Surviving the Information Explosion

Mark Ackerman, Christine Alvarado, David Karger & Jaime Teevan

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: A typical person deals with a great deal of electronic information, and it can be overwhelming to keep track of it all. The problems that people have in interacting with large amounts of information are well documented. In 1988 Lansdale outlined some of "the problems of classification and the role of memory" in information organization—in other words, the problems of storing and retrieving information[2]. Lansdale states that the main problem with information organization is that humans have too limited a memory to remember exactly where they file every piece of information, either in the physical world or on their computers. Instead, people remember certain pieces of their organization, such as visual cues or dates, and rely on these pieces to help guide them to the information they need. Unfortunately, existing information organization tools are not tailored to the aspects of electronic information that people recall easily when locating information.

Motivation: Haystack is a personal information management tool intended to provide users with centralized access to their personal information. With Haystack we seek to allow users to organize and access their information in a natural way. To allow for natural interactions, we must understand the cues that people use to store and retrieve their information. For example, when looking for a previously received email, a person might recall the email sender, but she may not remember specific words in the body of the email that could be used to generate keywords for a search. Thus keyword search for an email may not be natural, but accessing emails via an address book lookup might.

One way to understand what is important in information organization is to observe people dealing with large amounts of information and ask them to describe how they store and retrieve it (or how they would like to store and retrieve it). We are in the process of doing this by conducting a field study to uncover the features of electronic information that people use or would like to use to organize and retrieve that information.

Approach: In our field study we examined the organization and search[6] patterns of fifteen graduate students in Computer Science at MIT. Specifically, our study was designed to uncover the rationale behind how our subjects organize and search through their email, their personal files, and information on the Web.

With each subject, our study consisted of two parts, the first a series of short interviews and the second a longer interview. The goal of the short interview sessions is to get the subject to articulate the rationale behind any information accesses or searches he or she has performed in the recent past. These short interviews are conducted in the subject's office, twice per day over five consecutive days. We asked the subject to describe the following in as much detail as possible:

- His or her last email access.
- The last time he or she had to search for an email.
- His or her last access to a file.
- The last time he or she had to search for a file.
- His or her last Web access.
- The last time he or she had to search for information on the Web.

The goal of the long interview was to understand the rationale behind the subject's organizational structure, including how it had evolved to the current state. While the short interviews focus on how the subjects use their organizational structure, as well as what they remember about information they have worked with, the long interviews focus on what information was actually used to create the organizational structure. During each 45 minute

[6] Whenever we say search we simply mean "had to look for," and not necessarily "search for using traditional keyword search tools."

long interview sessions we asked the subjects to give us a guided tour of their email folders (including to discuss the messages in their inbox), their personal file folders and their Web infrastructure (such as bookmarks, homepage and hot links). We asked them to pay specific attention to how and why organizational structures had evolved over time and to describe the exact reasoning that when into the most recently created or deleted folder.

Previous Work: The design of our study has been strongly motivated by previous work in the field. A seminal field study in information organization and retreival was conducted by Malone[3] in 1983. Malone examined the way people organize the physical information in their offices. Another important recent study that strongly motivated our method was performed by Whittaker and Hirscherg [4]. They investigated people's attitude's toward paper information. Other researchers have attempted to address the issue of electronic information organization [1], but these studies have not yet uncovered specific aspects of electronic information people rely on for organization and retrieval.

Status: Currently, we have completed our interviews, made transcripts of the data and are now in the analysis phase. We have found several interesting results from our preliminary analysis of the data and have identified several exciting areas to analyze in more detail. We are examining the data using statistical analysis of quantifiable events, as well as anecdotal evidence, where we focus on how specific people access their information and what causes them to succeed or fail.

Although not yet analyzed for statistical significance, our first statistical results based on a subset of the data[7] illustrate important trends in user access patterns. Subjects rely heavily on bookmarks to access Web based information (57% of accesses), and type in the URLs slightly less often (20% of the time). Furthermore, 19% of the time subjects go to a page through a bookmark or by typing the URL, their final destination page is actually one or more clicks away from the original page. Three out of thirteen Web searches are for information people have seen before, and when searching for emails, people find them in their inbox 64% of the time.

In addition to such preliminary statistics, we have also found interesting trends that warrant further investigation within our data. For example, we have noticed that people engage in several distinct types of searches, such as contact information search, confirmation of existing information, and learning about a topic (browsing). Many current search tools do not support these different types of search interactions.

We are in the process of examining the data with several goals in mind. Our data will lend insight into when it is beneficial to incorporate personalization (automatic or maunal) into an interface, and when a standard technique will suffice across all users. For example, we find that some people rarely have to look for information in their email, while others perform email searches quite often. Thus, it is likely that a personlized email access tool would be beneficial. Also, we are looking to categorize the different searches that people actually perform for use in designing future laboratory studies of Haystack.

Future Work: We have yet to analyze the long interviews. Through studying these interviews, we will learn what information people use to construct their information organizational structure. Furthermore, when we gain a better understanding of the relationship between organization and search.

Finally, we will examine ways to tie our study results back to the construction of the Haystack system. A prototype interface has been developed by others in our group. The results from the study will guide the organization and search functionality available to the user in subsequent implementations of this interface. For example, rather than just supporting keyword search, the interface will align more closely with the different types of search discussed above.

Research Support: This research is supported by NTT, the Packard Foundation, Project Oxygen, the Authur P. Sloan Foundation and the National Science Foundation.

References:

[1] Deborrah Barreau and Bonnie A. Nardi. Finding and reminding: File organization from the desktop. *The SIGCHI Bulletin*, 27(3):39–43, July 1995.

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

[7] These results were compiled after 85 short interviews (out of a total of 150) had been transcribed.

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

[4] Steve Whittaker and Julia Hirschberg. The character, value, and management of personal paper archives. *Transactions on Computer-Human Interaction*, 2001.