

# Information Archiving with Bookmarks: Personal Web Space Construction and Organization

David Abrams

Perceptual Robotics, Inc.

1840 Oak Ave

Evanston, IL 60201

abrams@perceptualrobotics.com

Ron Baecker

Knowledge Media Design Institute

University of Toronto

Toronto, ON M5S 1A4

rmb@dgp.toronto.edu

Mark Chignell

Interactive Media Laboratory

University of Toronto

Toronto, ON M5S 1A4

chignell@mie.toronto.edu

## ABSTRACT

Bookmarks are used as "personal Web information spaces" to help people remember and retrieve interesting Web pages. A study of personal Web information spaces surveyed 322 Web users and analyzed the bookmark archives of 50 Web users. The results of this study are used to address why people make bookmarks, and how they create, use, and organize them. Recommendations for improving the organization, visualization, representation, and integration of bookmarks are provided. The recommendations include simple mechanisms for filing bookmarks at creation time, the use of time-based visualizations with automated filters, the use of contextual information in representing bookmarks, and the combination of hierarchy formation and Web page authoring to aid in organizing and viewing bookmarks.

## Keywords

WWW, bookmark, information space, survey, empirical study, design.

## 1. INTRODUCTION

The millions of documents on the rapidly expanding World Wide Web (WWW) further exacerbates the information overload problem documented by [8], [17] and others. The Web as we know it today has no classification system for Web sites, no formal indexing policy or controlled vocabulary, and no systematic system for naming Web pages or assigning authorship in a particular catalogue (except for domain names). Consequently, searching for specific information on the Web is a challenging and often frustrating task.

One strategy for dealing with information overload is to develop personal information systems consisting of focused subsets of information highly relevant to a particular user. Bookmarks are a simple tool for building these personalized subsets of information where interesting or useful Web pages (URLs) can be stored for later use. Users keep track of the resulting pointers to Web pages by creating a bookmark archive — a *personal Web information space*.

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

CHI 98 Los Angeles CA USA

Copyright 1998 0-89791-975-0/98/ 4..\$5.00

Bookmarks are very popular among users. In a 1996 survey of 6619 Web users [12], over 80% of respondents cited bookmarks as a strategy for locating information. Bookmarks were used slightly more than other navigation strategies: querying a search engine, referencing an index page, entering a known URL, and traversing links to a specific page. Over 92% of users had a bookmark archive and over 37% had more than 50 bookmarks.

Bookmarks serve as convenient shortcuts to frequently used Web pages as well historical pointers to useful information that may otherwise be forgotten. Bookmarks are file surrogates (aliases) pointing to original files in "tertiary storage," the massive distributed file system located in Web servers distributed around the world. Users create their own personal information space for the Web by making bookmarks, structuring the resulting collection, and managing its growth.

Personal information spaces have been studied in a variety of contexts (e.g., [6], [2], and [7]). Yet very little empirical research on Web users has been done and this research is the first in-depth empirical study of personal Web information spaces. Our goal was to uncover basic aspects of bookmarking behavior as a prelude to modeling large scale information archiving, thereby providing a scientific basis for the design of future bookmark management systems.

## 2. METHODOLOGY

Bookmark usage was studied using survey questionnaires, and through analyses of Web users' bookmark archives.

### 2.1 Survey of Bookmark Usage

The survey instrument was developed based on a pilot study of 12 university students with WWW experience done with informal interviews and pilot questionnaires. We used the results in the design of a formal survey questionnaire. More details on this questionnaire and its development (including the questions asked) appear in [1].

Approximately 450 of the questionnaires were distributed at The Internet Beyond the Year 2000 Conference (held in Toronto in early 1996). During the conference, registrants were asked to fill out a copy of the questionnaire. Respondents were instructed to identify themselves by their conference registration number, and were told that their data would be confidential and reported in the form of aggregations with questionnaire data from other participants.

The respondents consisted of University of Toronto faculty, staff, and students, and individuals from industry, government, and other universities. Most participants had an expressed interest in the Internet and the WWW.

322 completed survey forms were collected from the 450 distributed. 200 (62%) of the completed surveys were correctly filled in on every question. 76.5% of the 322 survey respondents were male and 23.5% were female; these proportions were roughly consistent across all sizes of bookmark archives. Age distribution was approximately normal with a mean of about 35 years. We estimate that approximately 80% of respondents were Netscape 1.x or 2.0 users, and there were very few Internet Explorer users at the time of the survey (we did not explicitly ask for browser type and version in the survey).

We stratified users by the number of bookmarks in our analysis. 6% of respondents had no bookmarks, 10% had 1-10, 24% had 11-25, 44% had 26-100, 14% had 101-300, and 2% had 300+ bookmarks. We selected these groups based on our pilot study which suggested distinct characteristics for each user segment. Respondents were asked to choose from this list of groups.

