

Keeping Found Things Found on the Web

William Jones

The Information School
University of Washington
Seattle, Washington 98195
+1 (206) 543-1794

jones@ischool.washington.edu

Harry Bruce

The Information School
University of Washington
Seattle, Washington 98195
+1 (206) 616-0985

harryb@u.washington.edu

Susan Dumais

Microsoft Research
Microsoft Corporation
Redmond, Washington 98052
+1 (425) 936-8049

sdumais@microsoft.com

ABSTRACT

This paper describes the results of an observational study into the methods people use to manage web information for re-use. People observed in our study used a diversity of methods and associated tools. For example, several participants emailed web addresses (URLs) along with comments to themselves and to others. Other methods observed included printing out web pages, saving web pages to the hard drive, pasting the address for a web page into a document and pasting the address into a personal web site. Ironically, two web browser tools that have been explicitly developed to help users track web information – the bookmarking tool and the history list – were not widely used by participants in this study. A functional analysis helps to explain the observed diversity of methods. Methods vary widely in the functions they provide. For example, a web address pasted into a self-addressed email can provide an important reminding function together with a context of relevance: The email arrives in an inbox which is checked at regular intervals and the email can include a few lines of text that explain the URL's relevance and the actions to be taken. On the other hand, for most users in the study, the bookmarking tool ("Favorites" or "Bookmarks" depending on the browser) provided neither a reminding function nor a context of relevance. The functional analysis can help to assess the likely success of various tools, current and proposed.

Keywords

Information retrieval, personal information management, human-computer interaction, World Wide Web use.

1. INTRODUCTION

The classic problem of information retrieval, simply put, is to help people find the relatively small number of things they are looking for (books, articles, web pages, CDs, etc.) from a very large set of possibilities. This classic problem has been studied in many variations and has been addressed through a rich diversity of information retrieval tools and techniques.

A follow-on problem also exists which has received relatively less study: Once found, how are things organized for re-access and re-use later on? What can be done to avoid the need to repeat the process by which the information was found in the first place? (If, indeed, it is possible to repeat this process.) We refer to this as the problem of Keeping Found Things Found or KFTF.

Our current research project focuses on the KFTF problem in the context of World Wide Web use. Follow-on projects will look at variations of the KFTF problem as these occur for email, electronic files and paper files.

We selected the World Wide Web for our initial study of the KFTF problem for several reasons. First, the Web is increasingly "everything" informational. It is not hard to imagine a time in the near future when virtually all information we need to manage our lives is available in some form on the Web. Second, the Web by its very nature engages us on many different dimensions and forces us to actively manage our information gathering experience. Third, much Web information is ephemeral. A hyperlink that works today may not work tomorrow or may point to very different content. We thus have no assurance that information found today can be found again later on.

Finally, because the computer mediates our Web experience, there is a tremendous opportunity to build tools that can help. A proper review of the many tools, current and proposed, that might help users to manage their experience of the Web is beyond the scope of this paper. The focus of the paper is rather on an effort to understand better an underlying problem – keeping found things found, or managing information for re-use – that many of these tools are meant to address.

2. RELATED WORK

2.1 Organizing Personal Files

Relatively few studies have focused directly on the strategies people use as they seek to manage information for re-use. Malone [11] conducted a study of the way people organize the papers in their offices. He observed that people tend to organize the papers in their offices into "piles" and files. Piles are located spatially around the office and serve as a kind of short-term memory. In particular, piles often provide a *reminding* function – a paper or document in view may remind the person of an action still to be performed. However, people have increasing difficulty keeping track of the contents of different piles as their number increases. One strategy for remedying this situation, observed in the Malone study, was to transfer papers to a much larger, longer-term storage where papers are placed into named folders or files.

Consistent with Malone's observation, Jones and Dumais [10] observed a rapid falloff in the ability of people to retrieve information by location cues alone as the number of information items (AP News articles in their study) increased beyond ten. Memory for names was much less subject to disruption as the total number of items increased even though, in their study, names were restricted to two characters in length.

Carroll [4] observed great creativity in the naming of computer-based files back in the days when file names were severely restricted in length (eight or fewer characters for the users in Carroll's study). When prompted with a file name, users were able to provide a fairly detailed, accurate description of contents on over 90% of the trials.

