

Once Found, What Then?: A Study of "Keeping" Behaviors in the Personal Use of Web Information

William Jones

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

Susan Dumais

Microsoft Research, Microsoft Corporation, Redmond, Washington 98052.
Email: sdumais@microsoft.com

Harry Bruce

The Information School, University of Washington, Seattle, Washington 98195.
Email: harryb@u.washington.edu

This observational study investigates the methods people use in their workplace to organize web information for re-use. In addition to the bookmarking and history list tools provided by web browsers, people observed in our study used a variety of other 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. Differences emerged between people according to their workplace role and their relationship to the information they were gathering. Managers, for example, depended heavily on email to gather and disseminate information and did relatively little direct exploration of the Web. A functional analysis helps to explain differences in "keeping" behavior between people and to explain the overall diversity of methods observed. People differ in the functions they require according to their workplace role and the tasks they must perform; methods vary widely in the functions they provide. The functional analysis can also help to assess the likely success of various tools, current and proposed.

Introduction

Finding useful, relevant information is often difficult. This classic problem has been studied in many variations and has been addressed through a rich diversity of information retrieval tools and techniques.

Keeping information, once found, is a problem in its own right, which has received relatively less attention. How is "found" information organized for re-access and re-use later on? 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. Subsequent projects will look at KFTF problems that arise in the use of email and the use of personal files – paper-based and electronic.

The Web poses special challenges with respect to the KFTF problem. The Web is increasingly everything informational. Its hyperlinked content engages us in many different aspects of our lives – and pulls us in many different directions. Within a session of web use we are likely to encounter much more information than we are able to analyze in the time we have available. We often look for ways to insure that we can return to a promising-looking web site for further study later on. And, once we have determined that the information of a web page is useful, we want to be sure that we can retrieve this information, in some form and with reasonable speed and ease, when a need for this information arises some weeks or months later.

The computer mediates our access to the Web and the computer can also help – through tools and the keeping methods these tools support. Facilities of web browsers such as bookmarking ("Bookmarks" in the Netscape Navigator, "Favorites" in the Microsoft Internet Explorer) or history lists are built with this in mind.

In our study of web use in workplace settings (Jones, Bruce, Dumais, 2001), we have observed a great variety of "keeping methods". Many of these methods are "outside the box" of a participant's web browser. For example,

several participants sent web addresses (URLs) to themselves along with notes regarding relevance and actions to be taken. Participants also pasted URLs into documents. For some participants these documents were internally structured, through headings and subheadings, into a scheme for the organization of web addresses and web information. For other participants, documents with web addresses were part of a larger external organization of folders and subfolders. For some participants, keeping methods went beyond the computer altogether. For example, one participant printed out web information with the intent of filing this information in a paper-based filing system. One person even volunteered that she jotted down URLs on “sticky notes” which she then pasted around the edges of her computer monitor!

In the current paper we report results of an expanded workplace study into methods for keeping web information. We also take a closer look at differences that appear to emerge between people as a function of their workplace roles, their overall orientation to information in the workplace and the tasks they must perform. Finally, we consider some of the implications that the results of our study may have for the design of supporting tools.

Related work

Just as the challenge to keep found things found can arise in many different areas of our lives, relevant research comes from studies in several domains. Here we selectively review studies into the organization and use of personal files, email and the Web.

Organizing Personal Files

In a study of the way people organize the papers in their offices, Malone (1983) observed that people tend to organize papers 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.

Jones and Dumais (1986), consistent with Malone’s results, 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 (1982) 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 (1995) observed a strong preference for what they termed “location-based” finding. Their study looked at 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, the user’s preferred strategy was first to guess 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.

Organizing Email

Many patterns in the observed organization of email correspond to those of file system organizations. Whittaker and Sidner (1996) observed that email systems are increasingly used for *task management* and *personal archiving* in ways that parallel the use of an electronic filing system. Filing decisions – which folders to create, what to name them, how to organize them, etc. – may have an essential difficulty whether the item is an email message or a personal file. Whittaker and Sidner note that 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 their study were sometimes much too large (containing hundreds of items) or too small (containing only one or two items) to have organizing value. Folder names, even though originally assigned by the users, were often not descriptive of folder contents and purpose. Also, users reported an “out of sight, out of mind” problem – items placed in a folder were sometimes forgotten until well after the period of their usefulness had passed.

