective bias. Nevertheless, as the captured data show nothing about what the user is trying to achieve or the tasks that they are performing, it is difficult to make any concrete statements about the reasons for the behaviour depicted in the logs.

In this paper, we propose a method that we believe assists with many of the limitations outlined above and has the potential to formally combine the findings from both kinds of study. By modelling the user's interactions with a system as a statistical process (we use Markov chains), we show that users' behaviour can be visualised in an intuitive way, allowing the experimenter to analyse behaviour retrospectively.

A Markov model is a discrete-time stochastic process which describes the state of a system at successive points in time. Markov Modelling (MM) has been applied in many domains for many purposes, e.g. speech, handwriting and gesture recognition. Modelling techniques have been applied to search behaviour before, particularly in the field of IR. For example, models have been used to improve retrieval algorithms based on prior behaviour e.g [11]. Other models have looked at trying to predict user behaviour such as queries they might

apply or results they may click on e.g. [4]. However, using such models to evaluate personal information management behaviour, particularly visually as we do here, is completely novel. We chose to start our work using Markov chains rather than more complicated processes, precisely because they are simple and we believed more suitable for the purpose of visualisation.

3. CONSTRUCTING THE MODELS

We examined the feasibility of the MM approach by using data collected from a user study investigating email re-finding behaviour. 21 participants consisting of a mix of undergraduate and postgraduate students, as well as research and academic staff from the University of Strathclyde each performed 9 re-finding tasks generated according to the method suggested in [7]. Each participant performed 3 tasks on each of the 3 experimental systems. However, for brevity and to simplify the explanation of the MM technique, in this paper we focus purely on one of the experimental systems that will be familiar to all of our readers – a folder-based email client[1]. The experiment included participants who had different quantities of emails (mean = 1938 , sd =2911), used email for different purposes, and who employed different filing strategies.[2]

We created the models by mapping the possible ways that a user could interact with the system to a set of states, using the interaction log data to count the number of times a user moved from one state to another and using the counts to calculate the transition probabilities in the model. We chose states to represent sorting the displayed emails by various attributes, opening a folder, and selecting an email to view its content. We also included a start and two end states – task completed and task abandoned. We present different models in Figures 1, 2, 3 and 4.

One of the advantages we foresaw with this approach was the ability to determine teleporting and orienteering behaviours [10] at a glance. We expected teleporting behaviours to be represented by few states and have high transition probabilities between the states and orienteering behaviours to have many states and low transition probabilities. To test this we derived models based on the interactions for 2 tasks for which we had noted these behaviours during the evaluation (Figures 1 and 2). You can tell that Figure 1 depicts teleporting behaviour, with the user going straight to the messages in the inbox, clicking in total 4 messages, before finding what he needed. This is contrasted with the behaviour in Figure 2, where the user was looking for clues in the messages. He sorted by sender, date and subject and selected 22 emails during the task. This is behaviour indicative of an orienteering strategy.

Figure 3 presents a model generated for all of the tasks performed on the folder-based system. This model provides a good overview of how the participants behaved with the folder-based interface. It shows, for example, that when the participants tended to start their search by sorting, 'sender'

[1]The interface was based on the Mozilla Thunderbird interface (http://www.mozilla.org)

[2]Due to space restrictions we are only able to provide minimal details regarding the experimental design. However, full details of the tasks and how they were created, properties of the participants and how the tasks and systems were rotated to create a balanced experimental design can be found in [6].

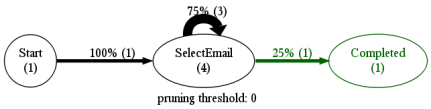


Figure 1: A model for user-id 2, task 3 – teleporting strategy

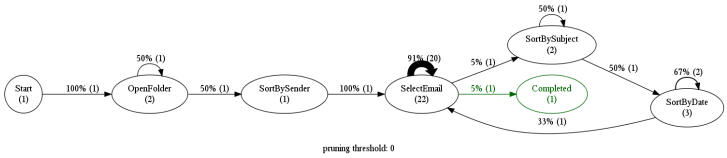


Figure 2: A model for user-id 25, task 1 – orienteering strategy

was the most frequently used attribute to sort on (~35% of all tasks), while 'subject' was used least often (~11% of all tasks). Considering emails were by default ordered by date, sorting by date was clicked on surprisingly regularly as a first interaction (~21% of all tasks). Although, if users were searching for older mails it makes sense for them to have reversed the order. In ~14% of tasks, the participants chose to open a folder as their first interaction. It seems, however, that participants weren't always sure which folder to search in as ~63% of folder openings were followed by opening another folder. The model also shows that the emails within folders were regularly (at least 13% of the time) sorted by sender. The third major strategy used by participants, after folders and sorting, was to look directly at emails in the inbox. This they did in ~17% of the tasks.

From Figure 3, we see that 'SelectEmail' is a 'hub' state with many in-links, but few out-links, the main out-link being to 'Completed'. 'SelectEmail' also has a very high percentage of looping transitions (~91% of email selects were followed by another). This suggests that after choosing to examine one email, the participants tended to continue to examine other emails until they found what they required or abandoned the task. Combining this observation with the fact that folders and sorting were used mostly at the start of tasks (with the exception of the start state, neither the open folder state nor any of the sorting states had many (if any) in-links), means that the model depicts an overall pattern of behaviour where the participants, firstly, narrowed the search space using sorting, folders or a mixture of both, and then followed this by examining the remaining emails. This is confirmed by following the paths in the model with transitions with the largest percentages. There are, however, two 'reverse' transitions ('SelectEmail' to 'OpenFolder', which represents 10 interactions and 'SelectEmail' to 'SortBySender', which represents 8 interactions). These transitions go against the flow of the narrowing and checking pattern described above, perhaps indicating a change of strategy mid-task. Thus, the main trend in the model is short, direct paths between start and end states, with very little interaction between the states, e.g. the participants did not transition between sorting states. However, the 'reverse transitions' show that the participants didn't always behave in this way. Further examination of the 'reverse transitions' revealed that all but 2 of the 18 interactions came

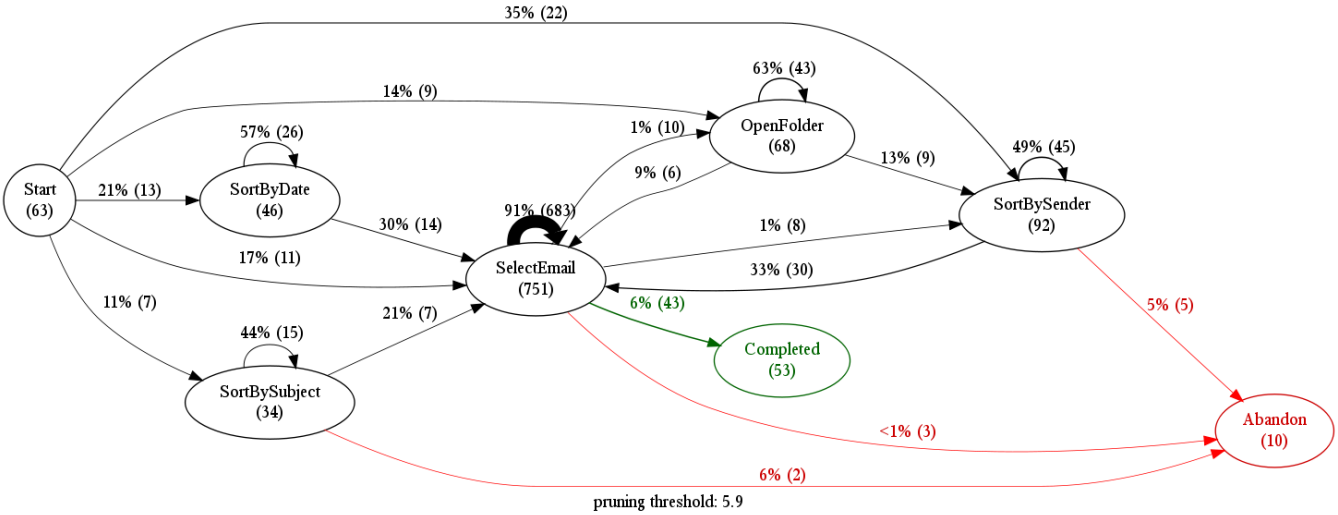


Figure 3: A model generated for all tasks performed on the folder-based system (# tasks = 63). The edges represent the frequency of choosing the target node as next action, the bracketed figures on the edges represent the total number of interactions between these states. The line thickness is also an indicator of interaction frequency with thicker lines representing more interactions. The bracketed figures within the states represent the total number of in-links to that state. We have removed edges representing <5.9 transactions with the exception of the transitions to 'abandon', which we felt helped the reader's understanding of the model and the behaviour it represents.

from sequences in which folders had previously been opened. This suggests that rather than changing strategy to folders or 'SortBySender' mid-search, it seems that the participants needed several attempts to find the correct folder and sometimes required to sort messages in folders by sender to detect this, even after looking at some of the messages in the folder.

The short, direct paths, depicted in Figure 1, are in contrast to Figure 4, which presents a model for the 10 of the 63 tasks that the participants failed to complete. Whereas the first model shows relatively little interaction between the states, the model for failed tasks shows much more interaction between states, with the transitions having lower percentages attached.[3] High interaction with low probabilites suggests longer interaction sequences. This is corroborated by the data. Completed tasks had on average 14.3 interactions, while incomplete tasks had on average 23.3. This is to be expected with incomplete tasks, as when an initial strategy failed the participants would have tended to try other tactics.

The 10 tasks used to generate this second model probably do not provide enough data to establish if the participants utilised different strategies when attempting these failed tasks, i.e. use sorting or folders more often. However, the transitions from the start state seem to be forming a similar pattern to those in Figure 3, suggesting that similar strategies may have been employed. Interestingly, Figure 4 shows that in half of the failed tasks, the participants used a sort by sender in a last attempt to find the required information.

[3]These transitions of course featured in the first model, however because our pruning algorithm removed transactions representing small numbers of interactions these were removed to increase the readability of the model and convey the main trends.

4. SUMMARY, FUTURE WORK AND CONCLUSIONS

The examples we have provided demonstrate how visualising users' interactions as Markov chains can allow experimenters to understand how users behave with a system by offering the opportunity to analyse behaviour visually. We were able to identify several aspects of behaviour, including how sorting and folders were used and recognise teleporting and orienteering behaviours. While visualising interactions in this way allows complicated datasets to be understood, one limitation is that this understanding cannot be gained simply by glancing at the models – they need to be studied in depth and this process requires no little creativity on the part of the experimenter. Nevertheless, the positive aspect is that the process of analysing the models forces the experimenter to ask questions he would not otherwise ask, leading to a better overall understanding of what is going on. A good example of this was examining the 'reverse transitions' as described above. Analysing the models can also lead to the generation of new research questions e.g. why was 'SortBySubject' a common last resort in failed tasks?

We also showed that it is possible to use the Markov chains as a means to visually compare behaviour in different situations. The example we provided compared all tasks with incomplete tasks, but the approach could be used, for example, to compare the behaviour of different types of user (e.g. experienced vs. novice users, filers vs. pilers, older vs. younger participants etc.), behaviour for different types of task (e.g. looking for older or newer information), or for different types of systems (e.g. browse-based vs. search-based). We are currently building these models and looking at ways in which interactions can be abstracted so that different systems can be compared. We are also exploring methods of mathematically comparing models that can be used to au-

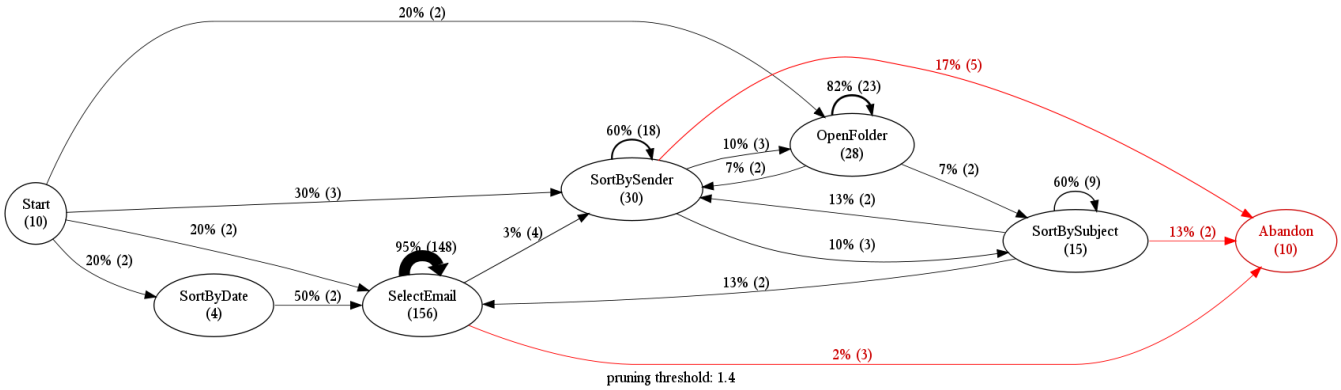


Figure 4: A model generated from all failed tasks on the folder-based system (# tasks = 10). To ease readability we have removed edges representing <1.4 transitions

tomatically detect the kind of features we observed when visually analysing the models (e.g. hub states, high or low interaction between states, probable paths etc.). This would help researchers identify behavioural changes and corroborate any observations made, as well as lessen the reliance on the experimenter's creativity when analysing the models.

We must mention some dangers with analysing data in this way. Pruning, for example, makes it easier to analyse the models and spot patterns, but it can be misleading. It is extremely important to verify hypotheses generated from the pruned model on the original un-pruned version. Another danger is that looping transitions can lead to misunderstandings. For example, although the transition from 'OpenFolder' to 'SortBySender' in Figure 3 has an associated probability of ~13%, actually, if you discount the looping transitions, this percentage would be closer to 40%. In other words, when the participants were satisfied that they had found the correct folder, they tended to sort by sender very often. This is not very clear from the model as it is somewhat disguised by the looping transition. It is important that experimenters are aware of such properties.

A limitation of Markov chains as we have presented them here is that they have no means to model temporal information, which has been shown to be useful in PIM and search behaviour [8, 4]. We plan to extend our work using different kinds of models to investigate the usefulness of temporal information in this context. However, first we plan to exhaust the potential with simple chains.

Although we have demonstrated the approach using data collected from a laboratory-based study, we believe it will offer even greater potential in the context of naturalistic studies. Naturalistic studies provide far larger quantities of data to work with, which would offer greater scope for patterns to emerge in the data and for mathematical analyses. We are in the process of planning a large scale log-based, naturalistic study of email behaviour and aim to use the MM approach to help analyse the data. A further benefit of the MM approach in this context is that it may allow the findings of the naturalistic study to be formally triangulated with those derived from lab-based studies. If we can find ways to mathematically compare sequences of interactions to models constructed from behaviour observed by experimenters it would go a long way to overcoming many of the limitations described in Section 1.

5. ACKNOWLEDGMENTS

This work was sponsered by the Alexander von Humboldt foundation, Germany (http://www.humboldt-foundation.de/)

6. REFERENCES

[1] D. K. Barreau and B. Nardi, *Finding and reminding: File organization from the desktop*, ACM SIGCHI Bulletin **27** (1995), no. 3, 39–43.

[2] R. Boardman and M. A. Sasse, *"stuff goes into the computer and doesn't come out": a cross-tool study of personal information management*, Proc. SIGCHI, 2004, pp. 583–590.

[3] E. Cutrell, D. Robbins, S. Dumais, and R. Sarin, *Fast, flexible filtering with phlat*, Proc SIGCHI, 2006, pp. 261–270.

[4] D. Downey, S. T. Dumais, and E. Horvitz, *Models of searching and browsing: Languages, studies, and application*, IJCAI, 2007, pp. 2740–2747.

[5] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D.C. Robbins, *Stuff i've seen: a system for personal information retrieval and re-use*, Proc. SIGIR, 2003, pp. 72–79.

[6] D. Elsweiler, M. Baillie, and I. Ruthven, *Exploring memory in email refinding*, ACM Trans. Inf. Syst. **26** (2008), no. 4, 1–36.

[7] D. Elsweiler and I. Ruthven, *Towards task-based personal information management evaluations*, Proc. SIGIR, ACM Press, 2007, pp. 23–30.

[8] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz, *Milestones in time: The value of landmarks in retrieving information from personal stores.*, Proc. INTERACT, 2003, pp. 184–191.

[9] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. van Dantzich, *Data mountain: using spatial memory for document management*, Proc. UIST, 1998, pp. 153–162.

[10] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger, *The perfect search engine is not enough: a study of orienteering behavior in directed search*, Proc. SIGCHI, 2004, pp. 415–422.

[11] J. Teevan, S. T. Dumais, and E. Horvitz., *Personalizing search via automated analysis of interests and activities*, Proc. of SIGIR, 2005.

Mental Workload at Transitions between Multiple Devices in Personal Information Management

Manas Tungare
Virginia Tech
manas@tungare.name

Manuel A. Pérez-Quiñones
Virginia Tech
perez@cs.vt.edu

ABSTRACT

Knowledge workers increasingly use multiple devices such as desktop computers, laptops, cell phones, and PDAs for personal information management (PIM) tasks. This paper presents the results of a study that examined users' mental workload in this context, especially when transitioning tasks from one device to another. In a preliminary survey of 220 knowledge workers, users reported high frustration with current devices' support for task migration, e.g. accessing files from multiple machines. To investigate further, we conducted a controlled experiment with 18 participants. While they performed PIM tasks, we measured their mental workload using subjective measures and physiological measures. Some systems provide support for transitioning users' work between devices, or for using multiple devices together; we explored the impact of such support on mental workload and task performance. Participants performed three tasks (Files, Calendar, Contacts) with two treatment conditions each (lower and higher support for migrating tasks between devices.)

Workload measures obtained using the subjective NASA TLX scale were able to discriminate between tasks, but not between the two conditions in each task. Task-Evoked Pupillary Response, a continuous measure, was sensitive to changes within each task. For the Files task, a significant increase in workload was noted in the steps before and after task migration. Participants entered events faster into paper calendars than into an electronic calendar, though there was no observable difference in workload. For the Contacts task, time-on-task was equal, but mental workload was higher when no synchronization support was available between their cell phone and their laptop. Little to no correlation was observed between task performance and both workload measures, except in isolated instances. This suggests that neither task performance metrics nor workload assessments alone offer a complete picture of device usability in multi-device personal information ecosystems. Traditional usability metrics that focus on efficiency and effectiveness are necessary, but not sufficient, to evaluate such designs. Given participants' varying subjective perceptions of these systems and differences in task-evoked pupillary response, aspects of hot cognition such as emotion, pleasure, and likability show promise as important parameters in the evaluation of PIM systems.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User interfaces—*Evaluation/ methodology*

Author Keywords

Personal Information Management, Mental Workload, Multiple Devices

INTRODUCTION

As we amass vast quantities of personal information, managing it has become an increasingly complex endeavor. The emergence of multiple information devices and services such as desktops, laptops, cell phones, PDAs and cloud computing adds a level of complexity beyond simply the use of a single computer. It is common for a lot of people to carry a laptop computer or a cell phone as they go about their everyday business, outside the usual contexts of an office or a home [7, 35], and to expect productive work output when mobile. However, the current state-of-the-art in information management solutions sends these users into a frenzy trying to locate the most current version of their slide shows, the documents they sent around for review, and the phone number of the person they need to call. In traditional single terminal computer systems, the majority of a user's attentional and cognitive resources are focused on the terminal while performing a specific task. However, in an environment where multiple devices require intermittent attention and present useful information at unexpected times, the user is subjected to different mental workload.

In previous work, we conducted a survey study [35] to understand the use of multiple devices in personal information and identify common tasks, activities, devices, patterns, device affinities, and problems in their use. Many findings were as expected: that users preferred laptop computers over desktops; several users owned and regularly used more than two computers, plus a cell phone, a digital camera, etc. However, a surprisingly high number of users reported chronic problems in using multiple devices together for managing their tasks. Synchronization issues between information collections on two or more machines were cited as the most common problem. Sprouting from this investigation, we decided to examine this problem deeper—whether the level of

PIM 2009, November 7 – 8, 2009, Vancouver, BC, Canada.

system support for such basic processes as information migration affects user performance and workload.

MOTIVATION

Content analysis of the survey responses revealed that many of the issues that users faced could be studied and understood within the framework of mental workload. For example, factors such as frustration level, mental demand and perceived ratings of own performance are all dimensions of the NASA TLX scale. It has been shown that an operator's task performance is inversely correlated with high levels of mental workload [24]. Prior work in mental workload measurement has established that physiological measures such as changes in pupillary diameter (known as Task-Evoked Pupillary Response [3]) can be used to estimate mental workload. Such continuous measures of mental workload can help locate sub-tasks of high task difficulty. Iqbal et al. [15] demonstrated that within a single task, mental workload decreases at sub-task boundaries. A fundamental goal of our research was to examine if their finding still applies when the second sub-task is performed on a different device than the first. Our contrary hypothesis was that mental workload rises just before the moment of transition, and returns to its normal level a short duration after the transition is complete.

The specific research questions were as follows:

RQ1. Mental Workload and Support for Multiple Devices

What is the impact of (1) different tasks and (2) different levels of system support for migrating information, on the workload imposed on a user? Certain tasks require more attentional resources than others, and may result in increased mental workload, while certain other tasks may be straightforward and may require fewer mental resources. What is the variability in the subjective assessment of mental workload for these tasks?

Systems differ in the level of support they provide for pausing a task on one device, and resuming it on another [31]. A goal of our research was to examine if mental workload at the point of transition was correlated with the level of system support available for the sub-task of transitioning. Miyata and Norman hypothesized [22] and Iqbal et al. [15] demonstrated that within a single task, mental workload decreases at sub-task boundaries. But when a sub-task is performed on a different device than the first, what are the changes in mental workload?

RQ2. Operator Performance & Levels of System Support

How is user performance impacted at differing levels of system support for performing tasks across multiple devices? To evaluate this, we simulated two conditions for each task; in each case, the L0 condition offered a lower level of support for migrating tasks between devices than the L1 condition. How does operator performance in condition L0 compare to that in condition L1? Several measures of task performance were used, on a per-task basis.

RQ3. Operator Performance and Mental Workload

Are subjective assessments of mental workload an accurate indicator of operator performance in this domain? Are both, subjective measures of workload (NASA TLX) and the physiological measure (pupil radius), sensitive to workload in PIM tasks? It is clear that workload does not stay constant during a task, but varies constantly. What are the types of changes that can be observed in workload during the execution of a task? How do the two measures of workload each correlate with task performance? Mental workload has been shown to be negatively correlated with several of these metrics in other domains [24, 1, 4]. Does the same (or a similar) relationship hold between mental workload and task performance in the PIM domain?

RELATED PRIOR WORK

Personal Information Management

This work overlaps three broad areas: Personal Information Management (PIM), Multi-Device Interfaces and Mental Workload Measurement. Studies in PIM include investigations of individual collections such as files [2], calendars [18, 27, 34], contacts [38], email [39, 12], bookmarks, etc. as well as users' information management practices [21], using a range of investigation techniques [32]. Issues such as information overload and information fragmentation [5] have also received attention. However, the issue of information fragmentation across multiple devices [17] looms larger as mainstream users increasingly have started to use portable devices such as cell phones, portable digital assistants (PDAs) and laptop computers for PIM.

PIM using Multiple Devices

In prior work [30], we explored the issues that arise in multi-device interfaces, especially when several devices are used together to perform a single task. The flow of information among a user's multiple devices has been likened to a biological ecosystem [28]. Several concepts in Personal Information Ecosystems are analogues of related concepts from biological ecosystems, and the metaphor helps construct a meaningful information flow among devices. While task migration is handled at the interface level, seamless data migration requires system support. The Syncables framework [36, 37] was developed in response to the need for being able to access data from any of a user's devices without extraneous task steps. It has been recognized widely that the mobile context is fundamentally different from the stationary context [25], and design must therefore account for the differences [29]. Dourish [9] refers to situated interaction as *"embodied interaction"*, and outlines several principles that designers must take into account for technology that, by its very nature, must co-exist in the environment that users use it in.

Holistic Usability in Multi-Device Environments

The origins of usability and human factors can be traced back to factories and environments where users performed specific duties at specific times. The goal of human factors specialists was to optimize operator performance and the fit between human and machine. Modern developments in the science of cognition have examined the relationship of the user in complex computing environments, and place greater

emphases on the situational aspects of human-computer interactions. Distributed cognition theory [14] extends the reach of what is considered cognitive beyond the individual to encompass interactions between people and with resources and materials in the environment. In multi-device computing environments, it is worthwhile to analyze the system as an integrated whole whose purpose is to assist the user in satisfying her information needs. Other recent theories such as Embodied Interaction [9] also support the notion that technology and practice are closely intertwined; they co-exist and co-evolve.

Hot Cognition Aspects in the Evaluation of PIM

Norman [23] argues that emotion plays a central role in our interaction and appreciation of the computing devices we use. But classic usability metrics fail to account for subjective factors such as emotional appeal, frustration, and likability. All these point to the necessity of bringing hot cognition aspects into the evaluation process: Jordan [16] advocates designing for pleasurability of the user, stating a hierarchy of needs for a computing system: functionality as the most basic, then usability, and finally, pleasure. Thus, usability is necessary but not sufficient to guarantee an optimal user experience. Kelly et al. [19] identify a shortcoming in PIM studies as well; quality of life measures (e.g. [11]) have received received little attention in PIM evaluations.

Mental Workload Assessment

Mental workload is defined as *"that portion of operator information processing capacity or resources that is actually required to meet system demands"* [24, 10]. It is task-specific and operator-specific (i.e., person-specific); the same task may evoke different levels of workload in different individuals. *Task complexity* is related to the demands placed on an operator by a task, and is considered operator-independent, whereas *task difficulty* is an operator-dependent measure of perceived effort and resources required to attain task goals [6]. Mental workload is considered an important, practically relevant, and measurable entity [13]. Several ways of measuring mental workload are used in practice: Performance-based Assessment Techniques; Subjective Workload Assessment Techniques, e.g. NASA Task Load Index (TLX) [13]; and Physiological Workload Assessment Techniques, e.g. task-evoked pupillary response [3, 20].

METHODOLOGY

Representative Tasks

From a content analysis of survey data, the following emerged as the most common tasks:

- **File Synchronization.** One of the most commonly reported frustrating tasks that emerged was synchronizing data (this echoes findings by others [7]). Users' responses to this question elicited a long list of problems and issues that they often encountered.

  Participants were asked to play the role of a consultant who worked with several clients, either at their own office on the desktop computer, or at one of the clients' sites, using their laptop. On each machine, an exact replica of a file system was provided, either deeply-nested, moderately-nested, or flat, based on participant preferences. Instructions were provided, one at a time, asking them to make certain specific edits to files. Mid-way, they were asked to wrap up their work and travel to a client site. In L0, they were provided USB drives and web-based email; in L1, a network drive allowed remote access to files.

- **Managing Calendars.** One of users' main motivations for using more than one device was to be able to access their calendar information when away from their desks. The use of paper calendars is widespread, even despite the availability of online calendars. It is not clear which of these methods is easier; almost equal numbers of participants reported preferring one over the other for several reasons [34].

  At the start of the calendar task, users were provided either two paper calendars labeled 'Home' and 'Work' (L0) or an online calendar program with two overlapping calendars in it, also labeled 'Home' and 'Work' (L1). During the task, participants were presented instructions that required them to consult or update their calendars. Different types of events included tentative, rescheduled, group events, events that required preparation, and conflicting events (details in [33]).

- **Contact Management.** Contact management on phones was identified as a frustrating task due to deficiencies in the phone interface, or a lack of features in the specific software they used, both on the computer as well as on the phone.

  Participants were described a scenario where they were a researcher attending a conference, and met several old acquaintances and made new contacts. They were allowed to access their laptop at some times, and their phone at other times, and both at some other times.

Experiment Design

In this experiment, we were interested in the impact of two factors—task, and level of support—on workload in participants. Since individual differences in work practices, task performance, and assessments of workload would display high variability across participants, a within-subjects design was used. Each participant was assigned to each cell, making this a complete block design (at 3×2 treatment levels). Each experimental task identified above was assigned to users to be performed in one of two sessions separated by at least two weeks, in order to minimize the learning effects associated with the first session. The order of tasks was completely counterbalanced. Figure 1 shows a graphical overview of the entire experimental setup.

Pilot studies were conducted with five participants. Training was provided in the form of demonstration videos, hands-on time, and required completion of a set of 10 familiarization tasks. Sample size estimation conducted after 6 participants had performed the experiment revealed that a medium to large effect was evident according to Cohen's d. The sample size chosen was 21, higher than that required to detect such an effect with a power of 0.8 at the α=0.05 level of significance for all three tasks, and to allow for experimental

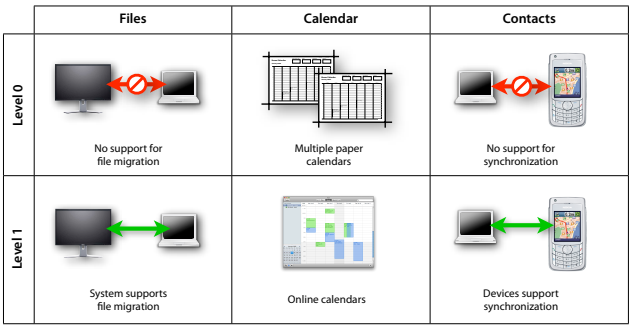


Figure 1. An overview of experimental tasks

mortality (since it was conducted in two sessions.) 3 readings had to be discarded due to scheduling conflicts, data collection issues, and a perceived risk of potential experimenter bias, respectively.

Participants were presented with a desktop computer, a laptop and a cell phone. Between the two computers, instructions were presented on a large 30-inch display. A custom web application was written to present instructions to the participants, one at a time. When the display changed from one instruction to the next, the app recorded the timestamp. This was later used to analyze sub-task-level changes in physiological measures of mental workload. Participants were requested to provide a subjective estimate of workload using the NASA TLX scale after each task. Pupil radius measurement was performed using a mobile head-mounted eye tracker. Illumination was carefully controlled to be the same for all participants and at all times of the day. The experiment was conducted in a closed room, and no external light was allowed to enter the room. The raw pupil data was extremely noisy and needed to be smoothened to isolate the signal from the noise, using the Savitzky-Golay filter. After smoothing, pupil radius data was adjusted to account for individual differences in pupil size. A baseline reading for pupil radius was obtained for each participant from the first 5 seconds of pupil activity data. During the first five seconds, participants were not assigned any specific task or provided any instructions to read, and was considered a period of minimal task-induced workload.

RESULTS

Results for Research Question 1

Research Question 1 explores the impact of (1) different tasks and (2) different levels of system support for migrating information, on the workload imposed on a user.

Subjective Metrics using NASA TLX

From an ANOVA of NASA TLX scores, Task was seen to have a main effect on Overall Workload (OW) ($F_{(2,102)}$=4.75; p=0.011). Post hoc analysis using Tukey's HSD showed that the Contacts task imposed significantly lower overall workload than the Files task (p=0.0074). Level of support for performing tasks across multiple devices (L0 vs L1) did not influence Overall Workload and there were no significant interactions.

This suggests that while NASA TLX ratings are able to discriminate between different tasks in the personal information management domain, the scale is not sensitive enough to detect differences in performing a task using two or more techniques. One reason for this could be that NASA TLX, being a subjective measure, can only be administered at the end of a task. It thus fails to capture variation in workload within a task, and provides only an aggregate per-task measure of workload.

Mean (SD)	Files	Calendar	Contacts
L0	41.11 (20.85)	36 (18.80)	30.89 (16.65)
L1	38.61 (18.92)	31.17 (18.91)	22.89 (11.49)

Table 1. Means (SDs) of Overall Workload ratings

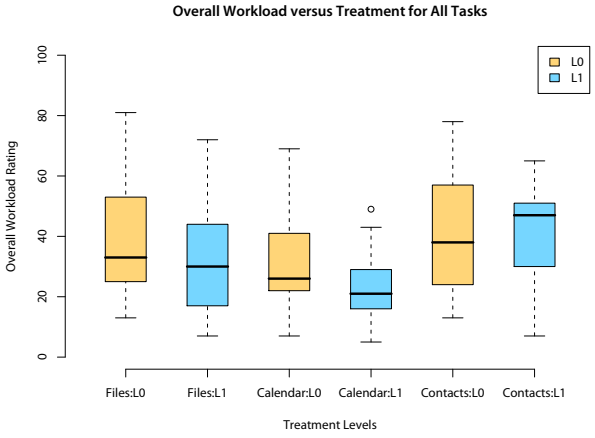


Figure 2. Overall Workload across Treatments

Similar effects were seen for three individual dimensions of the NASA TLX scale as well:

- **Mental Demand.** Task had a main effect on Mental Demand (MD) ($F_{(2,102)}$=6.69; p=0.0019). Post hoc analysis results for Mental Demand using Tukey's HSD revealed that the Files task imposed significantly higher Mental Demand than the Contacts task (p=0.0024), similar to the effect seen in case of Overall Workload.
- **Frustration.** Task had a main effect on subjective reports of frustration provided by participants ($F_{(2,102)}$=6.57; p=0.0021). Participants noted significantly higher frustration ratings for the Files task as compared to the Contacts task (p=0.0014, using Tukey's HSD). Differences among the other two pairs (Files-Calendar and Calendar-Contacts) were not significant.
- **Own (Perceived) Performance.** In this dimension, lower numbers indicate better performance. Participants rated their Own Performance differently for the three task conditions ($F_{(2,102)}$=3.37; p=0.038).

