

Don't Take My Folders Away! Organizing Personal Information to Get Things Done

William Jones, Ammy Jiranida Phuwanartnurak, Rajdeep Gill, Harry Bruce

The Information School
University of Washington
Seattle, Washington 98195
jones@ischool.washington.edu

ABSTRACT

A study explores the way people organize information in support of projects (“teach a course”, “plan a wedding”, etc.). The folder structures to organize project information – especially electronic documents and other files – frequently resembled a “divide and conquer” problem decomposition with subfolders corresponding to major components (subprojects) of the project. Folders were clearly more than simply a means to one end: Organizing for later retrieval. Folders were information in their own right – representing, for example, a person’s evolving understanding of a project and its components. Unfortunately, folders are often “overloaded” with information. For example, folders sometimes included leading characters to force an ordering (“aa”, “zz”). And folder hierarchies frequently reflected a tension between organizing information for current use vs. repeated re-use.

Author Keywords

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

ACM Classification Keywords

H.5.2. [Information Interfaces and Representation (HCI)]: User Interfaces—Evaluation/methodology; —User-centered design

INTRODUCTION

The use of folder hierarchies is problematic for a number of reasons. 1. Folders can obscure as well as organize. Information filed away is out of sight, out of mind and easily forgotten. Malone [10] noted the truth of this for paper documents over 20 years ago; it is true today for electronic information as well. 2. Today there are simply “too many” folder hierarchies [3, 4] supported by separate

applications for separate types information including electronic documents and other files, email messages and web references. 3. The hierarchy, as a representation, has basic limitations. In a strict hierarchy, an information item (email message, document, web reference, etc.) can go in only one place. Even if this restriction is relaxed, a hierarchy is poorly suited to represent certain collections of information. A collection of food recipes is an often cited example. Recipes have properties such as preparation time, ingredients, number of calories, ethnic background, etc. that have no inherent ordering and are better represented in a “placeless” fashion where property values can vary independently of one another [7] or, equivalently, in a faceted classification scheme [2]

Increasingly, desktop search utilities are available that support a fast, integrated search across electronic documents, email, recently viewed web pages and other types of personal information (e.g., [8, 12]). These utilities will only get better. In particular, utilities can be expected to support the customizable use of various forms of tagging (perhaps both in factorial combination and organized into hierarchies). It is tempting to conclude that the days of folders, as used to organize personal information for re-access, are numbered.

This may be so. But before we discard folders as an outdated relic it is important to understand better the purpose they serve. To be sure, folders organize information so that it can be found again later on (and sometimes they hide this information instead). What else?

This paper describes a study in which we looked at the role of folders in the organization of project-related information. Being alive and active means having projects – both professional and personal (“Get a job”, “Plan for my child’s college education”, “Buy a house”). The lifetime of a project can vary from several days to several years. How do people organize project-related information and how does this organization change over time?

STUDY METHOD

Fourteen participants (six women, eight men) were interviewed and snapshots were taken, using a digital camera, of various collections of information pertaining to a

selected project. Ten participants were employed by the University of Washington (four professors, two librarians, two support staff and two graduate students). In addition an electrical engineer, two software engineers, and one high-level manager, all male, not affiliated with the University of Washington, participated in the study.

Interviews were from one hour to 90 minutes in length. All but one interview took place at the person's place of work; the remaining interview took place in the person's home office.

The interview began with the completion of a background questionnaire. As part of this questionnaire, the notion of a "project" was introduced, informally defined and examples were given. Participants were asked to list projects in their own lives that they would be comfortable discussing with the interviewer and the to select one for the interview. Across participants an attempt was made to equally represent work and non-work related projects.

For the selected project, the participant was asked to give the interviewer a "guided tour" showing, in turn, how project-related paper documents, electronic documents, email messages, web references and any other project-relevant information types were organized.

The interview concluded with two questions. First, participants were asked why they created folders and what purpose created folders served. Then the interviewer described an ideal search utility and gave the participant what we terms the "Google option". "Suppose that you could find your personal information using a simple search rather than your current folders...Can we take away your folders? Why or why not?"

A note about methodology

Results described here are "late-breaking" as befits the short-paper format. Most of the study's data remains to be analyzed. The study itself is exploratory, free-form, situated (typically in the person's workplace) and follows an ethnographic style of inquiry. Analysis of study results continues and efforts are ongoing to collect survey data to help establish the prevalence of selected results.

RESULTS

Consistent with other research[3], file folder hierarchies were far more elaborate than were hierarchies for other types of information. Results presented here pertain to these file folder hierarchies. The main results are as follows: 1. File folder hierarchies are more than a means to an end -- the re-access of information items. Folder hierarchies are information in their own right. Folders, if only crudely, summarize as well as organize -- they represent an emerging understanding of the associated information items and their various relationships to one another. 2. The folders associated with a project frequently reflect a basic problem decomposition or, alternatively, a plan for project completion. 3. However, additional information is often

"squeezed" into folder hierarchies -- information that is not well-represented in a single hierarchy or is best represented through properties that cross folder boundaries.