Organizing Web Information

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

Bookmarks are a widely used web feature. In one survey of 6,619 web users (Pitkow and Kehoe, 1996) over 80% of the respondents cited bookmarks as a strategy for locating information. Abrams, Baecker and Chignell (1998), in another survey of 322 web users, reported that the size of a user’s bookmark collection grows steadily and roughly linearly over time. In their study, the average user’s

collection of bookmarks exceeded 40 after a year and more than 200 after two years.

Abrams et al. (1996) also observed that, as the number of bookmarks increases, there is a steady increase in the likelihood that a user will group bookmarks into a hierarchy of folders. Problems with these folders parallel those observed for email and file folders: Folders can obscure as well as organize. It takes effort to maintain a hierarchy of bookmark folders. If not maintained, the hierarchy can quickly get out of date. Users reported difficulty determining the folder location of a bookmark. Users also reported that the name of a bookmark, by default the title of the referenced web page, was often not very descriptive. Nevertheless, very few users in their study chose to re-name bookmarks.

Users can use other means to return to a web page. Within a session of web browsing, users may use the “Back” button. Tauscher and Greenberg (1997a, 1997b) 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 (Tauscher & Greenberg, 1997a, 1997b; Byrne, John, Wehrle & Crow, 1999) indicate that history lists facilitate page access in less than 1% of page opens.

In our study of web use in workplace settings (Jones, Bruce, Dumais, 2001) we observed no use of history lists (though one participant indicated during the interview that he occasionally used the history facility of his web browser). However, amongst the 11 participants in this study we observed many other methods of preserving web information for later web access and re-use including: 1.) Email to self. 2.) Email to someone else (and ask them for the web information or search through the sent mail folder for the reference). 3.) Print-out the web information. 4.) Save the web page as a file. 4.) Paste the web address into a document – which may or may not participate in a larger file system organization of web addresses. 5.) Add the web address to a personal web site. 6.) Do nothing – and count on finding the web page again later via a search service. 7.) Do nothing and count on direct entry of the web address later on or, more likely, entry of the first part of the address followed by selection of one of the browser’s suggested completions. 8.) Write the web address down on paper (or possibly a “sticky note” that can be affixed to the computer monitor). 9.) Bookmarks.

We observed at least one participant use each of the keeping methods listed above in the course of the half-hour observation. During this half-hour, participants worked to complete a work-related, web-intensive task (one among a list of three they originally listed in a questionnaire). Our intent was to get an accurate and detailed glimpse of the ways in which people actually organize web information for later re-access and re-use.

The Study

The expanded study that we describe in this paper follows the same methodology as used in Jones, Bruce and Dumais (2001). In the follow-on analysis we focus on two goals. The first goal is to understand better both the diversity of keeping methods and the underlying reasons for this diversity. How do methods compare, for example, with respect to the functions they provide? Are there important variations in the use of various methods? The second goal is to understand better how differences among participants – e.g., with respect to profession, role within their workplace, overall orientation to information and demands of the current task – can influence the choice of a given method.

The Participants

Participants were drawn from three distinct user populations – researchers, information professionals (including librarians) and managers. Members of each population 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:

- *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 more timely and efficient dissemination of information between geographically separated colleagues. How do researchers approach the re-use of information found on the Web?
- *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?
- *Managers.* Traditionally, managers have a preference for oral communication and depend heavily on colleagues and subordinates for their information (see, for example, Choo & Auster, 1993). Now managers can often, if they chose to, use the computer on their desktops to access work-related information from databases, intranets, and the Web. Has this ready availability of information changed the methods managers use to acquire and keep information?

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

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.

In this questionnaire, participants were also 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. During the subsequent observational session, one of these 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 think-aloud commentary.

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. 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 observational technique described above. Instead we conducted a brief (usually less than half hour) interview – either over the telephone or in person.

The Results

Observed keeping methods and apparent underlying strategies, as discussed in the follow-on interview, differed widely among participants. We can refer to a *practice of*

information re-use – a KFTF practice – as a combination of methods (involving supporting tools and facilities such as the bookmarking facility), strategies and habits. The practice for each participant appeared to be unique and differences between one practice and another appear to be non-trivial.

We can better appreciate these differences and the overall diversity of practices by looking at summaries of our sessions for selected participants from each participant group. Except where noted, all participants worked with the Microsoft Internet Explorer, version 5.0 or greater. Observed methods of special interest are highlighted.