Task-Evoked Pupillary Response

For the Contacts task, significant differences were found for each step between the two levels of system support in task

migration (synced versus unsynced conditions.) Graph 3 illustrates the means (SDs) and *p*-values for each step.

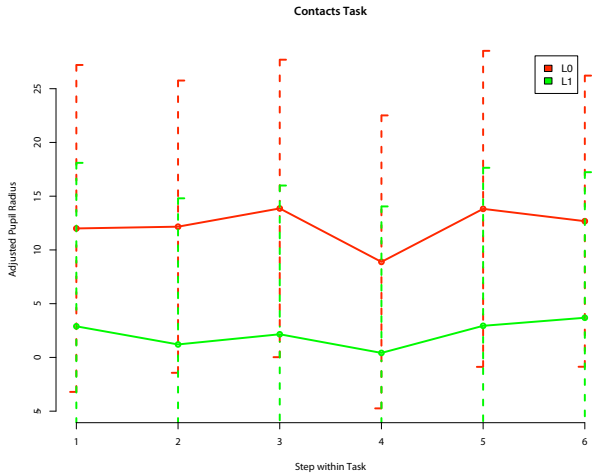


Figure 3. Adjusted pupil radius for each step of the Contacts task.

Differences in TEPR Between Steps in the Same Task

In the Files task, Level 0 (where participants used USB drives or email-to-self), significant differences were noted in the workload for the steps before and after the migration step ($F_{(8,136)}$=7.8835; p=1.12×10^{-8} using Tukey's HSD). This suggests that there is a distinct increase in workload before and after the migration step, when there is a lack of support for task migration. It is interesting to note that no significant differences were found in the L1 condition for the same task, suggesting that the file migration support has an effect on differences in workload before/after migration.

TEPR within Critical Sub-Tasks

Graphs 4 & 5 depict the task-evoked pupillary response for several participants for the Files task. These are time-series graphs (time in seconds on the X axis) against adjusted percent pupil radius on the Y axis. In the Files task, Step 5 was the critical task migration step, in which participants were required to pause their task on the desktop and to move to the laptop. As can be seen, the task-evoked pupillary response (TEPR) rises soon after the start of the critical step, and reaches a (local) maxima. In some instances, it progressively lowers, and in some, it stays at the new, higher level of workload until the end of the task. This provides support for the hypothesis that steps that involve transitions between devices lead to high mental workload.

Summary of RQ 1 Results

In NASA TLX scores, Task was seen to exhibit a main effect on Overall Workload, Mental Demand, Frustration and Own Performance, but not on the other three scales. There was no difference seen on any scale between two treatment levels of the same task. This suggests that NASA TLX is not very sensitive to changes in workload in the kinds of personal information management tasks tested in this experiment. Because of its lack of ability to discriminate between

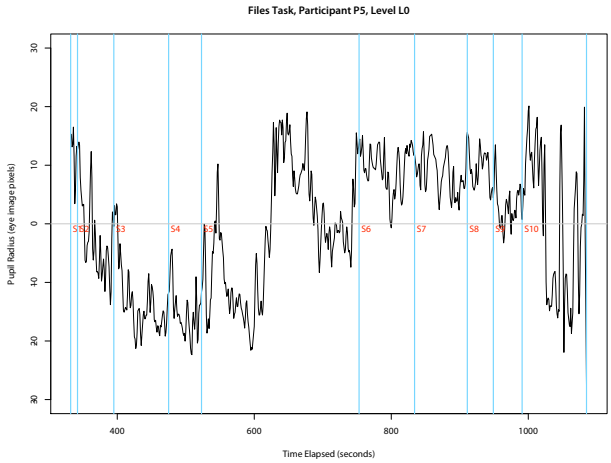


Figure 4. Task-evoked pupillary response, Participant P5, Files Task, L0

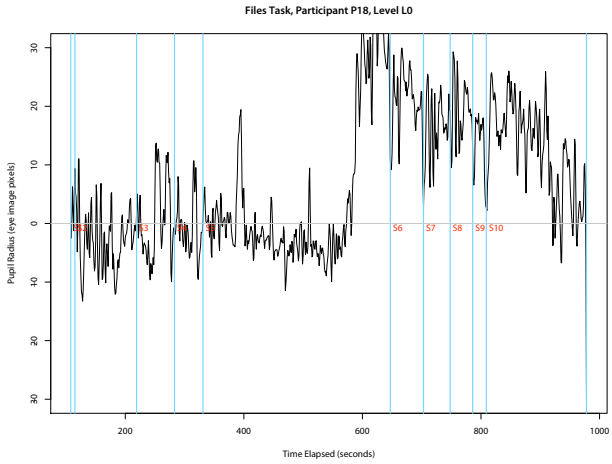


Figure 5. Task-evoked pupillary response, Participant P18, Files Task, L0

two or more ways of performing the same task, its validity and usefulness in PIM tasks cannot be established with the evidence obtained.

Task-evoked pupillary response, on the other hand, provided important insights into task migration. Specifically, it showed a significant difference for each step of the Contacts task between levels L0 and L1. Also, it showed significant differences between pre- and post-task-migration steps in the Files task. It was observed from the data that local maximas were attained during the task migration step. This points to the potential usefulness of task-evoked pupillary response as a continuous measure of workload in PIM tasks.

Results for Research Question 2

Research Question 2 seeks to explore the differences in operator performance, if any, between the L0 and L1 task conditions. The primary measure of operator performance used in this study (for all tasks) was time on task. Others, such as

number of errors, number of entries made, etc. were defined, measured and evaluated on a per-task basis. For the Files and Calendar tasks, no significant differences were found in the time taken to complete the task. However, for the Contacts task, participants completed the task significantly faster in the presence of synchronization support than without ($F_{(1,34)}$=4.72; p=0.037).

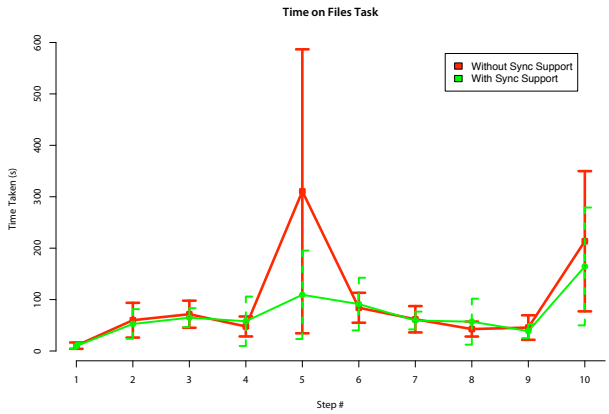


Figure 6. Time on task, per Step, in the Files task.

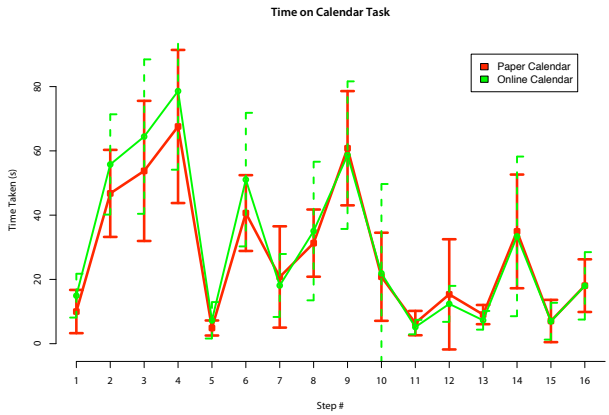


Figure 7. Time on task, per Step, in the Calendar task.

Significant differences ($F_{(1,34)}$=8.83; p=0.0054) were found for the transitional step in the Files task (Step 5) where participants were requested to pause work on their desktop computers and resume it on a laptop, taking their files with them, but not for any other step. This was expected; in fact, the lack of significant differences for steps that did not involve a transition from one device to another in the Files task confirms that the experimental setup did not lead to any biases in steps that were identical by design in both treatment levels.

For the Calendar task, two steps took significantly different times in case of the paper calendars versus online calendar ($F_{(1,34)}$=4.33; p=0.045). Both steps involved proposing a meeting time and scheduling it on the calendar. In both instances, participants took lesser time using a paper calendar than an online calendar. The ease of quick capture in paper calendars might explain why it is the tool of choice for several users despite the widespread availability of online calendars.

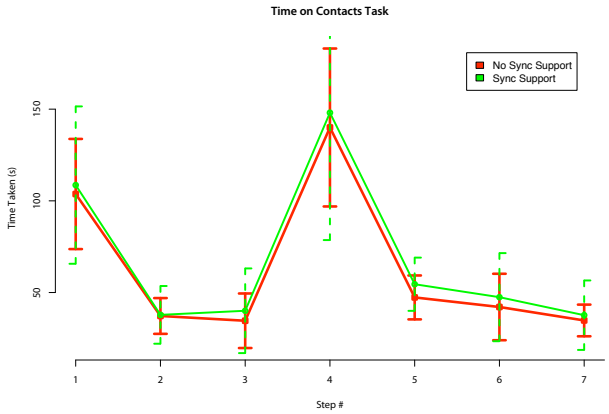


Figure 8. Time on task, per Step, in the Contacts task.

Participants correctly edited more files ($F_{(1,34)}$=5.52; p=0.025) in the condition with no support for file synchronization (Mean=6.40; SD=0.92 files) than in the condition with synchronization (Mean=5.22; SD=1.90 files) from a maximum of 7 files. This was an unexpected finding, disproving Hypothesis 2 (at least for one particular task metric) that task performance would be higher in the L1 condition.

In contact management, the number of entries made on the secondary device was significantly different in both treatment levels ($F_{(1,32)}$=15.86; p=0.00037): participants who managed contact information with syncing support made 4.71 entries on the other device, while participants without such support made only 1.00 entries. (If an instruction clearly required participants to add a contact record to a specific device (either the laptop or the phone), that device was termed the primary device. The other device (either the phone or the laptop, respectively) was termed the secondary device.)

Summary of RQ 2 Results

For the Files task, the time taken to perform the critical step in the Files task — moving from the desktop to the laptop — was significantly higher when there was a lack of system support for such migration (implemented in this experiment as a Network Drive). However, more files were edited correctly in the case where synchronization had to be performed using USB drives or email-to-self. For Calendars, there was no difference in any task metrics between the paper and online calendar conditions. In the Contacts task, more entries were recorded on secondary devices when synchronization was automatic. Thus, little to no support was found for Hypothesis 2, especially with the observation that more files were edited correctly with lower levels of support for task migration.

Results for Research Question 3

Research Question 3 examines if measures of mental workload may be used as predictors of task performance in personal information management tasks. Since time-on-task was the only performance metric that was (1) used for all three tasks, and was (2) not subject to any ceiling effects, further analysis of the correlation between performance and workload focuses on this metric. Mental workload was estimated via two methods; we consider them separately to examine whether either or both of them may be used as task performance predictors.

NASA TLX Ratings as Predictors of Operator Performance

Significant correlations were seen between NASA TLX subscales and time-on-task only in the following isolated cases: Overall Workload for Files Level L1 (p=0.01, r=0.57), Mental Demand for Files Level L1 (p=0.0071, r=0.61), Own (Perceived) Performance for Files L0 (p=0.05, r=0.47), Own (Perceived) Performance for Files L1 (p=0.02, r=0.54), Frustration for Files L0 (p=0.05, r=0.47), Frustration for Calendar L0 (p=0.51, r=0.17).

Task-Evoked Pupillary Response as a Predictor of Operator Performance

Workload estimated according to the Task-Evoked Pupillary Response was not found to be significantly correlated with Time on Task, using Pearson's product-moment coefficient (r). Table 2 shows the correlation coefficients and p-values for each task condition. It can be inferred that mental workload (measured via pupillary response) is not a good predictor of task performance.

TEPR × Time	L0	L1
Files	r=-0.062, p=0.46	r=0.15, p=0.063
Calendar	r=-0.11, p=0.078	r=-0.067, p=0.283
Contacts	r=-0.13, p=0.18	r=0.042, p=0.68

Table 2. Pearson's *r* for Task-Evoked Pupillary Response for each task condition.

Summary of RQ 3 Results

Neither NASA TLX ratings nor task-evoked pupillary response showed consistent correlation with task performance. Isolated instances of significant correlations were observed, but they do not support the use of workload measures as predictors of task performance. The lack of any meaningful correlation between performance-based metrics and workload metrics suggests that neither alone is sufficient to assess and describe highly contextualized tasks in the domain of personal information management. Thus, Hypothesis 3 was disproved in case of both metrics used in the measurement of mental workload.

Other Observations

While the preceding sections provide answers to the research questions posed at the start of this study, there were several interesting observations noted while participants performed the experimental tasks.

- **Lack of Preparation in Task Migration.** None of the participants performed any kind of planning tasks at the start of the Files task to prepare for migration. Since the means of task migration (USB drives, email access and network drive access) were already provided to them, it would have been possible for them to plan ahead by copying their files to the network, for example. However, none did so.

  This lack of planning has significant implications for those designing technologies for mobility: users cannot be expected to plan ahead or to prepare for a device transition [29]. Task migration technologies must take into account the opportunistic use of multiple devices without any pre-planning and must initiate any pre-migration activities without the need for explicit user intervention [30].

- **Maintaining Contextual Awareness in Calendars.** In the Calendar task, a few of the instructions provided to the participants mentioned the current date as a way to anchor them in temporal context. Since an entire week's worth of calendar events were presented in about 10 to 15 minutes, it was important to introduce the current day in order to preserve the hypothetical temporal unfolding of events in the experimental tasks. Participants adopted various techniques to maintain this temporal context while interacting with the calendars. Those who used the electronic calendar clicked the specified date in the calendar window, which would then highlight that day in the display. Such a visual representation helped as an external cognition aid so that the task of remembering the current day could be offloaded to the environment. Very few users who used paper calendars used similar techniques: those that did, marked each passing day with a dot or a cross towards the top of the day.

- **Capturing Information about Tentative Events in Calendars.** The scheduling of tentative collaborative events caused a high amount of confusion to users (noted via experimenter's observations; not statistically significant). Using multiple paper calendars, participants indicated the changes and rescheduling with an assortment of arrows, scratched lines, and other idiosyncratic annotation techniques. In electronic calendars, while participants could reschedule an event easily by dragging-and-dropping the electronic representation of the event to the rescheduled time, this did not solve the entire problem.

  The larger issue in tentative collaborative events is the ad hoc specification of attendees' constraints. Current calendar systems do not capture the set of constraints that lead to the tentative scheduling of an event. Hence, when such an event is to be moved to another time, the new start time must be evaluated against the complete set of constraints by consulting the originating source, e.g. email. The event record within an electronic calendar provides no way to indicate the justification behind the particular choice of time, and thus lacks an affordance for potential rescheduling. This is also a problem when adding a new constraint to the mix.

  While a few calendar systems do provide support for automatic multi-party meeting scheduling, the resulting artifact is a calendar event, not an expression of the constraints. This makes it difficult to add or remove con-

straints from the mix, to arrive upon a different time than originally scheduled.

DISCUSSION

Through the results of these studies, I found that specifics of the tasks and levels of support for task migration affected users' perceived workload ratings as well as task-evoked pupillary response in a variety of ways. These workload metrics were not the traditional usability metrics that are often used to evaluate computing systems, such as performance, efficiency, errors, etc. In fact, metrics such as whether users were able to answer questions correctly and time-on-task showed little to no difference with the different ways of performing a task, with and without support for task migration.

What this points to is that while both types of systems result in similar outcomes (and thus would be rated equally on traditional usability metrics), they do not evoke the same experiences in users. Frustration, mental demand, and workload: all are components of the entire user experience, but are not often captured by researchers and designers when assessing personal information ecosystems. This points to two separate, yet related, issues that warrant discussion: (1) evaluating usability using concepts from hot cognition that are more representative of user concerns when using multiple devices together, and (2) evaluating usability for a device ecosystem together instead of as disparate devices.

Evaluating Usability using Hot Cognition Aspects

Besides the need to measure traditional usability metrics, it is important to test whether we are, in fact, measuring the right metrics. Dillon notes [8] that in several tasks, efficiency may not be the user's priority. In particular, he highlights the inadequacy of traditional usability measures for many high-level, ongoing tasks such as information retrieval and data analysis. Other studies also have shown [26] that users' preferences for particular brands of devices have significant effects on their perception of usability of those as well as other devices. This shows that aspects of hot cognition such as affect, emotion, personal preferences, etc. play an important role in the user experience — perhaps an even greater role than purely objective metrics such as task completion times and feature comparisons.

Holistic Usability for Personal Information Ecosystems

Distributed cognition theory recognizes that actors in a system often rely on the use of external artifacts to augment their own cognition. Usability cannot thus be embedded into an artifact, but is distributed across an entire activity system. This is evident in this study in various ways: users performing the Calendar task kept track of the current day by highlighting that day in an online calendar, or by marking off corresponding days in a paper calendar. In the Files task, a few users kept modified files open in their respective editor programs as a means of tracking their changes. While these are just a few idiosyncratic examples, it points to the larger issue of systems and devices lacking explicitly-designed support for external cognitive tasks.

CONCLUSIONS

Pure performance-based measures are not sufficient to describe and assess highly contextual tasks in the domain of personal information management, and the inclusion of user perception in their assessment is important. Traditional usability metrics emphasize efficiency, effectiveness and satisfaction [ISO 9241], but they relegate metrics such as pleasure and emotion to the sidelines. This study describes that while performance metrics do not show much difference, mental workload (measured via the task-evoked pupillary response) shows a difference with/without support for synchronization (in the Contacts task).

Many devices that are intended to be used in collaboration with other devices are designed independently of one another. In some cases, it appears as if minimal attention has been given during the design process to understand the broader context of use and to situate the device in this context, offering support for the activities that are performed in real use scenarios. When evaluated for usability, many devices are often tested in pristine laboratory settings. Even if tested in real world scenarios, they may not be evaluated together with other interacting devices in the user's work environment. The lack of correlation in this experiment between task metrics and workload measures stresses the need for conducting holistic usability evaluations of such devices when they act together to fulfill a user's information needs.

ACKNOWLEDGMENTS

We would like to thank Tonya L. Smith-Jackson for instigating some of the ideas behind this project. Steve Harrison, Edward A. Fox, Stephen Edwards, Pardha S. Pyla, and Ranjana Mehta, all provided important insights that led to the design of this study in its current form. Thanks are also due to our participants who spent considerable time in completing both sessions of the experiment.

REFERENCES

1. J. Ballas, C. Heitmeyer, and M. Pérez-Quiñones. Evaluating two aspects of direct manipulation in advanced cockpits. In *CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 127–134, New York, NY, USA, 1992. ACM.

2. D. Barreau. Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, 46(5):327–339, 1995.

3. J. Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2):276–92, 1982.

4. D. A. Bertram, D. A. Opila, J. L. Brown, S. J. Gallagher, R. W. Schifeling, I. S. Snow, and C. O. Hershey. Measuring physician mental workload: Reliability and validity assessment of a brief instrument. *Medical Care*, 30(2):95–104, 1992.

5. R. Boardman, R. Spence, and M. A. Sasse. Too many hierarchies?: The daily struggle for control of the workspace. In *Proc. HCI International 2003*, 2003.

6. D. de Waard. *The Measurement of Drivers' Mental Workload*. PhD thesis, University of Groningen, Haren, The Netherlands, 1996.

7. D. Dearman and J. S. Pierce. "It's on my other computer!": Computing with multiple devices. In *CHI 2008: Proceedings of the ACM Conference on Human Factors in Computing Systems*, page 767, 2008.

8. A. Dillon. Beyond usability: process, outcome and affect in human-computer interactions. *Canadian Journal of Library and Information Science*, 26(4):57–69, 2002.

9. P. Dourish. *Where The Action Is: The Foundations of Embodied Interaction*. MIT Press, October 2001.

10. F. T. Eggemeier, G. F. Wilson, A. F. Kramer, and D. L. Damos. *Workload assessment in multi-task environments*, chapter 9. Multiple-Task Performance. Taylor and Francis, 1991.

11. J. Endicott, J. Nee, W. Harrison, and R. Blumenthal. Quality of life enjoyment and satisfaction questionnaire: A new measure. *Psychopharmacology Bulletin*, 29(2):321, 1993.

12. J. Gwizdka. Email task management styles: The cleaners and the keepers. In *CHI '04: Extended Abstracts on Human Factors in Computing Systems*, pages 1235–1238, New York, NY, USA, 2004. ACM Press.

13. S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, 1:139–183, 1988.

14. E. Hutchins. *Cognition in the Wild*. MIT Press, 1995.

15. S. T. Iqbal and B. P. Bailey. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *CHI '05: Extended Abstracts on Human Factors in Computing Systems*, pages 1489–1492, New York, NY, USA, 2005. ACM.

16. P. W. Jordan. *Designing Pleasurable Products: An Introduction to the New Human Factors*. CRC, 2000.

17. D. Karger and W. Jones. Data unification in personal information management. *Communications of the Association for Computing Machinery (CACM)*, 49(1):77–82, 2006.

18. J. Kelley and A. Chapanis. How professional persons keep their calendars: Implications for computerization. *The British Psychological Society*, 1982.

19. D. Kelly and J. Teevan. *Understanding What Works: Evaluating PIM Tools*, chapter 12, page 17. University of Washington Press, Seattle, Washington, May 2007.

20. J. Klingner, R. Kumar, and P. Hanrahan. Measuring the task-evoked pupillary response with a remote eye tracker. In *Eye Tracking Research and Applications Symposium*, Savannah, Georgia, 2008.

21. M. W. Lansdale. The psychology of personal information management. *Applied Ergonomics*, 19:55–66, 1988.

22. Y. Miyata and D. A. Norman. *The Control of Multiple Activities*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

23. D. A. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books, 2003.

24. R. D. O'Donnell and F. T. Eggemeier. *Workload assessment methodology*, chapter 2, pages 42/1–42/49. Handbook of perception and human performance: Vol. 2. Cognitive processes and performance. Wiley, New York, 1986.

25. A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti. Interaction in 4-second bursts: The fragmented nature of attentional resources in Mobile HCI. In *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 919–928, New York, NY, USA, 2005. ACM Press.

26. S. Park, A. Harada, and H. Igarashi. Influences of personal preference on product usability. In *CHI '06: Extended Abstracts on Human Factors in Computing Systems*, pages 87–92, New York, NY, USA, 2006. ACM.

27. S. J. Payne. Understanding calendar use. *Human-Computer Interaction*, 8(2):83–100, 1993.

28. M. Pérez-Quiñones, M. Tungare, P. Pyla, and S. Harrison. Personal information ecosystems: Design concerns for net-enabled devices. In *Proceedings of the VI Latin American Web Congress - LA-Web 2008*, 2008.

29. M. Perry, K. O'Hara, A. Sellen, B. Brown, and R. Harper. Dealing with mobility: Understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(4):323–347, 2001.

30. P. Pyla, M. Tungare, J. Holman, and M. Pérez-Quiñones. Continuous user interfaces for seamless task migration. In *Proceedings of the 13th International Conference on Human-Computer Interaction, HCII 2009*, 2009.

31. P. Pyla, M. Tungare, and M. Pérez-Quiñones. Multiple user interfaces: Why consistency is not everything, and seamless task migration is key. In *Proceedings of the CHI 2006 Workshop on The Many Faces of Consistency in Cross-Platform Design.*, 2006.

32. J. Teevan, R. G. Capra, and M. Pérez-Quiñones. *How people find information*, chapter 3, page 17. University of Washington Press, Seattle, Washington, May 2007.

33. M. Tungare. *Mental Workload in Personal Information Management: Understanding PIM Practices Across Multiple Devices*. PhD thesis, Virginia Tech, 2009.

34. M. Tungare and M. Pérez-Quiñones. An exploratory study of personal calendar use. Technical report, Computing Research Repository (CoRR), 2008.

35. M. Tungare and M. Pérez-Quiñones. It's not what you have, but how you use it: Compromises in mobile device use. Technical report, Computing Research Repository (CoRR), 2008.

36. M. Tungare, P. Pyla, M. Sampat, and M. Pérez-Quiñones. Defragmenting information using the Syncables framework. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*, 2006.

37. M. Tungare, P. Pyla, M. Sampat, and M. Pérez-Quiñones. Syncables: A framework to support seamless data migration across multiple platforms. In *IEEE International Conference on Portable Information Devices (IEEE Portable)*, 2007.

38. S. Whittaker, Q. Jones, and L. Terveen. Contact management: identifying contacts to support long-term communication. In *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, pages 216–225, New York, NY, USA, 2002. ACM.

39. S. Whittaker and C. Sidner. Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 276–283, New York, NY, USA, 1996. ACM Press.

No Room for White Space: The Personal Information Management of a Printing Company Employee in a Home Office

Leslie Thomson
Faculty of Information,
University of Toronto
140 St. George Street,
Toronto, Ontario M5S 3G6
(905) 259-4781
leslie.thomson@utoronto.ca

ABSTRACT

In library and information science (LIS) literature, information organization and use within traditional office environments has received much attention, while occurrences of these same phenomena within the intersectional home office space have gone largely unexplored. Similarly, no LIS literature has focused on the dimensions of information organization and use in the printing profession, though the volume of documents and resources necessary for job success in this realm make it an information-rich area, ripe for exploration.

This small-scale research study involved an ethnographic fieldwork outing into the home office of one printing company employee, and was guided by the question of what approaches to information organization and use, conventional or unconventional, arise in this environment. Data was gathered by way of diagrams, photographs, a guided tour, a semi-structured interview, and unobtrusive observation. Data analysis provided a preliminary glimpse into physical (paper-based) information organization and use as these occur within one specific home office context, and a means to begin theorizing about the effects of home office environments on such "meta-level" information activities (Jones, 2007, p. 464). This study will be expanded in an upcoming (2010) thesis that compares similar information practices across a larger sample of home office settings and devices.

1. INTRODUCTION

For forty-eight years, my father has worked in the printing profession. In his current position, he is a "sales representative," acting largely as a liaison between customers (most often publishing firms) and the printing plant facility that he assigns to handle and fulfill their orders for printed materials. He oversees the entire process from initial price quotations for "print runs" through to final, finished products.

Through the metatheoretical lens of ethnography, the aim of this study was to analyze my veteran printing company employee father's home office, a vast and ever-growing assortment of documents relating to his work and "samples" of the various items, mostly books, he takes part in creating. This paper provides a glimpse into the personal information management specific to this context by focusing on the organization and use of the print documents (papers or files) and print resources (books) contained within his office. Although undeniably a major factor in the home office studied, information residing within and exchanged via any technological systems was not included in the scope of this study. Investigation into the physical information phenomena in the home office was guided by the question, "What approaches to information organization and use, conventional or unconventional, arise in this environment?"

2. LITERATURE REVIEW

To date, no library and information science (LIS) literature has focused on the information dimensions of the printing profession, or of the commercial information production industry more broadly. Yet, the rapid accumulation and sheer volume of print documents and print resources necessary for job success in this realm situates it on a plane that is comparable to that of academic scholars—professionals who do have a large body of information-use literature behind them—in information richness.

Similarly, a small but valuable subset of LIS literature speaks directly to the 'real' office environment and to the organization and use of information artifacts therein, while the home office remains an unexplored terrain. The unique nuances of the home as a stand-alone concept have not gone unnoticed by LIS scholars, who recognize that information behaviours do not occur there uncompromised, or simply take on the same shape as they would in another setting. Rieh (2004) notes in her study of personal, at-home, web seeking that "information-use environments" (p. 3) do affect information behaviours, and that "home is considered to be a socially defined setting rather than merely a physical setting" (p. 2). Still, there has yet to be any study of how work-related information behaviours translate to, and transmute within, the home environment.

Both the early writing of Malone (1983) and the later work of Whittaker and Hirschberg (2001) surrounding information management in the general workplace, as present in the dispersal of piles, files, and papers across peoples' "desks,… tables, shelves, file cabinets, and other information repositories" (Malone, 1983, p. 100), underscore Kwasnik's (1991) argument that "situation attributes," or the contexts in which documents are received and required professionally, play key roles in determining their arrangement. Finneran (2007) has also proposed associations between information behaviour and peoples' cognitive and affective motivations to acquire, store, and manage documents. Building from this previous body of literature concerning the personal management of work-related information, the current study theorizes the effects that a quasi-professional, quasi-social home office environment might have on individuals' "meta-level" information activities, or how they "establish, use, and maintain a mapping between information and need" (Jones, 2007, p. 464).

3. RESEARCH DESIGN AND METHODS

In order to carry out this research study, two fieldwork outings into the home office of my father, the printing company employee, were scheduled, one to take place on a vacation day and the other on a typical work day.

The first time accessing the field was used as an opportunity for preliminary data gathering, including mapping, diagramming, and measuring of the overall home office space and its features (all shelving units, filing cabinets, and utilized flat surfaces). A photographic inventory was started, and note taking began. This rough first-round data was enhanced by a tour of the space guided by my father and a semi-structured interview with him later that day, when pertinent labels regarding information organization and prominent patterns of use were marked on the existing diagrams, further field notes were generated, and more photographs were taken. This outing lasted approximately two hours. The second time accessing the field was used as an opportunity for unobtrusive observation; as my father was working that day, I received a 'natural' picture of his information practices for approximately one and a half hours over two sessions.

4. THE FIELD

This study took place in a printing company employee's unheated, nine hundred square-foot suburban basement home office. The space is navigated via narrow paths that wind between looming filing cabinets and teetering six-foot high stacks of books, files, and papers. It is lined with floor-to-ceiling bookshelves that buckle under the weight they hold, illuminated with sporadic DIY lighting, and filled with continuous sounds of keyboard typing and computer printing. Figure 1 illustrates this basement home office space and the extent of the informational spread therein.

To better ground the diagrams in Figure 1 in their reality, Figure 2 provides a photographic overview of one part of the home office within the larger basement (it was captured from alongside the pool table, facing toward the desks). As well, Table 1 gives rough estimates of the number of print artifacts in the home office at the time of this study, indicating an approximate total of 8400 documents and resources, or files, papers, and books (3988.38 books in piles, 1760 books on shelves, and 2652 files in drawers).[1]

Figure 1: The basement home office space of a printing company employee, depicted in three images. The first (top left) shows the basement and its features. Those of the household include a pool table (green) and a couch (gold). Those of the home office include bookshelves, desks (all light brown), and filing cabinets (grey). The second image (top right) shows the "red thread" (Bates, 1999, p. 1048) of book samples from the home office as distributed across the basement. The third image (bottom) shows the "red threads" of books and files from the home office as distributed across the basement. (Red squiggles indicate items on shelves or within filing cabinets.)

Figure 2: The printing company employee at work in his home office, amidst overflowing shelves, papers, and piles of books that reach to the ceiling.

Table 1. Estimated Number of Print Artifacts in the Home Office

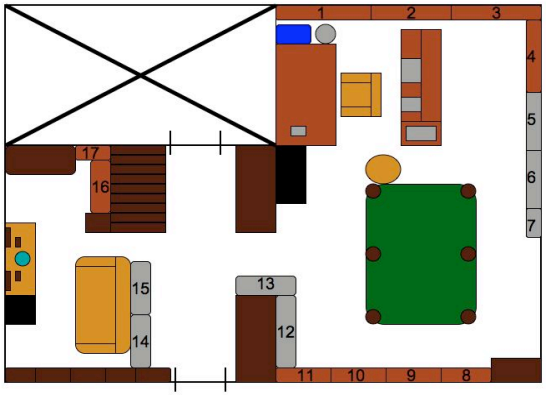
Books/ pile	Piles	Books/ shelf	Shelves	Files/ drawer	Drawers
60.43 (423 books/ 7 piles)	66	32 (128 books/ 4 shelves)	55	102 (306 files/ 3 drawers)	26

[1] This total number, however, is a minimum estimate, as it was difficult to derive the total number of piles with certainty (many were double- and triple-stacked), many shelves were rendered non-visible because of piles, and many flat surfaces were so covered that files thereon simply could not be counted. The total also excludes any files outside of drawers, which Figure 1 indicates there were many.

