

A Special Topics Course on Personal Information Management

Manuel A. Pérez-Quiñones, Manas Tungare, Pardha S. Pyla, Margaret Kurdziolek
Department of Computer Science
Virginia Tech
{perez, manas, ppyla, mkurdziolek}@vt.edu

ABSTRACT

Personal Information Management (PIM) is an important emerging area of study in Computer Science and Information Systems. During the Spring of 2006, we offered a special topics course in PIM at Virginia Tech. This paper presents some motivation of why studying PIM is important, the goals for the course, some sample material from the course, and a few student evaluations. The paper presents in detail an activity called “Day in the Life of My Information” that resulted in an interesting experience from both, educational and research points of view.

Categories and Subject Descriptors

H.4.m [Information Systems Applications]: Miscellaneous;
H.5.0 [Information Interfaces and Presentation (e.g., HCI)]: General;
K.8 [Personal Computing]

General Terms

Human Factors

Keywords

Personal Information Management

1. INTRODUCTION

Personal Information Management (PIM) is an emerging area of research in Computer Science and Information Systems that studies how people organize their information. It includes work from areas such as human-computer interaction, information systems, databases, information retrieval, and artificial intelligence, among others. PIM, as a field of study, is concerned with “the activities we, as individuals, perform to order our daily lives through the acquisition, organization, maintenance, retrieval, and sharing of information” [23]. What sets PIM apart from all of these computing sub-disciplines is the balance between academic research and the study of actual personal information management practices.

In this paper, we report on the experience of teaching an advanced special topics course in Personal Information Management.

The course attempted to strike a balance between discussion of the literature in the field and the coverage of personal experiences in information management. The importance of teaching such a course goes beyond the typical technical preparation of our students. As we learned in our course, all participants (students and the professor alike) learned and shared information management practices that made all of us more cognizant of how we organize and manage our information. It also highlighted how information overload is more than some unique personal problem each of us faces; it was almost therapeutical to share experiences and solutions to our information overload blues.

2. WHAT TO TEACH: BACKGROUND

PIM is an interesting area of study because it intersects with many sub-disciplines of Computer Science, Information Systems, and Library Sciences. Furthermore, it represents a challenge for researchers because many of the topics of study cannot be studied in isolation in laboratories. PIM research truly pushes us to study the users in the context of their work where activities, tools, environmental context, and personal preferences intersect and interact.

PIM is surprisingly difficult to study because the behaviors exemplified are spontaneous, opportunistic, demand driven, and highly individualized. As stated by Diane Kelly: ‘People create and access their personal information collections over long periods of time, executing a variety of information management tasks and exhibiting a range of behaviors that are often unique to their collections, tools, environments, preferences, and contexts’ [18].

Many attribute the initial vision of personal information management to Vannevar Bush [6]. In his 1945 classic paper, “As We May Think”, Bush described, among other things, the problems of managing personal information. The ideas of encountering information, cataloging it, and organizing it for future use are very much at the forefront of what his Memex machine was to be. Today, due to information explosion, we find ourselves overwhelmed by information at times, wishing we had Bush’s Memex available.

Research in PIM often studies how people find information, how they keep the information found [20, 17, 16], and how they use the information kept. An area of PIM also studies how people try to get back to information they have previously seen. This is known as refinding. Studies have focused on factors that affect refinding information [8], strategies used to refind information [22], and how web search engines are used to find and refind information [7]. There is a lot of research currently exploring how to make search engines more personal [9, 14, 12, 10, 13]. The goal is to try to make searching on the desktop as effective as searching the web.

One of the chief research problems that PIM attempts to solve, or at least mitigate, is the problem of Information Fragmentation. *Information fragmentation* is the condition of having a user’s data tied

to different formats, distributed across multiple locations, manipulated by different applications, and residing in a generally disconnected manner [24]. In current personal information management systems, information formats determine storage locations, means of access, addressing of individual pieces of information, and facilities to store or search collections.

The problem of information fragmentation has been widely studied in the literature. Bellotti *et al.* [3] explored the design of a PIM system iteratively by going back to the users for feedback. Boardman *et al.* [5] studied the organizational hierarchies created by users for bookmarks, files and emails, and noted a significant amount of overlap between the file and email folder hierarchies.

For example, Bergman *et al.* [4] describe the case of information fragmentation for a student who has her project-related data in three formats under three different hierarchies: documents, emails, and bookmarks. Since users associate information objects with their projects and tasks, rather than document formats [4], this represents a potential disconnect between information systems and users' needs.

Jones *et al.* [17] found that although some users used web-browser features such as bookmarks or history lists to preserve website addresses for later access, a significant number sent email to themselves with the URL and a brief note about their personal interest in that webpage. In addition to sending bookmarks over email, Whittaker *et al.* [26] observed that users often used email systems for purposes such as personal task management, task requests from collaborators, personal archiving, and asynchronous communication.