Researchers. The keeping methods observed for researcher participants were especially diverse:

*MC is a part-time lecturer and researcher 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. **She 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.*

*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 added URLs for web sites of proven value to his own personal web site.** UC also made heavy use of Favorites and invested effort to keep his subfolders up-to-date. During the observation, he added a subfolder named “Xpath” to reflect the change in terminology from “XQL” to “Xpath” in the XML field. New URLs to web sites of potential value were first 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.*

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.

Information specialists. Several of the information specialist participants made heavy use of bookmarking (Favorites or Bookmarks) and they took care to re-name and to organize bookmarks into a hierarchy of folders:

LK is a librarian at the University of Washington Library. Her task for the observation was to investigate the de facto standards used by the top academic library digitization programs. **LK made considerable use of Favorites, renaming them as needed to fit within her organizational scheme.** Favorites were organized into three tiers: 1.) At the top of the Favorites listing were the most frequently used Favorites. 2.) Next came a hierarchy of folders containing Favorites less frequently used but of "proven value". 3.) Last in the list were individual Favorites that LK intended to explore further when she had the time. She frequently "weeds" Favorites in this third tier – either deleting, moving to the first tier or a subfolder in the second tier. During the follow-on interview, LK indicated that she uses email to send web addresses to herself and others but did not do so during the observation. LK also said that she sometimes prints out web information of special importance to save in her paper-based filing system.

TC is a librarian with the Pacific Northwest Region of the National Network of Libraries of Medicine. Her task for the observation was to gather information regarding a dietary supplement/herbal drug. **She made heavy use of the personal toolbar** in Netscape Navigator as a way to keep links to her most commonly used sites at hand. Two times during the observation she expressed special interest in newly found web sites but did nothing overt to "keep" them for later access. During the follow-on interview she expressed confidence that she could easily return to these sites again using Google, her preferred search service. TC also indicated that she prefers to read articles on paper and so often prints out articles for this reason.

FI is a librarian in the University of Washington Health Services Library. Her task for the observation was to look for information regarding Linux's suitability as a desktop operating system. She also took time to review, re-organize and "weed" her collection of Favorites (which she indicated she did once every one or two weeks). **FI re-named Favorites during the observation, frequently giving "names" that served as reminders of actions to be taken** ("Check for relevance to project ..."). During the follow-on interview she indicated that she almost always gave her own names to Favorites (overriding the default name provided by the Internet Explorer).

DO is a 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. She was researching this topic for an upcoming conference presentation. **DO made heavy use of the copy-and-paste method. However, unlike TE, she 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.

MN is a children's librarian at a company that produces a children's literature web site. Her task for the observation was to research readability formulas. MN made heavy use of Favorites, placing newly created Favorites into an elaborate folder hierarchy. She created new folders in anticipation of and prior to beginning the task for the observation. **MN also wrote down notes to herself (including web addresses) on a note pad, which she kept next to her computer.** During the follow-on interview, she also indicated that she sometimes emails herself web addresses of interest and that she prints out web articles when she wants to read these in depth.

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 a 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.

QH is a “channel editor” at Microsoft. She maintains and regularly internal web sites for each of several topics of interest to a large internal audience. Updates for some topics are also sent out via email at weekly intervals to subscribing employees. QH’s task for the observation was to search the Web for new information relevant to one of the interest topics for which she was responsible. QH used Favorites on eight separate occasions during the observation in order to “jump” to web sites of relevance to the topic. Many Favorites pointed to such “resource” web sites and these Favorites were generally organized into folders according to interest topic. **QH also made repeated use of the “notes” facility of Outlook, pasting URLs along with comments into a note.** If her Favorites generally pointed to web resources of proven value over time, notes in Outlook often pointed to specific web pages for which some action – extract information, assess relevance, etc. – needed to be taken. QH associated a date with each note representing a date by which the action should be taken. Since Outlook listed a note on the task list for this date, QH was then reminded later on of the URL and the action(s) to be taken.

Managers. All managerial participants indicated that they made frequent use of email to send and receive web addresses and web information. Managerial participants differed from each other in the extent and nature of their direct use of the Web:

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 telephone. 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 an 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 telephone 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. 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 towards delegating 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”.