5. FINDINGS

The remainder of this paper is devoted to describing and analyzing the discoveries made while in the field, which follow from the study's initial research question about the approaches to information organization and use, conventional and unconventional, that arise in this particular home office environment.

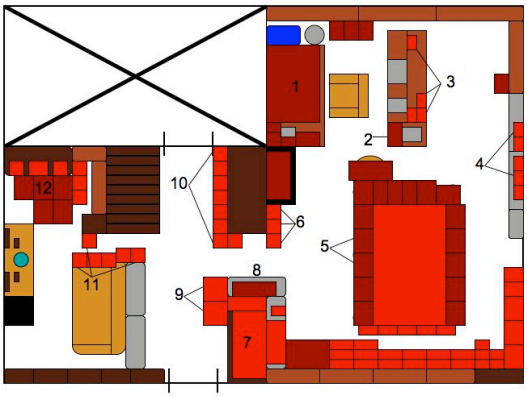
Surprisingly, the complex informational spread within the printing company employee's home office is organized according to the comparatively simple principle of "customer": all print documents relating to a single publisher's jobs are kept together in piles, filing cabinet drawers, or entire filing cabinets, and are clearly tabbed with the publisher's name and the related job (book) title. All sample resources of a single publisher are kept clustered together in larger piles, entire piles, on shelves, or on entire shelving units. It was revealed during the interview that this scheme of arrangement was always used for the communal storage units of the 'real' offices in which the printing company employee worked. Because he considers it the easiest system to maintain and the one that best facilitates use, he adopted it himself.



1. Bookshelf- McGraw-Hill Ryerson (MHR)
2. Bookshelf- Nelson
3. Bookshelf- Pearson Education (PE)
4. Bookshelf- Random House
5. Filing cabinet- various small groupings
6. Filing cabinet- Random House
7. Filing cabinet- Delmar
8. Shelf- various small groupings
9. Bookshelf- various small groupings
10. Bookshelf- various small groupings
11. Bookshelf- various small groupings
12. Filing cabinet- Wadsworth, Heinle
13. Filing cabinet- "in-production," back orders, and invoices
14. Filing cabinet- MHR, PE
15. Filing cabinet- PE, Nelson
16. Bookshelf- PE, MHR, McClelland & Stewart (MS)
17. Bookshelf- MHR, MS

Figure 3: Labeled diagram of the major organization scheme based on "customer," or publisher, in the printing company employee's home office.

The major exceptions to this organizational scheme are the printing company employee's placement of "active" (pending) information on or beside his desk, organized by assigned printing plant before customer, and his placement of "recently received," "in-production," and "current month's" information at pivotal points along the route to his desk, where they will inevitably fall into his line of sight as he approaches and passes by them. This seemingly anomalous organization therefore adheres to convention, as the "situation attributes" (Kwasnik, 1991) of these artifacts—being that they were required and referred to more frequently than other print artifacts within the home office at the time of this study—warrant their specialized placement. Also, as Jones (2007) notes, keeping items in view may aid the printing company employee's remembering of the tasks that still require his attention (p. 469).



1. Print documents- "active" quotations and corrections

2. Print documents- "active" orders

3. Print resources- Random House

4. Print resources- Random House

5. Print documents- "current month's" invoices

6. Print resources- Nelson

7. Print resources- for disposal

8. Print documents- "in-production" orders

9. Print documents/resources- "recently received in"

10. Print resources- for disposal

11. Print resources- for disposal

12. Print documents- for disposal

Figure 4: Labeled diagram of the co-existent organization scheme in the printing company employee's home office.

It might be said, then, that in large part, systems of information organization and patterns of information use carry over to the printing company employee's home office as they would have existed in a 'real' office. Still, slight adaptations to the home environment are visible.

Information practices in the home office both act on and are acted upon by the 'built' features of the home itself, which are, in this case, quite distinct from those of a typical workplace. Kwasnik (1991) writes that, "a person makes classification decisions within a context and for a purpose, but also within the constraints of physical objects and a physical environment" (p. 397). Because there is no strictly demarcated home office 'zone' within the printing company employee's overall basement space, patterns of organization and use must necessarily account for already existing structures and furnishings, such as decorative wood pillars, immovable islands, couches, and pool tables. Creative, ingenious ways to work around these features according to the needs of the home office must be employed, as demonstrated in Figure 5.

Likewise, information practices in the home office both act on and are acted upon by social relations in the home itself. The half-household basement, half-home office space, due to its hybrid nature, is "a site where the relations within [the] family can be played out and (re)negotiated" (Swan and Taylor, 2005, p. 4). Attempts to recreate 'seamless,' office-like information experiences across systems of organization and patterns of use lead to appropriations of spots formerly designated to household items alone. In turn, household objects end up re-imposed amidst the files and books of the home office so to reinforce prior routine, as shown in Figure 6. Information behaviours in the home office are no longer moderated through professional regulations or colleagues' judgments, as they may be in a 'real' office, but are, in fact, mediated by family members' sometimes vastly different expectations of what constitute acceptable and efficient information practices.

Figure 5: Information organization and use in the home office must incorporate 'built' features of the home itself, and occur within wooden framing, around pool table pockets, and overtop of couches.

Figure 6: Information organization and use in the home office modify based on social relations in the home itself. Everyday household objects such as vegetables (in the green basin) and travel coolers (in the top left picture), and Christmas decorations (on top of the filing cabinets in the top right picture,) are placed amidst home office documents and resources. In the bottom picture, another family member voluntarily takes on the role of 'managing' (shredding) documents that were deemed of no use in the home office.

6. REFLECTIONS

The home office of the printing company employee was set up and continues to evolve in accordance with professional needs and personal preferences, ensuring that it functions for its user like a finely tuned machine. During the observation periods in the home office, for example, the printing company employee never once left his desk chair: all of the information required to carry out his tasks was within arms reach, confirming what he said during the interview about the documents and resources on and surrounding his desk being those that are "not quite ready" to be filed or shelved farther away.

Still, the home office was far from a self-contained entity. Lee's (2003) "structure of users' information spaces" (p. 432) model provided a framework within which the home office was conceptualized in this study (shown in Figure 7). Like the academics Lee (2003) studied, whose information practices occurred mostly within "immediate" offices and occasionally within "adjacent" and "outside" spaces such as libraries, the activities of the printing company employee working from home dissipated outward from an "inner" hub of critical, independent information use to include less frequent, but nonetheless essential, interactions with the broader world.

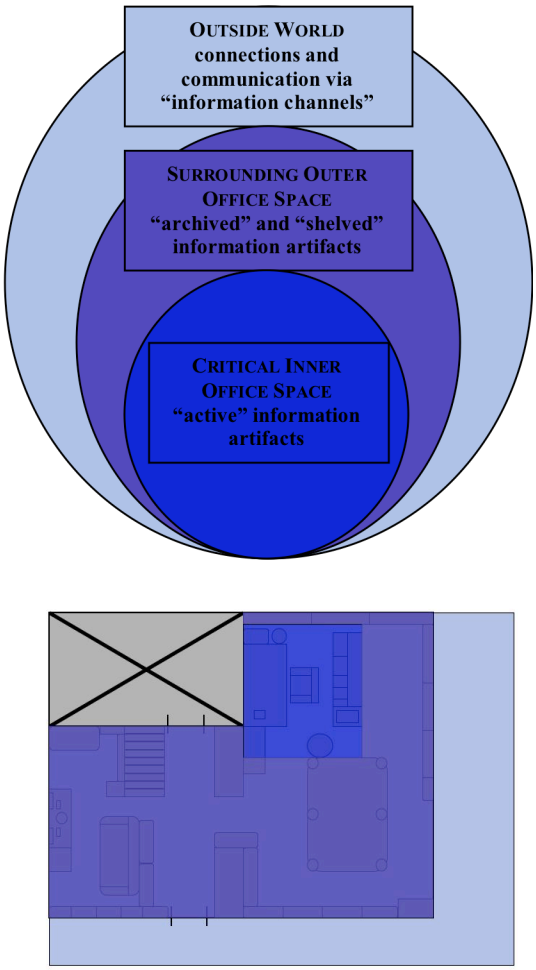


Figure 7: Model showing the information activities of the printing company employee in the home office, which disperse from a "critical inner office space" of "active" information artifacts, to an "outer office space" of archived and shelved ones, to finally include information interactions with an "outside world." Designations of "inner and outer spaces" and the "outside world" are correspondingly shaded on the diagram of the home office below.

7. CONCLUSION

This research study provides a basis from which to begin theorizing about "meta-level" personal information management activities (Jones, 2007, p. 464) as they occur in the home office space. While the printing company employee's professional concerns surrounding informational content, for example, remain unaltered and suggest an independence from matters of setting, his habits of information organization and patterns of information use have adapted and modified from both necessity and choice.

Building upon this study, future studies (including a thesis in 2010) will examine the degree to which information management practices that may be 'controlled' or constrained in 'real' offices are subject to personal idiosyncrasies in the home office. These will use a wider sample of home office spaces and consider a wider array of information and information devices. Bruce, Jones, and Dumais (2004) write, "effective information seeking and use continues to rest fundamentally with the individual and with his or her ability to create, manage, and use a personal information collection" (Conclusion). The printing company employee's information experience suggests that the home office may be an environment nourishing of diverse personal information management styles, where one is 'free' to act, for example, as an "informal librarian" (Whittaker and Hirschberg, 2001, p. 166), keeping information at their fingertips, or a "sentimental hoarder" (Finneran, 2007, p. 6), attached to the information they acquire.

8. REFERENCES

[1] Bates, M. J. (1999). The invisible substrate of information science. *Journal of the American Society for Information Science, 50*(12), 1043-1050.

[2] Bruce, H., Jones, W., & Dumais, S. (2004). Information behaviour that keeps found things found. *Information Research, 10*(1).

[3] Finneran, C. M. (2007). *Factors that influence users to keep and/or leave information items: A case study of college students' personal information behavior*. Excerpt from an unpublished doctoral dissertation, Syracuse University—New York.

[4] Jones, W. (2007). Personal information management. *Annual Review of Information Science and Technology, 41*(1), 453-504.

[5] Kwasnik, B. H. (1991). The importance of factors that are not document attributes in the organization of personal documents. *Journal of Documentation, 47*(4), 389-398.

[6] Lee, H. L. (2003). Information spaces and collections: Implications for organization. *Library & Information Science Research, 25*(4), 419-436.

[7] Malone, T. (1983). How do people organize their desks? Implications for the design of office information systems. *Proceedings of ACM Transactions on Office Information Systems, 1*(1), 99-112.

[8] Rieh, S. Y. (2004). On the web at home: Information seeking and web searching in the home environment. *Journal of the American Society for Information Science and Technology, 55*(8), 743-753.

[9] Swan, L., & Taylor, A. S. (2005). Notes on fridge surfaces. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1813-1816). New York, NY: ACM. Available online at http://research.microsoft.com/en-us/um/people/ast/files/swan_taylor_chi05.pdf

[10] Whittaker, S., & Hirschberg, J. (2001). The character, value and management of paper archives. *Transactions on Computer Human Interaction, 8*, 150-170.

"I Thought You Would Show Me How to Do It" - Studying and Supporting PIM Strategy Changes

Florian Evéquoz
Departement of Computer Science
Bd de Perolles 90
CH-1700 Fribourg, Switzerland
florian.evequoz@unifr.ch

Denis Lalanne
Departement of Computer Science
Bd de Perolles 90
CH-1700 Fribourg, Switzerland
denis.lalanne@unifr.ch

ABSTRACT

This article presents an unexplored perspective in personal information management (PIM): PIM strategy changes. In the first part of the paper, we introduce a naturalistic study we conducted in order to explore how people manage their information and how their strategies evolve. We propose a model for characterizing PIM strategy changes based on their cause, scope and action. Further, we present colateral user problems and needs found during the study. The second part of the paper presents WotanEye, a software tool making use of information visualization techniques for reflecting the usage, hidden structure and patterns of personal information, towards alternative ways of managing PI and reflecting on one's own PIM strategies.

1. INTRODUCTION

Personal information management (PIM) is a domain in constant evolution. New devices allow us to carry and access our PI everywhere. The democratization of high capacity storages and progresses in compression algorithms make it easier to collect and manage multimedia PI. The cheapness of storage also discourages any "spring cleaning" of our PI collections. To quote Jones in [18], we prefer buying a new mansion with plenty of rooms for piling up our personal information instead of cleaning our current home and throwing away our old stuff. Moreover, as portable devices' capacity increases, it seems we could in a near future be proud owners of a device comparable to Rincewind's *Luggage* in Pratchett's *Discworld* [23]: a multi-legged trunk with almost infinite storage for items of every kind and that follows us everywhere every time, for better or worse.

Not only does information pile up, but our needs evolve as well. As a consequence, we often have no other choice than to change the way we manage information, even if constantly changing our PIM habits has drawbacks. When we change our job for instance, we generally have to implement new strategies to adapt to the new context. We usually cannot stick once and for all to one strategy of PIM because the context constantly evolves. Instead, we sporadically adapt our strategy to fit with the new technical possibilities we have or simply because our current strategy has been pushed to the limit and does not respond any more to our needs.

The article is constituted of two somewhat disjoint parts, reflecting two different approaches: the first one is user-centered while the second is technology-driven. While the first part aims at understanding people practices and needs, the second part presents the current development of a prototype we implemented to support PIM through natural means. The first part of the article presents the results of a naturalistic study aiming at studing PIM practices and motivations to analyse and improve it. How do people analyse their PIM strategies? What motivates their changes? Are they interested in analysing their own practices and in learning better ones? And if yes, what is to be done to support them? The second part presents WotanEye, a tool we implemented with the hypothesis that people will better manage their information if manipulating meaningful and natural means such as social, temporal and thematic facets. Further, in connection with the first part, we argue that this tool could help people in adapting their PIM strategies when needed.

2. RELATED WORK

Barreau [5] and Nardi [4] pioneered the studies about digital PIM behaviours. They identified the role of the software and hardware tools in the files management practices of computer users and gave a preview of their PIM habits. One year later, Whittaker dived into people's emerging email habits [27] and proposed three categories of email management strategies: *no-filers*, *frequent filers* and *spring-cleaners*, the former category being further extended by Balter [2] into *folderless cleaner* and *folderless spring-cleaner*. Abrams [1] did a similar work for bookmarks management strategies. Barreau [3] defines five filing strategies that can be applied to whatever type of information (emails, files, bookmarks, etc.). However, despite the elegance of categorizations, it has been noted that people often fail to adhere strictly to one single category of PIM strategy.

More recently, Boardman [8] summarized recent research on PIM practices, with a particular focus on integration challenges. He notably emphasizes the facts that (1) users generally prefer to browse than to search through their personal information and (2) the personal email archive has a potential for being integrated with the personal files, as similarities are strong between personal files and filed emails, a conclusion also defended by Whittaker, Bellotti and Gwiz-

ASIS&T PIM Workshop 2009 Vancouver, BC, Canada

dka [26].

Many other researchers conducted studies directed towards the design of particular PIM tools or aiming at evaluating their suitability to user's needs [6, 9, 10, 13]. The different approaches they followed are thoroughly presented in [19].

Most (if not all) studies highlight the difficulties users have to face in order to organize and retrieve information items. Lansdale describes the psychological challenges associated with PIM [20] and brings out two key issues for organizing and re-finding personal information: (1) categorizing items is a cognitively hard task, and (2) people remember contextual cues to personal items that could complement retrieval procedures. It seems that PIM-tools developers have started to address these issues only recently.

Finally, this short state-of-the-art of PIM behaviours studies would not be complete without mentioning a most recent and interesting study conducted by Barreau [3]. She involved participants that already took part to her first study 10 years before and ask them the same questions about their PIM behaviour. Her conclusion is that, even if technological means or job responsibilities of the participants have evolved, some of their PIM behaviours have remained very similar across the years. Barreau's focus in this study is on tracking behaviours that remain the same accross the years, given "improvements in technology and expansion of the electronic information environment". Our focus is complementary to Barreau's. Instead of looking at conservative behaviours, we examine fine-grained PIM strategy changes occurring in a shorter time-scale, trying to identify the reason and context that make the change occur. In our understanding, Barreau's *behaviour* characterizes the way people behave in doing PIM with much hindsight, neglecting the specific details of which PIM software is used or which kind of information is managed. To describe it in a broader outline, we have the feeling that what Barreau calls PIM behaviour somewhat stems from indiviual personality traits which do not evolve much over time. In our work, we propose to use the term of *strategy* to characterize the way people manage their PI for a specific task and context (software, role, etc.). These fine-grained strategies are thus more subject to change than Barreau's *behaviours*.

3. PRESENTATION OF THE STUDY

Naturalistic studies are considered a complementary approach to quantitative studies in PIM, because they can provide a more holistic and contextual understanding of the actual practices and needs of people. Their final goal is to apprehend how they make sense of their personal information and the management strategies related to it [22]. Naturalistic studies rely on fieldwork and contextual observations, which leads to the formulation of hypothesis and knowledge modelling. As computer scientists, we of course have a lesser experience of the fieldwork than the experience a trained ethnographer would have. However, we are domain experts for digital PIM, which has the advantage of providing another insight into the matter [16]. The present section explains our goals, the context of the study and provides a discussion of its results.

3.1 Goals

Our initial goal was to get an overview of the strategies people employ to manage their personal information in the context of their work. We planned to cover more precisely the areas of task and calendar management as well as email and files management, but remained opened to other opportunities for PIM that we could have not envisioned, according to an holistic approach. During the fieldwork, we consolidated our understanding of PIM strategies. Among the many promising questions that arose, one caught our attention: do people analyse their own PIM practices and are they motivated to improve it? This is a topic not yet well covered in the literature and we saw our fieldwork as an opportunity to gain insight into it. We therefore reframed our goals in order to elicit the means people employ to analyse their own PIM strategies and the motivations behind their strategy changes.

3.2 Context

Twelve participants (coded P1 to P12) were recruited by convenience among the circle of friends and acquaintances of the authors, as it if often the case with naturalistic studies. Moreover, as this study deals with personal matters, we believe this also helps the participants to know and trust the interviewer in order to feel comfortable while discussing their PIM strategies. The participants were aged 26 to 49 (average 36), 7 of them were female, 5 male. Half of the participants had a computer science background, whereas the other half came from diverse fields like psychology, letters or commercial. Their professions range from full-time university professor to secretary, including research assistants and high-school teachers. Most of them have used a computer at work for more than 10 years, only two of them use one for less than 5 years.

The time and resources at disposal for the fieldwork were limited. In particular, we could not perform long-term observations of user behaviour, but would focus on a snapshot of their practice, combined with a semi-structured interview. We decided not to use recording devices that would be intrusive for the participants and require much time to transcribe and analyse. Instead, we used a notebook for taking live notes during the interviews. The raw field notes were taken on the right page while left pages were left empty. They would be used later to identify possible themes, add comments, notes and sketches as a first effort of synthesis, as advised in [15]. While note taking, the interviewer took care to write down not only what the participant was saying, but also relevant details he could notice, or what the participant was actually doing while talking.

The interviews took place at each participant's workplace, behind the computer they usually work on. The interviews lasted 1-1.5 hour on average, depending on the time the participant could invest. The interviewer used a "backpocket guide" to help him frame the interview. But, as it is the rule with naturalistic studies, open discussion was encouraged in order for the interviewer to discover unsuspected areas of interest. The backpocket guide was inspired from Barreau's questions set [3] and included specific directions towards the understanding of strategy changes:

- What do you do on a typical work day?
- What information do you have and use in your personal workspace?
- How do you organize information in your personal workspace?
- How do you typically go about finding information when you need it?
- What features do you wish were available for organizing

and retrieving information from your workspace that you do not already have?

- What are the main problems you have encountered managing your personal information?
- How have your organizing and retrieving strategies evolved over time?
- What motivates you to change a strategy?

The raw notes were then reworked by the interviewer in order to build an organized collection and interpretation of the participants' responses. These were published internally [14]. In the following we present selected results and discuss them.

3.3 Results

All 12 participants were able to talk about their PIM strategies and to explain at least one PIM strategy change. Moreover, some participants were able to say that the strategy they disclosed orally was ideal, and that exceptions happen in practice. Exceptions can be due to "lack of time" (P5) or "laziness" (P1) for example. Others tried to formulate a strategy, but evidences show that it was inconsistent with their actual practice. Indeed, the interview time was an opportunity for the participants to reflect on their PIM practices, which they had seldom done before. Some also became aware of strategy flaws or forgotten items that needed to be handled (P1-P3-P9-P10). Moreover, after the interview, one participant (P4) wanted to give precisions on her strategies and notify the interviewer that she noticed something new in her practice. This seems to emphasize the fact that the interview time made her think about her strategy and analyse it more precisely.

However, strategy changes are not always conscious and seem to "happen almost naturally" (P3). After people have explained us their strategy, we could sometimes notice evidences that they used another strategy in the past. For example, P1 explained that she manages her tasks with her calendar. While browsing through her emails, we noticed a folder called *Todo*. She could explain she planned to use it for managing tasks related to emails but forgot it in the end. Similar situations happened with other participants (P3-P10). Therefore, it seems that some people do not have a good mental model of their PI as a whole, and that they happen to forget portions of their PI collections and related strategies.

Other interesting results of the study shows that concurrent management strategies can coexist. In particular, archived items often adhere to a classifying scheme which is not used any more. For example, P12 mail archive contains past emails classified in directories. However, his current management strategy is that of a folderless cleaner in Balter's taxonomy [2].

We propose in the following paragraph a model for characterizing PIM strategy changes. Then, we classify the strategy changes that we observed during our study according to this model.

3.3.1 Model of strategy changes

From our empirical observations of PIM strategy changes, we attempt to formulate a simple model for characterizing changes. In this model, a PIM strategy change consist in a general *scope*, a *cause* and an *action*:

Scope The scope of a strategy change is either *specific* to a PIM tool or *longitudinal*, across tools but related to one meaningful activity. For example, using more folders to classify emails is a strategy change specific to the email client. Centralizing the task management that was scattered across many tools to a single one is a decision taken from a longitudinal scope.

Cause The cause of a strategy change is either *external* or *intentional*. External causes involve conformance to organizational requirements, for example rules to use collaboratively a filesystem or calendar. Intentional causes involve the user wanting to make its management better without external constraints, for example because she happens to have trouble re-finding emails or she takes too much time putting files into folders.

Action The action may be a decision to *simplify*, or *complete* the existing strategy. Depending on the scope, simplify and complete take different meanings. In a tool-specific scope, simplify means using *less categorization* (e.g. less folders, less labels, etc.), while complete means using *more categorization*. In a longitudinal scope, simplify maps to *centralize* (e.g. use one single tool for managing all tasks) and complete to *fragment* (e.g. keep copies of email attachments both in the mail archive and the filesystem). Note that we consider a change of strategy towards deletion of items a simplifying action, since its a kind of simplification or centralization (towards the trash). As well, a strategy which is abandoned altogether is considered being simplified to the extreme.

Table 1 details a selected sample of strategy changes that were discussed with participants during interviews. Some changes were mentioned by participants directly, others were asked by the interviewer when he noticed something unexpected according to a strategy disclosed by the participant. Of course stronger empirical evidences (e.g. alternate studies) would be necessary to validate our model attempt.

3.3.2 Motivation behind PIM strategy changes

People adapt their strategies. But the concrete reasons that make them change their strategies are many. For example, they may have discovered a new PIM tool that they want to give a try (P8). They have had a hard time re-asking for emails they had deleted and do not want to experience that situation again (P2). Their strategy for tasks management failed as they forgot appointments and tasks (P3-P8). They feel overwhelmed by tasks, because they use the same calendar for private and professional tasks (P10). They face collaboration problems for task management (P9). They have new job responsibilities (P8). They must respond to organizational requirements (P4-P6-P11-P12). They have learnt a new feature in a PIM program that they never used before (P6). They find that some features are missing in their current strategy (P12). They cannot cope any more with fragmentation (P9). Or they simply want to try a new idea inspired by a colleague's practice (P8).

Moreover, several people admitted being not satisfied with their strategy and consider it a makeshift solution, waiting for a better way of managing their information (see section 3.3.3). Thus, they invoke reasons for not changing their strategy. Some participants explained that changing would take too much time: "I prefer losing some time searching for a document than classifying everything perfectly" (P5).

Table 1: Types of some observed strategy changes

Property	Scope		Cause		Action	
Strategy change	Specific	Longitudinal	External	Intentional	Simplify	Complete
Stop filing emails into folders	×			×	×	
Stop deleting all emails, keep them in the inbox	×			×		×
Use naming conventions for files	×			×		×
Abandon the use of todo folder and do task management through emails		×		×	×	
Keep several copies of documents in different places (including attachments)		×	×			×
Stop using a folder for "source" documents	×			×	×	
Use a common filing strategy for documents on a collaborative platform	×		×			×
Stop tagging items in del.icio.us	×			×	×	
Add folders and subfolders in a systematic way for all projects		×		×		×
Use a single tool for task management		×		×	×	
Centralize collaborative tasks management		×	×		×	
Use two calendars for private and professional matters	×			×		×
Use less folders		×		×	×	
Centralize all tasks management in a specific tool		×		×	×	
Use a common calendar	×		×		×	

Other people also invoked a lack of education for justifying their strategy. For example, P10 feels overloaded by email, but she does not use any mail client, because "no one ever told [her] how to use them". The standard email web interface she uses is poor, slow and does not provide search tools. Thus it takes her much time and energy to manage emails. When asked to describe the strategies she was using for managing her personal information in her workspace, P6 expressed her disappointment: "I thought you were going to show me how to do it !". She further described herself as an "old-school secretary", educated to the use of real papers and binders and with a very basic knowledge of computer-supported information management tools, which somewhat frustrates her. Other participants also asked the interviewer for advice in their PIM or complained that they never have received lessons or hints for doing PIM.

3.3.3 Unsatisfied needs and user problems in PIM

During our naturalistic study on PIM strategies, some side problems were raised by participants. They often contribute to the decision of changing their strategy.

- *Classification*: One participant clearly expressed that keeping things organised takes too much time (P5). On the other hand, several people said that they are afraid of losing things which encourages them to classify information or even duplicate it (P4-P5). On a related note, one participant said that even if time-consuming, he considers classification calming and reassuring (P11).

- *Information fragmentation*: At least two participants (P5-P10) complained about the fragmented nature of their personal information, either because it stands on various materials (disks, servers, USB sticks, etc.) or because it is distributed over various applications (emails, calendar, documents, etc.).

- *Collaborative PIM*: Although shared information is not personal, various participants considered that it enters in the personal information sphere since information is sometimes shared among small groups of people. Various people said that they have problems to find things in collaborative setting (P4-P5-P6-P8-P9). Strangely, even when they consider that they have messy shared directories, they admit they do not have the will to change their organization because they cannot take decisions about shared data on their own (P4-P6).

Apart from that, people explicitely mentioned the following needs.

- *Browsing facets and information linking*: Various people use facets implicitely to organize and refind information. For example, P2, P4 and P5 classify some documents and emails in directories having the name of people from their social network. Others classify and/or access documents or emails by temporal indexes (P1-P3-P4-P8). Concerning the social facet, some told us that they would like to have their communication items (emails, chats) automatically clustered by people and further to have their broader PI automatically linked to the people concerned. Still, related to information linking, several persons expressed the need for tools to help them organize specific types of information. For instance a system administrator (P7) said she would like to be able to keep track of software and hardware problems-solutions pairs. Another expressed his need to link scientific articles (colleague of the authors, not included in the presented study).

- *Ubiquity versus fragmentation*: Although several people expressed their problems with the fragmented nature of information, on different disks or applications, as discussed in the previous section, they reversely expressed their desire to have information available and accessible from everywhere, for instance to have access

to appointments online or to have information available on the field (in a mobile task) (P3-P5-P9-P12).

- *Usage history*: This aspect is not only related to the temporal facet but also to the patterns of usage. For instance, P2 said that he would like to see the usage frequency and other attributes of folders without having to open them, which could be also used as a browsing facet. P5 explicitely said that he would appreciate to have the operating system taking care of the versioning of documents in order to be able to jump back to a previous version, without having to manually use a naming convention. Finally, some people said they would like to have the history of the tasks they performed, to observe how they invested time on specific tasks or projects (P2-P11, indeed P11 does it by hand at the time of the interview).

- *Education and support*: Various people (P6-P10) said that they would like to be more educated on the way to use the existing tools and would like to have support.

3.4 Conclusions

Although people say they do not like doing PIM for itself, because it is a painful and time-consuming task, the study presented above actually shows that people adapt their strategies to punctual changes in their activities or work contexts. They do it as a mean to an end: improve their efficiency and well-being at work. They change strategies for multiple specific reasons: a bad organization hindering to refind information, the discovery of a new technology (a search engine, a more complete or task specific tool, etc.), an excessive amount of documents making classification too time-consuming, the abandon of a useless or ineffective strategy, etc. Strategy changes are grounded, have a scope and require an action.

This study also suggests that what people say about their PI and strategies often differs from the reality of their data and practice. Finally, surprisingly, although people participating to our naturalistic study have spent one or two hours of their working day with us, they have all appreciated spending time to share and analyse their strategies, which counter balances the fact that they do not like doing PIM, as they say.

We present in the following section a software tool making use of information retrieval and information visualization techniques to provide a faceted view on personal information. This tool has been mainly designed to support personal information browsing and searching. Its design was based on a study we performed [7] focusing on how people remember what they have done in the past, where documents are located in their personal information and how they prepare further events. The result of this study motivated the usage of facets (social, temporal and thematic) as natural cues to information. Although the tool has not been designed primarily for this purpose, we believe it could also help to support PIM self analysis and strategy changes.

4. WOTANEYE: PIM AND REFLECTING PI USAGE

The naturalistic study presented above has confirmed the need for people to feel in control of their personal information and their difficulties to match their desire with the way they use and manage personal information in daily life situations. In particular, the study elicited the need for having a tool which reflect personal information usage and for highlighting usage patterns towards strategy evaluation. Further, in another survey we performed on the usage of PIM tools at work [7], it appeared that contextual cues are crucial for re-finding and forming a mental model of personal information, as Lansdale stated [20]. Three facets of personal information stood out from the survey: the temporal facet (when did the information item appear/was modified), the social facet (who is related to this information item), and thematic facet (what is the information item about).

Three observations from the above studies motivate in particular the potential usefulness of our tool: (1) people would like to organize and access PI using natural facets; (2) people would like to keep track of their PI usage history; (3) education and support is requested for PIM.

The tool we developed, called WotanEye, exploits those facets. This section first presents its general architecture and details the way we extract the data that helps us provide a faceted view of information. The section finally presents WotanEye's user interface and use cases.

4.1 Facets Analysis

WotanEye provides a faceted view of personal information. More precisely, temporal, social and thematic facets of information have been chosen because they reflect the way people naturally remember information. In order to build those facets, we need to extract metadata from raw information items.

Basic temporal metadata is easy to obtain for calendar entries (obviously), local files (the operating system stores creation, last modification and last access timestamps) and emails (sent date). Social and thematic metadata can be obtained more easily from emails than from other types of personal information and various studies tend to prove that email constitutes a representative subset of personal information [9]. Moreover, as indexing the whole personal information is a technically-solved problem, we leverage an existing indexing engine, like other researchers in the field do [11, 25]. In our case, we leverage the existing database maintained by Windows Desktop Search (WDS) [21] to keep track of information items. It can be accessed easily through an API using SQL-like queries. However, there is no mean to access the database structure directly; therefore it is not possible to read the detailed textual features extracted by WDS (inverted file along with $tf \cdot idf$ values[1], etc.). This is the reason why, in order to compensate this lack, we perform statistical analysis on textual content and addresses of emails to gather relevant features.

Firstly, temporal metadata can be obtained easily using operating system facilities for all kinds of personal information. In concrete terms, we retrieve the date of text files (last edition date), calendar entries and emails using Win-

[1]In text retrieval, $tf \cdot idf$ is a measure of terms' importance in documents. The better a term characterizes a document, the higher its $tf \cdot idf$ value is. See [24].

dows Desktop Search (WDS) queries constrained on time intervals only (i.e. without querying on a particular keyword).

Secondly, in order to automatically extract the social network of a person, we use all the email addresses that can be found in her email archive. We define a closeness measure between two contact's addresses that is proportional to the number of times these addresses appear together in the headers of emails, yet depends on the number of recipients of these common emails. Indeed if Georges and Nick belong to the same group of people and know each other well, there are chances that they often appear together on emails envelopes, including envelopes with few other recipients. Though, if Georges and Jack systematically appear together on envelopes along with many other recipients, chances are that they do not know each other well but are both members of some mailing-list. Our closeness measure takes these observations into account. Using the closeness measure between all pairs of contacts, we then build a social network graph, the edges of which are weighted by this measure. This social network is user-centered as it reflects the view the user has on her contacts. Using the prefuse [17] toolkit and its facilities for displaying graphs, we visualized this social network using a modified spring-embedder.