## 2.2. Bookmark Files of 50 Users

56 bookmark files were collected through electronic mail in late summer of 1996. 70% of the people volunteering bookmark files were from the Knowledge Media Design Institute (KMDI), a multi-disciplinary group of faculty and students in Toronto interested in new media. 20% of the bookmark files were from Computer Science students. The remaining 10% were from industry professionals. 95% of the sample consisted of Netscape users (identified by the file meta-tags). Six files were missing data. 50 bookmark files were then analyzed using the date/time that each bookmark URL was (i) created, (ii) last visited, (iii) last modified, and its location in the folder hierarchy.

## 3. THE PROCESS OF SELECTING BOOKMARKS

Users judiciously select useful Web sites to add to their bookmark archives. The comments made in the questionnaire illustrate why bookmarks are created and how they are used. Our analysis of the bookmark files identified patterns in the types of sites which users bookmark.

### 3.1 Why Bookmarks are Created

Bookmarks take very little physical interaction to create and they are "easy to make." The survey respondents employed five criteria to determine whether to bookmark a Web page: (i) general usefulness, (ii) quality, (iii) personal interest, (iv) frequency of use, and (v) potential future use.

Usefulness is an important factor in bookmark creation and in organizing and pruning a personal archive ("I typically only organize useful information."). Another user complained that "weeding ones [which are] no longer useful" takes too much time and cognitive effort. Since criteria for the usefulness of information changes over time [13], the usefulness of pages stored in a personal archive will tend to be re-evaluated from time to time.

Bookmarks also provide "quick access to key sites" and make it "easy to return to key pages." Users like being

"able to return to useful/interesting sites" and "the ability to quickly get to the URL of a site of personal interest." One respondent described his bookmarks as a "customized interest list." Thus bookmarking a set of Web pages results in a *personal Web information space*.

Since bookmark invocation is fast and easy, users bookmark pages they frequently return to. "I constantly use them — it's the best way to get back to frequently used sites. I'd be lost without them." Bookmarks provide "speedy access to 4-5 very frequently used sites."

Bookmarks are also created when people want to defer reading an interesting page until a future session, possibly because they are too busy dealing with a current problem. For example, one respondent uses bookmarks to "store all the useful sites that I would use for future use."

### 3.2 How Bookmarks are Used

Bookmarks (i) reduce the cognitive and physical load of managing URL addresses, (ii) facilitate the return to groups of related pages, and (iii) enable users to create a personal information space for themselves and others. Below is a summary of a taxonomy from [1]:

#### Reducing user load

- Avoiding managing URL addresses
- Aiding memory and keeping history

#### Facilitating navigation/access

- Speeding information access
- Finding Web information

#### Collaborating/publishing/archiving

- Creating a personal Web information space
- Authoring and publishing Web pages
- Collaboratively using Web information

Bookmarks reduce the cognitive and physical load of browsing hypertext [16]. They insulate users from the tedious task of typing, managing, storing and interpreting URL addresses (the "convenience of not having to retype a long Web site address.") The result is more fluid movement from one page to the next. Bookmarks let users "easily store an address without having to write it down." Representing Web pages by titles rather than by URLs reduces cognitive load and enables users to focus on the contents of the page. URLs are "cumbersome" and "tedious." Bookmarks serve as a "mnemonic device" for users, reminding them of important information they have found on the Web. One user wrote "they [bookmarks] free me from the exhaustive task of remembering everything." Bookmarks are a "memory replacement," external artifacts that mediate cognition [10].

Respondents associated bookmarks with the temporal sequence of browsing sessions. One user wrote, "I can keep track of what I was doing lately and a few weeks earlier" with bookmarks. A single bookmark may represent an individual (or discrete group of) browsing sessions. When reviewing a list of bookmarks, the user sees a time-ordered view of separate information foraging tasks (one respondent viewed bookmarked "pages as a history of using the Web."). However, when users organize their bookmarks they tend to lose this temporal sequencing. One user wrote that bookmarks "take me to the last site where I

*was not finished during the last session.*" This user associated tasks with individual browsing sessions, and also used bookmarks as an inter-session history mechanism, writing, *"bookmarks are essential to finding good sites again and remembering previous sessions."*

The temporal and task association of bookmarks is illustrated by users' suggestions for better Web browsers. One user wants to be able to find Web pages based on the *"last time I visited a site or used a bookmark."* Associating tasks with discrete browsing sessions is a key element defining the relationship between bookmarks and history mechanisms. *"I want a way to see and understand the history of my Web browsing for the last few weeks or more."* For example, *"I want to find a page I looked at two weeks ago but didn't think to bookmark."* Users do not always think to bookmark a page *"for future use,"* but require a means of accessing inter-session history. In the absence of such functionality, users are bookmarking pages to enable access to previous browsing sessions.

Bookmarks are also used to jump between Web localities [1]. Users create bookmarks to help them quickly find Web pages, reduce time spent foraging, and mark serendipitous sites found along the way.