More recently, Barreau and Nardi [2] observed some important similarities in file access among users of four different operating systems. Two of these systems, DOS and Windows 3.0, restricted file names to eight characters; the other two systems, OS/2 and Macintosh, had no such restriction. Regardless of operating system, users exhibited a strong preference for what Barreau and Nardi termed "location-based" finding. The user first guessed which folder a desired file was in. The user then generated a listing of files in this folder and attempted to recognize the desired file in this listing. The user sometimes sorted by name, date, file type or some other characteristic before making a choice. Users preferred not to enter the file name directly nor did they typically search on the file name. Users chose a full-text search only as a last resort.

2.2 Organizing Email

Many patterns in the use of a filing system are also found in email use. Whittaker and Sidner [18] observed that email systems are increasingly used for *task management* and *personal archiving* in ways that parallel the use of an electronic filing system. They note that filing decisions – which folders to create, what to name them, how to organize them, etc. – are fundamentally difficult regardless of the item being filed. Filing takes time and the folders that are created today may prove to be ineffective or even an impediment to the access of information in the future. Folders created by the users in the Whittaker and Sidner study were sometimes much too large (containing hundreds of items) or too small (containing only one or two items) to have organizing value. Subjects also reported difficulties determining a folder's contents and purpose from its name after some time had elapsed. In addition, subjects reported an "out of sight, out of mind" problem that items placed in a folder were sometimes forgotten until well after the period of their usefulness had passed.

2.3 Organizing Web Information

Like personal files and email messages, Bookmarks of the Netscape Navigator or the Favorites of the Microsoft Internet Explorer can be named and organized into folders. Throughout the remainder of this paper, we will use the generic term *bookmark(s)* to reference the comparable "file system-like" functionality provided in both the Bookmarks and Favorites.

Bookmarks are a widely used web feature. In one survey of 6,619 web users (Pitkow and Kehoe [13]) over 80% of the respondents cited bookmarks as a strategy for locating information. In another survey of 322 web users, Abrams, Baecker and Chignell [1]

reported that the size of a user's bookmark collection grows steadily and roughly linearly over time. The average user's collection of bookmarks in their study exceeded 40 after a year and more than 200 after two years.

As the number of bookmarks increases, Abrams et al. [1] also observed steady increases in the likelihood that a user will group bookmarks into a hierarchy of folders. Problems observed with the use of these folders are consistent with those observed for the use of folders to organize email messages and files: Folders can obscure as well as organize. Maintaining a hierarchy of folders takes effort. If not maintained, the hierarchy can quickly get out of date. In efforts to locate a bookmark after creation, users reported difficulty in determining which folder a bookmark was in. Users also reported that bookmarks themselves were often not descriptive enough to aid their recognition. At the same time, very few users in their study chose to re-name bookmarks.

Users seeking to return to a web page are not limited to bookmarks. Within a session of web browsing, users may use the back button. Tauscher and Greenberg [15, 16] observed frequent use of the back button – accounting for 30% of all navigational actions in their study. By contrast, history lists appear to be infrequently used. Several studies [3, 15, 16] indicate that history lists facilitate page access in less than 1% of page opens.

The use of bookmarks, the history list and the back button are methods of web re-access and re-use that are explicitly supported by the web browser (both the Microsoft Internet Explorer and the Netscape Navigator). However, in our own informal survey web users described many other methods for keeping web information for re-use that are not directly supported by browser tools. Several users indicated, for example, that they emailed web addresses to themselves. One user even volunteered that she wrote web addresses down on notes that she then affixed to the side of her computer monitor. The study described in this paper is a more formal attempt to catalog and analyze the methods by which people keep web information for re-use.

3. THE STUDY

We conducted an observational study of methods used in a workplace setting by users to manage web information for re-use. The study addressed the following questions:

- What methods do people use in a workplace setting to keep relevant or potentially relevant information for subsequent use?
- What considerations influence the choice of methods?
- How can selection considerations help us to assess the likely success of enabling tools, current and proposed?