Thirdly, in order to build a map of an individual's interests, we analyse the textual features of her emails. On one hand, the most salient themes in her emails express her main interests. On the second hand, discriminant words that appear in emails may help to relate those emails to other information items (documents, meetings, etc.). To extract the textual features of emails, we use the GATE toolkit [12] to remove stop words and perform stemming. Then we build an inverted file using $tf \cdot idf$ values and compute textual similarities between emails using a cosine similarity measure on $tf \cdot idf$ values, in order to be able to group similar emails. Thus the similarity between two emails is proportional to the number of words that appear in both, and strongly depends on the discriminancy of each word (a rare word, e.g. "Cappadoce", has more weight than a very common word, e.g. "meeting"). To group the emails by themes and preview the themes, we perform a hierarchical clustering based on their cosine similarity. The result of the clustering is somewhat average and unsatisfactory. Notably it often happens that a significant amount of emails from the same person are clustered together even if they do not treat the same topic. This can be attributed to several reasons: (1) emails signatures of this person may contain high-$tf \cdot idf$ words ; (2) the person may use regularly a particular vocabulary or slang words ; (3) copies of forwarded or replied-to emails also belong to the body of the new email, even of the new mail is about a new topic (common practice). In order to make the quality of themes extraction better, we plan on applying several methods. First, we could detect forwarded or replied-to parts of an email and lower the $tf \cdot idf$ score of their content. Second, and more important, we could balance the $tf \cdot idf$ scoring by taking into account all emails from a particular sender and filtering out highest-$tf \cdot idf$ words if they do not appear anywhere else in the corpus. This could notably lower the bias introduced by the signature and a too specific vocabulary.

4.2 Interface

As we said above, WotanEye provides a view of personal information based on facets. Our current prototype makes use of two facets already, the temporal and social facet. The thematic facet is still under development. Fig. 1 shows WotanEye main interface window. On the left part of the figure, information items are plotted according to their temporal facet in a calendar-like view. The view represents a full month, where each line shows a week from Monday to Sunday. The week line is divided into four horizontal axis which are used to display from top to bottom the user's appointments (squares), documents (circles), and emails (triangles pointing up are outgoing emails whereas triangles pointing down are incoming emails). A traditional search box is available on top of this temporal view. On the right part of the window lay two other views. On the top right, lays the document-list view. It lists the currently highlighted documents along with further details like the document's name, its author and timestamp. On the bottom right is the social facet view of information, in the form of a social network. People are displayed and clustered by communities. Selecting a group of users in the social network highlights pertaining document in the temporal view. Similarly, selecting documents in the temporal facet view highlights relevant contacts.

4.3 Use cases

WotanEye can be used for a large number of tasks such as finding information items, tasks management, document and email management, giving high/low priority to tasks, browsing information through personal facets, etc. In the following we illustrate the use of WotanEye with two scenarios of interest, the second one being especially relevant to the main topic of this paper.

4.3.1 Meeting preparation

According to our survey on the use of PIM tools at work and related to meeting [7], meeting preparation appears as one of the most important tasks for which people need support. When preparing a meeting, participants need to read related emails, documents, and other past meetings information, but often these information are not linked on their computer and they have to spend a couple of minutes to find them all individually. WotanEye aims at supporting meeting preparation by highlighting all the documents related to a selected event, email or document, as illustrated in fig. 1.

4.3.2 Strategy monitoring

In the study presented in section 3, we observed that people are not always fully aware of the strategy they follow to manage their emails, documents and other information. Some people do not use any folders in their mailboxes, others classify their documents and emails by projects, or by people, other even use temporal folders, etc. It also seems that what people say about their practices does not always match their real practices. Further, we observed that people do adapt their strategies. Not only do we expect WotanEye to support people in better managing their information, but also to help reflect what they are doing, how information items are related to one another using personal facets, and how they can improve their strategy.

For instance, through its temporal view, WotanEye enables to observe how we interact with our personal information in time. From the emails we receive and produce, the document authored or stored, the meetings we have, and their relationship with people, it is possible to observe pro-

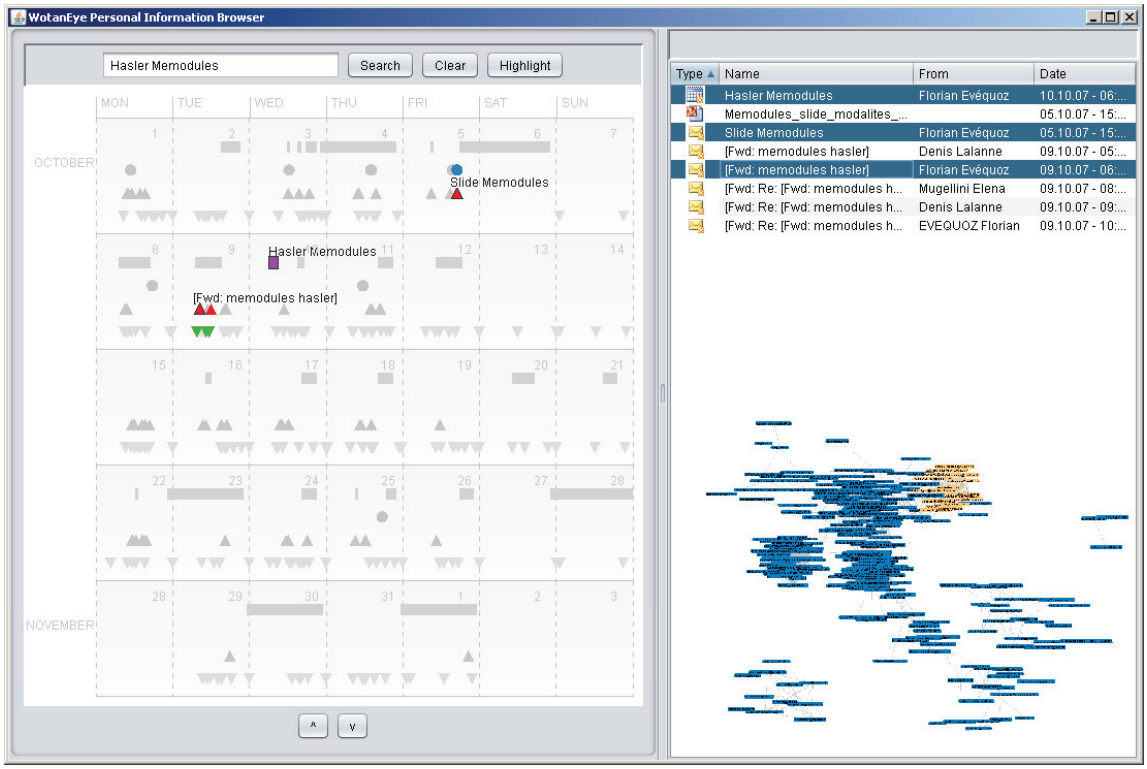


Figure 1: *WotanEye* browser window. The user selected an appointment "Hasler Memodules" (name of a project he is involved in) in the temporal view (left). Similar and corresponding items are highlighted in this view and listed on the top right. The social network view (bottom right) highlights relevant contacts.

ductive periods, repetitive patterns, and the proportion of time we allocate to a specific project, topic or related to specific contacts from our social network.

Currently, WotanEye relies on visualization to reflect PIM usage. It reflects the past activities but does not explicitely suggest better PIM strategies. We do not know at the time of writing if there are better strategies than others, depending on the number and type of documents or activities managed at a time, or on the role of people, etc. WotanEye could be improved to suggest strategies improvements or to give practical advices. However the exact nature of those hypothetical advices remains to be studied.

Furthermore, at the time of writing, no user evaluation of our tool has been performed. We plan to evaluate it with the same population we used for our naturalistic study, and observe if our tool has any impact on the way they manage their information and analyse their practices. The evaluation will take the form of a diary study in which we'll ask people to use WotanEye 5 minutes everyday, during a period of 2 weeks, and write down the things they learnt. The exact form of the diary study is still to be set, though.

5. CONCLUSION

As Barreau stated recently, "people manage their work in unique and creative ways" [3]. We found this assertion particularly true while performing our naturalistic study on PIM. Each participant seemed to have developed a different way to organize their information, in some way expressing not only their personality but also their role and function. Despite this great variety, we found some common patterns related to changes of PIM strategies. Each change has a cause: external, or intentional. Each change has a scope: tool specific or longitudinal (cross-tools). And finally, each change implies an action over the strategy: either to make it more simple, or to enrich it.

What we found particularly singular in our study is that people are not aware of their strategies or often not applying it as they would like. Various people even asked us to help them better organize or train them to do better. But what is the best way, since each way seems suited to a particular person, job or tasks? This is a challenging question yet unanswered.

Being not able to provide to each individual the best way for them to manage their personal information, we thought useful to develop a tool that could help them observe the way they use their personal information, and also making use of contextual browsing facets such as their social network or a temporal view. We present our prototype, WotanEye, aiming at bridging the gap between our mental map of personal information and the way it is accessible on our computer.

Future works are twofold: (1) work on the implementation to have a fully operational prototype of WotanEye to browse documents through temporal and social facets; (2) run a qualitative user evaluation of WotanEye to assess whether it helps people better manage and understand their personal information space or not.

6. ACKNOWLEDGEMENTS

The authors wish to thank the anonymous reviewers for their very helpful and constructive comments.

7. REFERENCES

[1] D. Abrams, R. Baecker, and M. Chignell. Information archiving with bookmarks: personal web space construction and organization. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.

[2] O. Bälter. Strategies for organising email. In *HCI 97: Proceedings of HCI on People and Computers XII*, pages 21–38, London, UK, 1997. Springer-Verlag.

[3] D. Barreau. The persistence of behavior and form in the organization of personal information. *J. Am. Soc. Inf. Sci. Technol.*, 59(2):307–317, 2008.

[4] D. Barreau and B. A. Nardi. Finding and reminding: file organization from the desktop. *SIGCHI Bull.*, 27(3):39–43, 1995.

[5] D. K. Barreau. Context as a factor in personal information management systems. *J. Am. Soc. Inf. Sci.*, 46(5):327–339, 1995.

[6] V. Bellotti, B. Dalal, N. Good, P. Flynn, D. G. Bobrow, and N. Ducheneaut. What a to-do: studies of task management towards the design of a personal task list manager. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 735–742, New York, NY, USA, 2004. ACM.

[7] E. Bertini and D. Lalanne. TotalRecall survey report. Technical report, University of Fribourg, August 2007. available at `http://diuf.unifr.ch/people/lalanned/publis.htm`.

[8] R. Boardman. *Improving tool support for personal information management.* Electrical and electronic engineering, Imperial College, London, 2004.

[9] R. Boardman and M. A. Sasse. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590, New York, NY, USA, 2004. ACM Press.

[10] D. H. Chau, B. Myers, and A. Faulring. Feldspar: A system for finding information by association. In *PIM 2008 - Proceedings of Personal Information Management, Workshop at CHI 2008*, Florence, Italy, 5-6 April 2008 2008.

[11] D. H. Chau, B. Myers, and A. Faulring. What to do when search fails: finding information by association. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 999–1008, New York, NY, USA, 2008. ACM.

[12] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, Philadelphia, July 2002.

[13] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 72–79, New York, NY, USA, 2003. ACM Press.

[14] F. Evéquoz. An ethnographically-inspired survey of PIM strategies. Technical report, University of Fribourg, DIUF, `http://diuf.unifr.ch/people/evequozf/wotaneye_ethno.pdf`, 2009.

[15] G. Fitzpatrick. Applying an ethnographic approach in design. Talk given at 3eme cycle romand in Computer Science (CUSO Winter School), January 2009.

[16] D. E. Forsythe. "It's just a matter of common sense": Ethnography as invisible work. *Comput. Supported Coop. Work*, 8(1-2):127–145, 1999.

[17] J. Heer, S. K. Card, and J. A. Landay. prefuse: a toolkit for interactive information visualization. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430, New York, NY, USA, 2005. ACM Press.

[18] W. Jones. *Keeping Found Things Found: The Study and Practice of Personal Information Management.* Morgan Kaufmann, November 2007.

[19] W. Jones and J. Teevan. *Personal Information Management.* University of Washington Press, Seattle, 2007.

[20] M. Lansdale. The psychology of personal information management. *Applied Ergonomics*, 19(1):55–66, March 1988.

[21] Microsoft. Windows desktop search 4. `http://www.microsoft.com/windows/products/winfamily/desktopsearch/default.mspx`, 2007.

[22] C. M. Naumer and K. E. Fisher. *Naturalistic approaches for understanding PIM,* in *Personal Information Management*, chapter 5, pages 76–89. University of Washington Press, 2007.

[23] T. Pratchett. *Discworld (series of novels).* Corgi Books, 1983-2009.

[24] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval. International Student Edition.* Computer Science Series. McGraw-Hill, Auckland, AU, 1983.

[25] G. Smith, M. Czerwinski, B. Meyers, D. Robbins, G. Robertson, and D. S. Tan. FacetMap: A scalable search and browse visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):797–804, 2006.

[26] S. Whittaker, V. Bellotti, and J. Gwizdka. *Everything through Email,* in *Personal Information Management*, chapter 10, pages 167–189. University of Washington Press, 2007.

[27] S. Whittaker and C. Sidner. Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 276–283, New York, NY, USA, 1996. ACM Press.

The Role Of Places And Spaces In Lifelog Retrieval

Liadh Kelly
Centre for Digital Video Processing
Dublin City University
Dublin 9, Ireland
lkelly@computing.dcu.ie

Daragh Byrne
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Dublin 9, Ireland
dbyrne@computing.dcu.ie

Gareth J.F. Jones
Centre for Digital Video Processing
Dublin City University
Dublin 9, Ireland
gjones@computing.dcu.ie

ABSTRACT

Finding relevant interesting items when searching or browsing within a large multi-modal personal lifelog archive is a significant challenge. The use of contextual cues to filter the collection and aid in the determination of relevant content is often suggested as means to address such challenges. This work presents an exploration of the various locations, garnered through context logging, several participants engaged in during personal information access over a 15 month period. We investigate the implications of the varying data accessed across multiple locations for context-based retrieval from such collections. Our analysis highlights that a large number of spaces and places may be used for information access, but high volume of content is accessed in few.

1. INTRODUCTION

Personal lifelog archives [8] can contain data from many diverse sources, e.g. personal photos, mobile phone SMSs, emails, IM chats, documents created, web pages viewed, etc, created within diverse information access spaces, including home, work, and social locations. Such collections are automatically and passively collected as a user goes about their day-to-day activities thereby offering rich insights into their lives. Given the volume and diversity of content within these archives, there is a clear challenge in the retrieval and location of important and relevant items in response to user queries, and also in presenting interesting data to a subject browsing through their archive. Any additional information which can assist in identifying important items is thus potentially very important. One potential source of useful information is contextual metadata which can be associated with the individual items within a long-term lifelog. One such contextual cue which can be applied is the location of the individual - the spaces in which they use this information. In this paper we explore the role of these spaces within information access as a cue to the potential utility of automatically captured geo-location context data within long-term lifelogs.

Personal Information Management: PIM 2009 ASIS&T 2009 Workshop, November 6-7, 2009, Vancouver, British Columbia

2. BACKGROUND & MOTIVATION

Lifelogs may contain many sources of automatically captured digital content - everything from items read, written, or downloaded; to footage from life experiences, e.g. photographs taken, videos seen, music heard, details of places visited, details of people met, etc, along with details of passively captured location and social context. This context data is important within lifelogs and as such shows its potential utility in retrieval scenarios [6], [14]. The notion of using context to aid retrieval in this and other domains is not new [12], [11] and existing work, such as [2], [5] and [17], have used context data such as location of file, actions performed on file, daylight status, weather and local time to aid file retrieval. Explorations which use manual annotation, in the form of *tagging*, have also been conducted to provide contextual metadata such as people present [4]. Tagging however places undue burden on the user, it is more desireable to employ wholly automatic techniques to provide such information. This has been outlined in our previous work [13].

Current studies in personal information retrieval tend to focus on one type of data, e.g. file, image or mobile data, retrieval. Rich personal lifelogs however are neither confined to the desktop, mobile, audio or image spaces, and are likely to contain data pertaining to the various facets of an individual's life, e.g. work and personal data. Our work emphasises the capture of multiple complementary sources in tandem and is discussed in the subsequent sections. We postulate that the types of context data which will prove beneficial to retrieval may vary across the many personal spaces contained in these rich multi-modal lifelogs. In particular in this paper we explore the utility of recalled geo-location in narrowing a lifelog search space.

3. EXPLORATION

The work presented here is an initial exploration of a single contextual channel, geo-location of the collection owner, and the potential value it adds to the lifelog. While location may be well recalled [14], it is our expectation that the value of context, and in particular location, within retrieval will vary across spaces. In this initial investigation we analyse the relationship between content in the lifelog and the location in which it was created, accessed or reviewed.

3.1 Collection Overview

For the past 15 months we have been engaged in continuous large scale multi-modal lifelog build-up for three subjects. These subjects (1 male, 2 females) are all post-

graduate students within our University. Subject 1 lives in a different region/county to the University. Subjects 2 and 3 live in close proximity to the University. All participants travel to some extent for leisure and research purposes. Our current archives of personal data annotated with rich sources of automatically generated context are much larger and more heterogeneous than typically created digital archives. These lifelogs contain data from a variety of complementary sources, including:

- Desktop Activity: All laptop and PC activity is monitored (every item (email, word document, web page, etc) accessed by the user, with time and duration of access, contents of item, path to item information, etc) for subjects using a combination of MyLifeBits [7], Slife [16] and in-house scripts.
- Passive Capture Media: Continuous passive image capture is enabled through the use of a small wearable device, the Microsoft Research SenseCam [9].
- Mobile Activity Data: Mobile phone activity in the form of call logs are recorded using a proprietary piece of software and SMSs are captured using an application developed in-house.
- Mobile Context Data: This data is captured on subjects' Nokia N95 mobile phones by constantly running the Campaignr software, provided to us by UCLA (USA) [10]. This provides location cues of GPS data, wireless network presence and GSM location data, from which placenames, light status and weather conditions can be derived, and co-present Bluetooth devices, from which people present can be uncovered [15], [1].
- Biometric Information: Physiological and heart rate readings taken from wearable biometric devices [3], which allow us monitor physiological conditions and infer emotional state were continuously captured for a one month period.

For this current preliminary investigation the exploration presented is restricted to PC, laptop and mobile phone activity - in particular SMS's sent and received[1], web pages viewed, emails created or accessed, and computer files created or accessed on laptop and PC for the 15 month lifelog capture period. While we acknowledge that people may consume other forms of textual data on mobile phones, SMSs were the only textual digital content consumed by our subjects. Additionally across the entire populous, other types of digital devices are used by individuals, for example PDAs. Our test subjects did not use such devices.

It should also be acknowledged that for geo-location logging due to device crashing[2], subjects' need to conserve battery life, subjects forgetting to turn on the device and subjects' occasional need for privacy some periods of the 15 month textual lifelogs are not annotated with geo-location. For Subject 1, 73% of their activity is geo-location tagged, for Subject 2 34% is geo-location tagged, and for Subject 3 83% is geo-location tagged.

[1]It was not possible to capture individual accesses to SMS messages using currently available software, hence only time of SMS sending or receiving is captured.

[2]During the first 8 months of the logging, the infrastructure and software was still naesent undergoing iterative change and improvement, in particular the mobile logging software. During this time device & software crashes occured intermittently and went undetected until the subject checked if the software was still running. Following the first 8 months a much more stable platform was available resulting in very occasional device crash.

Table 1: Total number of item creation/accesses captured by subjects on PC, laptop and mobile phone.

Item Type	Subject 1	Subject 2	Subject 3
Laptop			
Webpages	7105	3096	19894
Email	247	542	209
Word Document	2412	0	30
Excel File	2182	22	77
Powerpoint	886	0	18
PDF	663	155	98
Text File	692	1	184
XML File	0	166	18
Code / Language	4768	4	1562
Media	229	16	16
Other	13676	1359	3160
Laptop Total	*32860*	*5361*	*25266*
PC			
Webpages	3494	15781	78375
Email	554	618	11647
Word Document	992	2238	4885
Excel File	1122	406	1529
Powerpoint	537	558	677
PDF	454	1089	394
Text File	401	146	1407
XML File	82	180	685
Code / Language	4421	645	18450
Media	104	13	502
Other	16736	17149	26442
PC Total	*28897*	*38823*	*144993*
Mobile Phone			
SMS	3232	436	3023
Phone Total	*3232*	*436*	*3023*
Total	**64989**	**44620**	**173282**

3.2 Discussion - Collection Contents

As mentioned above, this evaluation is based upon the PC, laptop and SMS content collected within the lifelogs of three individuals over a fifteen month period. Over the course of this period, an individual will encounter a broad range of unique digital content which may be viewed or accessed in isolation or reviewed periodically. To better understand both the content of a lifelog and the prevalence of the various content types the users worked with, Table 1 provides a breakdown of information access across the available devices[3]. It illustrates that there is a huge dominance within the digital landscape of a lifelog for information creation/access to webpages, communication via email[4] (for Subject 3) and SMS messages (for Subjects 1 and 3) and code development (for Subjects 1 and 3).

Taking the information presented in Table 1, serves to

[3]Data type 'other' in Table 1 represents such things as file system accesses for example.

[4]Note for Subjects 1 and 2 little information on email accesses was captured owing to limitation in software.

highlight the different relationships the individuals have with the content housed within their lifelogs. We can see that these individuals have very different personal information management and access strategies. For example, participant 3 consumes high volumes of web pages and is more engaged in coding/development as compared with the other participants. In fact 69% of this participants computer activity is spent engaging in these tasks. While participants 1 and 2 consume more than double the volume of pdf documents of participant 3. Further to this we can also see that the devices employed by the participants have very different roles - with participant 1 favouring their laptop while participants 2 and 3 display a strong affinity to their desktop computer for content access. In particular, participant 1 greatly favours their laptop for Web page access and Word and Excel file creation or access.

3.3 Discussion - Geo-location Tagging

It is particularly important to consider the mobility of these devices as this is extremely pertinent to the utility of location-based context in narrowing the search space in such collections. For example, we can expect a desktop computer to be almost stationary in its location for all of its use (perhaps occasionally it might be moved but this is likely to be a rare event), while a laptop is mobile and therefore can be expected to traverse several locations in its use. In this data analysis we set out to examine the extent to which recalled geo-location can narrow a lifelog search space using the PC, laptop and mobile data in our subjects' lifelogs. The life styles of our subjects afford them certain levels of movement between geo-locations as described in Section 3.1. However, it is acknowledged that while the patterns of movement of our subjects are typical of those of many individuals, they do not represent the entire populous. This study seeks to form an initial exploration of the utility of geo-location in narrowing down the search space for individuals with relatively stationary lifestyles.

The nature of the device is likely to have implications for its relationship to the spaces in which it operates, this is illustrated within Tables 2 and 3. The tables present a list of the volumes of laptop, PC and mobile activity encountered in each 'place' that the subjects were determined to be in over the 15-month period for geo-location tagged items within each subjects lifelog. Here we consider a 'place' to be a unique region + county + city within a given country that the individual was located in. These 'places' were extracted from the available GPS data. Locations have been anonomysed, however encounters within the same country can be discerned by the use of the same prefixing letter.

We explored 'places' of the granularity of country + region + county + city + street. However, when we considered location at the granularity of street level several false locations were noted. Using GPS location alone poses issues with accuracy and achieving sufficient granularity to be useful. However, in the future using detected wifi and bluetooth signals may help overcome this.

As assumed a desktop will largely be confined to a single location as can be seen in Tables 2 and 3. For Subjects 2 and 3 the laptop is also used predominately in just a single location. In contrast to Subjects 2 and 3, Subject 1 has multiple dominant locations for laptop activity. For all subjects it does pentrate many more 'places' but to a lesser extent as shown in Tables 2 and 3. As might be expected, the mobile SMS content is most widely distributed across spaces for all subjects.

Given this information, what implications does it bear for retrieval using context? First the desktop should be considered. Large volumes of information are accessed in principally one static location. This suggests that in situations where an individual seeks to identify an item of interest from their PC that (location-based) context information may offer little assistance. Attempting to weight or filter by location in this circumstance is unlikely to sufficiently narrow the search space or clearly identify an item of relevance. The same will similarly be true for laptop activity for the subjects who have a dominant location for laptop activity. However, in situations where an individual does not recall the device the item was created or accessed on, recalling the dominant geo-location will narrow the search space, albeit not to the same extent as recollection of an item created or accessed in a non-dominant location. Conversely, given that some laptop accesses occur at infrequently encountered locations, when retrieving for such items context may offer real utility. The lack of one dominant location for laptop activity observed for Subject 1 suggests that geo-location will be of greater utility in narrowing the search space for this subject. Subjects 1 use of different spaces for laptop activity also serves to highlight the fact that variations in behaviour can be expected for different subjects.

Finally the spread of encountered geo-locations for sent and received SMSs suggests that location recall would be effective in aiding identification of required SMSs for all three subjects. This finding also suggests that geo-location offers general utility for any mobile phone content, for example photos, tweets, etc.

These observations suggest utility for the inclusion of geo-location context retrieval facilities in lifelogging retrieval systems. However, investigation of the volumes of geo-location recalled for retrieval scenarios and experiments to determine its 'real' utility in improving the detection of required items in lifelog search systems is required. The analysis carried out in this paper provides initial support for investing in such analysis.

4. CONCLUSIONS AND FUTURE WORK

This paper provides an initial exploratory investigation of the role of context data within a lifelog collection and its potential utility in the retrieval of content from these collections. The variation in types of content within a lifelog was illustrated. We then examined the relationship between this content and the 'places' in which it was accessed, through location-context gathered as part of the collection. The relationship between the space of access and the mode of access was further probed. This served to highlight the often implicit relationship between space, place, access medium and personal information. It is clear from the results presented that information access is often confined to one or two dominant spaces, however, mobile access - via laptops or phones - increases the spaces in which our personal media can penetrate. Space or place of access thus has implications on how we might later seek to retrieve that content, particularly should we employ context-cued or -aware approaches.

While we cannot generalise about the laptop, PC and mobile habits of entire populous from the observations made in this paper, the results give insight into the long term laptop, PC and mobile activity for individuals with relatively sta-

tionery lifestyles. They also highlight, even with our small subject set, that variation in the use of spaces is to be expected across individuals.

As part of our future work, we will undertake more detailed exploration to examine how the role of space can play a part within lifelog content retrieval and indeed how our relationship to that space should inform the retrieval place. We additionally plan to further explore how tacit and accurate location cues are to the individual in relation to the dominance of that space within the landscape of their lifelog content.

5. ACKNOWLEDGEMENTS

This work is funded by grants under the Science Foundation Ireland Research Frontiers Programme 2006 and under grant 07/CE/I1147. We would also like to thank the Irish Research Council for Science, Engineering and Technology.

6. REFERENCES

[1] B. Lavelle and D. Byrne and C. Gurrin and A.F. Smeaton and G.J.F. Jones. Bluetooth Familiarity: Methods of Calculation,Applications and Limitations. In *MIRW 2007 - Mobile Interaction with the Real World, Workshop at the MobileHCI07: 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2007.

[2] T. Blanc-Brude and D. Scapi. What do people recall about their Documents? Implications for Desktop Search Tools. In *IUI'07*. Honolulu, Hawaii, USA, January 2007.

[3] BodyMedia - http://www.bodymedia.com.

[4] E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, Flexible Filtering with Phlat - Personal Search and Organization Made Easy. In *CHI 2006: Conference companion on Human factors in computing systems*, pages 261–270, New York, NY, USA, April 2006. Montreal, Quebec, Canada, ACM Press.

[5] D. Elsweiler, I. Ruthven, and C. Jones. Towards Memory Supporting Personal Information Management Tools. *Journal of the American Society for Information Science and Technology*, 2007.

[6] M. Fuller, L. Kelly, and G. Jones. Applying contextual memory cues for retrieval from personal information archives. In *PIM 2008: Proceedings of Personal Information Management, Workshop at CHI 2008*, April 2008.

[7] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. MyLifeBits: Fulfilling the Memex Vision. In *ACM Multimedia '02*. Juan Les Pins, France, December 2002.

[8] J. Gemmell and H. Sundaram, editors. *CARPE 2004 The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, 2004.

[9] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. SenseCam: a Retrospective Memory Aid. In *Ubicomp 2006*, pages 177–193, 2006.

[10] A. Joki, J. Burke, and D. Estrin. Campaignr - a framework for participatory data collection on mobile phones. Technical report, Centre for Embedded Network Sensing, University of California, Los Angeles, October 2007.

[11] G. Jones. Challenges and Opportunities of Context-Aware Information Access. In *Toshiba Corporation Research and Development Center*. Kawasaki, Japan, 2005.

[12] G. Jones and P. Brown. The Role of Context in Information Retrieval. In *SIGIR 2004 - the 27th Annual International ACM SIGIR Conference, Workshop on Information Retrieval in Context*, 2004.

[13] L. Kelly. The information retrieval challenge of human digital memories. In *BCS IRSG Symposium: Future Directions in Information Access 2007*, 2007.

[14] L. Kelly, Y. Chen, M. Fuller, and G. Jones. A Study of Remembered Context for Information Access from Personal Digital Archives. In *IIiX 2008 - 2nd International Symposium on Information Interaction in Context*, 2008.

[15] V. Kostakos, E. O'Neill, and S. Jones. Social networking 2.0. In *Extended abstracts of the conference on Human factors in computing systems*, pages 3555–3560, 2008.

[16] Slife Labs - http://www.slifelabs.com.

[17] J. Teevan, C. Alvarado, M. Ackerman, and D. Karger. The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *CHI 2004*, pages 415–422. Vienna, Austria, April 2004.

Table 2: Total SMS, laptop and computer activity across geo-locations for subjects.

Location				Subject 1			Subject 2			Subject 3		
Country	**Region**	**County**	**City**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**
A	1	a	1							84		
A	2	a	1								1	
B	1	a	1		1							
B	2	a	1		1							
C	1	a	1								10	
C	1	a	2								6	
C	1	a	3								100	
C	2	a	1								3	
C	2	a	2								25	
C	2	a	3								2	
C	2	a	4								31	
D	1	a	1		23							
E	1	a	1					5				
F	1	a	1					1				
F	2	a	1					2				
F	2	b	1	2								
F	2	b	2	1								
F	2	b	3	6								
F	2	c	1		1			1				
F	2	d	1	258	13			4				
F	3	a	1					1				
F	4	a	1					1				
F	5	a	1								18	
F	6	a	1					1				
G	1	a	1	3	2						10	
G	2	a	1		1							
G	3	b	1		7							
G	3	b	2		1							
G	4	a	1	358	314	16918	337	15	12391	10212	1407	117712
G	5	a	1		3						2	18
G	6	a	1	11	3	48					4	
G	6	a	2	25	9	94						
G	6	b	1	4	18	82						
G	6	b	2	1	2	15						
G	6	b	3	18	13	136					11	
G	6	c	1					1				
G	7	a	1	10								
G	7	a	2	2								
G	7	a	3	2								
G	7	a	4	2								
G	7	a	5	42	5							
G	8	a	1	8	3							
G	8	a	2	9	2							
G	9	a	1	2								
G	9	a	2	2								
G	9	b	1				1					
G	10	a	1							46	7	
G	10	b	1								4	
G	11	a	1								2	
G	11	b	1								1	
G	11	c	1	3								
G	11	c	2	10	1							
G	11	c	3		8						10	
G	11	c	4	9065	549	2339				912	239	
G	11	c	5		1							
G	11	c	6	6460	451	323					1	
G	12	a	1		1							
G	12	b	1	1336	66	2668	71	14	2119	1853	49	5950
G	12	c	1							1570	85	3011
G	12	d	1							6		

Table 3: Continuation of Table 2 - Total SMS, laptop and computer activity across geo-locations for subjects.

Location				Subject 1			Subject 2			Subject 3		
Country	**Region**	**County**	**City**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**
G	12	e	1		1		1					
G	12	e	2				56					
G	12	e	3				11	3				
G	12	f	1		1							
G	12	f	2		2		2				2	
G	12	g	1		1							
G	12	h	1		1							
G	12	i	1	72								
G	12	j	1	238	11						10	
G	12	j	2	4765	237							
G	12	k	1				1					
G	12	l	1				2					
G	13	a	1				10					
G	14	a	2		3		5	3				
G	15	a	3		1							
G	15	a	4	1							2	
G	15	a	5								7	
G	16	a	1	2								
G	16	b	1	4	4							
G	17	a	1	2								
G	18	a	1				4					
H	1	a	1	35	18	8	1	3	103		8	36
I	1	a	1		15							
I	1	b	1		1							
J	1	a	1							718	8	
J	1	a	2							2	84	
J	2	b	1								1	
			Total	*22759*	*1795*	*22631*	*502*	*55*	*14613*	*15403*	*2150*	*126727*

The Relationship Between Research and Practice: What Are We Learning About Teaching Personal Information Management?