NF is the head of a large department at the University of Washington. NF prefers to delegate web research – either to his administrative assistant or to a relevant faculty member. He delegates either in person or via email (roughly 50% of the time for each). However, NF will occasionally access the Web directly to complete research if 1.) the time required is less than 30 minutes or so and 2.) the nature of the research task is difficult to describe to someone else. NF travels frequently and indicated that he has largely abandoned the work-related use of Favorites because these do not readily transfer to the different machines he uses (desktops at home and work, laptop when traveling). However, NF does opt to display references to frequently accessed web sites as “Links” which are visible on a tool bar of his web browser. (He uses this Links facility on all three machines.) Working with his assistant, NF also attempts to maintain the same folder organization for several different items: 1.) the files on his office desktop computer. 2.) his email messages. 3.) the paper documents in his filing cabinets 4.) the work-related files – paper and electronic – maintained by his assistant.

We mention one obvious caveat: The observations and interviews provide us with only small windows into the participants’ efforts to manage web information for re-use. Also, the methods we observed are influenced in part by the task selected for the observation. To take perhaps the most extreme example, TE’s task of preparing a research paper clearly influenced her decision to rely heavily on a method of pasting URLs into the “proto” reference section of the document she was preparing.

Nevertheless, the observed keeping methods and their diversity are real. For all participants, the follow-on interview reinforced the impression that these methods were a commonplace part of the participant’s workplace routine. Each of the following methods was used by at least one participant during the observation:

- **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). An instance of sending email containing a URL to someone else is not necessarily an instance of this method. For example, all information specialists indicated that they sent mail to others as part of their job (as information mediators). However, no one in this group indicated that sending mail to others was part of a strategy to keep found things found. On the other hand, we do count the email containing URLs that managers indicated they send to subordinates. Managers send this mail with the expectation of getting back information, albeit in processed, reduced form, later on.
- **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.**
- **Bookmark** the web page.
- **Write down the notes on paper** containing the URL and actions to be taken.
- **Copy to a “links” toolbar** so that the web address is always in view in the browser and can be quickly accessed.
- **Create a “note” in Outlook** that contains the URL and can be associated with a date. More generally, create an item in a task management or calendaring system that can then be associated with one or more dates.

Of course a keeping method implicitly comes with one or more companion methods for re-access later on. Keeping methods and methods for re-access combine to form a strategy for keeping found things found. For example, creation of a bookmark has a companion method of using the bookmarking facility later to locate the bookmark and, through the bookmark, the underlying web page. Sending a web reference to oneself via email has a companion method of periodically checking the inbox. Sending a web reference in email to someone else may have the companion method of searching through the “sent mail” folder if the need for the web reference arises later on. For at least one participant, this method also had the companion method of asking the email recipient for the reference later on!

With this understanding that keeping methods have companion methods of re-access, we can also list three variations of “doing nothing” as a keeping method. These variations are distinguished by the method of re-access by which participants indicated during the interview that they expected to be able to return to web information later on.

- Do nothing, and **search for** (find again) the desired web information later on.
- Do nothing and **enter the web address (URL) directly** later on. Or type in the first part of the address and then accept one of the browser’s suggested completions.
- Do nothing and **reach the web information later from a known point of access such as a web portal.**

The cells in Table 1 provide the percentage of participants within a group who use a given method. A method is counted as being used if we either observe its use or it is mentioned in the course of the interview (our only source of information for managers). We exclude variations of the “do nothing” method. Percentages should be regarded as minimums. For any participant, given more time (for a longer observation or a longer interview) it is quite possible that we would have uncovered additional methods.

Table 1. Use of keeping methods by participant group.

	Researchers	Information Specialists	Managers
# of participants	4	8	6
Email to self	75%	25%	0%
Email to others	25%	0%	100%
Print-out	50%	13%	17%
Save as file	25%	0%	0%
Paste URL in document	75%	13%	0%
Personal web site	25%	0%	0%
Bookmark (Favorite)	50%	88%	33%
Notes on paper	0%	25%	0%
Links toolbar	0%	13%	17%
Outlook "Note"	0%	13%	0%

Noteworthy Patterns in Keeping Behavior

It is worth pointing out a few additional patterns in participant keeping behavior.

- **Tested and untested references are often “kept” in different ways.** With respect to keeping methods, we note that several participants distinguished between references to new, interesting but as yet untested, unevaluated web information and references to web pages or sites of established value. Different participants used different methods or sometimes the same method in different ways to distinguish “tested” and “untested” web references. For example, UC placed untested references in his Favorites folder and tested references on his personal web site. On the other hand, QH placed promising but untested references into Outlook notes while placing references of long-term value into her Favorites collection. LK used Favorites for both. She placed frequently accessed, tested Favorites at the top of her Favorites listing, less frequently accessed Favorites within folders in the middle of her listing and, finally, untested Favorites at the bottom of this listing.