"Don't Take My Folders Away!" Folders as Information

When participants were first asked why they crated folders the most common answer given was a variation of "in order to get back to my files" (and other information items). All 14 participants gave this answer initially. Participants were then given the "Google" option -- "Suppose that you could find your personal information using a simple search rather than your current folders...Can we take away your folders? Why or why not?". Participants were permitted to stipulate additional features of this hypothetical search utility and the "folder-free" situation. Issues of control and storage would be handled in some other way. The search utility itself would be fast, effortless to maintain, secure and private (no personal information is communicated to the Web), etc.

Only one of the 14 participants answered "yes" -- he would be willing to part with his current folder organization. This participant did so with the "why" stipulation that a "time of last access" would be maintained that information items could be ordered by this property. Reasons for saying "no" ("I still want my folders") fell into one of three categories:

- **Trust.** "I'm just not willing to depend on search alone" (no matter what you say).
- **Control** (over the grouping of information items). "I want to be sure all the files I need are in one place".
- **Visibility/understandability.** "Folders help me see the relationship between things". "Folders remind me what needs to be done". One participant said that the very act of creating folders helped her to understand her information better.

The first two reasons -- trust and control -- may not be valid beyond the hypothetical situation posed. In the real world, people are not forced to choose between their folders and a search utility. All of the participants said they would be happy to have search utility that helped them to find their personal information better. And it might be possible to achieve sufficient control over the grouping of information items through support for an option to tag these items.

Visibility/understandability won't go away so easily as a reason to keep folders. Folders may represent, if only crudely, a person's emerging, often hard-won, understanding of the information items contained within, their relationships to each other, their important properties. Folders may be valuable information in their own right and not just a means of organizing information.

Folders as a Problem Decomposition

In the context of a specific project, folders can be very informative indeed. Consider the depiction in Figure 1 of one participant's folder structure for planning her wedding

(re-formatted and condensed to save space and remove identifying information).



Figure 1. A folder hierarchy for planning a wedding.

Note that many of the subfolders of the “Wedding” folder represent important components of the wedding -- and the associated decisions which must be made. Decisions had to be made concerning which invitation cards to use, what wedding dress to wear, where the reception was to occur, what kind of wedding cake, etc. Many subfolders came to represent sub-projects in their own right. For example, “wedding dresses” contains information relating to activities to select and fit a wedding dress that extended over a period of several weeks.

The participant’s comments make it clear that the folder structure of Figure 1 functioned as more than simply a way of getting back to files. Looking at the folders helped her to “see what needed to be done”. The folder hierarchy functioned as a kind of project plan – even though it lacked many properties features commonly associated with a formal project plan (e.g., due dates or % completed).

Snapshots of “wedding” folder contents at earlier points in time suggest that the decomposition of Figure 1 emerged, “bottom-up” over time rather in the “top-down” fashion often associated with a problem decomposition. Figure 2 shows a snapshot of the wedding folder’s top level taken some three months earlier when wedding planning had just begun. At this earlier point in time, the wedding folder contained only six subfolders, none of which had any child folders. The wedding folder directly contained over 30 individual files on a range of topics (e.g. “weddingdress”, “Hyatt-ballroom”). Most of these files were eventually organized into subfolders; the final “wedding” folder directly contained only four individual file.

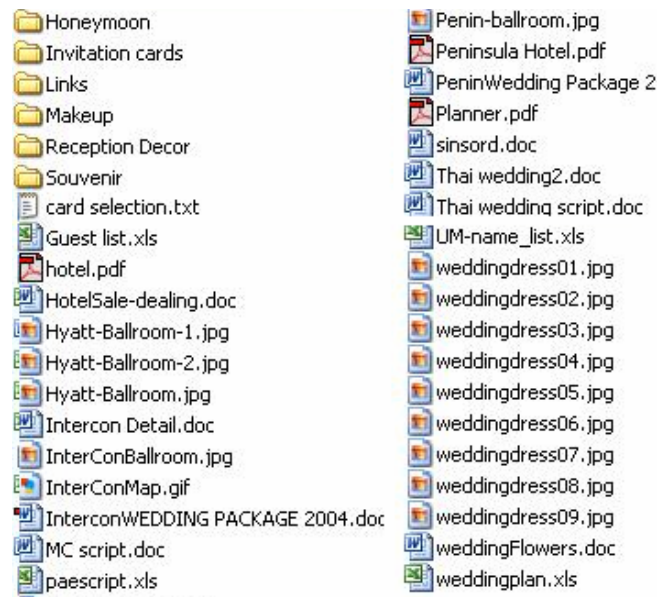


Figure 2. An earlier snapshot of the wedding folder.

