The Project Fragmentation Problem in Personal Information Management

Ofer Bergman
Tel Aviv University
Tel Aviv 69978, Israel
oferb@post.tau.ac.il

Ruth Beyth-Marom
The Open University of Israel
Raanana 43107, Israel
ruthbm@oumail.openu.ac.il

Rafi Nachmias
Tel Aviv University
Tel Aviv 69978, Israel
nachmias@post.tau.ac.il

ABSTRACT
The project fragmentation problem in personal information management occurs when someone who is working on a single project stores and retrieves information items relating to that project from separate format-related collections (documents, emails and favorite Web sites). This study was aimed to test empirically users' working habits in order to shed light on the project fragmentation problem. Twenty personal computer users participated in the study. Data collection tools included an interview, screen captures and a questionnaire. Results indicate that users tend to store and retrieve project-related information items based on different formats in one project folder when the interface design encourages it. However, they store and retrieve project-related information items in different folders (documents, emails and favorite Web sites) when the design encourages such fragmentation. Two types of attempts to solve the project fragmentation problem are reviewed and a new possible solution is suggested.

Author Keywords
Personal information management, projects, fragmentation, integration, folder hierarchies, documents, email, favorites

ACM Classification Keywords
H.5.2 User Interfaces: Evaluation/methodology, User-centered design.

INTRODUCTION
Personal computers are often used for Personal Information Management (PIM). In PIM a single person stores his\her information items (e.g., document files, emails and favorite Web sites) in order to retrieve them later on. Currently information items are stored in separate collections depending on their formats: documents are saved in a documents' folder hierarchy (e.g. in My Documents folder), emails in a separate mailbox hierarchy, and favorite Web sites in another browser hierarchy (these will be referred to as the *three hierarchies*). This hierarchy separation, also known as "information fragmentation" [9] has several negative outcomes: In addition to being time consuming, managing three different hierarchies generates cognitive load in trying to maintain some sort of consistency between the hierarchies and in using three different applications with inconsistent interaction designs [5, 10].

This article focuses on an additional negative outcome: the fragmentation of information into different collections forces a person who is working on a single project to store and to retrieve information items from different locations with no structural connection between them [3, 5, 12]. Take for example Jane, a chemistry student, who has a *Chemistry* folder in each of the three format-dependent hierarchies (documents, emails and Favorites). Her chemistry project is fragmented between these three collections and so when she works on chemistry she needs to navigate among these separate folders, which can be quite cumbersome (see Figure 1). We chose to name this problematic situation *the project fragmentation problem*.

This paper consists of two parts: the first reports a study which tests the extent of the project fragmentation problem (as a part of a larger PIM study), and the second reviews previous solutions to the problem and suggests an alternative one.

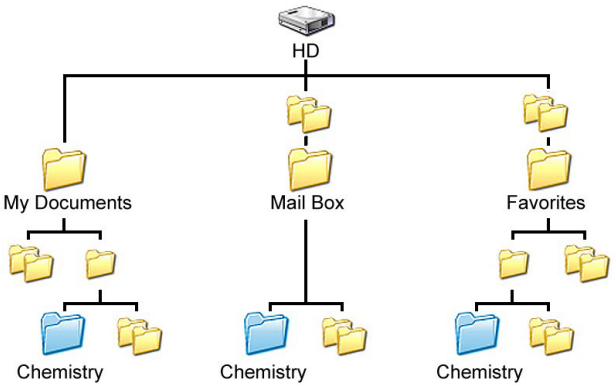

Figure 1. Example of the project fragmentation problem: information related to a chemistry course fragmented into separate collections

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
Copyright 2006 ACM 1-59593-178-3/06/0004...$5.00.

STUDY

This research was aimed to test empirically users' working habits in order to shed light on the project fragmentation problem. More specifically, we were interested in knowing whether personal computer users tend to work with their information items according to the items' formats (as suggested by current PIM systems design) or according to the items' projects (as we assumed).

Research Questions

1. How do personal computer users tend to talk about their information organization - in terms of technological format, or in terms of projects?