- **Some participants went to great lengths to maintain a single hierarchy.** KS maintained a single top-level, “documents” hierarchy on her desktop computer into which she attempted to place all work-related material. Documents, web addresses (pasted into documents) and key email messages were all saved inside this hierarchy. NF, with the aid of his assistant, attempted to maintain essentially the same hierarchy in different places – to organize his electronic documents, his email messages, his

paper files and the work-related paper and electronic files of his assistant. Other participants mentioned problems with maintaining – and maintaining consistency between – several different organizational schemes – one for bookmarks, one for email, one for electronic files and yet another for paper files.

- **KFTF practices appear to vary according to job position and a participant’s general relationship to web information.** Managerial participants in our study rarely use the Web directly to conduct research. Instead, they receive their information “second-hand” through communications with others. Email is used extensively in this communication.

Differences between the remaining two groups – information specialists and researchers – were less clear but some interesting patterns emerge. The bookmarking facility (Favorites or Bookmarks) was heavily used by seven of the study’s eight information specialists. By contrast, only one of four researchers was observed to use bookmarking – and then primarily as a holding area for potentially relevant, but as yet untested, web addresses. (Several research participants expressed the feeling that their bookmarks were a “mess” in need of a clean up.) Is it possible that information specialists, perhaps by inclination or by virtue of their training, are more likely to have the discipline required to maintain a growing list of bookmarks?

Observed differences in the use of the Web and in associated keeping behaviors are consistent with an argument advanced by Grudin (2002). Grudin argues that the users of a software application – especially if the application mediates collaborative work or supports information exchange between people – often fall into distinct groups. In a workplace, from three to five groups typically emerge according to job position, activity pattern, and incentive structure. Each group approaches the software application with its own distinct set of needs and requirements. The needs and requirements of managers, for example, are very different than those of individual contributors.

A Functional Analysis

We can gain some explanation for the observed diversity of methods through an analysis of function. Based on observations and interviews, 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 filed paper printouts into an existing paper-based filing scheme.

- **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.
- **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 the colors of a web page is also lost. 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).
- **Currency of information.** It is often 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 a web address? Do they know when and why it should be accessed? 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. 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.

With possibly one exception, we did not observe nor did participants indicate during their interviews, that bookmarks were routinely checked. Consequently, we assigned a low rating to the reminding value of bookmarks. The possible exception was the information specialist LK who indicated that she regularly checked and “weeded” her list of Favorites.

- **Ease of integration.** Does the method help users to integrate new information or new references with ongoing projects and 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.
- **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, as currently implemented in both the Internet Explorer and the Netscape Navigator, 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 for selected keeping methods. The ratings are summarized in Table 2.

Table 2. 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?
Bookmark	Low	Low	Low	Med	High	Low	Low	Low	Low	Low

Table 2 should be regarded primarily as an illustration of the functional approach. It is not intended to be complete, either with respect to keeping methods or functions of possible value. Clearly, the ratings – especially those with a question mark – are subject to qualifications and debate. How portable, for example, is web information that is saved as a 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. 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.

Whether or not we agree on all of the answers, the functional analysis is useful for the kinds of questions it encourages us to ask. Does a keeping method have a reminding function? Does it enable us to provide a context for the web information? Does it integrate well with our existing organizational schemes? Does it allow us to access the web information we wish to keep from home and on the road as well as at work?

In addition to its use in the analysis of existing methods and supporting tools, the functional analysis can help to guide efforts to build new and improved tools. And the analysis can help to set expectations for the likely success of these 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 (Hightower et al., 1986; McEneaney, 1996; Wexelblat & Maes, 1999; Cockburn et al., 1999; Cockburn & Greenberg, 2000). Other research has explored the use of a virtual 3D environment for the organization of bookmarks (Robertson et al., 1998; Czerwinski, et al., 1999).

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. People who work in several places (e.g., home and office) may continue to ignore a history or bookmarking facility that is machine-specific. People 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.

Follow-on Work

The current study affords 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 this study are being used to help structure a larger investigation into the methods and supporting tools that people use to manage web information for re-use.