Bookmarks provide *"fast access to information"* because the time to select a bookmark is very short compared to browsing for a page, entering the URL or using a search engine. They provide a *"most spontaneous response"* to users because of the *"direct access to important pages."*

Users collect bookmarks so that they can create their own personal information space and share it with others. This personal Web information space represents their most critical information resources on the Web. For example, *"I like being able to create my own classification system."* Creating a personal Web information space means users *"can create an organized, logical format for getting fast and easy access to a site."*

Users create Web pages out of bookmarks by authoring an HTML file with the bookmarks as hypertext links. They can thereby easily add annotations and create customized views.

Bookmarks are used to share Web resources with third parties. A group of users working on the same project will mail each other bookmarks in order to collaborate. Users share bookmarks based on individual expertise. A user that regularly manages a set of bookmarks in one domain is a precious resource to other members of the group. One respondent uses bookmarks *"in my course to provide students with initial navigational markers."* A librarian wrote that she *"collects [bookmarks] for clients."*

Bookmarks are used in presentations. One user wrote *"I mostly use them while giving lectures,"* using them to move quickly from one site to the next.

### 3.3. Bookmarking Metaphors

Survey respondents mentioned a number of different metaphors for describing how they used bookmarks. We observed four major metaphors: *identification*, *collection*, *movement*, and *episodes*.

The identification metaphor conceptualizes bookmarks as small tags or distinctive labels that are placed on information. For example, one user wrote that bookmarks are an *"extremely easy method to mark information."* Another respondent uses bookmarks to mark a spot and come back to it later.

The collection metaphor is based on the notion that the user is stationary and he/she is pulling specific information out of the vast information space of the WWW. One respondent likes *"the ability to quickly retrieve sites"* with bookmarks.

Many users employ notions of location and movement in describing the process of browsing. They perceive themselves to be traveling through a vast space of information. The traveling metaphor implies destinations, landmarks and paths. For example, bookmarks tell *"where I was."* They are an *"easy way to find places I've been."* Bookmarked pages stand out as landmarks, and as temporal and navigational guides. A subset of the traveling metaphor is information foraging which encompasses an active process of finding or searching for a piece of information [11] (e.g., [3] is an information foraging tool).

Bookmarks also represent a chronological list of episodes. One user wrote that bookmarks tell *"what I was doing over a period of many browsing sessions. I can keep track of what I was doing lately and a few weeks earlier"* and they describe *"my history"* of navigating the Web.

### 3.4. What is Bookmarked

The users that we studied tended to create only one bookmark for each distinct Web site. On average, 86% of the bookmarks in a user's archive pointed to distinct domain names. The proportion of bookmarks that pointed to distinct domain names dropped off only slightly with larger archives, ranging from a high of 93% distinct domain names for small archives (1-25 bookmarks) to 81% distinct domain names for the largest archives (over 300).

Over 95% of users with 20 or more bookmarks had at least one search engine or index site in their archive. When bookmarking a query engine (e.g., Lycos) users bookmarked the page for entering queries. In contrast, they tended to bookmark selected parts of the hierarchy when bookmarking a meta-index (e.g., Yahoo). They rarely bookmarked search results pages.

When bookmarking on-line news sites (e.g. CNN, NY Times), users tend to bookmark the newspaper-like front page with headlines and links to articles. Most publishers regularly update this page with fresh content. Users rarely bookmark Web pages of individual news articles. [1] presents a taxonomy and details on what users bookmark.

### 3.5. Rate of Growth of Bookmark Archives

Time-series analyses of the bookmark files of 50 users showed that archives grow incrementally in spurts and that many bookmarks go unused for months.

To examine the rate of growth in bookmarks over time, the bookmark files of 50 participants were combined in an aggregate file tracing one year's growth. For each bookmark file, we calculated the creation date of every bookmark relative to date that the archive began. The

average number of bookmarks was highly correlated with time ( $R = .996$ ,  $P < 0.001$ ). The best fitting linear regression equation is:

$$N = 7.45 + 0.18 * T$$

where  $N$  is the number of bookmarks and  $T$  is number of days. Bookmarks in our aggregated sample were added at a fairly constant rate of about one every five days.

Over 94% of users surveyed created an average of five or fewer bookmarks during each browsing session. We analyzed bookmark files from individual users to identify patterns in their growth. Figure 1 illustrates this, showing the time course of bookmarking for a particular user by plotting the elapsed number of days each bookmark was added to the archive relative to the date when the first bookmark was created. Vertical segments in the figure show clusters of bookmarks added on the same day.