2. To what extent do users work on projects involving information items of different formats?

3. How much overlap is there between the three folder hierarchies*?*

4. Do users tend to classify their information according to format or projects?

5. To what extent does interface design affect the project fragmentation problem?

6. What are users' attitudes towards integration of the different hierarchies?

Method

Participants were 20 computer users (10 men and 10 women), professionals in various occupations. Their ages ranged from 30 to 53 years, and their experience with computers ranged from 4 to 20 years. All participants were sophisticated computer users with a relatively large amount of daily information to manage. Seventeen of the participants used a PC, one was used a Macintosh, and two participants used both computer types.

Participants were interviewed on an individual basis for about 90 minutes using a semi-structured technique in which each participant gave the interviewer a "guided tour" of her or his computer. Participants and interviewer viewed participant's computer while she or he explained the way in which his\her personal information was organized. The interview was taped, transcribed and analyzed (see results section for details). During the interview a number of screen captures were taken. After the interview the users completed a 1 to 5 Likert-type questionnaire. This article will report on findings from the interviews, snapshots and two questionnaire items.

Results

How do personal computer users tend to talk about their information organization - in terms of technological format, or in terms of projects? To examine this question an analysis of the interview transcripts was performed. Two independent judges were asked to determine whether each of the interviews paragraphs contained reference to formats, projects, both, or neither. The judges analyzed only spontaneous paragraphs, excluding paragraphs in which the participant answered an interviewer's question. Only those judgments on which both judges agreed (88.16% of all judgments) were included in the analysis. On average, 70.52% of the paragraphs referred to projects (*SD*=16.35), while only 28.26% referred to formats (*SD*=15.22). As expected, participants referred to projects more than to formats ($t(19)= 8.88$, $p<0.01$), confirming the notion that personal computer users are thinking about their personal information organization in terms of projects more than in terms of format.

To what extent do users work on projects involving information items of different formats? Screen captures of participants' documents, emails and Web pages which were used the day before the interview, were taken (using *Recent Documents*, and *Web History* functions). Next to each information item (regarding documents, emails and Web pages) participants wrote the project it was related to. An overlapping information item was defined as one that has another information item relating to the same project located in a different format collection. Overlap was measured by the percentage of overlapping items among all previous-day items for each participant. Results show an average overlap of 55.57% of the information items (*SD*=32.61%). This indicates that participants used different information-item formats to work on the same projects.

How much overlap is there between the three folder hierarchies? Screen captures of the three folder hierarchies at root level were analyzed. These printouts showed 968 folders: 544 document folders, 261 email folders and 163 Favorites folders. Overlapping folders were defined as folders of different hierarchies relating to the same project. For example an M.D. participant had a folder called *Diabetes* both in her My Documents and in her mailbox directories. Results showed that an average of 19.79% of the folders were overlapping (*SD*=19.38%). In other words, nearly one fifth of the folders had another folder relating to the same project in a different hierarchy. Although validity of these results might be limited because only root level folders were analyzed, they confirm the findings in [5].

Do users tend to classify their information according to format or projects? The 968 folders in the three different hierarchies mentioned in the above paragraph were classified according to their names. This was done separately for each participant. The proportion of project folder names (*M*=79.94%, *SD*=11.91%) was significantly higher ($t(19)=5.12$, $p<0.01$) than the proportion of format folder names (*M*=6.16%, *SD*=7.3%). These results indicate that users tend to classify their information according to projects more than to formats.

To what extent does interface design affect the project fragmentation problem? Our research assumption was that users' tendency to store information items of different formats in the same folder depends on interface design: When interface design encourages it, participants will ignore the format and store according to project. My

Documents is the default storage location for most documents regardless of their file format (e.g., Word, Excel) allowing project filing. However, when interface design doesn't allow such easy joint storing, participants will sort items according to format. They won't store emails and favorite Web sites in documents folders. In the questionnaire, participants were asked about their storage habits. Participants indicated that they mix documents of different formats in the same folders (M=4.4, SD=1.61), but only rarely save emails and favorite Web sites in these folders (M= 1.75, SD=1.73). As expected a paired t-test shows significant difference between results $t(19)=11$, $P<0.01$.