3.1 The Participants

Participants were drawn from three distinct user populations – managers, information professionals (including librarians) and researchers. Members of each population are "high-end users" who depend heavily upon the timely availability of information for the performance of their jobs. However, the traditional orientation to information and information gathering differs between the three groups:

- *Managers*. Traditionally, managers have a preference for oral communication and depend heavily on colleagues and

subordinates for their information (see, for example, Choo and Auster [5]). However, ready accessibility of the Web may be changing managerial patterns of information access.

- *Information professionals including librarians and corporate information specialists.* Information professionals make information available to others (including managers and researchers). How has the ready availability of information via the Web changed the job of the information professional? How has it changed their practices for the re-use of information?

- *Researchers.* Researchers have traditionally been direct consumers of information (in large quantities). Now, much of the research that once required a trip to the library can be done via the Web. The Web also enables a much earlier dissemination of information between geographically separated colleagues. How do researchers approach the re-use of information on the Web?

In total, four researchers, three information specialists and four managers participated in the current study.

3.2 The Procedure

Information professionals and researchers were observed and interviewed in their own workplace in a session lasting approximately an hour. A few days prior to a session, participants completed an email questionnaire designed to establish background information concerning education level, job responsibilities, experience with computers and the Web, etc.

One question was especially important. Participants were asked to list at least three work-related, web-intensive “free-time” tasks they might like to work on over the next week should they have a half-hour or more of unscheduled time. Participants also filled in a table of other web tasks they might expect to perform in a typical work-week – whether or not work-related.

During the subsequent observational session, one of these “free-time” tasks was selected, by agreement between the observer and the participant. The participant then spent the next 30 minutes working on this task. Participants were instructed to *think aloud* while performing the task. An “over-the-shoulder” video recording was made to capture screen contents (at very coarse resolution), the participant’s hand movements and the participant’s verbalizations.

Participants were asked to handle office interruptions (phone calls, visitors, etc.) as they normally would. Participants were also encouraged to do what they would normally do in the face of serendipitous discoveries (e.g., web pages of relevance to other aspects of their lives such as upcoming vacations, purchases, health insurance, child care, etc.). The observer did not speak except to answer questions of procedure or, as needed, to remind the participant to continue to think aloud.

In a follow-on debriefing, the observer reviewed the participant’s actions and the participant was asked to “fill in the gaps” of certain actions. Participants were asked to discuss other KFTF methods they might use in other situations of web use. Participants were also asked to discuss the pros and cons, the estimated costs (e.g., execution time) and benefits (e.g., success rate) of various methods.

We quickly learned that the procedure described here would not work well for managers. We found several interested, willing

managerial participants but scheduling an hour of their time proved difficult in the first place. More important, each of our managerial participants indicated that they seldom, if ever, spent a half hour in a predominantly web-based task. Participants indicated they might read email, return phone calls or work on a report. But a half-hour web task would be somewhat contrived.

We suspected that these difficulties in the scheduling of managerial participants might be more than simply a matter of logistics. Difficulties might, in part, reflect the different way in which managers approached the Web and information gathering in general. Consequently, for managerial participants, we abandoned the observation. Instead we conducted a brief (usually less than half hour) interview – over the phone or in person.

3.3 The Results

Methods and the underlying strategy for selection of methods differed widely among participants. A brief consideration of the results for each subject is illuminating. Except where noted, all participants worked with the Microsoft Internet Explorer, version 5.0 or greater.

DT is a graduate research assistant at the University of Washington. One of her primary jobs is to research topics provided by professors at the university. Her task for the observation was to search the Web for statistics on Internet usage. She made use of several on-line databases. DT printed articles of potential relevance for later review. Articles of relevance were then passed on to the requesting professor. DT also created several instances of the IE browser during the observation to represent separate topics of inquiry and to separate different searches, on different databases, under the same topic. DT rarely used Favorites. (DT explained this by saying that she did not have her own dedicated computer and moved from one computer to the next.) DT occasionally emailed URLs to her friends and family but rarely to others in her workplace.