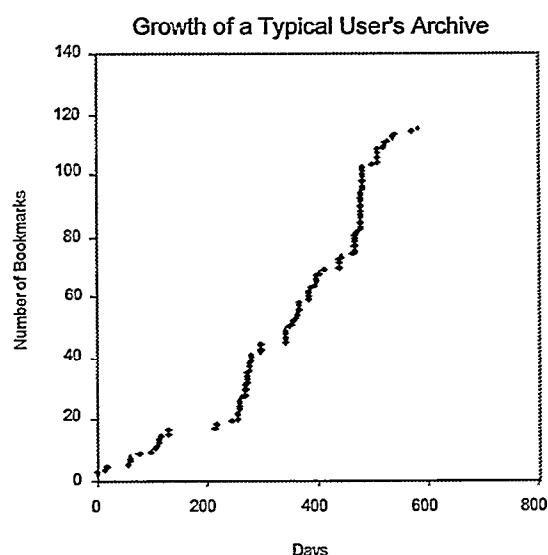


Figure 1. Growth rate of a typical user's bookmark archive.

It can be seen that the user adds bookmarks sporadically with periods of intense bookmarking separated by periods where few if any bookmarks are added.

#### 4. ORGANIZATION OF A BOOKMARK ARCHIVE

Users organize their bookmark archive to fight entropy and add structure [1]. This helps reduce retrieval time, and enables the user to collaborate with other users by sharing the structured archive. Large bookmark archives become unwieldy when they are not organized. Organizing bookmarks is one of the top three Web usability problems reported by 6619 survey respondents in [12].

Users must continually tradeoff the cost of organizing their bookmarks and remembering which bookmarks are in which folders versus the cost of having to deal with a disorganized set of bookmarks. They optimize the cost structure of their information environment to make a small amount of information available at very low cost, and larger amounts available at higher costs [11]. Many users cost-tune their archives by expending the least amount of effort

needed to build up enough structure to support fast retrieval of their most useful bookmarks.

#### 4.1. Methods of Organizing Bookmarks

Our 322 users employed seven methods for organizing bookmarks (Table 1, Figure 2).

Approximately 37% of respondents checked "I don't organize bookmarks — they stay in the order in which I created them." Organizing bookmarks is labor intensive, requires extensive time, and is difficult to do. One respondent noted "I'm just getting started. I didn't even consider this [organizing bookmarks] yet." Users with fewer than 35 bookmarks tended not to organize them.

Table 1. Methods for Organizing Bookmarks
<i>No Organization:</i> bookmarks stay in the order created.
<i>Ordered List:</i> users manually re-arrange a list.
<i>Set:</i> users create folders to categorize bookmarks.
<i>Hierarchy:</i> users create folders within folders.
<i>External:</i> users export bookmarks to a separate program.
<i>Web Page:</i> users create Web pages out of their bookmarks.
<i>Other:</i> users search, sort and use unusual methods.

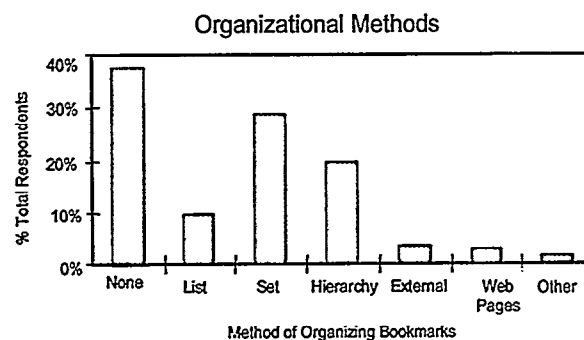


Figure 2. Percentage of survey respondents who reported using each of the seven methods to organize bookmarks.

Relatively few users managed a list without also using folders. The strategy of using an ordered list was most likely to occur for users who had relatively few (less than 25) bookmarks.

Many users organized bookmarks within a single-tiered set of folders. An analysis of variance indicated that the self-reported frequency of creating folders depended on the reported total number of bookmarks the user had, ( $F[4,257] = 6.24$ ,  $P < 0.001$ ). The use of a set of folders peaked with the 101-300 bookmark group and dropped considerably when users had 300 or more bookmarks. The 300+ bookmark users reported that they preferred to use a multi-level hierarchy to manage their bookmarks.

Users with 26-100 bookmarks and with 101-300 bookmarks created folders within folders 28.7% and 24.1% of the time, respectively, in contrast to a corresponding figure of 44.4% of users with 300+ bookmarks. The reported frequency of use of sub-folders (i.e., a hierarchy) differed significantly with the reported total number of bookmarks ( $F[4,249] = 3.63$ ,  $P < 0.01$ ). More hierarchies were used by the group with the largest number of bookmarks (300+) as indicated by a post hoc comparison ( $P < 0.05$ ).

Relatively few (2.2%) of the survey respondents reported using Web pages for archiving bookmarks. Users wrote "I put links on my bookmarks page," and "I create Web pages once I get more than 10 bookmarks." This provides an expressive medium for managing bookmarks, which can include the use of comments and embedded hierarchies: "I organize certain categories into home pages." I organize "by section in sub-pages of my home page."