3. A COURSE ON PIM

This course was a graduate course focusing exclusively on personal information management. The course was organized as a readings course with weekly reading assignments from research papers. Upon completion of the course, the students were expected to be able to:

- identify particular characteristics that make PIM difficult;
- design and implement solutions to PIM problems that are based on understanding of individual work habits and knowledge of the relevant literature;
- read and understand the research literature in the field.

The course had a software-prototype-related research paper due at the end of the semester. This project had several deliverables throughout the semester to encourage the students to make steady progress towards the final outcome. For several of the deliverables, we did peer-review activities on the students' work. These helped the students know what others were doing, provided the students an opportunity to receive feedback, and provided them experience on the publication process.

The second major part of the student deliverables was the "Day in the Life of My Information". The next section describes this assignment and shows some samples of students' presentations on that assignment.

4. A DAY IN THE LIFE OF MY INFORMATION

The goal of the "Day in the Life of My Information" project was to have students present their personal experiences in managing some of their information. In spite of great tools that address the management of information and methods recommended by experts,

PIM is still highly individualized. Lots of personal experiences, contextual cues, and other characteristics shape how we manage our information. Students were asked to prepare a 15-minute presentation on how they managed their own personal information. They were recommended to present one of their 'types' of information, for example, to-do lists, email, calendar, file management, music management, etc.

They were encouraged to focus not just on how they managed their information, but also on how they came to manage it in that particular way, over the years. They were instructed to indicate what works well and what does not, in their current setup. The following questions were used as guidelines on what to present and discuss.

- Describe how you manage your information today.
- Which other approaches have you tried? Why did you move away from them? Would you try them again if things changed? What would have to change for you to switch?
- What do you particularly like about your current approach? What do you not like?
- Based on what you know about HCI and cognitive psychology, and what we have read and discussed in this class, which part of your approach is in agreement with existing theories? Which part of your approach works well, in spite of what the theory or common practice tells you?
- What tools have you used? Commercial? Freeware? What tool do you currently use? Have you written a program to help you manage your information? Do you still use that program?

The results were fascinating in several ways. First, each presentation confirmed what we already know, namely that we are struggling to keep our heads above water in the ways we manage our personal information. Some practices were surprisingly simple, others extremely sophisticated. But on all accounts, the students reported not being completely satisfied with their solution. Most reported having problems at times with forgotten meetings, deadlines, or struggling to find information they knew they had with them.

Second, the sessions had a bit of therapeutical value. It was amazing to see how many times students listening to another student would say "*oh, that happened to me*". Often, it was in the other direction, students providing alternative (e.g., "*have you tried this tool?*"). Given that PIM can be highly individualized and deeply personal, there were a lot of common aspects to the problems, strategies, and solutions that students employed.

Third, it was interesting to the students (and no surprise to the professor) to find all different types of strategies and behaviors described in the literature as examples in the students' PIM practices. The next four sections provide a glimpse of some of these strategies and behaviors observed.

4.1 Filer vs. Pilers

Malone [19] observed that some people organize information into piles and use the layout of the piles to identify where information is stored. Others, according to Malone, use files, where the actually organization of the file system helps people find their information. A similar effect has been observed as users piling documents on the desktop of their computer [2]. They often do so as a way to remind themselves about needs to process these files in some form. In studying users' email-related behaviors, Whittaker



Figure 1: Paper files

and Sidner [26] observed that users leave messages in their inbox as reminders for future processing. In all, ‘piling’ served a valuable reminding function.

In the class, we saw examples of both. One student did not organize his notes; he simply kept piles of papers and post-its all over his apartment. The location of the notes (e.g., desk vs. kitchen) served as a reminding function.

Another student used a filing scheme, which he claims, he ‘perfected’ after many changes over the years. He explained that he wanted a scheme that would let him file all the papers he read in a way that would let him grab all the papers on a particular topic as well as any individual paper by a unique ID. This student used hanging folder bins as shown in Figure 1 where each hanging folder was labeled with a particular topic. Examples of such topics included affordances, SE-Agile, Standish group, standards, etc. He used a numbering rubber-stamp to imprint a number on a sticky note and attached that note to each paper he read. These numbers were unique and were incremented with each paper.

The student then entered the citation of the paper into the bibliographic software, EndNote, and used one custom field in that software to store the unique number of the paper and another to store the name of the topic. So in essence, he created a filing system that allowed him to identify a particular paper by searching for it in EndNote, getting the paper’s unique number in one of the custom fields, finding the topic it was filed under in the second custom field, and looking for that bin in his file folders.