MC is a part-time lecturer at the University of Washington. Her task for the observation was to locate web materials that might relate to a lecture she was preparing on the use of Microsoft PowerPoint. MC made frequent use of email. MC mailed several URLs to herself – each in a separate email along with comments. On two occasions MC also emailed URLs to colleagues along with comments regarding potential relevance. MC maintained an elaborate organization of folders and subfolders in her email application (Microsoft Outlook) and expressed confidence that she could quickly locate an old email when needed. MC uses Favorites from time to time but declared that “it is a mess” because it hadn’t been organized recently. MC explained that she is reluctant to invest much time organizing Favorites since these are lost each time her computer is upgraded. Also, she is unable to access her workplace Favorites from home. Incoming email, by contrast, could be accessed from both home and work. MC made extensive use of the Back key during the observation to return to a starting point after exploring an interesting path. MC worked within only one browser window and expressed frustration when, on occasion, clicking a hyperlink resulted in the creation of a separate browser window.

EC provides research, advice and other support relating to the development and licensing of intellectual property at the University of Washington. Her task for the observation was to identify seed management companies in the Seattle area. **EC printed out web pages of interest.** She indicated that she sometimes found it useful to have the printouts of two or more web pages for “side-by-side comparison”. On two occasions, EC also sent an email containing an URL for follow-on exploration – once to herself and a second time to an assistant. EC declared that her Favorites were a “mess” and that she rarely used this facility in her research. EC made heavy use of the Back key in order to return to a starting point after exploration of a path. EC also opened several separate browser windows to represent various lines of inquiry for the topic she was researching. When a line of inquiry was complete she would either close the window or “send it home” by clicking on the Home icon.

UC is a professor at the University of Washington. His task for the observation was to use the Web to learn more about XML. **UC made heavy use of Favorites and invested effort to keep subfolders in Favorites current.** During the observation, he added a subfolder named “Xpath” to reflect change in terminology from “XQL” to “Xpath” in the XML field. UC also used his own site (whose home page he used as “home” for IE) as a jumping off point. New URLs of interest were stored as Favorites. URLs that had “proven themselves” to have value over time were eventually added to UC’s personal web site. UC also made heavy use of the Google search service as a way of returning to relevant web sites. During the post-observation interview, EC reported that he generally did not print out web pages nor did he email URLs.

TE is a senior researcher at Microsoft. Her task for the observation was to continue research for a conference paper she was preparing. TE used the reference section of paper to contain URLs she wished to track. **TE simply pasted URLs into the document** along with comments regarding relevance, actions to be taken, questions to pursue, etc. TE made no use of the Favorites facility during the observation and reported afterwards that Favorites was used, if at all, as a kind of holding bin to reference potentially relevant web pages to be explored further as time permitted.

KH is a senior researcher at Microsoft. His task for the observation was to locate the on-line versions of articles he had in paper form. **KH saved web articles, once found, as files on his hard drive.** KH also sent two emails containing URLs to colleagues. During the debriefing, KH reported that he often sent email containing URLs to colleagues to accomplish two things in one action: 1.) Maintain a reciprocal information-sharing relationship with the email recipient. 2.) Keep the email with URL in the Sent Mail folder for later access. On some occasions, KH might even ask the email’s recipient for the web page reference and related information later on (“remember that email I sent you with information on....”). In this way, colleagues served as a kind of extended memory for KH. KH made no use of the Favorites facility during the observation and reported afterwards that Favorites was used mainly to reference web pages of personal rather than work-related interest. KH reported occasionally using more than one method to insure re-access to an especially important web page. For example, KH might save the web page

as a file, send an URL for this page to a colleague and even create a Favorite for this web page.

DO is a reference librarian at the University of Washington Law School library. Her task for the observation was to locate rules and guidelines proposed by various states for web site accessibility. Research was being done for an upcoming conference presentation. DO made heavy use of the paste method. However, unlike TE, DO placed web addresses in a separate document, not in the reference section of her conference paper. Addresses were organized by state. For each address, DO was careful to insert a comment and a title for the referenced web site.

DH is a third-level manager at Boeing. He travels frequently and is rarely in his office for an hour at a time. DH was interviewed over the phone. **DH rarely accesses the Web directly for workplace matters.** When he does use the Web, the task is nearly always limited in time and scope. For example, he may occasionally use the Web to look up contact information for someone or to confirm a flight. DH depends heavily on email – from colleagues, his subordinates and other managers to whom he is responsible in one way or another. Email may occasionally contain URLs referencing a web page. However, DH rarely visits the referenced web page. Instead, he depends upon a accompanying “executive summary”. If DH requires more information on the topic, he forwards the email on to a subordinate with instructions to investigate and report back. DH also frequently forwards on email containing web references as an FYI to colleagues, subordinates and to his manager.