Deborah Barreau
University of North Carolina at Chapel Hill
100 Manning Hall, CB# 3360
Chapel Hill, NC 27599-3360
barreau@email.unc.edu

Laura O'Neill
University of North Carolina at Chapel Hill
100 Manning Hall, CB# 3360
Chapel Hill, NC 27599-3360
loneill@email.unc.edu

Amanda Stevens
University of North Carolina at Chapel Hill
100 Manning Hall, CB# 3360
Chapel Hill, NC 27599-3360
acsteven@email.unc.edu

ABSTRACT

Research should inform practice as well as guide developers in designing better tools. Much of the emphasis on Personal Information Management (PIM) research has been developing better tools for finding information among expanding work spaces. This research attempts to look at PIM from a pedagogical perspective – what we have learned that may inform our teaching. Web sites were examined and scholars were surveyed. Our findings suggest that PIM is not yet achieving significant attention in the information and library science (ILS) curriculum.

Author Keywords
Personal information management; User education.

ACM Classification Keywords
H.1.2 [Information Systems]: User/Machine Systems: Human Information Processing

INTRODUCTION

Research into the PIM behaviors of individuals in various contexts can be traced from the early 1980's (Cole, 1982; Lansdale, 1983; Malone, 1983). Much of the research has been conducted and used by software developers and computer scientists to design and implement useful tools for personal computers (Dumais et al., 2003; Bellotti et al., 2004). A search of the ACM Digital Library using the keywords "personal information management" returns more than 500 papers and some 100 authors who have explored aspects of PIM. A search of the information and library science literature results in more than 100 papers, many overlapping with the ACM study, but many from different sources as well. The information and library science field has a stake in this research, but with different questions and different concerns. Information professionals, including archivists, information architects, and librarians are concerned with human information behavior. Helping individuals to manage, process, use, and preserve their information for the various contexts in which they work is at the heart of PIM from the ILS perspective.

The purpose of this study is to examine the status of PIM in information and library science programs. PIM is still an emerging area of inquiry, but it is mature enough after more than twenty-five years to have some influence on the ILS curriculum. If we define PIM broadly, as Jones (2007) does, emphasizing "the organization and maintenance of personal information collections in which information items, such as paper documents, electronic documents, email messages, web references, handwritten notes, etc., are stored for later use and repeated re-use," then its connection to the information and library science field is clear.

BACKGROUND

PIM researchers have studied behaviors in a variety of settings, yet we have little to recommend in terms of formula or procedures that will ensure more effective PIM. One explanation for this lack of convention is simply that we are so diverse, our needs so varied, and our situations and styles so unique that one solution simply cannot fit all. We are left with information about PIM behavior that is of sociological and psychological value, but offers little to inform future information architects, analysts, archivists, and librarians.

A second reason for our lack of progress in defining and prescribing successful PIM practice is the emphasis on finding. Research has shown that people are reluctant to develop and maintain formal organization system for their personal information. Consequently, data stored in our personal workspace is growing exponentially and without good search tools, much of this data may become quickly inaccessible. However, finding is not the only function related to PIM – we interact with information in a variety of ways, including being reminded of things that we need to do, encountering things that we have forgotten, rereading or reviewing past work, converting and transferring data to new hardware or software environments, and even organizing and purging directories of unwanted files. When people are interviewed about their behaviors, they often ask about how they might work more effectively.

The only study of PIM in the university curriculum that we could locate was undertaken and reported in 1985. At that time five library and information science schools offered PIM courses to undergraduate students, and four schools incorporated aspects of PIM in other classes. Goals for those courses included "the role information plays in society, how to identify information resources within and outside the library, and the use of the personal computer as an information resource both for their college studies as well as for life-long learning" (Jahoda & Brockmeier, 1985). The questionnaire in that study was sent to the deans of library schools. While many recognized the value of such classes both for the students and for potential recruitment of those students to graduate programs, the researchers found there was little demand for such a course and some schools, which previously had offered the course, dropped it due to low enrollment.

In 1985, computers were becoming more accessible to individuals, particularly in university libraries and computer labs, but in the home as well. Research into personal information management behaviors, including studies of human-computer interaction, emerged at this time. Twenty years later, technology has advanced to the point that our cell phones have more memory and processing capability than the personal computers of 1985, and we have the ability to make use of a variety of devices for storing and using information from wherever we are. As technology advanced, needs to improve the usability of these devices and to understand their benefits and consequences for our personal and collective behaviors have stimulated research. An international research community has emerged, led by William Jones, holding workshops approximately every 18 months since 2005 to share ideas and raise questions. One goal of *this* study is to assess the impact of this research community on professional education.

METHODS

The purpose of this study is to examine the state of PIM in the curriculum. To achieve our purpose, we decided to look at what is being taught related to PIM and to touch base with researchers to assess their feedback on pedagogical issues. This research involved the following steps:

(1) Examine the Web sites of the information and library science schools, and the schools of PIM researchers in other fields, to identify courses devoted to PIM or that cover PIM issues.
(2) Gather related syllabi and analyze them for target audience, topical content, and skill set.
(3) Identify PIM researchers from colleges and universities who also have teaching responsibilities and survey them concerning how they address PIM in their teaching.

There is some precedence for these methods. Various aspects of the ILS curriculum have been studied in the past including such areas as cataloging, digital libraries, knowledge management, and digital preservation (Turvey & Letarte, 2002; Hsieh-Yee, 2004; Gracy & Croft, 2006). The purpose of these studies included assessing trends and objectives, and in some cases, making practical recommendations. Our study utilizes similar data collection strategies to assess how PIM research is informing ILS teaching and practice.

Although we focus primarily on schools of information and library science, we cast our net a bit wider, looking at schools where PIM researchers were found (some in computer science departments, for example) as well. We could not approach these departments systematically because we did not have a comprehensive list of those programs as we did for the information schools, but we have included analysis of courses in those departments when we found them.

Research Questions

- Is there evidence that the LIS curriculum has been informed by PIM research?
- What is being taught currently with respect to PIM?
- Can we identify PIM-related concepts, topics, and findings that should be included in professional education?

RESULTS

Examination of the School Web Sites, Catalogues, and Course Syllabi

Examination of the information and library science school Web sites (and the Web sites of the researchers identified in our sample described below) resulted in the identification of six programs with a PIM course, and nine programs offering some aspect of PIM within the broader context of other courses. The PIM courses appear to be special topics courses that are not part of the regular curriculum. Two of the courses conform to the 1985 focus for PIM, a class designed to help undergraduates acquire, organize, and present information. Four of the classes take a broad approach, providing opportunities for students to explore and conduct their own research or develop PIM tools. The latter courses are designed for advanced (mostly graduate) students.

The three institutions offering research-oriented courses in PIM are, not surprisingly, institutions where there are five or more scholars conducting PIM research. The institutions include Virginia Tech, the University of Washington, and the University of North Carolina at Chapel Hill. Of the nine institutions where PIM was clearly identified as a sub-topic within a broader course, only two (Rutgers and the University of Hawaii) have faculty engaged in PIM research. Courses that cover some aspect of PIM include information retrieval, collection development and management, knowledge management, basic computer skills, information organization, and computer-mediated communication. Relying upon the Web for this analysis is problematic, for reasons that will be discussed later.

Identification and Survey of Researchers

A search of the PIM literature in Library and Information Science Abstracts, the ACM Portal, the ISI Web of Science, and prior PIM Workshops identified 237 researchers (authors and co-authors) who have published papers related to personal information management, and 140 of these are teaching faculty.[1] Researchers represent 15 countries and 54 academic communities, and cover such disciplines as information and library science, computer science, psychology, education, management, and engineering. We verified institutional affiliations and electronic mail addresses, prepared a survey, obtained approval from our Institutional Review Board for Human Subjects Research, and conducted the survey in early March 2009. Responses identified courses that our Web analysis missed and provided additional information on topics covered, resources used, and insight into why PIM may not have been offered. We received 48 responses (34.3%) of the 140 surveyed.

Results confirm the findings from the Web analysis and identified additional courses that address PIM in some manner. Table 1 illustrates responses to questions about course offerings. The PIM courses are not required – all are special topics or seminars that are offered irregularly. However, 6 of the courses that include aspects of PIM are required by their respective departments.

	Yes	**No**	**Total Responses**	**No response**
Teach a course on PIM	6	42	48	0
Include PIM in a course	16	25	41	7
Plan to offer a course on PIM	3	13	16	32

Table 1. Respondents who address PIM in some way.

Respondents were asked why they thought PIM courses were not offered at their respective institutions. Three

[1] Many PIM researchers can be found in the research and development institutes of major corporations, including Microsoft Research, Xerox Parc, IBM Watson Labs, and Google. These researchers were not included in our study.

indicated that there is not enough interest in the subject, 7 responded that PIM is not perceived as important enough to be the focus of a course, 2 indicated that there were not enough resources to offer the class, 7 responded that there are not enough faculty to cover PIM as an area of study, and 7 indicated other reasons. Responses that fell into the "other" category were varied. In fact, two persons indicated that their programs have a PIM class, but they do not teach it. Another participant said, "I think it's spread out over various classes. Also, the main principles are taught but are not called PIM (for example building a database for a personal collection, indexing a personal collection, etc.).

Among the courses that participants identified as covering some aspect of PIM are:

- Advanced Seminar in Interactive Information Retrieval,
- Artificial Intelligence II,
- Computer Privacy and Security,
- Datamining in Education,
- Digital Libraries,
- Human-Computer Interaction, Design, People and Security,
- Human Information Behavior,
- Information Retrieval,
- Internet in Higher Education,
- Introduction to Human-Computer Interaction,
- MSC HCI (use for research/design projects),
- Multimedia Content Production,
- Organization of Information,
- Personalized Information Delivery: Information Filtering,
- Public Libraries,
- Retrieving and Evaluating Electronic Information,
- Software Engineering: Advanced Topics in Software Systems, and
- Survey of Human-Computer Interaction Research.

We asked instructors to identify the important readings assigned in their PIM courses. Although most were aware of PIM-related texts, including *Personal Information Management*, and *Keeping Found Things Found*, instructors typically assigned a variety of articles, and two of the instructors shared their entire reading list.

Finally, we asked participants what they thought students should know about PIM. This question is central to understanding the relationship between research and practice. We have grouped responses into three broad categories: (1) structure and use of information, (2) social and psychological framework for PIM study, and (3) information technology. These are discussed below.

Structure and Use of Information

Information structure and use is largely the concern of the information and library science field. These answers are characterized by a concern for differentiating between Information Management broadly, and Personal Information Management specifically, and for understanding the relationship between them. Responses included the following:

- "Understand what is meant by personal information, and how PIM differs from traditional IR/management/learning techniques by taking into account the particular relation of users and their information, that has special meaning to them. They should also be aware of the issues regarding the design of PIM applications, and the challenges this poses regarding indexing and access to the data, browsing, searching and visualizing personal information."
- "Understand information and the various ways in which it is used. Understand various approaches and strategies in PIM and support for them (e.g. piling, filing, searching, organizing, etc.)"
- "Finding, re-finding, weeding/pruning, tagging and other socially derived naming and categorizing (PIM in community)"
- "List the types of information that must be organized and managed, the sources of overload, and our own information management challenges."
- "In teaching public libraries, the goal of introducing PIM is to encourage students to think about ways in which libraries can connect with individuals in their personal information space in order to support information literacy education in the area of PIM and to create connections between personal information and community creation. Explain 'insider' comments. Ensure that your whole audience understands any reference whose meaning you do not describe (e.g., do not assume that everyone has used a Macintosh or a particular application)."
- "PIM as part of user modeling; implications for the design of information filtering systems, recommender systems; design of personal digital libraries and personal websites."

- "Identify the characteristics of personal collections that complicate PIM (distributed collections, collections fragmented by form, size, etc.) and the implications for both current refinding and long-term keeping. Draw connections between their professional interests and PIM (for example, interface design and opportunities for building better PIM tools; digital archives and issues of personal digital collections; librarian (various flavors) and possible bibliographic instruction, service, and resource design issues to help people assess/do PIM). Think critically about PIM systems – their own and those of others. What works? What doesn't? What is good enough? What works now, but might be problematic in the future? What, given the realities of work and life, could be improved?"

Social and Psychological Framework for PIM Study

Those who identified social and psychological frameworks for study emphasize the importance of research and indicate such perspectives as "serious leisure" and "everyday life information seeking" as possible frameworks for PIM analysis. The connection between PIM and other disciplines, including computer science, psychology, education, information and library science, sociology, and business was offered as a possible framework for PIM study. Some respondents were interested in exploring the history of PIM as a matter of popular and public interest to get a better understanding of its utility as an area of inquiry. Others suggested that a communications approach is useful to understand how people share information, or a consumer approach to understand how people identify and use information.

Most of these responses focus on the importance of communication and information sharing as being the critical concern for why PIM matters. Information is shared in the workplace, in social networks, or in our close, personal relationships, and these contexts may offer their own frameworks for addressing issues of portability, security, and rights management.

Information Technology

Many of the responses focused on the importance of designing, selecting, configuring, evaluating PIM tools. One respondent indicated that it is important to "design and develop a prototype to solve a specific problem in PIM."

DISCUSSION

There are always limitations to studies such as this one. Analyzing Web sites will fail to identify information that is not reported or will misidentify information that is out of date. Additionally, course descriptions are often incomplete, course syllabi are not always available, international sites may not be translated to English, and schedules often fail to list the frequency with which a course is offered. Similarly, surveys restrict options so that respondents are forced to select an answer that may not precisely reflect their situations. Any discussion or analysis of the above results must consider these limitations.

Our study posed three questions. The first was whether information and library science education has been informed by PIM research. While PIM-related issues are clearly recognized and discussed within a variety of classes, the idea of PIM as a significant area of inquiry has not been widely embraced. This is not to suggest that it should – only that we as researchers should be asking, and perhaps promoting, lessons learned that may benefit others if incorporated in our curricula. Early on, the focus of PIM was the design of better tools. The tools have improved, but our practices have not. In fact, better tools may have contributed to poor practice by making it easier to find things no matter how badly one has organized them and by providing the illusion that backing up one's files is somehow preserving them. We will continue to build better tools, but these results suggest that we also need to focus on the management of personal information as well.

The second question addressed what we are currently teaching. The information and library science curriculum has not addressed PIM systematically. Courses are offered in programs where researchers are concentrated, but PIM-related readings, PIM-focused projects, and topical lectures are included in a variety of programs and courses.

The third question concerned the concepts, topics, and findings that should be included in the curriculum. Three perspectives were identified as key PIM-related approaches: the structure and use of personal information, social and psychological perspectives of PIM, and information technology to support PIM. Although there is no consensus from our participants concerning what specifically should be addressed, there is considerable overlap among assigned readings, suggesting some commonality in what is studied. Researchers working in

fields as diverse as digital preservation and collaborative information retrieval have begun to relate the PIM literature to their course content. Examining the assigned readings in the context of the course objectives and syllabi may provide a better indicator of the specific, valued concepts and relationships.

SUMMARY

Our assessment of PIM in the curriculum raises questions about the relationship between research and teaching. It has always been true that programs vary in part on the basis of strengths and interests of faculty. Those of us with the responsibility for training information professionals have a broader concern, specifically what we are learning from our research that can and should inform our students. After more than 25 years of inquiry, it is time that we articulate the message, if there is one, about how PIM should be addressed in the curriculum.

Future work will include a follow-up conversation with teaching faculty to assess specifically how PIM fits in the broader curriculum, and a study to assess whether there is a relationship between PIM behaviors of students and academic performance. It would be helpful to expand this work to other fields as well. In the information and library science field, we have seen movement from mediated information seeking toward self-serve applications. It may be even more important than ever to understand what people are doing with information, and how they are doing it, to design more effective information products and services.

REFERENCES

Bellotti, V., Dalal, B., Good, N. , Flynn, P., Bobrow, D.G. & Ducheneaut, N. (2004). "What a to-do": Studies of task management towards the design of a personal task list manager. *Proceedings of CHI 2004*, 735-742.

Cole, I. (1982). Human aspects of office filing: Implications for the electronic office. *Proceedings of the Human Factors Society*.

Gracy, K.F.and Croft, J.A. (2006). Quo vadis, preservation education? A study of current trends and future needs in graduate programs. *Library Resources & Technical Services, 50 (4*):274-288.

Hsieh-Yee, I. (2004). Cataloging and metadata education in North American LIS programs. *Library Resources & Technical Services, 48(1):*59-68.

Jahoda, G. and Brockmeier, K.C. (1985). Personal information management courses offered by library schools. *ASIS Proceedings 1985*, 217-18.

Jones, W. (2007). *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Morgan Kaufmann.

Lansdale, M. (1983). The psychology of personal information management. *Applied Ergonomics, 19:*55-66.

Malone, T.W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems, 10:*99-112.

Turvey, M.R. and Letarte, K.M. (2002). Cataloging or knowledge management: Perspectives of library educators on cataloging education for entry-level academic librarians. *Cataloging and Classification Quarterly, 34 (1/2):*163-185.

Appendix. Survey Questions

1. Do you teach a course on Personal Information Management (PIM)?
   ___Yes
   ___No (skips to question 5)

2. At what level is this course taught?
   ___Undergraduates only
   ___Advanced undergraduates and graduates
   ___Graduates only
   ___Continuing education
   ___Other (Please indicate: _____)

3. Is this required coursework or an elective?
   ___Required
   ___Elective

4. When you cover PIM in your teaching, what readings do you assign? Please list the textbook and/or the most important assigned readings. (Open-ended response, text box provided)

5. Do you incorporate PIM into any courses that you teach?
   ___Yes
   ___No (skip to question 7)

6. Into what other courses do you incorporate PIM? Please list courses.

7. Do you teach full-time or part-time?
   ___Full-time
   ___Part-time

8. If your institution does not offer courses in PIM, what are the likely reasons? (check all that apply)
   ___Not enough student interest
   ___There is not enough content to warrant a separate course on PIM
   ___Lack of resources (teaching, classroom space, or funding)
   ___Not relevant to the school's degree programs
   ___Other (specify)

9. Does your institution plan to introduce PIM coursework in the near future (1-3 years)?
   ___Yes
   ___No

10. Please indicate in the space below what students should learn, or should be able to do, as result of taking a course on Personal Information Management. (Open-ended response, text box provided)

11. May the investigators of this study contact you or a representative of your institution again if there are any follow-up questions?
   ___No, I would prefer not to be contacted again.
   ___Yes, I (or a representative of my institution) may be contacted if there are follow-up questions. Please contact _____ at the following address, phone number, and/or e-mail:_____
   _____

Metaphor Analysis as a Method for Holistic PIM Research

Christina M. Finneran
School of Information Studies
Syracuse University
Syracuse, NY
cmfinner@syr.edu

1. Abstract

This paper introduces metaphor analysis as a potential method for studying personal information management, particularly as the nature of "personal information collections" (Jones, 2008) is becoming more complex with Web 2.0 and cloud computing. Metaphor analysis discovers underlying metaphors, and potentially multiple perspectives, of a phenomenon. For research, the evaluation of metaphors is primarily how effective they are as an analytical tool. An exploratory study is described to demonstrate the use of metaphor analysis for PIM research.

2. Introduction

The growing personal information management (PIM) research challenge is understanding PIM across the multitude of information channels and applications in which the user interacts (Jones & Bruce, 2005). Clifford Lynch (2003) encourages research "on the individual user, workgroup, or intellectual community of practice 'in the abstract' as a free-standing entity that chooses to employ (or not employ) a range of systems and sources (rather than in relation to a specific system)."

This is not a new idea, though the environment may be. Zweizig's 1973 dissertation asserts that we "have too long focused on the user in the life of the library. We need instead to focus on the library in the life of the user." The difficulty of conducting holistic research of users' behavior with information, particularly information as amorphous as personal information, has precluded more progress in this area.

We need to seek out new methods to investigate PIM from the user perspective. One potential method, which may prove fruitful, is metaphor analysis.

3. Metaphor Analysis

Metaphor analysis has been applied in organizational science and information systems literatures. It is particularly useful when a researcher seeks to provide a rich understanding of a phenomenon. Metaphor analysis enables the study a complex phenomenon as a coherent whole.

The essence of metaphor analysis is seeking to identify underlying metaphors of the phenomenon. A metaphor projects one schema from source domain onto another schema from the target domain (Indurkhya, 1992); the target domain is the researcher's interest. Thus, the researcher tries to identify source domain schemas that are applied to the phenomenon of interest. By discovering the metaphors the researcher increases her/his understanding of the phenomenon and the multiple perspectives involved with the phenomenon.

One of the unique strengths of metaphor analysis is that it enables, and even promotes, multiple perspectives of a single phenomenon. None of the metaphors are truer than the other. The metaphor lens merely highlights one aspect and hides others while the researcher tries to understand the one in focus. Most phenomena will have multiple metaphors. These metaphors help individuals, and the organizations they are a part of, understand complex phenomena. For example, if a group understands the systems development life cycle as a game, this game metaphor helps them conceptualize their role. They will try to win by following the rules and may think of the users as the opponents (Kendall & Kendall, 1993).

Koch and Deetz place interpretive techniques, such as metaphor analysis, within the context of inquiry. "Interpretive research methods do not add new facts to a cumulative base of knowledge. Rather they situate or contextualize bases of knowledge by explicating the implied possibilities inherent in current situations and endeavors (Koch & Deetz, 1981)." It is this context and rich understanding that make metaphor analysis a potentially exciting tool for researchers, especially in the area of PIM where context is so essential to understanding.

4. Metaphors

Metaphor is defined as, "an unconventional way of describing (or representing) an object, event or situation (real or imagined) as another object, event or situation (Indurkhya, 1992)." According to Turbayne (1970), a metaphor has three life stages. The first stage is when a metaphor is first introduced; it is rejected because literally it does not seem congruent. Stage two of a metaphor is when users suspend their disbelief and draw useful comparisons between the source and target schema. As the metaphor is adopted and used more, it will progress to stage three. In stage three the metaphor takes on a literalness, where the users do not consciously compare the schemas. The source and target are now associated unconsciously.

Lakoff & Johnson (1980) discuss the metaphor of "argument is war" which has reached stage three. "Our conventional ways of talking about arguments presuppose a metaphor we are hardly ever conscious of. The metaphor is not merely in the words we use – it is in our very concept of an argument." In their seminal book, *Metaphors We Live By*, Lakoff & Johson (1980) explain how metaphors are ubiquitous not only in our language, but, more importantly, in our conceptualizations.

Metaphors focus on particular aspects of a phenomenon. To use Morgan's (1997) well-developed metaphor for an organization, when we conceptualize an organization as a machine, we will think of efficiency as the goal and departments as interlocking parts of the centralized whole. However, this view is hardly complete. Lakoff & Johnson (1980) state, "In allowing us to focus on one aspect of a concept…a metaphorical concept can keep us from focusing on other aspects of the concept that are inconsistent with that metaphor." Clearly, there is no one complete metaphor, rather many metaphors should be considered and evaluated by if they are useful or if they are dominant within a particular subculture.

An empirical study on management problem-solving showed that depending on which metaphor the organization used to describe itself, subjects formulated dramatically different problems and solutions from the same problem statement (Boland & Greenberg, 1988). Subjects that were introduced to the organization as an organism focused on environment, growth, and decentralization, whereas those introduced to the organization as a machine focused on centralization and controlling growth.

The metaphors themselves do not illustrate the similarities between a source and a target domain, but actually create the similarities in a person's mind (Lakoff & Johnson, 1980). The person, thus, becomes aware of the similarities that may have always existed. Further, they ignore dissimilarities that would weaken the comparison (Morgan, 1997).

The power of metaphors, within both poetry and theory, is that they are vivid, compact, and they overcome the inexpressibility of unfamiliar phenomena (Ortony, 1975). When a researcher states that technology is considered to be like magic in subcultures of an organization (Kaarst-Brown & Robey, 1999), the image is vivid and lasting. Also, it is a rather compact way to portray the complex interactions between that subculture and technology. If a college student, who had never worked in an organization before, heard that the organization treated technology like magic, s/he would be able to understand the interaction even though s/he had no direct experience working.

Poetry uses metaphors to generate a gestalt with fused images, associations and emotions. Contrarily, with theory, metaphors provide a rational and somewhat reductionist understanding (Inns & Jones, 1996). With poetry it is common enough, and sometimes welcome, if readers have their own unique interpretations that are inconsistent with others'. However, when building theory, we need to ensure a degree of mutual understanding to have reliable findings (Inns & Jones 1996). Metaphor analysis in research does not aim to generate metaphors for the sake of inspiring creativity, but instead aims to discover metaphors that exist underneath the surface to enhance understanding.

Not every metaphor is useful. Evaluating metaphors within research is primarily how effective they are as an analytical tool. Does it increase understanding for the theorist? Does it provide insight for the people being studied? Literalizing a metaphor too much will backfire; a metaphor loses its power when it is forced.

Metaphors can serve the following functions within qualitative data analysis: data reducing devices, pattern making devices, de-centering devices, and theory development aids (Miles & Huberman 1984). Metaphors reduce data by enabling the researcher to say in a single word or phrase what one book of field notes might indicate. Metaphors can help the research find patterns by abstracting each particular situation into a similar conceptual understanding. The de-centering occurs when the researcher distances her/himself from the immediate phenomenon and tries to find a metaphor that may be completely novel to his/her past conceptualizations. Lastly, as a tool for abstraction, metaphors can help researchers develop theories, by combining reason with imagination (Miles & Huberman 1984).

5. PIM Exploratory Study

With the growth of Web 2.0 applications and cloud computing, limiting the study of PIM to the "personal information collections" (Jones, 2008) that an individual has complete control over has become more problematic. In real-world projects, individuals have a cadre of information items which they may consider their "own" at moments in time, though they are controlled partly or wholly by others (e.g., Google docs, wikis, project websites, general web). Metaphor analysis can help researchers develop new theories in this changing PIM environment and uncover user's conceptual thinking about the "personal information collections" they use.

This exploratory PIM study seeks to understand how students use a web-based course site (Blackboard), which they have minimal control over, to do their work for a college course. If the syllabus, assignments, and readings are on the Blackboard course site, do students merely

access the information on demand when they need it? Do they download the materials to manage these materials in their own "personal information collection" (in digital or paper form)? By using metaphor analysis, the researcher is able to identify the multiple perspectives the students have toward Blackboard and the PIM behavior shared by those who have the same metaphor.

Three college courses from different disciplines were selected and 5 to 6 students from each of the courses were recruited for a total of 16 participants. The students were interviewed twice during the semester and asked to give a guided tour of their personal information collections, including the Blackboard course site. The transcripts of the interviews were analyzed for metaphors-in-use and overarching metaphors that related to Blackboard.

The shared Blackboard metaphors that emerged are described below. Though sets of students shared these metaphors, the affective feeling toward the metaphor was not always the same. With the "Self-service Station" and "Tutor" metaphors, the PIM behavior depended on whether the student had a positive or negative association with the metaphor.

Self-service Station: Students felt that they now had to go to Blackboard once or twice a week to fill up and print materials for their own "personal information collections." They appreciated the autonomy, where they could decide which information items to access, and the anonymity of not having to ask the professor for something if they lost it. The majority tended to miss "full service" where the instructor would print and handout the material for them in class. The degree of loss varied from those who were resigned to it because it made the instructor's life easier to those who felt like the extra work was a "nuisance." The PIM behavior of these students was to acquire the material from Blackboard and integrate it into their personal paper systems (e.g., binders, folders). A minority of students who exhibited this metaphor had purely positive associations with the "self-service." These students preferred to use the digital versions of materials and were more organizing-neutral.

Day Planner: Students used Blackboard to identify what was due when, to see what was coming up, and to manage the multiple aspects of the course. These students tended use Blackboard in lieu of their paper syllabi. This metaphor was more common in courses where the instructor had a very clear structure to the course and his/her Blackboard site. Also, the students tended to be those who were organizing-neutral and had less of their own "planning" systems.

Tutor: Students saw Blackboard as a tutor to help them with assignments. They visited Blackboard when an assignment was due or exam was pending, and felt that Blackboard had all of the materials needed for them to succeed in the course. A few students had negative reactions to this extra content. They felt Blackboard was like a **Helicopter Tutor**, overwhelming them with so much material that they felt an added pressure that they "shouldn't screw up" and they were disempowered from "rigorous engagement" with the course content. These students "skimmed" the material online and did not acquire many items for their personally-managed collections. Other students had positive associations with Blackboard as a Tutor and incorporated the additional study materials into their own collections.

Insurance: Students did not actively use Blackboard, or large portions of the Blackboard course. Nonetheless, they felt less anxiety knowing it was there in case they needed some information, especially at 2am when they could not reach a friend or professor. These students did not feel that it mattered that they did not use Blackboard.

Several students had more than one metaphor for the particular Blackboard course site, demonstrating that multiple perspectives may exist within a single "personal information collection." The affective association with the metaphors and their distinct PIM behavior was a new discovery.

6. Summary

When trying to understand the complex nature of PIM in a constantly connected world with multiple information channels and applications, metaphor analysis can be a worthwhile methodological tool. Metaphors can serve as a data reducing device while retaining richness with holistic user-based studies. Further, metaphors can be used to spark new theories.

7. References

Boland, R.J. & Greenberg, R.H. (1988). Metaphorical structuring of organizational ambiguity,. In *Managing ambiguity and change*, ed. L.R. Pondy, R.J. Boland, & H. Thomas, pp. 17-36. Chilchester: Wiley.

Indurkhya, B. (1992). *Metaphor and cognition: An interactionist approach*. Boston: Kluwer.

Inns, D.E. & Jones, P.J.(1996). Metaphor in organization theory: Following in the footsteps of the poet? In *Metaphors and organizations*, ed. D Grant, C Oswick, pp. 110-26. Thousand Oaks, CA: Sage.

Jones, W. (2008). *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Burlington, MA: Morgan Kaufman.

Jones, W., & Bruce, H. (2005). *A Report on the NSF-Sponsored Workshop on Personal Information*

Management, Seattle, WA, 2005. Retrieved, from the World Wide Web: http://pim.ischool.washington.edu/final PIM report.pdf

Lynch, C. (2003). Colliding with the Real World: Heresies and Unexplored Questions about Audience, Economics, and Control of Digital Libraries. In N. A. V. House & A. P. Bishop & B. P. Buttenfield (Eds.), *Digital Library Use: Social Practice in Design and Evaluation* (pp. 191-216). Cambridge, MA: The MIT Press.

Kaarst-Brown M.L. & Roby D. 1999. More on myth, magic and metaphor: Cultural insights into the management of information technology in organizations. *Information Technology & People* 12: 192-217

Kendall J.E., & Kendall K.E. (1993). Metaphors and methodologies: Living beyond the systems machine. *MIS Quarterly* 17, 37-47.

Koch S. & Deetz S. (1981). Metaphor analysis of social reality in organizations. *Journal of Applied Communications Research* 9, 1-15.

Lakoff G. & Johnson M. (1980). *Metaphors We Live By*. Chicago, IL: The University of Chicago Press.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Morgan, G. (1997). *Images of Organization*. Thousand Oaks, CA: Sage Publications.

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory* 25, 45-53.

Turbayne, C.M. (1970). *The myth of metaphor*. Columbia, SC: University of South Carolina Press.

Zweizig, D. (1973). *Predicting amount of library use: An empirical study of the role of the public library in the life of the adult public.* Unpublished Dissertation, Syracuse University, Syracuse, NY.

Position Paper

Why Personal Information Management (PIM) Technologies Are Not Widespread

And What to do About It

Karl Voit
Institute for Software Technology (IST)
Graz University of Technology
Karl.Voit@IST.TUGraz.at

Keith Andrews
Institute for Information Systems and Computer Media (IICM)
Graz University of Technology
kandrews@iicm.edu

Wolfgang Slany
Institute for Software Technology (IST)
Graz University of Technology
Wolfgang.Slany@tugraz.at

ABSTRACT

Users of computer systems create and store valuable personal information in files, email folders, and bookmark collections. For decades, the main principle of interacting with files, emails, and bookmarks has remained unchanged: hierarchical directory trees with standard (Windows Explorer style) browsers.

Users often have problems both in classifying new items and maintaining a classification hierarchy as such. With files, emails, and bookmarks, users often end up maintaining three parallel classification hierarchies, one in each tool. Over the past thirty years, a number of alternative personal information management (PIM) tools have emerged, but the typical user is still faced with hierarchical directory structures.

This position paper addresses some of the reasons why modern PIM tools are not widespread and proposes a set of eight requirements for future PIM tools.

Keywords

information retrieval systems, file storage, hierarchical structures, dynamical structures, tagging, faceted search.