2.9% of users in our survey created bookmarks and then off-loaded them to a separate application. This was in spite of the fact that such tools as were available at the time of the survey were difficult to use, were not powerful enough to manage a large number of bookmarks, and possessed a steep learning curve. Offloading to a separate application was generally done by respondents who had a large number of bookmarks (e.g., 44.4% of users with 300+ bookmarks offloaded them to separate programs). Users stored bookmarks in simple text files or databases, or used bookmark management tools.

Poor scalability makes visualization, retrieval, and browsing difficult. For example, it is "hard to get an 'overall' view of bookmarks in long lists," "I can't see them all at once." The multi-tier pull-down menu is difficult to use. "The menu system is awkward and folders are in the way when browsing." Semantic hierarchies break down easily, as one user wrote "bookmarks are unstable if they are not kept in meaningful categories, which takes a lot of time." In addition, "part of the problem is software the other is creating stable categories in one's mind."

#### 4.2. Organizational Methods and Experience

The sophistication of the organizational method used increased with the experience level of the user (Figure 3).

Experience Level of Organizational Methods

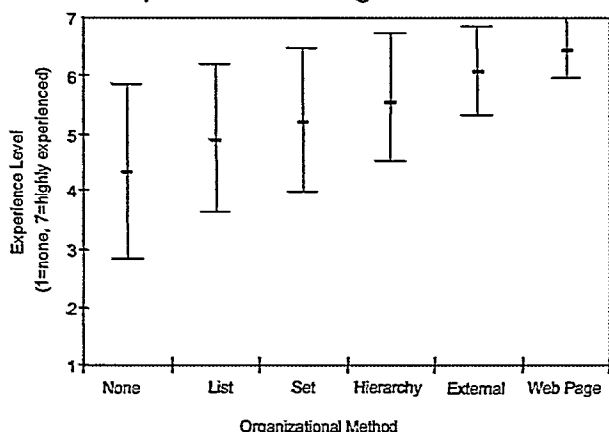


Figure 3. Self-reported experience level from 291 respondents, showing means and error bars (plus or minus the standard error).

There was a significant correlation (Spearman's rho) between level of experience and organization method used ( $r[291] = 0.3333$ ,  $P < 0.001$ ). Users who authored Web pages with bookmarks had a mean experience level of 6.50 (on a scale with 7 as the maximum), while those who do not organize their bookmarks had a mean self-reported experience level of 4.36.

#### 4.3. Organizational Habits of Bookmark Users

Users file bookmarks at different times and rates. The filing habits we observed in bookmark users (Table 2, Figure 4) and are similar to the way users manage electronic mail archives [15].

Table 2. Bookmark Filing Strategies

<b>No Filers:</b> users who never organize bookmarks. The bookmarks stay in the order in which they were created.
<b>Creation-time Filers:</b> users who store a new bookmark in the appropriate category when it is first created.
<b>End-of-session Filers:</b> users who organize all their new bookmarks at the end of the session.
<b>Sporadic Filers:</b> users who organize bookmarks occasionally and schedule cleanup sessions.

Organizational Habits

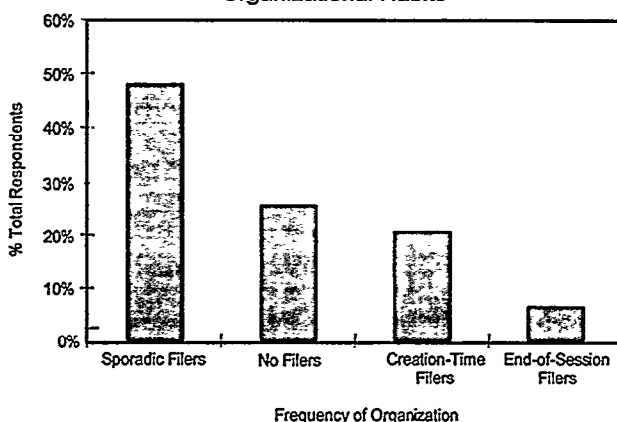


Figure 4. Bookmark filing habits of 299 respondents, showing the percentage of total users in each category

Roughly half of the respondents organized their bookmarks sporadically. 26% of respondents never organized their bookmarks, and 23% of them stored each new bookmark in its proper place at the time it is created. Only 7% of users organized bookmarks at the end of every session.

Sporadic filers chose when to optimize their work environment. They started a special "spring-cleaning" session to organize their bookmarks. As a result of these decisions to clean up their bookmarks, they tended to "organize 2-3 times/week" or "organize once a week."

Creation-time filers categorized a new bookmark and stored it in a folder at the time when the bookmark was created. These users avoided a pile-up of unorganized bookmarks because they "organize often at the time of creation." Overall, 23% of users surveyed reported filing at creation time; however, a much higher proportion (67%) of users with over 300 bookmarks reported filing at creation time.

Very few users (7%) organized bookmarks at the end of every browsing session. End-of-session filers must set aside a special period of time to organize after each session. Most users, 94% of all respondents, create relatively few (at most five) bookmarks per WWW browsing session. There is therefore no large accumulation of unorganized bookmarks at the end of each session which forces them to file.