RR is a third-level manager at Microsoft. RR, like DH, travels frequently and is rarely in his office for an hour at a time. RR was interviewed over the phone regarding his use of the Web. Like DH, RR depends heavily on email – from colleagues, subordinates and from other managers to whom he is responsible in one way or another – to keep informed. Also like DH, RR will often forward on an email as an FYI or with a request for more information. RR reported greater direct use of the Web than did DH. For example, RR frequently accesses information concerning the work of other groups within Microsoft via the corporate intranet. RR reported using Favorites to keep track of useful web sites. He also depended heavily on his group’s web site as a jumping off point to useful web information.

NL is a third-level manager at Microsoft. Like DH and RR, NL travels frequently and is rarely in his office for an hour at a time. NL was interviewed in person. NL’s methods for keeping web information were very similar to those reported by RR. NL will often forward on an email as an FYI or with a request for more information. NL frequently accesses information concerning the work of other groups via the corporate intranet. NL uses Favorites to keep track of useful web sites. He also uses his group’s web site as a jumping off point to useful web information. In addition, NL reported that he frequently used an intranet search facility to re-access information. NL was unique among our participants in reporting that he regularly uses the history facility of the Internet Explorer.

KS is the head of a department at the University of Washington and was interviewed over the phone. KS still makes occasional direct use of the Web (to locate background information on someone she will be meeting with, for example) but expressed a resolution to move away from direct access and to delegate web

research to her support staff. KS uses the Netscape Navigator. However, she does not use bookmarks at work because these cannot be accessed from home. KS does use bookmarks on her home computer for personal reasons but noted that she frequently forgets about a bookmark until after its usefulness has passed. KS occasionally prints out web information to be read during “dead times” (while waiting for a meeting to begin, for example). **KS sends email with work-related web addresses to herself because these can be accessed from home and work.** KS also sends web addresses to her support staff for further research. **KS also pastes work-related web addresses into a file that resides in a top-level “documents” folder on her work computer.** In general, KS tries to keep all work-related information within this folder, either at the top-level or organized into subfolders representing different projects. KS will even save important email messages into the “documents” folder in order to have “everything in the same place”.

Some obvious caveats apply to the interview and observational data of this study. The observations and interviews provide us with only small windows into the participants’ efforts to manage web information for re-use. We have nothing approaching a complete picture of a participant’s overall practice of managing for information re-use. Clearly, methods observed are heavily influenced by the selected task. For example, the task for one of the study’s participants was to complete a paper for a conference. For another participant, the task was to locate, and eventually deliver in printed form to a professor, articles on a particular topic. It should come as no surprise that pasting URLs was a preferred “keeping” method for the first task and that printing was a preferred method for the second task.

These caveats notwithstanding, two important points can be reasonably made from the data. First, we see a great diversity in methods and supporting tools to manage for re-use. Each of the following methods was directly observed in our study:

- **Send email to self**, with URL referencing web page.
- **Send email to others** that contains a web page reference (and then search the Sent Mail folder or contact recipients to re-access the web information).
- **Print out the web page.**
- **Save the web page as a file.**
- **Paste into a document** the URL for a web page.
- **Add a hyperlink into a personal web site.**
- **Search for** (find again) the desired web information.
- **Enter the web address (URL) directly.** Or type in the first part of the address and then accept one of the browser’s suggested completions.
- **Bookmark** the web page.

A second point we make more tentatively: Two methods directly supported by browser tools – the history and bookmarking facilities – were among the least frequently used by participants in our study. We observed only one participant using bookmarks. (In the interview, other participants also indicated that they used bookmarks.) Several participants expressed the feeling that their bookmarks were a “mess” in need of a clean up. We observed no

use of the history facility (though one participant indicated in the interview that he used the history facility).