Follow-on work is already underway:

- We are conducting a survey of keeping methods for a much larger sample of users from each of the three groups – researchers, information specialists and managers – represented in the current study. We will add a fourth group, college students, to this survey. Our goal is to acquire quantitative data on the relative popularity of different keeping methods to complement the qualitative data of the current study. We are also gathering data that should help us to determine the relative importance of different functions for different participant groups.
- We are conducting a study involving many of the participants of the current study in order to gain a more complete picture of KFTF practices used by people. Just as the current study looks at keeping behavior, this second study looks at “re-finding” behavior. How easily can people return to a useful web site later on? Using what methods? What kind of problems do they encounter?
- We are also completing work on a simple prototype that extends the current Add Favorite facility of the Microsoft Internet Explorer. Our intent is to support a kind of “super” keeping method in which additional functions, such as reminding, the provision of context and an improved ease of integration, are supported in a single keeping method.

Longer range, 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.

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. The study also suggests that people may differ systematically in their “keeping” practices according to their job position and their relationship to web information.

A functional analysis helps to explain the diversity of methods observed. The analysis can also serve as an aid in the analysis of existing tools and in the development of new and improved tools to help people keep found things found.

Note

The KFTF project is supported by a three year National Science Foundation grant (#0097855,

<https://www.fastlane.nsf.gov/servlet/showaward?award=0097855>).

References

- Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: personal web space construction and organization. *Proceedings of CHI '98*, 41-48.
- Barreau, D., & Nardi, B.A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bulletin*, 27, 3.
- Byrne, M.D., John, B.E., Wehrle, N.S., & Crow, D.C. (1999). The tangled Web we wove: a taskonomy of WWW use. *Proceedings of CHI '99*, 544-551.
- Carroll, J.M. (1982). Creative names for personal files in an interactive computing environment. *International Journal of Man-Machine Studies*, 16, 405-438.
- Choo, C.W., & Auster, E. (1993). Environmental scanning: acquisition and use of information by managers. *Annual Review of Information Science and Technology*, 28, 279-314.
- Cockburn, A., & Greenberg, S. (2000). Issues of page representation and organisation in a web browser's revisitation tools. *Australian Journal of Information Systems*, 7 (2), 120-127.
- Cockburn, A., Greenberg, S., McKenzie, B., Jasonsmith, M., & Kaasten, S. (1999). WebView: A graphical aid for revisiting Web pages. *Proceedings of the OZCHI '99 Australian Conference on Human Computer Interaction*, 15-22.
- Czerwinski, M., Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S., & van Dantzich, M. (1999). Visualizing implicit queries for information management and retrieval. *Proceedings of CHI '99*, 560-567.
- Grudin, J. (2002). Emerging Norms: Feature Constellations Based on Activity Patterns and Incentive Differences. Microsoft Research Technical Report MSR-TR-2001-91. Microsoft Corporation, One Microsoft Way, Redmond, WA 98052. Retrieved Jan. 2002 from <http://www.research.microsoft.com/research/coet/Roles/TRs/01-91.doc>.
- Hightower, R.R., Ring, L.T., Helfman, J.I., Bederson, B.J., & Hollan, J.D. (1998). Graphical multiscale web histories: a study of PadPrints. *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*, 58-65.
- Jones, W., Bruce, H., & Dumais, S. (2001). Keeping found things found on the web. *Proceedings of CIKM'2001*, 119-126.
- Jones, W., & Dumais, S. (1986). The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Transactions on Office Information Systems*, 4, 42-63.
- Malone, T.W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems*, 1, 99-112.
- McEneaney, J. E. (1999). Visualizing and assessing navigation in hypertext. *Proceedings of the Tenth ACM Conference on Hypertext and Hypermedia*, 61-70.
- Pitkow, James E., & Kehoe, Colleen M. (1996). Emerging trends in the WWW user population. *Communications of the ACM*, 39 (6), 106-108.
- Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D. & van Dantzich, M. (1998). Data mountain: using spatial memory for document management. *Proceedings of UIST '98*, 153-162.
- Tauscher, L., & Greenberg, S. (1997). Revisitation patterns in World Wide Web navigation. *Proceedings of CHI '97*, 399-406.
- Tauscher, L., & Greenberg, S. (1997). How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47, 97-137.
- Wexelblat, A., & Maes, P. (1999). Footprints: history-rich tools for information foraging. *Proceedings of CHI '99*, 270-277.
- Whittaker, S., & Sidner, C. (1996). Email overload: exploring personal information management of email. *Proceedings of CHI '96*, 276-283.