An analysis of variance indicated that the average time spent per session varied significantly with when bookmarks were organized ( $F[3,278] = 3.5237$ ,  $P < 0.05$ ). Post hoc analysis indicated that subjects who organized their bookmarks at the end of the session spent significantly more time per session than did subjects who either sporadically organized or never organized ( $P < 0.05$ ). They also used folders extensively, with 80% of end-of-session filers using folders in some manner to organize bookmarks.

One user said that he organizes bookmarks "when they no longer fit in my drop down menu." Beyond the 35 bookmark threshold for easy visibility, users created folders in direct relation to the number of bookmarks in their archive (Figure 5). A linear regression equation ( $F[1,26] = 38.52$ ,  $p < .001$ ) describes the relationship between the number of folders and the number of bookmarks:

$$\text{Folders} = 1.14 + 0.14 * N$$

Use of Folders

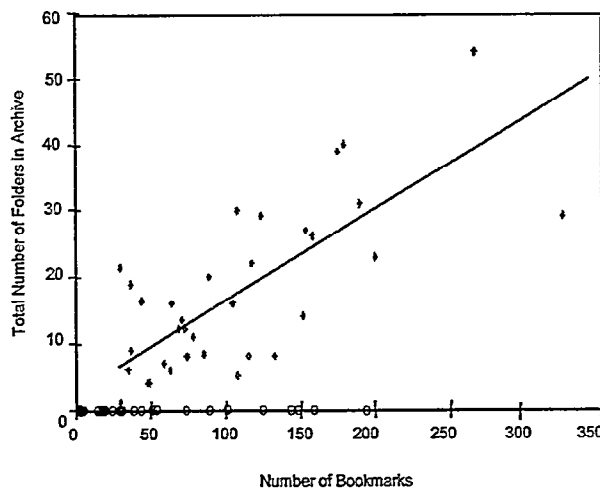


Figure 5. Use of folders in 50 users' bookmark archives, showing the number of folders in each archive as a function of the size.

As shown in Figure 5, users with less than 35 bookmarks have no folders. Above 35 bookmarks, folders grow linearly with bookmarks.

## 5. RETRIEVAL OF BOOKMARKS

Users create many bookmarks that they subsequently do not use. When bookmarks are used, they may be stored temporarily (for short term use) or archivally (for use over the long term); they may be published or used collaboratively. One of the survey items asked respondents to rate the importance of these four different types of bookmark usage (temporary, archival, publishing, collaborative) on a seven-point scale (1=not important, 7=very important). There was a significant difference in the degree of importance assigned to the four different types of bookmark use ( $F[3,1064] = 2.61$ ,  $P \leq 0.05$ ). Respondents reported using bookmarks as archives (for long-term storage) rather than as caches for frequently used information. Respondents assigned the archival use of bookmarks a mean importance rating of about 6, as

compared with mean ratings of about 3 for each of the other types of use (temporary, publishing, and collaborative).

Tauscher [14] found that users employ history to revisit pages they have just visited. They access only a few pages frequently and browse in very small clusters of related pages. [1] includes a detailed analysis of the navigation patterns which emerge after bookmark use. Bookmarks are usually a launching point for exploring clusters and a means of jumping between different clusters.

Our analysis of bookmark files substantiates the archival use of bookmarks. The cumulative plot (aggregated over users) for the number of days since the last time each user visited each bookmark is shown in Figure 6.

There is a median time of about 100 days since the last visit to a bookmark. Roughly half of bookmarks in the study had been visited in the preceding three months; 67% of the bookmarks had been visited in the last 6 months; and 96% of the bookmarks had been visited in the past year. Bookmarks were used infrequently, but almost all of them had been used at least once in the previous year.

Retrieval from Bookmark Archives

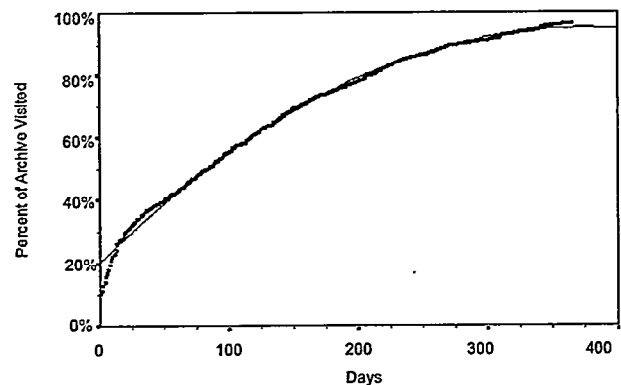


Figure 6. Aggregate retrieval rate of bookmarks pooled from 50 users (the best fitting second order regression curve is also shown).

Figure 7 shows the archival retrieval patterns of a typical user by rank ordering each bookmark by the number of days since its last use.