All but three of the participants selected projects for which there was an associated file folder structure with the following characteristics: 1. the root folder represented the project as a whole. 2. this folder contained at least four subfolders and 3. at least $\frac{3}{4}$ of these subfolders represented subprojects of the main project (as confirmed by the participant).

Some (Additional) Problems with Folder Hierarchies

One problem with folder hierarchies has been noted elsewhere: There are too many of them [3, 4, 9]. Folder hierarchies are separately used to organize electronic documents, email messages and web references. This was true for the project information in the current study – though, consistent with previous research [3], folders were most highly elaborated for electronic documents and other files.

Three additional problems were observed for file folder hierarchies in the current study:

No support for ordering. The file folders for a project sometimes included leading characters in an effort to force an special ordering. Folders were given leading characters such as “aa” to force them to the top of a listing. Folders were numbered.

A tension between organization for current use and later re-use. For several participants folders with names like “images”, “references” and “articles” were scattered throughout their file folder hierarchy. One instance was top-level and designated as a repository for the repeated re-use of associated documents. Other instances occurred within the context of a specific project. For example, one participant had a top-level “images” folder and then multiple instances of “images” in the context of courses he was now teaching or had taught.

No support for the re-use of structure. For many participants, the same structure was essentially repeated again and again across projects. Examples included taking a course, teaching a course, planning a conference trip, and the testing and release of a software product. One participant reported making a complete copy of the folder for one project. He re-named the copy to represent the new project and then carefully deleted most of the files contained within. He said that the folder structure for the old project was a useful guide for the new project and he wanted to avoid re-creating this structure from scratch. Another participant created a file-free “xxx-xx-Course Name” template folder structure which he then copied and instantiated for each new course he took.

DISCUSSION

Previous research suggests that search alone, no matter how good it becomes as a way of finding personal information again, is likely to remain a second choice to more incremental, stepwise methods of re-access [11, 13]. Initial results of the current study make an additional statement that the re-access to personal information is not necessarily the sole or even the primary purpose of a folder organization.

In recognizing that the folder structure for a project is frequently a problem decomposition, we open up a potentially rich set of connections to psychological research on real-world planning and problem-solving (see, for example, [1, 5, 6]).

The recognition that a folder structure (as a problem decomposition) is sometimes used as a basic project plan also raises the intriguing possibility that personal information management and the management of personal projects are “two sides of the same coin”. Can a single representation serve both kinds of activities?

If, more generally, folders help people to understand and “see” their information better, it is reasonable to ask “can we do better?” What about better representations that make it easier for people to order folders (as they would like to order the components of a project)? Why not better support for the use and re-use of folder structure as a first-class object? (Perhaps we can start by supporting a “Paste Structure” option.). These and other questions naturally arise from the study of how people organize information to get things done.

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation (#0097855).

REFERENCES

1. Barsalou, L. Deriving categories to achieve goals. *The Psychology of Learning and Motivation*, 27. 1-64.

2. Bates, M.J. After the Dot-Bomb: Getting Web Information Retrieval Right This Time. *First Monday*, 7 (7).
3. 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 *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*, (2004).
4. Boardman, R., Spence, R. and Sasse, M.A., Too many hierarchies? : the daily struggle for control of the workspace. in *HCI International 2003 : 10th International Conference on Human-Computer Interaction*, (Crete, Greece, 2003), p. 616-620.
5. Catrambone, R. The Subgoal Learning Model: Creating Better Examples So That Students Can Solve Novel Problems. *Journal of Experimental Psychology: General*, 127 (4). 355-376.
6. Chen, Z., Mo, L., Honomichl, R. Having the Memory of an Elephant: Long-Term Retrieval and the Use of Analogues in Problem Solving. *Journal of Experimental Psychology: General*, 133 (3). 415-433.
7. Dourish, P., Edwards, W.K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D.B. and Thornton, J. Extending document management systems with user-specific active properties. *ACM Transactions on Information Systems*, 18 (2). 140-170.
8. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. and Robbins, D., Stuff I've seen: a system for personal information retrieval and re-use. in *26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003)*, (2003), 72 - 79.
9. Jones, W., Dumais, S. and Bruce, H., Once found, what then? : a study of "keeping" behaviors in the personal use of web information. in *65th Annual Meeting of the American Society for Information Science and Technology (ASIST 2002)*, (Philadelphia, PA, 2002), American Society for Information Science & Technology, 391-402.
10. 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.
11. Nardi, B. and Barreau, D.K. "Finding and reminding" revisited : appropriate metaphors for file organization at the desktop. *SIGCHI Bulletin*, 29 (1).
12. Pogue, D. Google Takes on Your Desktop *New York Times*, 2004.
13. Teevan, J., Alvarado, C., Ackerman, M.S. and Karger, D.R., The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. in *the ACM Conference on Human Factors in Computing Systems (CHI '04)*, (Vienna, Austria, 2004).