What are users' attitudes towards integration of the different hierarchies? The questionnaire included the following question: "To what extent do you want to save documents, emails and favorite Web sites in the same folders, so that you will have only a single folder hierarchy?". Participants' average answer was 3.74 (SD=2.24).

Discussion

While working on projects participants often use information items of different formats. They tend to store and retrieve project-related information items of different formats in one project folder when the interface design encourages it (i.e. with different document formats). However, they store and retrieve project- related information items in different folder hierarchies (documents, emails and Favorites) when the design encourages such fragmentation. When talking about their information organization participants refer to projects more than to formats. Moreover, most of their folders have project names and there is some overlap between the names of their folder hierarchies.

These results reveal the problem with project fragmentation: The users tend to relate to their information items in a certain way (i.e. according to projects), however, current design discourages them from doing so (by suggesting format-related storage). The project fragmentation problem has already been mentioned in the User Subjective approach to PIM systems design [3], but hasn't been empirically studied so far.

The next section discusses possible integration solutions to the project fragmentation problem.

INTEGRATION SOLUTIONS

There have been several attempts to confront the project fragmentation problem (for a more comprehensive literature review see [9]). In this section we will attempt to categorize these solutions it into two groups – *integration through search* and *integration through an additional structure*. Finally, we will present a new third solution: the *single hierarchy*

Integration through Search

Several search tools address the project fragmentation problem by enabling users to search for items related to the same project but in different formats in the space of a single query. Such tools are *SIS* [7], *Enfish Personal*, *PC Data Finder*, *80-20 Retriever*, and *Google Desktop*. Such a feature is included in Macintosh's new search engine *Spotlight*, and in the next release of MS Windows code named *Longhorn*. Other more radical systems such as *Lifestreams* [8], *Presto* [6], and *SwiftWare* omitted folder hierarchies altogether and rely heavily on search tools. The ability to search across multiple formats is certainly a positive feature of search tools, however their effect on the fragmentation problem might be limited, as users appear to prefer retrieval by location-based navigation to search [1, 4].

Integration through Additional Structure

Additional-structure tools allow the user to create projects in a structure distinct from the three existing format-dependent hierarchies. This structure usually contains "shortcuts" to relevant information items in these hierarchies. Several experimental systems employ this strategy, such as *Raton Laveur [2]* and *UMEA* [13] as well as commercial software such as Aladdin System's *DragStrip* and MS *OneNote*. While the additional structure solution allows users to work in an integrated project environment, it requires managing yet another structure, and may cause more cognitive load because of the two locations associated with each item.

The Single Hierarchy Solution

Taking the user-subjective approach [3] we propose a different kind of solution – the *single hierarchy solution* - in which all project-related information items are stored in the same folder regardless of their format (see also [11]). This solution is also derived from the reported empirical data: users tend to manage their personal information in a project-oriented way whenever design suggests it (documents of different formats in the same project folder), but they store items related to the same project in different locations when interface design reinforces such fragmentation.

One possible implementation of the single hierarchy solution is *ProjectFolders*. In this suggested system all the project-related items irrespective of their formats (such as documents, emails, favorite Web sites, tasks and contacts) are stored together[1] but are separated by tabs (see Figure 2). This will allow the users to work in the context of their projects and retrieve all their information items from a single location.

[1] Technically email messages will not be stored in these folders only pointers to these emails in the mailbox database, however this will be transparent to the user.