4. A FUNCTIONAL ANALYSIS

It is difficult to gauge the extent to which user experience and habit influence the choice of method. However, we can gain some explanation – both for the observed diversity of methods and for the relative lack of use for bookmarking and history tools – through an analysis of function. Based on observation and the interview comments of participants, several functions appear to influence the choice of method. We first describe the key functions observed in the study and then we compare methods of re-use with respect to these functions.

- **Portability of information.** Can users take the information with them wherever they go? Paper is especially portable. Notepad and palmtop computers may eventually make electronic copies of a web page equally portable. Two participants indicated that printouts made it possible to work through the information during “dead times” (for example, while commuting on a bus or waiting for a meeting to begin). One participant also indicated that paper was somehow more “real” and provided a safer medium for preservation of information.
- **Number of access points** – related to but not the same as portability. Can users access the information from multiple places? From their laptop as well as their desktop computer? From home as well as work? The bookmarks of one machine cannot typically be accessed from another machine. Similarly, the history list of one machine cannot be accessed from another machine. On the other hand, email – especially if the server is web-based – can often be accessed from several machines. The search and personal web site methods will generally work from any machine with access to the Web. Direct entry of a web address can be done from any machine with access to the Web. However, the browser’s ability to suggest completions to a partially entered web address is machine-specific.
- **Persistence of information.** Will a web page still be there tomorrow? ... using the same URL? Will the page still have the same content? Printing a web page or saving as a file insures that the information, in some form, persists. Creating a bookmark for a web page or emailing its URL does not insure persistence.
- **Preservation of information in its current state.** A printout cannot preserve the interactivity of a web page. On black-and-white printers, the information content in a web page’s color is lost as well. Perhaps the best method available to the user for preserving information in its current form is to save the web page as a file. (However, this method does not make it possible to track updates of the web page – see the “Currency” function). The preservation value of a bookmark, or of other methods that depend upon a web address, is harder to assess. The user can hope to see the same information again if the address is still valid and if the content if the page is unaltered.
- **Currency of information.** If currency of information is important (e.g., for weather or stock prices) then it is important to keep a representation of the information that can be refreshed to reflect updates in underlying content. All methods that use a web address score high with respect to currency. On the other hand, printed or saved web pages cannot be readily updated.

- **Context.** Do users know why they saved the web information (or a reference to this information)? Do they know when it should be accessed and for what? Participants were able, through comments and subject line, to establish a context for a web address sent in an email message. The same was true for web addresses pasted into a document. On the other hand, the ability to provide a context for a bookmark is limited. At best, users can attempt to provide some context through the name and folder location of a bookmark.
- **Reminding.** Saving information does little good if we don't remember to use this information later on and in the right situation. The "reminding" value of a method depends upon a person's habits. For users who routinely check their email inbox, the reminding function of email is likely to be high. But if an inbox is routinely crammed with hundreds of new email messages each day, the reminding function of email may be compromised. Likewise, paper printouts of web information may have a good reminding function – unless the user's office is already cluttered with paper. We did not observe nor did participants indicate during their interviews, that bookmarks were routinely checked. Consequently, we give a low rating of reminding value to bookmarks.
- **Ease of integration.** Does the method help users to integrate new information or new references with ongoing projects? Existing organizational schemes? An URL sent as email can be readily saved into the same organizational structure that is used for other email messages. Similarly, a web page saved as a file can

be incorporated into the existing file structure. On the other hand, bookmarks cannot directly participate in these organizational schemes. Any organization the user imposes on bookmarks must be maintained separately from the organization imposed on files or email messages. One participant expressed frustration with the number of different organizational schemes he needed to maintain in his life – one for email, one for electronic files on his hard drive, one for his paper files and, yet another, for his bookmarks.

- **Communication and information sharing.** Some methods make it easier to share information with others. Emailing a web address (or the web page itself) is clearly one effective way to share information. For several of our participants, paper printouts of web information are also effective. On the other hand, bookmarks cannot be readily shared.
- **Ease of maintenance.** Whether maintenance tools provided for email or for personal files are better or worse than those provided for bookmarks is open to discussion. But, people typically have to maintain their personal files and email regardless. It is unlikely that sending web addresses through email or saving web pages as files adds appreciably to this maintenance burden. Other methods, such as relying on one's memory for the address of a web page or one's ability to find the web information again via a search service, are essentially maintenance free (though they may be more likely to fail than other methods).