Retrieval from a Typical Bookmark Archive

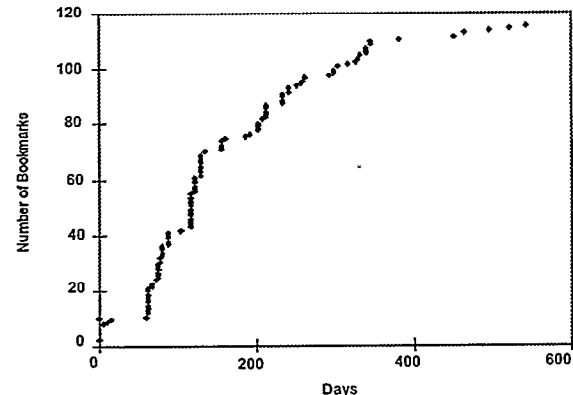


Figure 7. Typical user's retrieval pattern from an archive of 115 bookmarks.



Only 6% of this user's bookmarks had been visited in the previous month, and only 30% were visited in the previous 3 months.

### 5.1. Retrieval from a Hierarchy of Bookmarks

The comments from survey respondents showed that bookmark users find it difficult to manage a large number of bookmarks. The bookmark hierarchy is created incrementally and somewhat informally over time, but once in place it enforces a rigid structure. Users must then remember where bookmarks are stored within this structure. Finding an item in a deeply nested hierarchy is often challenging for users [5].

Labeling of bookmarks also tended to be a problem, because *"bookmarks are not descriptive enough"* and they *"aren't great describers of the actual content"* of the Web page. 276 respondents reported on a scale of 1 (not a problem) to 7 (very significant problem) that descriptiveness of titles is a problem (mean=4.2, std.=2.1). Yet very few users reported that they actually change the name of bookmarks. This may be due to lack of good tools and traditional difficulties of naming items [4].

Since a bookmark is a pointer to a Web page which can be modified at any time, the content can change without warning and URLs can become invalid. One user wrote, *"I cannot easily tell when a bookmark has become obsolete because the URL has changed or the page has disappeared (I would like a link checker)."*

## 6. DESIGN IMPLICATIONS

Our interpretation of the results is that bookmarking is a rational process that transfers knowledge in the head to knowledge in the world [9]. Users have to deal with a number of tradeoffs and concerns in bookmarking, including:

- A. Is it worth making a bookmark when it will probably be used infrequently?
- B. Is the time taken to organize bookmarks (a current cost) worth the future benefit?

Bookmarking behavior seems to change as the number of bookmarks increase. The use of folders begins at about 35 bookmarks, while use of multi-level hierarchies becomes prevalent in archives of over 100 bookmarks. 44% of users with over 300 bookmarks offload to a database and many invent their own creative organizational schemes. Users with archives of over 300 bookmarks tend to be motivated to organize because they share the resulting index of URLs with others.

For many users who cannot expend the significant effort required to maintain a growing archive the value of the user's personal information space breaks down. Entropy creates disorder until the archive finally becomes a *complex information space* [1]. As the WWW becomes the dominant global information resource and new compelling Web sites emerge, users will need bookmark management systems that scale-up to manage their growing archives. Consideration of these results leads us to design recommendations in four fundamental areas: organization, visualization, representation and integration.

### Organization

Users need scalable tools because *"bookmarks pile up too fast and become unmanageable."* They should minimize the effort needed to organize bookmarks and build hierarchies. *"I cannot place it in the prescribed folder easily at the mouse click."* Systems should provide users with an immediate filing mechanism so that bookmarks can be filed at creation-time, thereby avoiding the buildup of disordered bookmarks. Since many users *"hate having to sort them [bookmarks] into folders,"* systems should provide automated sorting capabilities and let hierarchies be organized within the context of the browsing session. *"I'd like to have another metaphor for managing bookmarks than folders."* Bookmarks could be organized based on usage patterns so users can easily cost-tune their archives. For example, users could order bookmarks based on frequency of use. The most commonly used bookmarks could be stored within the pull-down menu for navigation, and archival bookmarks could be stored separately for long-term retrieval.

### Visualization

Poor visualization hampers users with large archives. *"My problem is that I cannot get all the bookmarks on the screen."* Bookmark management systems must provide ways of visualizing large numbers of bookmarks to aid retrieval. Our research showed that visualization directly affects organizational habits (i.e. they create folders once the pull-down fills the screen limit of approximately 35 items). For example, one user wrote *"folders are in the way when browsing."* Visual clutter results from pollution of a growing archive. *"I end up with many bookmarks that I don't use; it clutters my list."* Designers should investigate time-based visualization with automated filters to display large numbers of bookmarks [18].