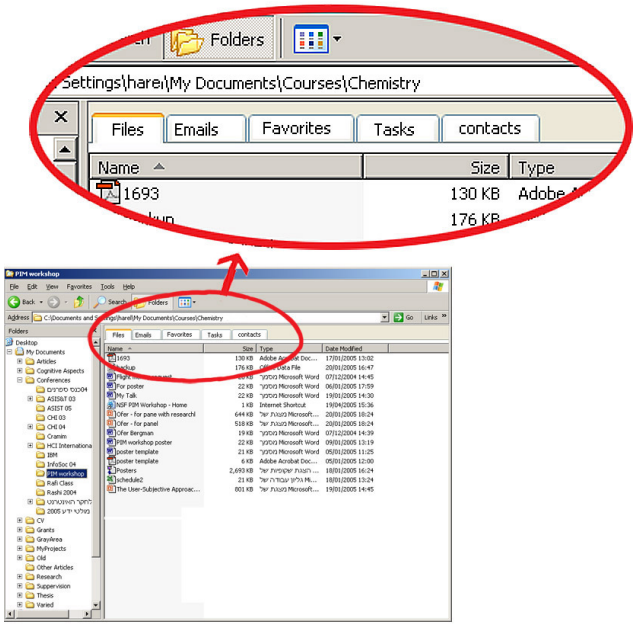

Figure 2. *ProjectFolders* stores project-related items of different formats separated by tabs

Notice also that there is no additional structure in this solution as the single hierarchy is designed to replace the standard multiple hierarchies. When operating an application only related information items will be presented. (emails for the mailbox and favorite Web sites for the browser).

Although participants showed a positive attitude towards this solution, there is no guarantee that *ProjectFolders* or any other single hierarchy solution will be accepted by users and improve usability. Research has shown that users have different strategies to manage each of the three collections [4]. This observation is also somewhat supported in our study which showed an average overlap of only 20% between folders of different format hierarchies. However, interface design often determines subjects' preferences and strategies; changing interface might change subjects' behavior and improve usability. In the early 90's when each document application suggested a separate storage location, users might have been shown to employ different storing strategies for the different document formats. This might have avoided the creation of a common storage place (e.g. My Documents). Thus, it is our intention to develop and test such a single hierarchy system.

ACKNOWLEDGEMENTS

We thank the participants for their time, R. Boardman, and the paper reviewers for their comments, and the Sacta Rashi foundation for its part in funding the research.

REFERENCES

1. Barreau, D. and Nardi, B.A. Finding and reminding: file organization from the desktop. *SIGCHI Bull 27,* 3 (2000), 39-43.
2. Bellotti, V. and Smith, I. Informing the design of an information management system with iterative fieldwork. In *Proc. DIS 2000,* ACM Press (2000), 227-237.
3. Bergman, O., Beyth-Marom, R. and Nachmias, R. The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology 54,* 9 (2003), 872-878.
4. 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 *Proc. CHI 2004,* ACM Press (2004), 583-590.
5. Boardman, R., Spence, R. and Sasse, M.A. Too many hierarchies? The daily struggle for control of the workspace. In *HCI International 2003*, Lawrence Erlbaum Associates (2003), 616-620.
6. 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 (2000), 133-161.
7. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. and Robbins, D.C. Stuff I've seen: a system for personal information retrieval and re-use. In *Proc. SIGIR 2003,* ACM Press (2003) 72-79.
8. Fertig, S., Freeman, E. and Gelernter, D. Lifestreams: an alternative to the desktop metaphor. In *Proc. CHI 1996,* ACM Press 1996, 410-41.
9. Jones, W. and Bruce, H. A Reporet on the NSF-Sponsord Workshop on Personal Information Management. NSF (2005), 29-32.
10. Jones, W., Dumais, S. and Bruce, H., Once found, what next? A study of 'keeping' behaviors in the personal use of web information. In *ASIST 2002*, American Society for Information Science & Technology (2002), 39-402.
11. Jones, W., Munat, C.F. and Bruce, H., The Universal Labeler: Plan the Project and Let Your Information Follow. In ASIST *2005,* American Society for Information Science & Technology (2005).
12. Jones, W., Phuwanartnurak, A.J., Gill, R. and Bruce, H. Don't Take My Folders Away! Organizing Personal Information to Get Things Done. In *Proc. CHI 2005*, ACM press (2005), 1505-1508.
13. Kaptelinin, V. UMEA: translating interaction histories into project contexts. In *Proc. CHI 2003,* ACM Press (2003) 353-360.