For each of the functions listed above, we've attempted to give a simple high/medium/low rating to each of the keeping methods under discussion as summarized in Table 1.

Table 1. A functional comparison of different methods of keeping Web information for re-use

	Portability	Number of access points	Persistence	Preservation	Currency	Context	Reminding	Ease of integration	Communication	Ease of maintenance
Email to self	Low	High	Low	Med	High	High	High	Med	Low	Med
Email to others	Low	High	Low	Med	High	High	Low	Low?	High	High
Print-out	High	High	High	Low	Low	Low	High	Med	High	Med
Save as file	Med?	Low?	High	High	Low	Low	Low	Med?	Low	Med
Paste URL in document	Low	Low?	Low	Med	High	High	High?	High?	Low	High
Personal web site	Low	High	Low	Med	High	High	High?	High	Med	High?
Search	Low	High	Low	Med	High	Low	Low	?	Low	High
Direct entry	Low	High	Low	Med	High	Low	Low	?	Low	High
Bookmarks	Low	Low	Low	Med	High	Low	Low	Low	Low	Low
History	Low	Low	Low	Med	High	Low	Low	Low?	Low	?

Clearly, these ratings – especially those with a question mark – are subject to qualifications and debate. How portable, for example, is web information that is saved as file? If the file is transferred to a palmtop, then the method of saving the web page as a file affords a high degree of portability. Otherwise, the method would not score high with respect to portability. Likewise, if the web page is saved to a file that can be accessed over a network, then the user might have a number of access points to the information. But if the file is kept on a local machine with no access over a network, then the user has a single point of access to the web information. As various technologies of wireless access to the Internet grow, we can expect that the portability of some methods for re-accessing web information – search, direct entry or the use of a personal web site, for example – will increase. But this day has not yet arrived for the participants observed in our study.

The functional analysis summarized in Table 1 is illuminating nonetheless. We note, for example, that bookmarks and history score low across most functions and have only a single high rating – for currency. On the other hand, the email methods and the printout method score high with respect to a number of functions. For example, for our participants, a self-addressed email could typically be accessed from both home and work. Participants could give the email a context through the choice of a subject line and through the text of the email. And when the email arrived in the inbox, it served as a reminder.

We believe a functional analysis of keeping methods can help to guide efforts to build new or improved tools. This analysis can also help to set expectations for the likely success of tools and tool improvements. For example, several research projects have explored the use of graphics as a way to make history lists more used and usable [6, 7, 9, 12, 17]. Other research has explored the use of a virtual 3D environment for the organization of bookmarks [8, 14].

These projects indicate that a better use of graphics can improve the usefulness of bookmarking and the history facility. Graphical representations might, for example, improve the ease with which bookmarks or history items are recognized. However, the functional analysis suggests that graphics alone may not be enough to insure a tool's long-term usefulness and popularity. Users who work in several places (e.g., home and office) may continue to ignore a history or bookmarking facility that is machine-specific. Users may be reluctant to invest time in the creation of bookmarks until these bookmarks can be integrated into existing organizational schemes (such as the hierarchy of an email or filing system). And until bookmarks include a reminding function and a better ability to describe relevance and actions to be taken, many users may continue to send web references through email instead.

5. FOLLOW-ON WORK

The study reported in this paper provides a qualitative glimpse of the diversity of methods for keeping Web information as well some of the factors influencing the choice of methods. The results of the study are being used to help structure a larger investigation into the methods and supporting tools people use to manage web information for re-use. Our investigation will include in-depth longitudinal studies of individual users, usage surveys and modeling in order to address the following questions:

1. How do users organize web-based information for re-use? What tools and techniques do they use? How do users assess the costs and benefits of various tools and techniques?
2. What kinds of problems do users encounter when they attempt to re-access web-based information? How do they overcome these problems? How serious are these problems?
3. What do users remember about the web-based information that they want to re-access?
4. What tools and tool improvements would users like to see?

Data from all studies will be incorporated into models that describe individual and aggregate behavior. We expect that the modeling, especially with respect to points of failure in various methods, will help to suggest supporting tools and tool improvements. We also hope to identify individual practices of web information re-use that appear to work especially well today given the currently available tools.