1. INTRODUCTION

Much progress has been made in the development of new personal information management (PIM) tools and ideas over the past three decades. However, many of these have failed to make it beyond the research laboratory and onto the PCs of typical computer users. A number of factors may lie behind that.

1.1 How Users Organise

Several studies have investigated user behaviour when organising information in paper-based offices. Malone [15] studied user behaviour regarding information in a physical office environment. He identified *files* (explicitly titled and logically arranged collections of information, for example in folders or binders) and *piles* (untitled piles of information arranged by physical location) as the main schemes employed by users to organise their information.

Lansdale [13] makes the point that users make use of piles to compensate for the difficulty of classifying (filing) things: "To avoid the process of classification, ..., he puts objects in a particular place. With this he forgoes the opportunity to retrieve the document by any simple classification-based search." [13, page 56]. As both Malone and Lansdale argue, and will be seen later, it is extremely difficult to create and maintain a neat, intuitive hierarchical classification scheme (or taxonomy) for documents.

Regarding computer-based information management strategies, Barreau interviewed seven managers [1]. She identified four common sub-activities in their management of information: (1) *acquisition* of items, (2) *classification* of items, (3) *maintenance* of the collection, (4) *retrieval* of items. She also found that users developed highly personalised strategies for organising their information and documents and that, broadly, three types of information could be identified: *ephemeral* (temporary), *working*, and *archived* (dormant).

In a similar study done around the same time, Nardi et al [19] interviewed 15 Macintosh users about their information management behaviour. The combined analysis of both studies [3] showed that users (1) liked to arrange resources by location (for example by grouping icons on the desktop), (2) avoided elaborate filing schemes, and (3) archived relatively little information. Macintosh users tended to use subdirectories to organise information, whereas DOS users did not.

Barreau re-interviewed four of the seven managers ten years later [2]. The four continued to leave most of their documents in a catch-all directory (such as My Documents) and still rarely grouped or classified documents into folders or directories (this was the case for all 7 managers in the first study).

Whittaker and Sidner [26, page 280] reported three basic behaviours in the personal management of email: *no filers* (no use of folders), *frequent filers* (folder users who file messages daily), and *spring cleaners* (folder users who file messages only periodically).

In an interview study of 10 users, Boardman [6] looked at user behaviour in organising and maintaining three separate hierarchies for files, email, and web bookmarks. Five of the ten users attempted

to maintain parallel hierarchies, with varying degrees of success. In a later study, Boardman and Sasse [7] again looked at cross-tool organisational strategies relating to file, email, and web bookmark data. Organising strategies varied significantly between the three types of data: files were the most extensively organised, with deeper hierarchies and fewer unfiled items compared to email and bookmarks.

In a more recent study, Bergman et al [4] surveyed several hundred users in a series of studies of personal computer users and asked them (among other things) to estimate the percentage of their file retrievals performed via search (desktop search), navigation (folders), shortcuts (desktop links), recent documents lists, and other mechanisms. Users strongly preferred naviation through a folder hierarchy (56–68% of retrievals) to search (only 4–15% of retrievals). Users often searched only when they could not remember the location of a file in the folder hierarchy.

1.2 To Classify or Not To Classify

The preceding studies indicate a clear reluctance on the part of users to invest time in advance to file (classify) documents, even if they would then be easier to retrieve later. Why is this? First, it is extremely hard to create category names which are unambiguous. Second, it is hard to find category names which divide up the parent category into mutually exclusive sub-divisions, so categories invariably overlap to some degree. Third, the child categories should completely partition the parent category, so that the user does not feel like a category is missing [23, page 3]. Fourth, information in the real world often falls into several categories. Taking an example from Morville and Rosenfeld [18, page 55], a tomato may be considered to be a vegetable, a fruit, or a berry, depending on the context. Fifth, the classification scheme may become unbalanced, with too many items in one category, and too few in another [24].

Consider the case of filing documents in a hierarchical file system. A file `bobs-ideas-on-XY.txt` contains ideas from Bob about a project XY. Should it be placed in a sub-folder for colleagues `people/bob/` or a sub-folder for projects `projects/XY/`? A decision has to be made. Putting copies into both places will lead to inconsistencies, as soon as any modifications are made.

The file systems of common operating systems such as Windows, OS X, and Linux already provide mechanisms (called shortcuts, aliases, and symbolic links, respectively) to make the same file visible in multiple places in the file hierarchy. Windows shortcuts are, in fact, only special text files rather than a feature of the file system itself. If an application does not know how to process shortcuts, it cannot access the linked information. Microsoft Windows does not actually make use of the available file system level link technologies (NTFS Junction Points, NTFS hard links, and NTFS symbolic links). Moreover, such linking mechanisms seem to be rarely used by users in practice anyway [12]

1.3 Location-Based Spatial Layouts

Several studies have indicated that users like to arrange files by placing icons into groups on the desktop [3]. In effect, spatial location is being used as an aid to memory. Windows Explorer and OS X Finder also support this behaviour by allowing users to position items in its Icon View (and remembers their positions for next time) at any level in the file hierarchy, not just on the desktop. However, spatial layouts ultimately suffer from lack of space. There is simply a limit to the number of items which can be organised effectively in this way.

1.4 Tagging Systems

The basic idea behind tagging has been around for a while: add a few descriptive keywords (tags) to an item so that you can find the item again later by searching for one or more of the keywords. Tagging is cognitively much easier than categorising (classifying), because it only involves users making local conceptual observations [24]. However, tagging also suffers from people using different words or variants to describe the same characteristic [14]. Weinberger [25, page 95] describes the advantages of shared social tagging in communities such as Delicious, but concedes that there will always be ambiguity when tags are assigned by (millions of) different people. Indeed, much current research has focussed on social and collaborative tagging rather than on tagging by individuals in a personal setting. Dourish et al [9] describe the use of tags (called properties) to support the concept of Placeless Documents in a system called Presto.

1.5 Faceted Browsing

Faceted classification was invented by Ranganathan in 1933 [22]. Whereas in tagging users are free to select any words to be tags, in faceted classification a restricted set of words (isolates) are available for use in each of a set of facets to describe the item [25, page 80]. For example, `epicurious.com`, an online recipe web site, characterises its recipes along 8 facets: recipe category (5), dietary consideration (12), cuisine (27), meal/course (12), type of dish (12), season/occasion (18), preparation method (18), and main ingredient (31). The number in parentheses indicates the number of valid tags for that facet.

The retrieval process, faceted browsing, proceeds through progressive refinement. The user can select a value from a first facet (say cuisine = Irish) to receive 92 recipes with that characteristic. Then further selecting, say, main ingredient = beef further restricts the number of matching recipes to just 7. There are no dead-ends in faceted browsing: combinations having 0 matches are not offered to the user. Feldspar [8] is a system which works in a similar way to faceted browsing. Document attributes are progressively refined until the intended document is located.

1.6 Desktop Search

Desktop search engines such Google Desktop and Copernic Desktop Search are increasingly common among average users. A desktop search engine indexes the full text content and various metadata attributes of documents, email, and bookmarks stored in the local file system. Items can be retrieved by typing in appropriate search terms, just like in a web search engine. Desktop search engines are extremely useful to users, but supplement rather than replace tagging systems and folder hierarchies [4].

2. REQUIREMENTS FOR PIM TOOLS

Modern technology offers far more possibilities for users than the mental model of a physical desktop: information in the real world can only occupy one location at the same time. A physical folder has its one physical place. In the digital world, information can be located in many virtual places at the same time [25]. The file from the previous example `bobs-ideas-on-XY.txt` can be made findable both by browsing through projects and also by browsing through people.

The metaphor of the physical desktop, although handy for novice users migrating from a paper-based environment, should no longer be used as the dominant mental model in the virtual desktop environment. Once users are liberated from the limitations of the

classical desktop metaphor, they can experience a variety further benefits of the digital world.

Computer environments today are not like the computer environments of even a decade ago. Hardware has become much more powerful, software has become more capable (and complicated), and much more data is being processed from far more sources. There needs to be a shift of metaphor to meet today's computer environments rather than those of the last century.

Based on the previous discussion, eight fundamental requirements are proposed for future PIM tools, with their main focus on the retrieval process in a local file system. These requirements are not a final nor a complete set of requirements. Additional requirements will be developed as PIM research continues to loosen the limitations of the metaphors previously introduced from the physical world. Some requirements are obvious, but are not implemented well enough in current systems. Some requirements can not yet be found in current systems.

2.1 Be Compatible with Current User Habits

Users are comfortable with their application environment and want to keep it that way. Any new software solution has to integrate into the current environment as smoothly as possible. Any tool which covers only a subset of applications [10] will fail to satisfy a broader user population, because they do not want to be limited to a subset of applications.

Special file browsers were developed to provide more power to the user for the retrieval process [8, 16]. Unfortunately, most of these solutions require a different and sometimes confusing user interface which the typical user might reject. Lack of integration with preinstalled applications is a crucial issue. The user should not be locked-in to a special interface for browsing and searching.

The file system level seems to be a good level to achieve compatibility, since all existing applications share this interface level. Gifford et al [11] and Bloehdorn et al [5] propose promising approaches, although they require special (new) file systems and sometimes special file servers. However, a typical user is unwilling to install a special file system or file server software, particularly if it is not guaranteed to be compatible with their familiar operating system tools. In the long term, future operating systems will have to provide some kind of information retrieval features even in the lower layers of the file system. In the mean time, informations retrieval software solutions have to compensate for the missing support within current file systems.

Many PIM solutions are based on databases [10]. Here again, users are seldom willing to run a specialised file storage database on their computers. They may be unable to make backups with their familiar tools (backup often means simply copying to external storage media), backing up a database is a very different procedure. Users may have to learn a new interface, may be locked-in to the new interface, and the new interface is often poorly integrated into existing applications.

2.2 Minimal Interference

Any new software solution requires some kind of additional user interface. It is essential to keep the learning effort as small as possible. Any interaction step which the users have to make should be absolutely necessary to the process. Optional features should be hidden behind an optional (advanced) interface. In contrast to popular belief, snazzy graphic displays do not automatically result in usable and efficient information retrieval interfaces [13, 16].

2.3 Support Multiple Contexts

A user searching for information always has some kind of mental context. This mental context depends on the current situation and is typically different from the context the user was in when performing the storage process [13, 21]. Good PIM software supports different mental contexts with multiple browsing paths [16].

Considering the file example from the introduction, the user should be able to find the file from Bob about project XY using the people context and/or using the project context. Users want to be able to file information under different categories such as task-related, topic-related, time-related, provenance-related, and form-related [2].

That means that information should be able to be found in multiple places rather than only in one specific location. Tagging seems to be a promising approach [21, 24], although two recent studies comparing tagging with classification reported inconclusive results [14, 20].

2.4 Support Browsing

Studies show that over the years users still prefer browsing over teleporting [4, 2, 8]. When browsing a classification hierarchy, users can see the choices available at each level and choose the most promising.

A great deal of effort was invested into developing improved search engine technology. Although these advances were important and resulted in more capable desktop search engine products, users still prefer browsing in a directory hierarchy to searching with a desktop search engine [4]. Thus it would make sense to invest some future energy and effort into radically improving current hierarchy browsing mechanisms.

2.5 No Unnecessary Limitations

Since large numbers of computer files define our everyday lives, any PIM software solution should scale well to a large number of files and should not affect the efficiency of the browsing process. Even with a large number of files, users must be able to locate their data as quickly and easily as possible.

Some special retrieval tools handle only a small set of file types. Such systems — although very popular in the form of music or photo management software — are not a general solution to the underlying shortcomings of current file browsing tools. Some features provided in specialised management tools would be of great help for other file types.

For example, OS X Finder has the feature of smart folders. A smart folder is a stored search query which dynamically shows any results matching the search criteria. When the user "opens" the smart folder, the content is updated instantly. With this feature, it is very easy, for example, to create a smart folder showing all text files modified within the last two days without having to repeatedly define a search query every time. This is very similar to what iTunes offers for music collections, but other software do not yet offer this feature, say, for image files. Future file browsing solutions should provide enhanced methods for all kinds of file types.

2.6 Transparency

One major aspect of good PIM solutions is transparency to the user. User have built up knowledge of their software environment: a set of experiences, expectations, and standard processes concerning file storage and retrieval. For example, an existing backup process should not be affected by a new PIM system. Users should know where their files are located and what happens to them.

Approaches which require the installation of unfamiliar underlying software introduce complexity and opacity to the system. Users do not trust database systems for metadata or file storage. Ordinary users do not know about database management, database structures, and binary large objects. They do not know how to get their files out of a database system again. Users lose confidence in the software environment, if they are confronted with software that they do not understand.

2.7 Provide for Expiry Dates

Studies show that, with progressively cheaper storage, users tend to keep files over a longer period of time or do not delete them at all [2]. This compounds the information overload problem. There are increasing calls for "forgetting" to be recognised as an important feature in the digital age [17].

During the storage process, users often have an idea of how long the file might be of interest, but this information is forgotten once the file has been stored. Giving the user a method to explicitly define expiry dates — even if they are in the far future — can diminish data overload over time. Providing an expiry date offers the user to explicitly define information as ephemeral, which is an important need as user studies [3] suggest.

In addition, users might be allowed to hook into the process of handling expired data. A user could, for example, automatically move files which are no longer of interest from the "working area" into explicit "archive areas" to remove clutter from current work. Due to the enormous amount of data, users can no longer afford the time to screen all data for archiving.

However, all of this requires an expiry date be attached to the user's information, which in most cases only the user can define.

2.8 Add Metadata While Storing

When a file is stored the user should be given the option to manually add metadata and contextual information to the file. Manual and semi-manual tagging can offer an effective solution for a better retrieval method. Other metadata can (and should) be added automatically, such as a timestamp for time-related retrieval.

Automatically extracted metadata (alone) is often of less use for the purpose of retrieval through browsing. Desktop search engines handle the entire content as metadata and so provide this additional means of access.

Allowing users to explicitly add metadata supports the creation of a user-defined vocabulary (intentionally or subconsciously), which can strongly support subsequent browsing.

3. CONCLUSIONS

The previously proposed set of requirements are intended to spark discussion and serve as a basis for the development of future PIM tools.

Such tools should not seek to radically change user behaviour in one stroke, but rather to bring to pass a gentle evolution. Special interfaces and special software layers requiring additional user interaction are not being accepted by ordinary users.

Modern desktop search engines are a great help to some users, but most users prefer browsing to their files within a hierarchical directory. Thus the browsing process needs to be revisited by PIM researchers and interface developers.

4. REFERENCES

[1] D. Barreau. *Context as a Factor in Personal Information Management Systems*. *Journal of the American Society for Information Science*, 46(5):327–339, June 1995. ISSN 0002-8231. doi:10.1002/(SICI)1097-4571(199506)46:5<327::AID-ASI4>3.0.CO;2-C.

[2] D. Barreau. *The Persistence of Behavior and Form in the Organization of Personal Information*. *Journal of the American Society for Information Science and Technology*, 59(2):307–317, January 2008. ISSN 1532-2882. doi:10.1002/asi.20752.

[3] D. Barreau and B. A. Nardi. *Finding and Reminding: File Organization from the Desktop*. *SIGCHI Bulletin*, 27(3):39–43, July 1995. ISSN 0736-6906. doi:10.1145/221296.221307. http://www.sigchi.org/bulletin/1995.3/barreau.html.

[4] O. Bergman, R. Beyth-Marom, R. Nachmias, N. Gradovitch, and S. Whittaker. *Improved Search Engines and Navigation Preference in Personal Information Management*. *Transactions on Information Systems*, 26(4):1–24, September 2008. ISSN 1046-8188. doi:10.1145/1402256.1402259.

[5] S. Bloehdorn, O. Görlitz, S. Schenk, and M. Völkel. *TagFS – Tag Semantics for Hierarchical File Systems*. In *Proc. 6th International Conference on Knowledge Management (I-KNOW 06)*, pages 304–312. September 2006. http://triple-i.tugraz.at/blog/wp-content/uploads/2008/11/37_tagfs.pdf.

[6] R. Boardman and M. A. Sasse. *Multiple Hierarchies in User Workspace*. In *Proc. 19th SIGCHI Conference on Human Factors in Computing Systems (CHI 2001) Extended Abstracts*, pages 403–404. ACM, March 2001. doi:10.1145/634067.634304. http://www.iis.ee.ic.ac.uk/~rick/research/pubs/workspace-chi2001.pdf.

[7] R. Boardman and M. A. Sasse. *"Stuff Goes into the Computer and Doesn't Come Out": A Cross-Tool Study of Personal Information Management*. In *Proc. 22nd SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, pages 583–590. ACM, April 2004. doi:10.1145/985692.985766. http://www.iis.ee.ic.ac.uk/~rick/research/pubs/boardman-chi04.pdf.

[8] D. H. Chau, B. Myers, and A. Faulring. *What to do When Search Fails: Finding Information by Association*. In *Proc. 26th SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, pages 999–1008. ACM, April 2008. doi:10.1145/1357054.1357208. http://www.cs.cmu.edu/~dchau/feldspar/feldspar-chi08.pdf.

[9] P. Dourish, W. K. Edwards, A. LaMarca, and M. Salisbury. *Using Properties for Uniform Interaction in the Presto Document System*. In *Proc. 12th Annual ACM Symposium on User Interface Software and Technology (UIST'99)*, pages 55–64. ACM, November 1999.

doi:10.1145/320719.322583. http://www2.parc.com/csl/projects/placeless/papers/uist99-presto.pdf.

[10] J. Gemmell, G. Bell, and R. Lueder. *MyLifeBits: a personal database for everything*. *Communications of the ACM*, 49(1):88–95, January 2006. ISSN 0001-0782. doi:10.1145/1107458.1107460. http://research.microsoft.com/pubs/64157/tr-2006-23.pdf.

[11] D. K. Gifford, P. Jouvelot, M. A. Sheldon, and J. James W. O'Toole. *Semantic File Systems*. In *Proc. 13th ACM Symposium on Operating Systems Principles (SOSP 1991)*, pages 16–25. ACM, October 1991. doi:10.1145/121132.121138. http://cgs.csail.mit.edu/history/publications/Papers/sfs.ps.

[12] D. J. Gonçalves and J. A. Jorge. *An Empirical Study of Personal Document Spaces*. In *Proc. 10th International Workshop on Design, Specification and Verification of Interactive Systems (DSV-IS 2003)*, pages 46–60. Springer LNCS 2844, June 2003. doi:10.1007/b13960. http://virtual.inesc.pt/dsvis03/papers/05.pdf.

[13] M. W. Lansdale. *The Psychology of Personal Information Management*. *Applied Ergonomics*, 19(1):55–66, March 1988. ISSN 0003-6870. doi:10.1016/0003-6870(88)90199-8. http://simson.net/ref/1988/Lansdale88.pdf.

[14] S. Ma and S. Wiedenbeck. *File Management with Hierarchical Folders and Tags*. In *Proc. 27th SIGCHI Conference on Human Factors in Computing Systems (CHI 2009) Extended Abstracts*, pages 3745–3750. ACM, April 2009. doi:10.1145/1520340.1520565.

[15] T. W. Malone. *How do People Organize Their Desks?: Implications for the Design of Office Information Systems*. *Transactions on Information Systems*, 1(1):99–112, January 1983. ISSN 1046-8188. doi:10.1145/357423.357430.

[16] G. Marsden and D. E. Cairns. *Improving the Usability of the Hierarchical File System*. In *Proc. Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology (SAICSIT 2003)*, pages 122–129. South African Institute for Computer Scientists and Information Technologists (SAICSIT), September 2003. 1581137745. http://pubs.cs.uct.ac.za/archive/00000190/01/saicsit2003-dec.pdf.

[17] V. Mayer-Schönberger. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press, October 2009. 978-0691138619.

[18] P. Morville and L. Rosenfeld. *Information Architecture for the World Wide Web*. O'Reilly, Third edition, November 2006. 0596527349.

[19] B. A. Nardi, K. Anderson, and T. Erickson. *Filing and Finding Computer Files*. Technical Report 118, Apple Computer, Department of Computer Science, 1994.

[20] R. Pak, S. Pautz, and R. Iden. *Information Organization and Retrieval: A Comparison of Taxonomical and Tagging Systems*. *Cognitive Technology*, 12(1):31–44, 2007. http://business.clemson.edu/Catlab/pubs/pak-pautz-iden-2007.pdf.

[21] D. Quan, K. Bakshi, D. Huynh, and D. R. Karger. *User Interfaces for Supporting Multiple Categorization*. In *Proc. 9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT '03)*, pages 228–235. IOS Press, September 2003. 1586033638. http://www.idemployee.id.tue.nl/g.w.m.rauterberg/conferences/INTERACT2003/INTERACT2003-p228.pdf.

[22] S. R. Ranganathan. *Colon Classification*. The Madras Library Association, 1933.

[23] E. M. Rasiel and P. N. Friga. *The McKinsey Mind*. McGraw-Hill, September 2001. 0071374299.

[24] R. Sinha. *A Cognitive Analysis of Tagging*. Rashmi's Blog, September 2005. http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/.

[25] D. Weinberger. *Everything Is Miscellaneous: The Power of the New Digital Disorder*. Times Books, May 2007. 0805080430. http://www.everythingismiscellaneous.com/.

[26] S. Whittaker and C. Sidner. *Email Overload: Exploring Personal Information Management of Email*. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 1996)*, pages 276–283. ACM, April 1996. doi:10.1145/238386.238530. http://dis.shef.ac.uk/stevewhittaker/emlch96.pdf.

Guidelines for the Design of Personal Document Management User Interfaces

Sarah Henderson
Department of Information Systems
and Operations Management
University of Auckland
Private Bag 92019, Auckland, NZ
s.henderson@auckland.ac.nz

ABSTRACT
Personal document management describes the activities performed by an individual in creating, acquiring, organizing and maintaining collections of their documents. A study involving 10 in-depth interviews and a survey of 115 participants was conducted in order to better understand the approaches people take to document management in order to inform the development of better user interfaces. These were used to develop an understanding of issues and concepts in personal document management, and a description of three major approaches to personal document management: a piling strategy, a filing strategy and a structuring strategy. From the findings, some general guidelines are proposed for the development of personal document management user interfaces, along with specific user interface guideline to support each of the three identified approaches to personal document management.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation (e.g. HCI)**]: User Interfaces

General Terms
Design, Human Factors.

Keywords
Personal document management, personal information management, document management strategy.

1. INTRODUCTION
Personal document management is the activity of managing a collection of digital documents. The unit of analysis in personal document management is an individual user and the collection of digital documents he or she owns. The process of document management incorporates the creation/acquisition, retrieval, organizing and maintenance activities described above, provided they are performed by the document owner. Personal document management is an activity that is performed intermittently, embedded in the daily life of users.

Most people store their documents in the hierarchical file system provided by their computer's operating system, and manage these documents through a hierarchical file browser (such as Windows Explorer) [9]. These file browsers were intended to allow a systems administrator to manage files on a computer (at a time when there were generally only a few hundred files). Additionally, when these were developed, computers were not used by the general public, but by highly trained technicians with a thorough understanding of computer technology. The basic paradigm of the tool has not changed in the decades since its introduction, although the user interface to it significantly improved with the widespread introduction of graphical user interfaces in the Macintosh and Windows operating systems. Despite these improvements, the user interfaces of these systems were not designed for modern document management tasks.

A basic principle of user interface design is that the design of a tool should be thoroughly grounded in an understanding of how the users work, what tasks they perform and how those tasks are carried out. However, with personal document management, very little research has been done into how people are managing their documents and what the requirements are for document management tools. This knowledge gap needs to be addressed before better tools can be developed.

2. BACKGROUND
The seminal work in the field of personal information management is Tom Malone's 1983 study titled 'How Do People Organize Their Desks?' [13]. He studied how people used paper files in their offices and identified two distinct strategies: 'neat' and 'messy'. In a neat office, the person tried to designate a category for every document and place it the location corresponding to that category. The location may have been a folder inside a filing cabinet, a paper tray, or a named pile. In the messy office, the person would tend to pile up documents over time, in a less structured way. In both offices, files and piles are the basic building blocks of paper document management.

Several studies have attempted to classify styles of email use in a similar way to Malone's 'neat' and 'messy' classifications. One of the earliest was Mackay [12], who identified 'prioritizers', 'archivers' and 'requesters and responders'. The requesters and responders use email for task delegation; prioritizers concentrate on managing incoming messages while archivers use email to archive information for future use. Whittaker and Sidner [18] also looked at organizing behavior in email, identifying 'no filers', 'frequent filers' and 'spring cleaners'. The 'no filers' were the email equivalent of pilers, allowing all their email to pile up in the inbox, while the filers attempted to place all their emails into folders. The spring cleaners occupied a middle position between the other two groups, using a 'no-filing' strategy most of the time,

Conference'04, Month 1–2, 2004, City, State, Country.

but periodically attempting to put their documents into files. Without the folders that others use to aid retrieval, 'no filers' rely on full text search and temporal ordering to retrieve their information. This categorization was extended by Bälter [2] to subdivide 'no filers' in to 'folderless cleaners' and 'folderless spring-cleaners' depending on how often they deleted information from their inbox. A more recent study of email behavior identified two major approaches: 'cleaners' and 'keepers' [11]. Cleaners have specific times for dealing with email, and don't keep events or to-do items in their email. Keepers read email constantly, allowing tasks to be interrupted by email. They keep events and to-do items, and search their email archives.

Studies of organizing approaches taken with respect to web bookmarks have found similar results to the studies of email, identifying 'no-filer', 'creation-time filer', 'end-of-session filer' and 'sporadic filer', depending on whether and when the user saved web bookmarks during a browsing session [1].

Another more recent study to look at digital documents was conducted by Richard Boardman [5]. He analyzed information behavior across three collections: documents, email and web bookmarks with the intention of analyzing difficulties people had in managing their information collections across tools. He found that people could be categorized as either 'pro-organizing' or 'organizing neutral', but that people didn't always adopt the same strategy across all collections. People were more likely to be 'pro-organizing' in their document collection and email than they were in their web bookmarks.

Table 1: Classifications of organizing strategies

Reference	Information Type	Classifications
Malone [13]	paper documents	neat, messy
Mackay [12]	Email	prioritizers, archivers, requesters and responders
Whittaker & Sidner [18]	Email	no-filers, frequent-filers, spring-cleaners
Bälter [2]	Email	folderless cleaners, folderless spring-cleaners, cleaners, spring-cleaners
Gwizdka [11]	Email	cleaners, keepers
Abrams, Baecker & Chignell [1]	web bookmarks	no-filer, creation-time filer, end-of-session filer, sporadic filer
Boardman & Sasse [6]	documents, email and web bookmarks	pro-organizing, organizing neutral

3. RESEARCH DESIGN

This study into personal document management practices consisted of a field study and a survey. In the field study, 10 knowledge workers were interviewed about their personal document management practices. The participants were all employees of a large university environment, and included researchers, teachers and professional staff. Such an environment is particularly helpful for work of this nature because it encompasses a wide variety of usage situations coupled with a good mix of individuals with varying requirements. All participants were using the Windows XP operating system.

The interviews were largely unstructured and took place in the participant's offices so their document management practices could be seen in their natural context. Participants were asked to give a tour of their documents, and the interview was centered on the participant's practices. Participants were encouraged to demonstrate their structures and processes during the interview. This technique of interviewing participants in their offices and using their computers as a questioning point for the interview has been used many times in investigation of related aspects of personal information management [8, 13, 17, 18], and was used in prior studies of personal document management [3, 4]. These interviews were analyzed using thematic analysis, and an initial conceptual model of document management concerns was developed.

In order to validate this conceptual model, a questionnaire was used in a survey of knowledge workers designed to gather more generalized data about personal document management practices. The questions were derived from the conceptual model, and were delivered as a web-based survey. The sample frame was the staff of the commerce faculty of the university. The survey was completed by 115 participants (out of 490 people invited).

In addition, a snapshot of each participant's file system was taken so that their document structures could be quantitatively analyzed. All field study participants and 72 survey participants provided a file system snapshot. These were analyzed and a number of metrics were calculated to describe the overall shape of the structure. The metrics included:

- Overall size (number of files and folders)
- Tree characteristics (depth, breadth and balance)
- Duplication (of file and folder names)
- Top level files and folders

The field study indicated that there were three primary strategies that the participants adopted in order to manage their document: piling, filing and structuring. The three strategies seemed to differ in the following attributes:

- Overall level of organization (self-assessment)
- When folders are created (self-reported)
- Preferred retrieval strategy (self-reported)
- Preferred document view (self-reported)
- Use of tree (self-reported)
- Depth of structure (from snapshot)
- Breadth of structure (from snapshot)
- Unfiled documents in top level (from snapshot)
- Folders in top level (from snapshot)

In order to validate this finding with a wider population, a K-means cluster analysis was performed on the survey data to see if particular combinations of these attribute values tended to group together. This analysis resulted in three distinct clusters.

Analysis of variance indicated that several metrics were not contributing to discrimination between any clusters. These included the questions on when folders are created, retrieval strategy for old files, use of tree and the breadth of the structure. These were removed one at a time and the cluster analysis

repeated until all remaining variables differed significantly across the clusters. Table 2 below shows the resulting variables and the typical values for each cluster.

4. RESULTS

The findings from the studies are grouped into four main areas: general **attitudes** to personal document management, approaches to **finding** documents, issues surrounding **creating** folders and documents and **strategies** for document management. The following sections elaborate on the finding in each of these areas, integrating the findings from the field studies and the survey.

4.1 Attitudes to document management

4.1.1 People want to be "organised" not "messy"

Many of the participants' responses indicated that they felt that being organized was a desirable state. Those who consider themselves organized expressed pride in their file structures, for instance, one participant spoke with pride of her colleagues being "*surprised that they can come into my office and they can ask for an article and I will know where it is*," adding "*thank God for the power of computers*." In contrast, several participants referred to their documents as being messy, or a mess. Others mentioned that they find they tend to organize some documents but not others, with more effort being put into managing documents perceived as more important.

Many participants seemed to have an idea of a hypothetical 'perfect organization' against which people measure themselves. Trying to attain that level of organization was seen to be a good thing. The opposite end of the spectrum was 'messy', and people do not want to be considered messy. This was confirmed in the study with over 90% of people agreeing that they think it is important to have well organized documents. Since it was noted in the field study that people often compared themselves with others, or were interested in learning how other people performed document management, it seems that people may feel less satisfied with the document management structures simply because they perceive it to be messy or poorly organized, even if the actual amount of time or effort they spend organizing their documents isn't impacted.

4.1.2 Hierarchies are intuitive

The folder hierarchy is intuitive to many people and reflects the way they think about their documents. One participant said that folder structures are "*second nature, and I probably don't feel things that someone who is new to them would find puzzling and annoying. I'm reasonably happy with this hierarchical tree structure of Windows Explorer*." Another noted that "*it seems to fit in with my mindset*." This finding from the field study was reinforced in the survey, with the number of people in free-form comments mentioning they liked ability to create a folder structure and appreciating the flexibility to create their own organisation scheme within it.

4.1.3 Sense of file ownership

People need to feel they have control and ownership over the folders and files in the collection. One participant had a particularly strong conception of file ownership. She related a past experience where she was assigned a computer that previously belonged to someone else and still had the previous user's files and folders on it. She was very careful not to move, rename or delete or in any way interfere with that person's files. Windows XP automatically creates a number of folders for each user, including folders for Application Data, Local Settings, Templates, and known network computers and printers. Several times she referred to these as being someone else's folders that she didn't touch, seemingly not knowing they were system created. She mentioned being very careful not to use them or touch them. She also feels that she doesn't have the right to rename files that she didn't create herself. Any files she saves into her folders that were emailed to her or that she downloaded from the web always keep their original names. She explains that even though they are on her computer in her folder structure, she doesn't feel she has the right to rename them because she didn't name them and she doesn't own them.

This theme was reinforced in the survey with people commenting that one reason why they didn't like the system provided My Documents folder was that they didn't create it and therefore didn't have full control over it.

4.1.4 Variable willingness to change practices

Most people are willing to change their document management practices in order to be more organized, however some people are resistant to change. One participant said he would like to change "*if you could tell me after this research what is a good way or better way to organize files, that means is easy to name it and easy to retrieve it*." This was confirmed in the survey, with approximately three-quarters of the respondents agreeing they would be willing to change if they were shown a better way to doing things.

Habit is a very powerful force, with several people giving habit as the reason for various document management practices they engage in. Once someone has a reliable way of doing something, they are comfortable with that and may not be willing to change unless there is a compelling reason to do so (or unless they are forced to by the change of a system). One participant mentioned that she was quite annoyed at having to open My Computer and then drill down through folders from My Computer to C drive and down to her document folders every single time she opened a document. At the end of the interview, the interviewer mentioned that it was possible for her to create a shortcut to her documents folder and put it on the Desktop for fast access with a single-double click. She politely said thank you, but was not interested in creating a shortcut, explaining that she was used to doing things a certain way and wanted to stick to the methods she was used to as she knew they were reliable.