He also went on to describe that sometimes when he was working on a paper or presentation and needed to refer to all the papers he read on a particular topic, he just had to grab that bin from his file folders and quickly glance at all the papers in that topic. When questioned on how he managed when he was away from his filing cabinets in his office, he described his ‘backup’ system: an electronic version of his filing scheme that was modified to leverage the limitations and affordances of the virtual world. He used a naming scheme shown in Figure 2 to file the electronic versions of the papers he read. Each file was labeled using the <Unique ID>-<topic>-<author> template.

4.2 Filing for Remembering

As seen before, some people file information to help them in refinding information at a later date. One student had an interesting



Figure 2: PDF documents

approach to filing. He used very complicated file names, based not on what was stored in the file, but instead based on how he felt he would need to refind the information. Straight from his classroom presentation:

When filing, I think:

- *In what context will I need this file?*
- *At that time, what will I be thinking of when I think of this file?*
- *Put those keywords into the filename.*

Some example of file names that he showed were: “*Visa Appointment Confirmation US Embassy Consulate December 2003.pdf*”, “*Kingston 1 GB Compact Flash Card Rebate from Amazon.com.pdf*”, and “*Insurance Claim Form Dentist Fortis Dental DHA.pdf*”.

This is quite an interesting filing approach, influenced by the idea of prospective remembering [21]. The emergence of keyword-based search tools such as Google Desktop [15] and Apple Spotlight [1] has enabled faster ways to get at information than the traditional hierarchical navigation needed to traverse disk directory structures. By incorporating likely search keywords into the filename itself, this then becomes a method to ‘tag’ files with metadata at the time of creation or filing.

4.3 No filing, Just Tag and Search

We saw examples of heavy users of search engines; one student uses Google Mail and does not do any filing of messages. He simply relies on tags applied by rule systems and searches his email archive to find things. He reported being quite comfortable with this approach and appears to have his email management under control. This is similar to findings in the literature and general research approaches where advanced tools are making organization less and less important. There are those that believe that better search tools will eliminate much of the need to organize your personal information [10].

4.4 Calendar

A common theme in the “Day in the Life” presentations was how life changes triggered changes in the way students managed their calendar information. Students often expressed that they relied on their parents to inform them of important dates when they lived at

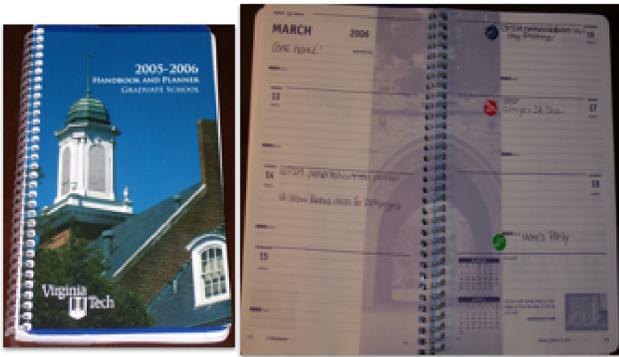


Figure 3: Personal calendar showing stickies to mark events



Figure 4: Use of a phone event scheduler and alarms as a portable To Do list manager

home, but when they moved away, they had to adopt a calendar system of their own. Some students started off with “primitive” means of keeping track of their calendars, such as marking to-do’s on their hands or scrap pieces of paper. As their to-do lists became longer and their schedules busier, the students had to find another method for managing their information. Some students discovered that the university provided free paper calendars they could use. Figure 3 shows a paper calendar used by a student, note how stickies are used to highlight events.

Other students discovered management technologies such as PDA’s and calendar software such as Apple’s iCal. Some students expressed that they now had to travel more than they used to, and opted for systems that could migrate with them; online calendars and to-do lists such as HipCal (<http://www.hipcal.com>) and calendar systems provided on their cell phones.

One student used a paper calendar but switched to using his phone’s event scheduler and alarms as a to-do list manager. He would enter in this phone events and things to do and assign times when the alarm will remind him of to do them. Figure 4 shows a sample of his daily schedule.

5. COURSE EVALUATION

The course was evaluated following the typical university evaluation procedure. This includes a typical evaluation of the course plus an open-ended survey about things that students found most valuable about the course. Below are some of the comments for the

course, each bullet line is a different student:

- “The material taken from the class and applied to my own life.”
- “How to manage personal information.”
- “Appreciation of the problems in the PIM area.”
- “Discussion of tools and PIM concepts.”
- “Opportunity to think about an important area in Computer Science. Great discussions. Think about design of cutting edge tools of the field.”
- “Lots of good PIM resources.”
- “Good content area, an emerging area that is becoming more and more important.”

Based on these comments we can conclude that the students were satisfied with the course. More importantly the comments show a variety of areas of strength in the course, from reading of the literature, discussion about tools, design of systems, to even new techniques applicable to their own personal lives.