6. CONCLUSIONS

The study reported in this paper reveals a great diversity in the methods people use to organize web information for re-access and re-use. Of these methods, two that are explicitly supported by tools of the web browser (in both the Microsoft Internet Explorer and the Netscape Navigator) were relatively underused. Only one participant was observed to use bookmarks. No participant was observed to use the history list.

A functional analysis helps to explain the diversity of methods observed. Based on the observations and the interview comments, several functions appear to influence the choice of method. Does a method make the web information portable? Can the web information be accessed from several places? Will the web information be around the next time the user needs it? In what form? Can a user make notes to explain relevance or actions to be taken? Does a method remind the user of a web page's relevance later on? Can the web information be integrated into existing organization schemes? Can it be shared with others? Can the method still be effectively used over time as the amount of web information and the number of web references continues to grow? The functional analysis can also help us to gauge, at least qualitatively, the likely success of various proposed tools and tool improvements.

People exhibit great flexibility and creativity in their choice of methods and in their overall practice of information re-use. We begin to glimpse this flexibility and creativity only when we move away from a study of individual tools and their use and towards a study of what the user, by whatever means, is trying to accomplish. It is our belief that a broader study of user actions to manage for information re-use, to “keep found things found”, will yield results that prove very useful in the development of new tools and tool improvements.

7. REFERENCES

- [1] Abrams, D., Baecker, R., and Chignell, M. Information archiving with bookmarks: personal web space construction and organization. In Proceedings of CHI'98 (1998), 41-48.

- [2] Barreau, D. and Nardi, B.A. Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27, 3 (1995).
- [3] Byrne, M.D., John, B.E., Wehrle, N.S., and Crow, D.C. The tangled Web we wove: a taskonomy of WWW use. In *Proceedings of the CHI '99* (1999), 544-551.
- [4] Carroll, J.M. Creative names for personal files in an interactive computing environment. *International Journal of Man-Machine Studies*, 16 (1982), 405-438.
- [5] Choo, C.W. and Auster, E. Environmental scanning: acquisition and use of information by managers. *Annual Review of Information Science and Technology*, 28 (1993), 279-314.
- [6] Cockburn, A. and Greenberg, S. Issues of page representation and organisation in a web browser's revisitation tools. *Australian Journal of Information Systems*, 7, 2 (2000), 120-127.
- [7] Cockburn, A., Greenberg, S., McKenzie, B., Jasonsmith, M., and Kaasten, S. WebView: A graphical aid for revisiting Web pages. In *Proceedings of the OZCHI'99 Australian Conference on Human Computer Interaction* (1999), 15-22.
- [8] Czerwinski, M., Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S., and van Dantzich, M. Visualizing implicit queries for information management and retrieval. In *Proceedings of CHI '99* (1999), 560-567.
- [9] Hightower, R.R., Ring, L.T., Helfman, J.I., Bederson, B.J., and Hollan, J.D. Graphical multiscale web histories: a study of PadPrints. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia* (1998), 58-65.
- [10] Jones, W.P. and Dumais, S.T. The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Transactions on Office Information Systems*, 4 (1986), 42-63.
- [11] Malone, T.W. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems*, 1 (1983), 99-112.
- [12] McEneaney, J. E. Visualizing and assessing navigation in hypertext. In *Proceedings of the tenth ACM Conference on Hypertext and hypermedia* (1999), 61-70.
- [13] Pitkow, James E. and Kehoe, Colleen M. Emerging trends in the WWW user population. *Communications of the ACM* 39, 6 (1996), 106-108.
- [14] Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D. and van Dantzich, M. Data mountain: using spatial memory for document management. In *Proceedings of UIST '98* (1998), 153-162.
- [15] Tauscher, L. and Greenberg, S. Revisitation patterns in World Wide Web navigation. In *Proceedings of CHI '97* (1997), 399-406.
- [16] Tauscher, L. and Greenberg, S. How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47 (1997), 97-137.
- [17] Wexelblat, A., and Maes, P. Footprints: history-rich tools for information foraging. In *Proceedings of CHI '99* (1999), 270-277.
- [18] Whittaker, S. and Sidner, C. Email overload: exploring personal information management of email. In *Proceedings of CHI '96* (1996), 276-283.