A related theme that came up several times in the survey was participant's lack of knowledge. Many suggested the addition of features in Windows XP that were in fact already available. And several indicated that they hadn't availed themselves of available view options because they hadn't known it was possible. People don't tend to receive any training in personal document management. They are generally left to themselves to figure it out. For instance, the university has run professional development courses teaching people how to deal with their email, manage tasks and projects and manage their time, there are no courses teaching people how to use their documents. Basic computing courses teach the basics of creating folders, saving and opening documents, but usually don't discuss any more advanced features like changing views, sorting, advanced search options, and how to change the Desktop to enable spatial layout of items. Very few people would consider it worth spending time investigating the topic themselves, since people are generally more concerned with getting on with their tasks.

4.2 Finding documents

4.2.1 Browsing more common than searching

The majority of the participants reported that if they need to locate a document, they would browse to it in their folder structures. This browsing technique is also known as location based search. As one participant puts it: "*I usually know where I put stuff.*" The survey confirmed this predominance of a tendency to browse rather than search. This cannot be construed as a clear preference in all cases, since many people weren't familiar with the ability to do full text search in Windows XP, and others complained about how slow search is.

A minority reported using search as their primary means of finding a document, with keywords from filename being the most common way of trying to locate it. 62% of respondents reported experiencing search failure. Most believed the file was on their computer somewhere and they just couldn't find it..

When asked about their use of a search tool, the majority of respondents said they would use a search tool only if they hadn't quickly found their document through other means. More than a quarter said it would be a last resort, while those who would search first or never search were a small minority.

4.2.2 Sorting is an important search technique

Sorting proved to be a very important was of locating documents, either in search results or in folder views while browsing. As one participant explains: "*If I'm looking for a particular file here, sometimes it's useful to sort by type, cause I know it's a text file and I can go to text files and then find it immediately, [foo].txt for example. Sometimes I know that it's the most recent file so I scroll down to the bottom and there it is. Sometimes it's by name and sometimes it's the biggest file that I'm looking for, so I can know relative what's the most efficient way for me to find it.*"

Sorting by anything other than name is only possible in the details view (which is the mostly commonly used view). Changing between names, date and file type sorts were very common, with size being much less common. This fact was confirmed in the survey. Sorting can be viewed as a sort of 'quick and dirty' way of searching or filtering within a folder. Users often change the name of their documents to force a specific sort order inside a folder. For example: "*That's one thing I am careful with though, because it's a 12 week course, I always put the zero in [Module 01 not Module 1] so they actually stay in order.*" Other participants also mentioned using certain prefixes to force a certain sort order within their folders.

4.2.3 Tree view useful for overview and navigation

Many people in the field study use the tree view when navigating, although the perceived time taken to click down through the levels is an annoyance for some. The survey confirmed this, with 70% of respondents reporting they use the tree to navigate. 65% report using the 'Up one level' button to go to the parent folder of the current folder, and 50% use the back and forward buttons to navigate between folders. Some participants in the field study had the tree visible, but didn't usually use it for navigation, instead double-clicking through files in the details or list view in the right hand side.

Those that didn't use the tree tended to have very shallow file systems with a large number of files in each folder. One of these participants describes himself as "*tree averse*", pointing to his My Documents folder which contains 32 folders and 170 files and saying "*that's ridiculous, how can any sane person possibly cope with that? That much vertical stuff.*"

4.3 Creating folders and documents

4.3.1 Document creation is application-centric

The field study observed that people mainly create documents through the appropriate application rather than through their file system. This was confirmed in the survey, with the majority of people opening the application to create a new document. Less than 10% of respondents create their files directly in Windows Explorer. Although using Windows Explorer is still important for the 27% who use it to locate an existing document to open and reuse, the majority of people name and place their files through the Save dialog boxes of applications. This is important since it means that designing a document management interface doesn't only involve creating a file management interface, but also a coherent set of Open/Save dialog boxes. This also means that any unified interface can easily be disrupted if applications are able to use their own custom dialog boxes.

4.3.2 Periodic reorganizations are common

Many participants spoke of cleaning up, organizing or reorganizing their files. It is frequently done on a periodic basis (such as every semester or annually), but may also be done in response to rising level of mess, or continually.

Several participants mentioned they clean up every six months or so, with one saying "*I look at all the stuff that hasn't been saved into a folder, and I figure out if I need to delete or move it or whatever.*" Others say they wish they could reorganize their documents but never have the time to do so. One participant says she sometime starts and gets partway through but always something else comes up that prevents her from finishing. She describes it as a constant guilt that she knows she should do something but never does. She does say that maybe once a month or so something will annoy her about a certain folder and she'll try to delete old items or move things to better locations, however she usually never finishes.

4.3.3 Implicit limits on folder structures

Many participants remarked that reorganization activities such as splitting a folder into multiple folders or creating subfolders were prompted by a folder reaching some limit. One participant says "*I'm reacting to the fact that it's building up and I'll think well I'll subdivide at this point. And that could be something in the order of ... and again, it will depend upon the topics that are there. No point in differentiating them if there's only one topic. If there are two quite distinct topics, you might think that. And that might be something on the order of 10 to 15 I guess.*" This is supported by the fairly consistent and low average number of files and folders people keep in their folders.

4.3.4 Three folder creation tactics

Folders are created for a number of different reasons. They can be created before there are files to be placed within them, created ad hoc to contain files needing to be saved, or created in order to clean up and move existing documents. Most participants reported using multiple folder creation tactics depending on the circumstances. In-advance creation sometimes involves the creation of entire folder structures, often similar to or duplicating existing folder structures. The survey confirmed this, with just-in-time creation being the most prevalent (reported by 56% of respondents). Folder creation in response to cleanup was reported by 28% of respondents, with the remaining 16% creating

in advance. Note that as suggested in the field study, it is quite likely that many people use a combination of these techniques at different times. The survey asked which technique they would usually employ. The survey also found that those who create in advance tend to be happier with their file system overall.

4.4 Approaches to document management

From the field study and survey data, three distinct clusters of strategies have been identified. Following previous researchers, these have been named piling, filing and structuring. The piler strategy identified here is analogous to messy, no-filers, keepers, and organizing neutral strategies identified by other researchers. Filer and structurer are variants of the pro-organizing, frequent-filer and keeper categories identified by others but have some distinct features that mean they are likely to require different user interfaces for optimal support.

Table 2 summarizes the results of the cluster analysis:

Table 2: Summary of quantitative features of personas

Metric	Piling	Filing	Structuring
Self reported level of organization	Not very organized	Somewhat organized	Somewhat organized / very organized
Use of search	Last resort	Second choice	Second choice (sometimes first)
Preferred view	List/Details	List/Details	Details/List
Number of Top Level Folders	Medium	High	Low
Number of Top Level Files	High	High	Low
Average depth	Low	Medium	Medium/High

The following sections briefly summarize the main characteristics of each of these three strategies, combining the quantitative data from the survey and the qualitative data from the field studies.

4.4.1 Piling

The piling cluster perceive themselves as relatively disorganized, preferring a list view, with a medium number of top level folders and a high number of top level files and relatively shallow system.

A person adopting a piling strategy doesn't really file his documents; he just lets them pile up in various convenient locations. Folders are usually created in order to dump a large group of old documents that are no longer needed. Because folders are rarely created, the folder structure tends to be fairly shallow, with many folders and files at the top level of the structure. Because recently used files are always easily available, they are retrieved through browsing, with sorting often used to locate the most recent document. A piler may make periodic half-hearted attempts to delete things or organize them into folders, but more because he feels this is how he is supposed to do it than any perceived usefulness. It's peer pressure. Someone adopting a piling strategy tends to be a high Desktop user, since one of the key concerns is least effort and maximum availability. Minimizing visual clutter isn't really an issue, nor does he feel any need or desire to organize documents in order to get an overview of his stuff.

4.4.2 Filing

The second cluster is perceived as more organized, with just in time folder creation, combination of browsing and searching only as a last resort. The structure is medium in depth and width and has a moderate number of unclassified top level folders.

Someone adopting a filing strategy creates folders in order to split up collections of documents. They split folders up if the number of documents grows so large that they cannot easily spot items within them anymore. They tend to create folders either during cleanups or just-in-time as they need to save a folder that doesn't fit an existing category. They do have a hierarchy, although it is moderately broad and not particularly deep. They are likely to have some files in the top level (pending cleanups), and quite a few folders as well, resulting in a tree of moderate depth but high breadth. There is no particular preference for view, but they are much more likely to locate files by browsing their structures than searching. They would generally consider themselves to be relatively organized.

4.4.3 Structuring

Members of the third cluster have high depth, low level of unclassified files, in-advance or just-in-time creation and consider themselves to be fairly organized.

Someone adopting a structuring strategy intensively organizes their files, creating deep and meaningful document structures, often before there are documents to put in them. Related folders are typically grouped together into more levels of nesting, in order to hide complexity and indicate their relationship. This results in a fairly narrow and deep tree, often with fewer than 3 or 4 top level folders and very few or no files at the top level of their folder structures. They are more likely to browse through their structures although because there are so many folders to inspect, if they can't remember where something is they will readily search, particularly for older files. Browsing is often done using the tree, since the tree gives them an overview of how everything fits together. The parent folders give context to the subfolders. They get frustrated with views that don't show them the full context. For instance, search that only shows them the file name is very irritating. Showing the parent folder is even better, but they really would prefer to see the full path for context. Folders are often created in advance, as soon as a new responsibility, project, course or something appeared on their horizon, to have a place to store the documents. They tend to consider themselves very well organized.

5. USER INTERFACE GUIDELINES

The following table lists some general user interface guidelines based on the findings described in the previous section. The following sections describe these in more detail.

Table 3: Summary of user interface guidelines

General Guidelines
Provide usable, fast, powerful full text search
Integrate Open/Save dialog boxes into the UI
Support flexible sorting and custom sort order
Do not mess with the user's folders or documents

Guidelines for Piling strategy
Do not require containment
Support a time-based interface
Provide optional tagging
Guidelines for Filing strategy
Support containment
Provide a cleanup interface
Guidelines for Structuring strategy
Support hierarchies with multiple classification
Support dynamic containers
Provide relationships between items
Provide optional tagging and color coding.

5.1 General User Interface Guidelines

5.1.1 Provide usable, fast, powerful full text search

All users rely on search tools to sometimes locate documents and thus need a very fast and robust full text search. Although users of piling and filing strategies don't rely heavily on search tools for accessing their documents, they do use it sometimes, particularly to find old documents or documents in their archives.

5.1.2 Integrate Open/Save dialog boxes into the UI

All users interact with their document collection through applications' Open/Save dialog boxes, and filers perform most creation, acquisition and locating activities this way. Thus, these need to be considered first class citizens in a personal document management user interface. They should present the same interface the user would normally use to access their files, including preferences for views, sorting or other customizations.

For filers and structurers this dialog will probably need to be much larger than they currently are in order to provide a useful view of the file system. For adopters of a piling strategy, the dialog should be as minimal as possible with perhaps simply a field to specify the filename (which ideally should default to something sensible suggested by the document). Since pilers don't usually specify a place, there is no need for a large view of a folder structure for them to select one. There should be an option to switch views, since most users do not operate exclusively according to type.

5.1.3 Support flexible sorting and custom sort order

Sorting is a very important mechanism used to locate documents, and sorting on any visible attribute should be easy to accomplish. In addition, it should be possible to specify a custom sort for a folder or container, in which the user can reorder folders and documents to appear as they wish. This should be remembered so that if the user switches to another sort order, they can switch back and have their custom sort presented again. This should prevent people from using file naming techniques to force documents to sort in particular ways.

5.1.4 Do not mess with the user's stuff

Users need a sense of ownership over their files, and so the general principle is that the system should not interfere with their structures unless absolutely necessary. For instance, the system should not create the pseudo-folders My Music and My Pictures. Rather, the user should be allowed to create as many folders for their pictures and music as they want, wherever they want and name them however they want. They should be able to select a custom view (pictures view or music view) for those folders, and this custom view could also be reused in displaying search results for the appropriate type of file. Likewise, the system should not move documents around or take any actions without the user's knowledge and consent. User settings related to the operating system and applications should be stored elsewhere, either a designated settings folder for each user or in the Registry. These should not be intermixed with user folders and documents.

5.2 Piling Strategy Guidelines

5.2.1 Do not require containment

Pilers don't need a folder-like containment mechanism in order to group their documents, since they are interested in expending as little up-front effort as possible. This doesn't mean that folder or a grouping mechanism needs to be completely absent, just that if present, it should be optional. It should be entirely possible to use the interface without ever having to think about where to put something or in what to contain it.

The attempt to take literally the piling paradigm to create a user interface that supports piles is misguided when it comes to the personal document management piler. Electronic implementations of piles (e.g. [14]) are a containment mechanism just like folders. Conceptually, they operate exactly as folders although with a slightly richer visual representation, one which folders views could easily match (and with picture folders starting to show thumbnails of contents, this is getting closer). The nature of the piling strategy is that followers don't really want to group and organize things. They adopt piling because it involves the least initial effort. This doesn't mean that a containment or grouping or folder needs to be completely absent, just that if present, it should be optional. It should be entirely possible to use the interface without ever having to think about where to put something or what to contain it in.

5.2.2 Support a time based interface

Time based retrieval is more important to users of a piling strategy than users of other document management strategies. The piler naturally has (or maybe is forced to have) some sense of chronology, since their pile stacks up in order of creation/acquisition. While they don't need to remember absolute times or time spans, they need to have a relative idea how far back through the Desktop stack to look, or how many cleanup folders back to look for something. An interface such as Lifestreams [10], provided it had very strong search support, would probably suit the piling strategy very well.

One way of leveraging this tendency is to ensure the default document view shows all recent files ordered by either when they were most recently used or when they were created. The Desktop could potentially use the same view, making the view easier to access. This view should be dynamic, rather than the static view currently offered by the Desktop. Items that have not been used recently should just disappear from view. Thus, the default view might show an item that was added a month ago but which was

used three days ago, while an item added two weeks ago but not used since may not be visible.

Rather than having items disappear after a certain time, the view should simply show as many recent documents as possible. This takes advantage of the common practice of sorting by date to find the most recent document, and eliminates the need for periodic cleanups or dumps of files. There should be an option to 'jump back' or scroll back to show earlier sets of documents as well, giving this interface something in common with the TimeScape software [15], although without the spatial element.

All dates and times should be shown as relative times by default (although the option of switching to absolute times should be available), since few people have sufficiently good recall to pinpoint exactly when they created or worked with a document. Examples of relative times include '30 minutes ago', '5 hours ago,' 'yesterday' and '2 weeks ago.'

5.2.3 Provide optional tagging

If someone adopting a piling strategy wants to do any kind of categorization at all in order to make sure that he is more easily able to retrieve stuff, the easiest way to support this would be to allow tags to be specified when saving the document (or added later). These can be free-form comma separated tags in which he can just type additional keywords that he might want to use to search for it but that don't appear in the document itself. This provides a way of being able to group related documents without the containment semantics, since it is easy to create a view of all documents sharing the same tag or tags. The advantage of tagging is that it lets the user add words they associate with the documents, but which might not appear within it. This makes future searching more effective.

5.3 Filing Strategy Guidelines

5.3.1 Support containment

Users adopting a filing strategy need a containment mechanism in order to group their files into manageable locations. The standard folder metaphor would probably work very well, although there are many other ways of implementing containment semantics, which would also work. Different means of visualizing containers should be explored. One place to start would be developing views that allow more of the hierarchy to be seen at once, since a common complaint is the time taken to click down the levels. Within containers, items should be able to be viewed with or without details, since name is probably the most important dimension. If other dimensions are used, file type, date created and date last used would be the most useful.

While the ability to change sorting is important, there should also be a custom sort or user defined sort. In this way, filers could organize things into the exact order they wanted and know things wouldn't change. This creates a sense of stability and permanence and makes finding items through known paths easier and more reliable. It also obviates the need to change the 'common sense' file names in order to force a sort order.

5.3.2 Provide a cleanup interface

While most of the time a user of this strategy is interacting with documents through Open/Save dialog boxes, they do want a larger view of their file structure when doing a cleanup. During a cleanup, they are going through files in a temporary location (e.g. top level folder or Desktop) and placing them into their permanent folder home. To do this effectively they need to be able to see the list of files they are cleaning up, as much of their folder structure as possible (expanded tree view), and ideally a preview, in case they need to be reminded what the document is before they can decide where to put it. In this view it must be easy to create new folders and to reorder folder contents in the tree.

5.4 Structuring Strategy Guidelines

Followers of a structuring strategy need the ability to express containment just as filers do, but they also need richer containment semantics.

5.4.1 Support hierarchies with multiple classification

Systems must provide the ability to create hierarchies of containment, since many people appreciate the ability to create folder structures. Multiple classifications enable a document to live in more than one location. Previous means of approximating this such as shortcuts or copies are not sufficient – the document actually needs to have one location but appear in multiple locations. Regardless of the location from which the file is viewed and accessed, any changes to the document or its metadata should be immediately effective in all locations. When a file is deleted, if it exists in multiple locations the user will need to be prompted whether the file should be deleted from that location only or from all locations. A user interface should support collapsing or hiding of levels of information, to enable the ability to see an overview and drill down to detail on demand.

5.4.2 Support dynamic containers

Providing dynamic containers is another way of providing some of the same functionality as multiple classifications. Dynamic containers don't have a predefined set of contents, but rather display the contents based on a search. The containers in the Presto system [7] are an example of this, as are Outlook 2003's Search Folders. For instance, an expense report could be stored in a folder with other trip information, but a dynamic folder could be created to view all expense reports together. The dynamic folder can be organized into folders like any other folder.

5.4.3 Provide relationships between items

To a structurer, the file system is more than simply a place to store things; it is a representation of the structure of his information. For this reason, the ability to make arbitrary relationships between things would be a useful extension. This can be partly automatic and partly manual. For instance, the system could track which documents are opened with other documents or emailed together with other documents and therefore infer relationships between documents. This could be presented by having a 'Related items' panel that displayed the other documents related to the currently selected document, enabling them to be quickly accessed. In addition, there should be an ability to manually create relationships between items, thereby choosing the items that appear in the 'related items' view.

5.4.4 Provide optional tagging and color coding

Other methods to provide the structuring filer with richer abilities to organize files include allowing the ability to tag documents or files with keywords (as described for Nathan), and to color code files and folders. These should be entirely optional but if used are entirely user-generated. The organizer can use any colors they want, and can assign an optional descriptive label to the color, or just simply use the color.

6. DISCUSSION

Since some people are using their document management tools sub-optimally due to lack of knowledge, one valid question is whether or not we need to change the tools or whether we merely need to train the users to use them more effectively? However principles of usability would suggest that a good software tool would not require extensive training in order to be effective – it should either be designed so it is effective without training, or it should incorporate training of the user as they use the system.

Some results from the classification model differed from the strategies described based on the field study. For example, it was anticipated users of a piling strategy would make greater use of search tools to compensate for their lack of folder structure. However, it is possible that their piling strategy means that most of the time they can browse through their top level documents, assisted by sort options until they find their target document. In this way, they are predominantly relying on a browsing technique rather than search. In contrast, adopters of a structuring strategy were not expected to be heavy users of search, since the effort they expended in structuring their folders should pay off by providing more effective browsing. However the survey results showed that structurers were more likely to search in their own documents. This result has also been independently observed in a study of email [16].

It is unclear whether more frequent searches mean the document management system is less effective. It is possible that the folder hierarchy makes the search much more useful through being able to search only a related subset of the documents, and because the metadata provided by the folder path makes recognizing found documents easier. More research would need to be done examining the amount of time spent in document management activities by adopters of the various strategies before a determination can be made.

7. CONCLUSION

This paper has presented some of the key findings from a study of personal document management, including an identification of three approaches to document management: piler, filer and structurer. From these findings, suggested guidelines for personal document management user interfaces have been developed, along with guidelines specific to each of the approaches.

It is necessary to remember that although these strategies and the personas that illustrate them are useful tools to guide user interface development, people do not necessarily neatly fit these three strategies all the time. People will at times adopt one or the other depending on the circumstances, although there is usually a dominant preference. These three categories collectively cover the spectrum of personal document management behavior observed in this study and therefore an interface that can accommodate all three should be useful to everyone.

8. REFERENCES

[1] Abrams, D., Baecker, R. and Chignell, M., Information Archiving with Bookmarks: Personal Web Space Construction and Organization. in *CHI'98 Conference on Human Factors in Computing Systems*, (Los Angeles, California, USA, 1998), ACM Press.

[2] Bälter, O. Strategies for Organising Email *Proceedings of HCI on People and Computers XII*, Springer-Verlag, 1997.

[3] Barreau, D.K. Context as a Factor in Personal Information Management Systems. *Journal of the American Society for Information Science, 46* (5). 327-339.

[4] Barreau, D.K. and Nardi, B.A. Finding and Reminding: File Organization from the Desktop. *SIGCHI Bulletin, 27* (3). 39-43.

[5] Boardman, R. Improving Tool Support for Personal Information Management *Dept. of Electronic and Electrical Engineering*, Imperial College, London, England, 2004.

[6] Boardman, R. and Sasse, M.A., "Stuff Goes into the Computer and Doesn't Come Out" A Cross-tool Study of Personal Information Management. in *CHI'2004 Conference on Human Factors in Computing Systems*, (Vienna, Austria, 2004), 583-590.

[7] Dourish, P., Edwards, W.K., LaMarca, A. and Salisbury, M. Presto: An Experimental Architecture for Fluid Interactive Document Spaces. *ACM Transactions on Computer-Human Interaction, 6* (2). 133-161.

[8] Ducheneaut, N. and Bellotti, V. E-mail as Habitat: An Exploration of Embedded Personal Information Management. *Interactions, 8* (5). 30-38.

[9] Faichney, J. and Gonzalez, R. Goldleaf Hierarchical Document Browser. *Australian Computer Science Communications, 23* (5). 13-20.

[10] Fertig, S., Freeman, E. and Gelernter, D. "Finding and Reminding" Reconsidered. *SIGCHI Bulletin, 28* (1). 66-69.

[11] Gwizdka, J., Email Task Management Styles. in *CHI'2004 Conference on Human Factors in Computing Systems Extended Abstracts*, (Vienna, Austria, 2004), 1235-1238.

[12] Mackay, W.E., More than just a communication system: diversity in the use of electronic mail. in *CSCW'88 Conference on Computer-Supported Cooperative Work*, (Portland, Oregon, USA, 1988), ACM Press, 344-353.

[13] 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). 99-112.

[14] Mander, R., Salomon, G. and Wong, Y.Y., A 'Pile' Metaphor for Supporting Casual Organization of Information. in *CHI'92 Conference on Human Factors in Computing Systems*, (Monterey, California, USA, 1992), 627-634.

[15] Rekimoto, J., Time Machine Computing: A time-centric approach for the information environment. in *UIST'99 Symposium on User Interface Software and Technology*, (Asheville, North Carolina, USA, 1999), 45-54.

[16] Teevan, J., Alvarado, C., Ackerman, M.S. and Karger, D.R., The Perfect Search Engine Is Not Enough: A Study of Orienteering Behaviour in Directed Search. in *CHI'2004 Conference on Human Factors in Computing Systems*, (Vienna, Austria, 2004), 415-422.

[17] Whittaker, S. and Hirschberg, J. The Character, Value, and Management of Personal Paper Archives. *ACM Transactions on Computer-Human Interaction, 8* (2). 150-170.

[18] Whittaker, S. and Sidner, C., Email Overload: exploring personal information management of email. in *CHI'96 Conference on Human Factors in Computing Systems*, (Vancouver, Canada, 1996), 276-283.

Using mobile phone contextual information to facilitate managing image collections

Jakob Eg Larsen Maciej Luniewski
Technical University of Denmark
Department of Informatics and Mathematical Modeling
Richard Petersen Plads, Building 321
DK-2800 Kgs. Lyngby, Denmark
jel@imm.dtu.dk, maclun@gmail.com

ABSTRACT
In this paper, we describe a prototype application that utilizes the embedded sensors in advanced mobile phones to infer meaningful contextual information, with the potential to support the users in managing their personal information. Contextual information such as time, location, movement, surrounding networks, devices, people, and application data is used to semi-automatically annotate information in our current proof-of-concept prototype. The application allows the derived contextual information to be annotated as tags to available content and thereby facilitating the processes involved in personal information management. We hypothesize that information inferred from embedded mobile phone sensors can offer useful contextual information for managing personal information, including the domain of interest here, namely image collections. This has potential for individuals as well as groups managing shared image collections or other types of information.

Keywords
Personal information management, context, context-awareness, metadata, image collections, mobile phones, embedded sensors

INTRODUCTION
Personal information management can require tedious efforts by the person managing the information even in modern interactive tools aiming to support the activities involved. This involves activities such as to create, store, organize, retrieve and use information to handle daily activities. For instance organizing the information into hierarchical structures in email and file systems. The concept of tagging has become popular in Web 2.0 social networks characterized by user generated and organized content. By labeling the content with one or more tags different ways of navigating and retrieving the information is made possible. However, manually labeling content with tags requires efforts from the user. This introduces a dilemma of storing and retrieving personal information. If less effort is spent on storing the information (e.g. a piling strategy) more effort is required to retrieve the information and vice versa [7]. By choosing a personal filing strategy individuals implicitly choose a balance of effort spent on storing and retrieving information respectively. Thus a challenge for tools aiming to support personal information management could be to minimize, to the possible extent, the effort needed for both storing and retrieving information items.

We introduce a proof-of-concept application implemented for modern mobile phones, which allows the user to semi-automatically annotate information with contextual information [4] derived from the multiple embedded sensors. In the specific application prototype facilitating the management of images captured using the embedded camera has been the focus. The semi-automatically annotated tags can be used for storing, retrieving and managing collections of images. By presenting the derived contextual tags to the user the aim is to support the users in recalling and recognizing their personal information organization when later retrieving the information. The importance of such contextual information in personal information management was emphasized by Barreau [1].

Even though the focus here has been on image collections, we speculate that the principles of semi-automatic metadata annotation (tagging) can apply to any information type dealt with as part of everyday personal information management. The collective sources of sensor information provide descriptions of the multiple contexts in which the user of the device is part of on a daily basis. The rich set of sensor information available on modern mobile phones provide multiple ways of potentially inferring contexts such as location, places, people, environment, activities and movement. When tags are explicitly chosen by the user it also serves as implicit feedback to the system in order for similar contextual situations to be recognized at a later stage and thus over time improve the likelihood of suggesting useful tags.

CONTEXT GALLERY APPLICATION
We have constructed a proof-of-concept prototype application Context Gallery for navigation and semi-automatic annotation of images by means of contextual information available through a mobile phone. The application contains two parts. The first is a gallery that provides multiple interfaces for the user to navigate, browse, and retrieve images by means of the annotated tags. The second part of the application contains modules that continuously read the multiple sensors available and establish the current user's context. Each module can be switched on and off by the user.

The prototype application is developed for Nokia N95 mobile phones, which provide the following set of embedded sensors:

- GPS – for location discovery (accuracy in meters)
- GSM cellular information – for place discovery (accuracy of meters to kilometers)

- WLAN – for discovery of places
- Bluetooth – for device and people discovery
- Application data – calendar information (activities)
- Phone specific data – state, profile (state information)
- Light sensor – available when capturing images with the built-in camera.
- Accelerometer – to determine movement patterns (not included in the present prototype)
- Microphone – could be used to establish noise level and activities (not included in present prototype)

Devices with additional sensors will simply enable more contextual information to be derived. Most of the raw sensor data is not directly usable for the end-user of the application. For instance, the coordinates acquired by the GPS sensor or the GSM network information has little meaning to most people. Therefore, a set of online services[1] are used in order to infer and translate the sensor data into human readable information, that is, meaningful descriptions and labels. The mentioned GPS and GSM information [10] is translated into place descriptions (city, region, country, etc). Weather information (temperature, humidity, clouded) for the particular region is also acquired. Bluetooth is used to discover hand held devices and through user provided feedback translated into people present. Finally, time is an obvious parameter [5], which is also translated into descriptions including season, month, weekdays, weekend, morning, afternoon, and light condition (dusk, dawn) etc. Time is also used to group images into events, dividing by levels of the time span in between the pictures taken. This is supported by calendar events (if present). Images taken within a time span is classified as belonging to an event and associated to a corresponding collection.

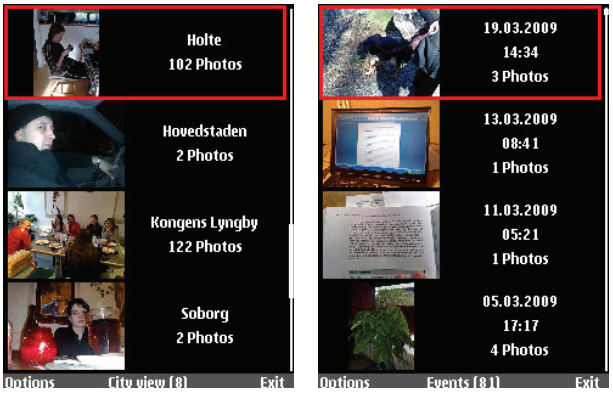


Figure 1. Screenshots from the prototype mobile application with "view by place" and "view by event"

Beyond these system derived contextual labels, the user has the option to add his/her own textual labels simply by manual input of one or more tags to be annotated to the information item (image) or collection. As mentioned above, such input can serve as feedback to the system in terms of not only annotating the particular information item, but also annotating the corresponding context in which the information is used.

Based on the multiple annotations of the images the information stored in the application repository is indexed in order to allow retrieval to be performed quickly and to establish different views of the content. Views are organized as lists of photo collections, in which images are presented in grids of a chosen size, as shown in Figure 1. The current prototype allows the user to view the information by time (days, weeks, and months), places, events, and similar images based on similar context or other user-defined search criteria. Searching by means of the annotated tags is possible by means of an interface where one or more tags can be indicated. Figure 1 contains screenshots from the prototype mobile application showing "view by place" and "view by event".

Initial experiments with the proof-of-concept prototype application have been carried out. Three test participants used the application in the greater Copenhagen area in Denmark for four weeks in order to gather contextual information. It was found that the information that could be obtained from the sensors could be translated into meaningful and useful context descriptions for annotating the images allowing multiple ways of navigating and retrieving the images. Furthermore, the prototype application was performance tested with 1700+ existing images annotated with contextual information (temporal annotation and some with location), which showed the capabilities with a high quantity of images.

DISCUSSION AND CONCLUSIONS

Related image browsing applications exist [2,3,6,9,11,12,13], where the emphasis has been on photo management and less on the general aspects of and implications for personal information management as well as facilitating information management by means of derived contextual information. Our focus is personal information management facilitated by establishing user contexts by means of multiple mobile phone sensors and enhancing this by also including the personal and manual annotations provided by the user. This provides useful contextual feedback to the system. Thus when a similar context is detected later the manually provided tags in that context could be used along with the tags derived from the embedded sensors. In the present work we have focused on contextual information about the environment inferred by means of the embedded sensors. However, additional information could be acquired either by analysis of the content itself [9], (e.g. face recognition) or from descriptions derived from tags generated in social networks. In future versions the derived context information could be shared among several users, thus enabling them to potentially benefit from contextual labels added by other users for instance in a social network. ZoneTag is an example of such an application that utilizes social networking in the image annotation process [8]. Users can share tags for specific locations (Cell ID or GPS based), that is, user contributed tags are propagated to other users in the social network, in terms of suggestions in similar locations. Context Gallery utilizes a wider range of embedded sensors and the spatial context is translated automatically (both GPS and Cell ID) with use of web services, which results in decrement of user's involvement in storing images. Contrary ZoneTag, the inference process primarily takes place on the mobile phone in Context Gallery. As a result, in situations where the context is already known (the second picture

[1] http://geonames.org/, http://opencellid.org/, http://wigle.net/, http://world-gazetteer.com/, and http://maps.google.com

taken in the same location), the location description is retrieved from the local storage without connections to web services.

Our initial prototype and experiments has shown that it is possible to infer meaningful contextual information from embedded mobile phone sensors, such as, GPS, Bluetooth, cellular and WiFi network information, and other low-level sensors along application information, such as the calendar. The contextual information can potentially be used to facilitate management of personal information such as in the information domain of interest here, namely image collections. Contextual information has the potential to provide multiple ways of searching and retrieving the information at need, depending on the recall by the user. However, further experiments are needed in order to establish to what extent this approach apply for other types of information in order to support the activities involved in personal information management. In particular usability tests are needed in order to establish the usefulness in recall based on use of the semi-automatic annotations provided.