A second way that the course was very successful is in publication of some of the class papers. The course had seven groups. Out of the 7 groups, two of the papers were submitted and accepted for participation at the SIGIR 2006 Workshop on Personal Information Management [24, 27], and a third paper was published at an international conference [25].

6. CONCLUSIONS

Personal Information Management is an important area of study. Digital information reaches us via email and instant messages at rates never before experienced. Information is available for us to consume on the web, blogs, digital TVs, downloadable content, etc. We are inundated with information. Soon, we will have personal devices that will capture digital memories [11] of all of our lives. Managing this overload of information is important if we are to be effective workers in the workplace.

But, studying this area as a research area is also important. As has been pointed out by many (e.g., [18]), there are as many PIM approaches as there are people doing PIM. We need to understand the basic principles behind PIM (e.g., organizing information, filing information, refinding information, etc.) as well as the unique practices (e.g., filing vs. piling).

The course described here was very effective at conveying to the students the importance of PIM research and of good PIM practices.

7. REFERENCES

- [1] Apple, Inc. Mac OS X Spotlight. <http://www.apple.com/macosx/features/spotlight/>, 2004.
- [2] D. Barreau and B. A. Nardi. Finding and reminding: file organization from the desktop. *SIGCHI Bull.*, 27(3):39–43, 1995.
- [3] V. Bellotti and I. Smith. Informing the design of an information management system with iterative fieldwork. In *DIS '00: Proceedings of the conference on Designing interactive systems*, pages 227–237, New York, NY, USA, 2000. ACM Press.

- [4] O. Bergman, R. Beyth-Marom, and R. Nachmias. The project fragmentation problem in personal information management. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 271–274, New York, NY, USA, 2006. ACM Press.
- [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] V. Bush. As we may think. *The Atlantic Monthly*, 1945.
- [7] R. G. Capra and M. A. Pérez-Quiñones. Using web search engines to find and refine information. *IEEE Computer*, 38(10):36–42, October 2005.
- [8] R. G. I. Capra and M. Pérez-Quiñones. Factors and evaluation of refining behaviors. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*, 2006.
- [9] E. Cutrell, S. Dumais, and R. Sarin. New directions in personal search ui. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*, 2006.
- [10] E. Cutrell, S. T. Dumais, and J. Teevan. Searching to eliminate personal information management. *Commun. ACM*, 49(1):58–64, 2006.
- [11] M. Czerwinski, D. W. Gage, J. Gemmell, C. C. Marshall, M. A. Pérez-Quiñones, M. M. Skeels, and T. Catarci. Digital memories in an era of ubiquitous computing and abundant storage. *Commun. ACM*, 49(1):44–50, 2006.
- [12] S. Dumais. PSearch: An interface for combining personal and general results. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*, 2006.
- [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 information retrieval*, pages 72–79, New York, NY, USA, 2003. ACM Press.
- [14] S. Dumais, E. Cutrell, and H. Chen. Optimizing search by showing results in context. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 277–284, New York, NY, USA, 2001. ACM Press.
- [15] Google, Inc. Google Desktop. <http://desktop.google.com/>, 2004.
- [16] W. Jones and H. Bruce. Keeping found things found: The project as a unit of analysis in pim. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*, 2006.
- [17] W. Jones, H. Bruce, and S. Dumais. Keeping found things found on the web. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 119–126, New York, NY, USA, 2001. ACM Press.
- [18] D. Kelly. Evaluating personal information management behaviors and tools. *Commun. ACM*, 49(1):84–86, 2006.
- [19] T. W. Malone. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112, 1983.
- [20] C. C. Marshall and W. Jones. Keeping encountered information. *Commun. ACM*, 49(1):66–67, 2006.
- [21] S. J. Payne. Understanding calendar use. *Human-Computer Interaction*, 8(2):83–100, 1993.
- [22] 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. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, New York, NY, USA, 2004. ACM Press.
- [23] J. Teevan, W. Jones, and B. B. Bederson. Personal information management. *Commun. ACM*, 49(1):1–1, 2006.
- [24] M. Tungare, P. S. 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.
- [25] S. Wahid, S. Bhatia, J. C. Lee, M. Pérez-Quiñones, and D. McCrickard. Cabn: Disparate information management through context-aware notifications. In *Proceedings of the 16th International Conference on Computer Theory and Applications, Alexandria, Egypt*, page 4 pps., 2006.
- [26] 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.
- [27] X. Yu, M. Alkandari, P. Liu, and M. A. Pérez-Quiñones. Visualizing a personal social network of email archives for re-finding. In *Proceedings of the 2nd Invitational Workshop on Personal Information Management at SIGIR 2006.*, 2006.