ACKNOWLEDGMENTS

This research is supported by the Danish Strategic Research Councils through the project "Measurement Systems for Ethical Capital in the Experience Economy". The authors would like to thank the participants which took part in the initial experiments with the prototype application and Forum Nokia for the equipment used in the experiments.

REFERENCES

[1] Barreau, D. Context as a factor in personal information management system. Journal of the American Society for Information Science, 1995. Vol. 46(5): 327-339.

[2] Davis, M., House, N., Towle, J., King, S., Ahern, S., Burgener, C., Perkel, D., Finn, M.,Viswanathan, V., Rothenberg, M. MMM2: Mobile Media Metadata for Media Sharing. CHI '05 extended abstracts on Human factors in computing systems. New York, USA : 13th Annual ACM International Conference on Multimedia , ACM Press, 2005. pp. 1335-1338.

[3] Davis, M. and Sarvas, R. Mobile Media Metadata for Mobile Imaging. 2004 IEEE International Conference on Multimedia and Expo (ICME). 2004, Vol. 3, pp. 1707-1710.

[4] Dey, A.K., Abowd G.D. Towards a Better Understanding of Context and Context-Awareness. Lecture Notes in Computer Science. 1999, Vol. 1707, pp. 304-308

[5] Graham, A., Garcia-Molina, H., Paepcke, A., T., Winograd. Time as Essence for Photo Browsing Through Personal Digital Libraries. Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries. 2002, pp. 326-335.

[6] Kim, S.K., Lee, J.W., Lee, R., Hwang, E.H., Chung, M.G. User-Friendly Personal Photo Browsing for Mobile Devices. ETRI Journal, 2008. Vol. 30(3): 432-440.

[7] Larsen, J.E.: NEXUS – A unified approach to personal information management in interactive systems, Ph.D. thesis, Technical University of Denmark, 2005.

[8] Naaman, M., and Nair, R. ZoneTag's Collaborative Tag Suggestions: What is This Person Doing in My Phone? IEEE Multimedia 2008, 15(3):34-40.

[9] O'Hare, O., Lee, H., Cooray, S., Gurrin, S., Jones, G.J.F., Malobabic, J., O'Connor, N.E., A.F., Smeaton, Uscilowski, B. MediAssist: Using Content-Based Analysis and Context to Manage Personal Photo Collections. Springer Berlin / Heidelberg. 2006, pp. 529-532.

[10] Trevisani, E. and Vitaletti, A. (2004). Cell-ID location technique, limits and benefits: an experimental study. Sixth IEEE Workshop on Mobile Computing Systems and Applications, WMCSA 2004. pp. 51–60.

[11] Tvarozek, M. and Bielikova, M. Personalized Faceted Navigation for Multimedia Collections. Washington, DC, USA : Proceedings of the Second International Workshop on Semantic Media Adaptation and Personalization, 2007. pp. 104-109.

[12] Wilhelm, A., Takhteyev, Y., Sarvas, R., Van House, N., and Davis, M. Photo Annotation on a Camera Phone. Vienna, Austria : ACM Press, Conference on Human Factors in Computing Systems, 2004. pp. 1403-1406.

[13] Yee, K. P., Swearingen, K., Li, K., Hearst, M. Faceted Metadata for Image Search and Browsing. Ft. Lauderdale, Florida, USA: Conference on Human Factors in Computing Systems, 2003. pp. 401-408.

Integrating Place and Time with Tasks: Adding Transportation to PIM

Stephen Lackey
University at Albany, SUNY
Albany, NY 12222
SL239263@albany.edu

ABSTRACT
Personal Information Management research has examined the development of applications and data structures to overcome human limitations in memory and cognition. Transportation researchers have struggled with analyzing transportation activity, an intermediate good derived from fulfillment of scheduled appointments and task completion. Recent work has explored the promise of mobile computing to improve task efficiency by bridging the gap between needed information and the physical location of the user. This paper argues that spatial-temporal extensions to personal information data document are essential to improve the efficiency of task completion and coordination.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval.

General Terms
Design, Human Factors, Theory

Keywords
Personal Information Management, Activity Analysis, Task management, Time, Personal Workflow management

1. INTRODUCTION
This proposal presents a framework that builds on research in Personal Information Management, Mobile Computing, and transportation Activity Analysis. Extensions of familiar PIM tools will improve management of tasks, time, and transportation resources to improve awareness of the travel and costs of commutes. Interfaces to external data sources enable opportunities for resource discovery and savings to improve the efficiency of task completion.

Computing tools have been seen as an effective way to overcome the limitations of human memory and manage the complexity of modern life [1]. The emphasis on Personal Information Management (PIM) has focused on its role as memory extension- to replace the task of remembering information with merely remembering how to find it. These tools have remained comparatively unchanged since their origin with Personal Computers in the early 1980's. PIM tools and methodologies have sought to improve the capture, organization, and retrieval of information deemed useful to further the goals of the individual.

The classic PIM applications have merely extended traditional work patterns [2,3]. Components typically include a list of addresses, a list of calendar items (appointments), a Task List, and a list of individual, free-form text notes. In electronic form, these are treated as separate applications, with separate data stores. These information sources are fragmented into "information islands", duplicated between multiple applications or devices, with minimal connections between these islands. [4,5]

This separation is further exacerbated by the differing contexts of information use; user information needs are not limited to the availability of stationary computing resources, but while mobile as well [6]. While this has advantages of simplicity, the reusability of personal information is limited at best. Data portability between PIM applications has proven limited due to a lack of open standards between applications.

PIM tools have focused storage and retrieval tasks within these "islands", where the data stored there serves as reminders of tasks and appointments. As few as 14% of needed "to-do" tasks may be organized as lists of any sort, and completed less than 80% of the time [7]. Routine, essential tasks tended to become encoded in habit, rather than recorded. The result is an excessive reliance on memory, which can quickly become overburdened. Habit can become inefficient as the user's operating environment changes. The task of traditional PIM has been to balance the task of merely extending memory without overloading it, a balance that has proven to vary by individual and work context.

2. RELATED WORK
2.1 Managing Personal Tasks and Workflows
Research has begun to focus on task management as a deficiency in traditional PIM tools. Task lists have proven to be effective reminders of critical tasks [7], with appropriate cues effectively limiting errors [8]. The occurrence of errors in task completion recognizes the divisibility of tasks into subtasks, and the need to explicitly manage the workflow of discrete tasks to reach desired objectives.

A primary shortcoming of PIM tools is their inability to accommodate the process-orientation of daily activities [9]. Though addressed in Project Management and Workflow Management processes [10], this capability remains virtually unknown in PIM

tools. Though appointments, tasks, and addresses remain essential items to record and retrieve, the realization of relationships between these PIM items remain an essential yet missing component in task management [11].

Workflow management in the completion of larger tasks emerged from Computer Supported Collaborative Work (CSCW) research, but focuses on multiple actors in a workplace setting, as embedded in business functions. Personal workflow management research has emerged from directions as diverse as Grid Computing and Mobile computing. These projects share a common thread in the realization of tasks as aggregates of separate functions that require the coordination of external resources under constraints. However, work to associate tasks with locations has been limited.

2.2 Activity Analysis

Research in Personal Information Management and Task Management has focused on reminding the individual about tasks and appointments, while the transportation literature has examined the roles of personal activities in generating trips. Transportation is primarily a derived good; demand is based on its role in fulfilling external objectives. Unlike PIM efforts, activities are seen as highly location-dependent, constrained by time.

Hägerstrand [12-15] first discussed the importance of "time bundling" to find common corridors for transportation in his urban studies. He further stresses the limitations on activities based on time and location via the "time-space prism" model. Therein, location and transportation mode constrain the range of possible activities that can be completed within a fixed amount of time. The constraint of time on activity has reemerged in the 1990's within transportation research with the emergence of Information Technology tools that promised the ability to manage the complexity of coordinating and scheduling activities.

Time is a finite resource. Though it cannot be controlled, tasks and goals can be ordered to better use time and other resources towards greater efficiency. Currently, tools such as Personal Information Management programs, and appliances such as Personal Digital Assistants or Smart Phones have been created to augment human memory, to act as reminders of appointments and tasks prior to deadlines. However, these tools serve as reminders of a linear sequence of fixed events, rather than aid in the sequencing of flexible activities under constraints.

The concept of a Travel Time Budget (TTB) has emerged as a pool of time that travelers have been willing to allocate to travel in the course of making lifestyle choices in housing and work locations, though various demographic and employment factors constrain this budget. Since initial formulations in the early 1980's [16], the support for Travel Time Budgets has remained consistent. The existence of said budgets reinforces both the notion of a "time-space prism" model and the need to optimize activities within these constraints.

2.3 ATIS and Activity Management

More recent developments in Advanced Traveler Information Systems (ATIS) have sought to incorporate current developments in Information and Communication Technologies (ICT) in improving the effectiveness of trips [17]. These tools have begun to investigate opportunities to incorporate ICT to replace travel. ATIS seeks information about travel in progress to reduce the time between activities, allowing a greater number of activities within a fixed time period.

ATIS research has a limited focus of ICT to alter travel behavior [18]. Its focus on "telecommunications" may limit its understanding of the applications of ICT in not only voice and e-Commerce, but applications in coordination and collaboration which permit opportunities to gather and share information. The substitution of travel with ICT use remains unclear; though ICT use occupies a growing share of personal time, the link between activity and physical location has diminished [19], and some ICT use has no physical counterparts.

Incorporation of ICT with personal task management and external information such as ATIS or public transportation schedules hold the promise of clustering activities by location and travel corridors [17]. Clustering of activities would permit the reduction of vehicle trips and the overall efficiency of the transportation network.

2.4 Mobile Computing and Communications

Personal Digital Assistants (PDA) and Smart Phones have been widely embraced as means to help manage our daily affairs, despite their oft-cited limitations [20]. Activity and location have been found to comprise up to 70% of mobile communications, with discussions about activity twice as likely as those about location [21]. Mobile communications differ by urgency or spontaneity, and brevity [2].

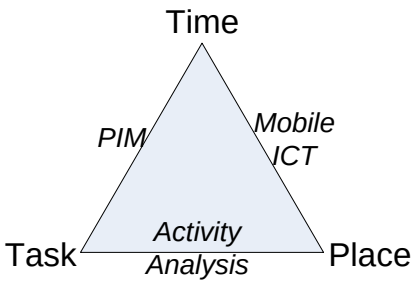
Mobile ICT research in optimizing travel effectiveness have focused on the use of task reminders. Through mechanisms such as a "Geofence", events such as approaching a location are designed to trigger reminders to perform an associated task [22,23]. Subsequent findings have held that location-based reminders were perceived as less useful than time-based reminders [24]. Time reminders remain a standard feature of PIM Calendars, where travel time may be implicit in the setting of the ubiquitous "Alarm".

These findings conflict with ATIS and Activity Management efforts in using location as a means of clustering tasks. Where ATIS and Activity Management focus on locations and the activities available, the Task or To-Do list focuses on deadlines alone. This indicates the difficulty in task decomposition, or the inability to relate place and time with tasks.

3. Conceptual Framework

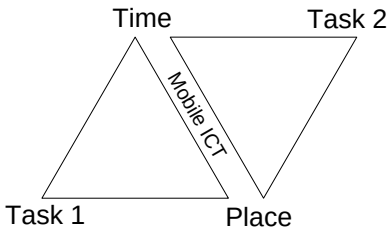
The simplified conceptual framework proposed here attempts to reconcile the contrasting goals of PIM, Activity Analysis, and Mobile ICT. Notably, the PIM focus on Time and Task overlaps the Activity Analysis focus on Task and Location. Mobile ICT in part focuses on delivering services based on time and location.

Figure 1: The Proposed Conceptual Framework



Though imperfect, these linkages between PIM, Mobile ICT, and Activity Analysis had lacked explicit treatment, despite a growing body of research which implicitly overlaps these areas. In a sample Task modeled in Figure 1, attributes of time and place are included. Tasks may have a range of locations and times where they can be performed. However, tasks as currently defined may be part of a workflow, which unifies tasks set in separate places and times. Location and time assignments to tasks would enable the realization of the goals of Activity Analysis, in merging tasks by location to minimize travels.

Figure 2: Tasks arranged by time/place overlaps



Within the framework, it is considered that Mobile ICTs that facilitate the management of tasks within times and places would be needed to further manage the workflows that intersect there. Research tools that enable "geofencing" [22,23] implicitly suggest tasks from separate workflows to combine with the one currently being followed, using opportunism to improve task efficiency rather than explicitly planning for it.

4. CONCLUSION

This work attempts to bring together divergent research in Mobile ICT, Activity Analysis, and Personal Information Management within the context of Transportation. The addition of location and time attributes with tasks has received brief reviews in the context of Mobile ICT. Standardization of these attributes may prove an essential component in Personal Information Management.

Research into Personal Workflow Management has emerged as a means of enabling workers to make sense of recorded information in Mobile ICT, owing to their greater constraints. Improving task management capabilities would offers greater responsiveness to changing circumstances during task execution. Though traditional PIM software has limited ability to manage workflows, assisted management of tasks and schedules via Mobile ICT remains a source for substantial improvement.

Transportation research has long understood that travel is based on its role in meeting external objectives. Despite a growing interest in traveler activities as a source of travel demand, users have lacked the infrastructure and ability to integrate transportation considerations into the individual decision processes. Spatially-encoded PIM data holds the promise of reducing travel demand and increasing the effectiveness of trips taken.

Personal Information Management has long been seen as burdensome and application-specific, particularly in the context of Mobile ICT. Replacing today's "information islands" with data interchange standards, or even a centrally-shared data structure, would permit greater data sharing between various PIM tools. The addition of time and location attributes would improve task and time management within PIM systems, while permitting greater responsiveness during task execution. Real time monitoring of time and location would enable the goal of access to the right information at exactly the right time and place, improving effectiveness in task completion.

5. ACKNOWLEDGMENTS

This work has emerged from a dissertation in progress, to help understand information barriers faced by travelers in utilizing mass transit despite stated preferences. I am grateful to Catherine Lawson, Eliot Rich, Sandor Schuman, and James Looby for their insights.

6. REFERENCES

[1] D.C. Engelbart, *Augmenting Human Intellect: A Conceptual Framework. Summary Report AFOSR-3223 under Contract AF 49 (638)-1024, SRI Project 3578 for Air Force Office of Scientific Research*, Stanford Research Institute., 1962.

[2] G. Schrott and J. Glückler, "What makes mobile computer supported cooperative work mobile? Towards a better understanding of cooperative mobile interactions," *International Journal of Human-Computer Studies*, vol. 60, May. 2004, pp. 737-752.

[3] J. Gwizdka and M. Chignell, "Individual Differences in Personal Information Management," *Personal Information Management*, University of Washington Press, 2007.

[4] W.P. Jones and J. Teevan, *Personal information management*, 2007.

[5] D. Elsweiler, "PhD: Thesis - "Supporting Human Memory in Personal Information Management"," University of Strathclyde, 2007.

[6] P.L. Mokhtarian and C. Chen, "TTB or Not TTB, that is the Question: A Review and Analysis of the Empirical Literature on Travel Time (and Money) Budgets," *Transportation Research Part A*, vol. 38, Dec. 2003, pp. 643-675.

[7] V. Bellotti, B. Dalal, N. Good, P. Flynn, D.G. Bobrow, and N. Ducheneaut, "What a to-do: studies of task management towards the design of a personal task list manager," *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienna, Austria: ACM, 2004, pp. 735-742.

[8] P.H. Chung and M.D. Byrne, "Cue effectiveness in mitigating postcompletion errors in a routine procedural task," *International Journal of Human-Computer Studies*, vol. 66, Apr. 2008, pp. 217-232.

[9] S. Hwang and Y. Chen, "Personal Workflows: Modeling and Management," *Mobile Data Management*, 2003, pp. 141-152.

[10] H. Tarumi, K. Kida, Y. Ishiguro, K. Yoshifu, and T. Asakura, "WorkWeb system--multi-workflow management with a multi-agent system," *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, Phoenix, Arizona, United States: ACM, 1997, pp. 299-308.

[11] E. Falke, "The associative pda 2.0," *CHI '08 extended abstracts on Human factors in computing systems*, Florence, Italy: ACM, 2008, pp. 3807-3812.

[12] T. Hägerstrand, "Reflections on "what About People in Regional Science?"," *Papers in Regional Science*, vol. 66,

1989, pp. 1-6.
[13] F. Snickars, *Dynamic Allocation of Urban Space*, Saxon House, 1975.
[14] K. Ellegård, "A time-geographical approach to the study of everyday life of individuals – a challenge of complexity," *GeoJournal*, vol. 48, Jul. 1999, pp. 167-175.
[15] T. Hägerstrand, "Space, time and human conditions," *Dynamic allocation of urban space*, 1975, pp. 3–12.
[16] B.P. Bailey and S.T. Iqbal, "Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management," *ACM Trans. Comput.-Hum. Interact.*, vol. 14, 2008, pp. 1-28.
[17] K. Vaughn, M. Abdel-Aty, and R. Kitamura, "A framework for developing a daily activity and multimodal travel planner," *International Transactions in Operational Research*, vol. 6, 1999, pp. 107-121.
[18] *Managing Demand Through Travel Information Services*, U.S. Department of Transportation Federal Highway Administration, 2004.
[19] S. Shaw and H. Yu, "A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space," *Journal of Transport Geography*, vol. 17, Mar. 2009, pp. 141-149.
[20] J.A. Fails, A. Druin, and M.L. Guha, "Collocated mobile collaboration," *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, Boston, MA, USA: ACM, 2009, pp. 3495-3496.
[21] F.R. Bentley and C.J. Metcalf, "Location and activity sharing in everyday mobile communication," *CHI '08 extended abstracts on Human factors in computing systems*, Florence, Italy: ACM, 2008, pp. 2453-2462.
[22] P.J. Ludford, D. Frankowski, K. Reily, K. Wilms, and L. Terveen, "Because I carry my cell phone anyway: functional location-based reminder applications," *Proceedings of the SIGCHI conference on Human Factors in computing systems*, Montréal, Québec, Canada: ACM, 2006, pp. 889-898.
[23] N. Marmasse and C. Schmandt, "Location-Aware Information Delivery with ComMotion," *Handheld and Ubiquitous Computing*, 2000, pp. 361-370.
[24] A. Kessell and C. Chan, "Castaway: a context-aware task management system," *CHI '06 extended abstracts on Human factors in computing systems*, Montréal, Québec, Canada: ACM, 2006, pp. 941-946.

Writing New Stories for the Same Old File System

William Jones
University of Washington
Seattle, Washington 98195,USA
williamj@u.washington.edu

Dawei Hou
University of Washington
Seattle, Washington 98195, USA
daweiuw@gmail.com

Bhuricha Deen Sethanandha
Portland State University
Portland, Oregon 97207, USA
bhuricha@cs.pdx.edu

Eric Sheng Bi
University of Washington
Seattle, Washington 98195, USA
ericshengbi@gmail.com

CHI 2009, April 4 – April 9, 2009, Boston, MA, USA
ACM 978-1-60558-247-4/08/04.

ABSTRACT
Planz enables the creation of a document-like overlay to the file system. Headings and subheadings correspond to file system folders and subfolders. The notes under a heading can point to files – and also to email messages and web pages. Headings and notes can express the structure of a personal project. This structure, in turn, provides a basis for organizing the various forms of information needed for project completion.

INTRODUCTION
The Personal Project Planner [6] (Planz for short) addresses a basic challenge of our modern lives: Our digital information comes in many different forms (files, email messages, read-only and transactional Web pages of all kinds) and is often scattered in many different ways.

Tools meant to help can actually make things worse. Special purpose tools such as OmniPlan[1] or even a general-purpose tool such as Microsoft OneNote[2] tend to "box us in" by imposing a fixed structure of dialog boxes, tabs and pages that don't always fit with the more fluid, informal nature of our everyday planning. Also, the organizations of different tools are usually

[1] http://www.omnigroup.com/applications/omniplan/.

[2] http://office.microsoft.com/en-us/onenote/default.aspx.

incompatible with each other and with the organizations we've created already to manage files, email messages and web pages.

How can we organize our information—and ourselves—to stay on track? How can we do this without spending all of our time on organization? Planz addresses each of these issues by adhering to two basic principles:

1. **Plan first, let the organization follow.** The Plans you create in Planz are part of a single document that looks like the draft document you might create in a word processor or the notes you might scribble to yourself on a notepad. You plan by typing your thoughts freehand. Then link your thoughts to files, web pages, and email messages. You organize your thoughts under headings and subheadings over time and only according to your needs. Access your Plans and the information you need to complete these Plans through a single integrated document.

2. **No new organization.** The Plans you create in Planz are essentially an alternate way to work with your existing files and folders[3]. The headings and subheadings of a Plan represent folders and subfolders in your file system. Use these to organize not only files but also e-mail messages and web references. Change your Plans as easily as you might change a document in a word processor.

WHY THE "SAME OLD FILE SYSTEM"?

Features of Planz can be found in other tools. However, Planz is unique in its efforts to integrate various features of word-processing and linking as an overlay to a folder-based hierarchical file system. Planz represents an effort to explore new modes of interacting with information organized into the files and folders of a conventional file system.

Why the file system? The file system – especially that portion local to a personal computer -- is still where people are most likely to feel a sense of ownership and to invest in efforts to organize their information [2, 3, 5].There may be many good reasons to believe that people will increasingly entrust their information to the Web (e.g., for easier maintenance and anywhere access). However, if and as this transition occurs, there are also good reasons to believe that the hierarchical folder model will follow[4]. For example, even as support for tagging and search continues to improve, an established preference for browsing or location-based finding [1] endures [2]. Moreover, people use their folder hierarchies for many other things besides the enablement of browsing. People report, for example, that folders help them to represent and track tasks to be done and to "see" their information better [5].

To be sure, a strict hierarchy suffers from a basic limitation that while an information item can only appear in or be referenced from one location (e.g., a folder), we may want to access and work with this item in several different contexts. After noting several variations of this basic limitation, Dourish et al [8] go on to propose a radical property-based, "placeless" solution.

By contrast, the approach taken in the design of Planz is modest with respect to changes envisioned. What

[3] Planz currently works with Microsoft XP, Vista & 7. However, the approach in PLanz readily extends to other file systems.

[4] An irony, of course, is that the hierarchical file system is already "there" as an underlying basis for organizing the information of the Web.

can be accomplished through small changes realized in a light-weight overlay to an existing file system? Can this be done in ways that preserve and even enhance a sense of digital place -- a sense of knowing "where to go" – that is already evident the preference for browsing vs. search?

Planz is designed to work like a basic "no frills" word processor (see Figures 1 & 2). But, as noted above, the document created in Planz is an overlay to the file system. Headings in the document correspond to folders in the file system. The notes under a heading point to files under a corresponding folder. Some of these files are "local". Others may be shortcuts pointing to files, folders, email messages or web pages that are elsewhere (in the file system, in storage managed by Outlook or on the Web).

KEY FEATURES OF THE DOCUMENT OVERLAY

Through its maintenance of a document overlay, Planz supports several key features:

- ***Drag & link***. Select a file, an email message or (via its address) a web page and drag into Planz. The selected item stays where it is. But a new note with a link to this item is created in the Planz document. Or select text from the item, drag and drop into Planz. The selected text is now included as a new note in the Planz document but with a link pointing back to the source information item. Drag & link is, as one user noted, a "kind of super-shortcut maker". An item is stored in only one place (e.g. as managed by the file system, Outlook or a web server) but can easily appear in any number of places.[5]
- ***Annotate***. Simply type – anywhere – in the Planz document to add reminders, explanations, elaborations, etc. Support for annotation is motivated by anecdotal comments and fieldwork observations [7] that people often devise elaborate workarounds to annotate their files and folders including, for example, naming conventions or the provision for a companion "documentation document" to explain a folder's contents.
- ***Order***. Move the elements (headings and notes) of a Planz document up or down just as you might move the sections or paragraphs of a document using a word processor. Move elements in Planz as an alternative to moving files and folders more directly through the use of a conventional file manager. Or make smaller moves to establish a local ordering. The importance of order is evident in the naming schemes people often devise to insure that files and folders are listed (by name) in a certain order or to insure that key files or folders appear first in a listing [5].
- ***Create*** a new document or create and send a new email message "in-context". The item is created as it would be normally (e.g., in separate windows managed by a supporting application such as Microsoft Word or Outlook). However, Planz creates a new note pointing to this item near the insertion point in the Planz document.

A STORY OF PLANZ USE

In the following story, Bob is beginning to think about the re-model of his house. He uses Planz (figure 1) to record his initial thoughts just as he once used a word processor.

[5] Links (shortcuts) are represented by a small shortcut component in a note's icon. Initial evaluation suggests that people understand the distinction between a note icon representing an item (e.g., a file created locally) and a note icon for a link. The messages to confirm a deletion also are also distinct for these two cases. However, the possibility for confusion between item and its links remains a focus of ongoing evaluation.

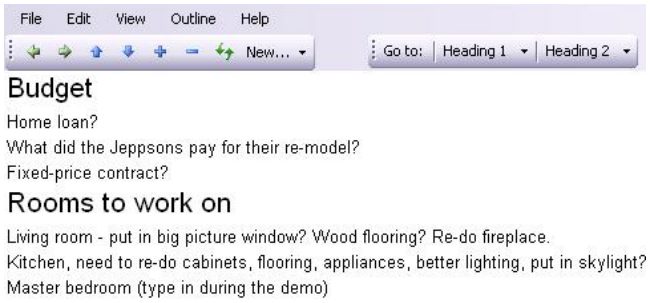


Figure 1. An early view of a "house re-model" in Planz.

Bob continues to think about the re-model and he continues to write his thoughts down. Several weeks later, the re-model plan is much larger. At this point, Bob may decide to switch to an outline view in Planz for greater control over the ordering and presentation of components in the re-model plan (figure 2).

The headings in figure 2 correspond to folders; the subheadings to subfolders; the icons represent files or file shortcuts (e.g., to web pages, email messages or remote files).

Notice then, that the re-model plan also provides a basis for grouping together project-related information in various forms, from various sources.

WHAT'S NEW?

The key features of the Planz document overlay, as listed above, were previously described in connection with an earlier version of Planz (referred to as "the Planner")[6]. At that point the prototype provided only

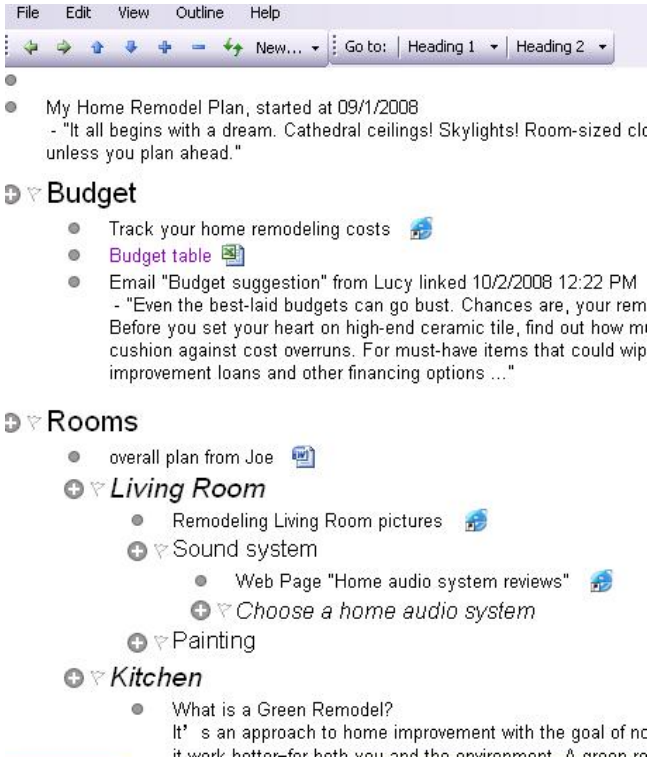


Figure 2. A later, outline view of the "house re-model".

partial for support for these features and could only be used for short periods of time in a directed, "hands-on" demo. In its current version, Planz provides full support for these key features and is available for free download[6]. Planz has a small but dedicated base of

[6] http://kftf.ischool.washington.edu/planner_index.htm.

users and has been used for periods of seven to fourteen days as part of a controlled evaluation. [7]

The first answer, therefore, to the question "What's new?", is that Planz is now "real" and can be (is being) used for the daily management of personal information. Three new features are also noteworthy:

1. **Two-Click Access** –from any document, email message or web page, click once (Windows-o) to jump back to a note referencing this item in Planz. Click a second time, on the icons for a notes nearby, to access related information.

2. **In-place Expansion**. Link to any folder, anywhere; promote the resulting note to be a heading; expand to see the folder's contents. In this manner, information can stay where it is (e.g., on a network share) but still be included for access and manipulation in the current view.

3. **Consolidated Support for Task & Time Management**. Click to the left of a heading to create a task or an appointment elsewhere (Figure 3). Click again to check a task as completed.

[7] This evaluation delivered good and bad news. On the plus side, all eight participants expressed appreciation for the "at-a-glance" visibility of information and project status afforded by Planz. Three participants expressed an intention to continue using Planz beyond the period of the evaluation. On the minus side, three (other) participants indicated that they came to appreciate Planz only late in the evaluation and that they would not consider using Planz unless it was more "product-like" (faster, smoother, with more features). The results of this evaluation are the subject of another article (under preparation) and are not a focus of this extended abstract.

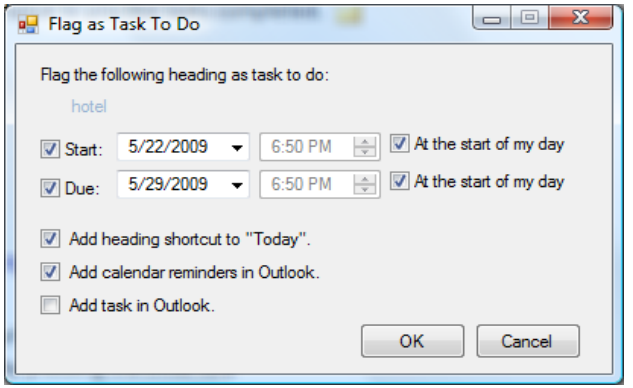


Figure 3. Any heading can be flagged to appear as a task or an appointment in Microsoft Outlook.

PLANZ ARCHITECTURE

Behind the scenes, Planz is able to support its more document-like outline view by placing an XML fragment in each folder. Planz assembles fragments on demand to present a coherent view of headings, notes and links.

The distribution of XML fragments in association with folders has more general application in support of a variety of views into personal information including mind-mapping, decision tree, workflow and tabular views.

SUMMARY

Planz begins with the notion that describing projects and project structure should be as easy as typing thoughts, free-hand, into a document. Represent projects as major headings in the document. Use subheadings (to any level) to represent the areas of a project or individual tasks or anything else of relevance to the project. The hierarchy of headings corresponds

to a folder hierarchy in the file system. This hierarchy represents not only projects and their structure but can also be used to organize project-related documents, email messages and web pages.

Work on Planz continues. Next steps include:
1. Completion of a longer evaluation during which participants will use Planz for 3 to 8 weeks.
2. Conversion of Planz to make use of Windows Presentation Foundation (WPF)[8] for a much improved user experience.
3. Web-based and mobile versions of Planz that further explore the Planz approach of dynamically generating coherent views from an assembly of XML fragments.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. 0097855 and 0534386

CITATIONS

[1] Barreau, D. K. and B. Nardi (1995). "Finding and reminding: file organization from the desktop." SIGCHI Bulletin **27**(3): 7.

[2] Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., and Whittaker, S. 2008. Improved search engines and navigation preference in personal information management. ACM Trans. Inf. Syst. 26, 4 (Sep. 2008), 1-24.

[3] Boardman, R. and M. A. Sasse (2004). "Stuff goes into the computer and doesn't come out" A cross-tool study of personal information management. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004), Vienna, Austria.

[4] Jones, E., H. Bruce, Klasnja, P, & Jones, W. (2008). "I Give Up!" Five Factors that Contribute to the Abandonment of Information Management Strategies. 68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2008). Columbus, OH.

[5] Jones, W., Phuwanartnurak, A. J., Gill, R. & Bruce, H. (2005). Don't take my folders away! Organizing personal information to get things done. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2005), Portland, OR, ACM Press.

[6] Jones, W., Klasnja, P., Civan, A. & Adcock, M. (2008). The Personal Project Planner: Planning to Organize Personal Information. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2008), Florence, Italy, ACM, New York, NY.

[7] Voida, S., & Greenberg, S. (2009). WikiFolders: augmenting the display of folders to better convey the meaning of files. In *Proceedings of the 27th international conference on Human factors in computing systems* (pp. 1679-1682). Boston, MA, USA: ACM.

[8] Dourish, P., Edwards, W., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., et al. (2000). Extending document management systems with user-specific active properties. *ACM Transactions on Information Systems*, *18*(2), 140-170.

[8] http://en.wikipedia.org/wiki/Windows_Presentation_Foundation