### Representation

Bookmarks are represented by texts which defaults to the title of the respective Web site. Consequently, *"bookmarks aren't great describers of the actual content [of the Web page]."* Systems should allow users to easily rename a bookmark when it is created because it is *"difficult to change the name to something more descriptive."* Multiple representations would help users visualize a large archive. Currently, the *"limited information in the top level listing of bookmarks"* makes it difficult to find bookmarks. One user suggested to *"make a short descriptive title available at the top level."* Designers should also investigate the use of contextual information in representing bookmarks (e.g. based on the bookmark's surrounding hypertext cluster).

### Integration

*"There is no really good bookmark organizer that doesn't demand that you exit the browser."* An integrated bookmark management system should fit naturally into the Web browser. One user suggested a tight coupling between bookmarks and search engines: *"I tend to collect and organize resources in batches. An interface between the search queries and my bookmarks would be nice."* Since bookmarks are used to collaborate with other users and publish information, tools should combine hierarchy

formation and Web page authoring to aid organizing and viewing bookmarks. Users with more than one computer need ways of integrating multiple lists. For example, "I cannot reference a single bookmark file across multiple platforms. I need NFS-like networkable bookmarks."

## 7. CONCLUSION

Bookmarks serve as starting points for hypertext exploration. They are created and stored for archival purposes, and often not visited for months. Users must weigh the costs of organizing bookmarks against the expected gains. Thus bookmarking takes place within the context of the users' ongoing information requirements and their assessment of how important current bookmarks will be to them in the future.

The advantage for users who maintain a few bookmarks in a single list is that they can see all of their bookmarks on the screen at the same time. This strategy also minimizes current effort. However, as the number of bookmarks increases, users typically employ more sophisticated organizational strategies such as hierarchies of folders. Our results indicate that a sizable proportion of users organize their bookmarks only when they have to, i.e., when a backlog of unorganized bookmarks is accumulated.

Proactive organization of bookmarks tends to occur when people have large numbers of bookmarks. Many users with 300+ bookmarks tend to file at the time of creation because they must keep an up-to-date archive in order to manage so many bookmarks. For these users the benefits of an up-to-date organizational structure outweigh the distraction from the browsing that is required to file a bookmark.

Our study used a questionnaire and an analysis of bookmark files to obtain an initial assessment of how bookmarks are archived and why. Many questions remain to be answered: What happens when people try to manage many hundreds or even thousands of bookmarks? How do these results depend upon the bookmark management capabilities of the browser? Do our results hold with larger samples of users taken from broader populations of users? What other functions do bookmarks serve? For example, a mental maps study of 27 users in [1] suggests that users may create bookmarks to personalize the Web and structure the "cloud of unmapped resources" in terms of familiar landmarks, but further study is needed.

Given the ever increasing importance of the Web and its role as a general repository of information, understanding the bookmarking process and developing appropriate tools for organizing large numbers of bookmarks are likely to become pressing issues. This research [1] lays an empirical foundation for further study of Web user behavior and the design of bookmark management systems.

## ACKNOWLEDGMENTS

Bruce Homer was instrumental in analyzing the data set. We are also grateful to Colin Furness and Jade Rubick for

assistance, and to the Information Technology Research Centre of Ontario and the Natural Sciences and Engineering Research Council of Canada for financial support.

## REFERENCES

1. Abrams, D. *Human Factors of Personal Web Information Spaces*. MS Thesis, Department of Computer Science, University of Toronto, 1997. Also available as <http://www.dgp.toronto.edu/~abrams>.
2. Burton, H.D. (1987). FAMULUS revisited: Ten years of Personal Information Systems. *Journal American Society for Information Science* 32, 440-443.
3. Card, S.K., Robertson, G.G., and York, W.M. (1996). The WebBook and the Web Forager: An Information Workspace for the WWW. *Proc. CHI'96*, 111-117.
4. Carroll, J.M. (1982). Creating Names for Personal Files in an Interactive Computing Environment. *Int. Journal of Man-Machine Studies* 16, 405-438.
5. Lansdale, M. (1983). The Psychology of Personal Information Management. *Applied Ergonomics* 19, 55-66.
6. 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.
7. Nardi, B., Barreau, D. (1995) Finding and Reminding: File organization from the Desktop. *ACM SIGCHI Bulletin*, 27(3), 39-43.
8. Nielsen, J. (1990). *Hypertext and Hypermedia*. Academic Press.
9. Norman, D.A. (1988). *The Design of Everyday Things*. N.Y.: Basic Books.
10. Norman, D. (1993). *Things That Make Us Smart*. Addison-Wesley
11. Pirolli, P. and Card, S. (1995). Information Foraging in Information Access Environments, *Proc. CHI'95*, 118-25.
12. Pitkow, J. (1996). GVU's 5th WWW User Survey. [http://www.cc.gatech.edu/gvu/user\\_surveys/survey-04-1996](http://www.cc.gatech.edu/gvu/user_surveys/survey-04-1996).
13. Schamber, L., Eisenberg, M.B., and Nilan, M.S. (1990). A Re-examination of Relevance: Towards a Dynamic Situational Definition. *Information Processing and Management* 26, 755-776.
14. Tauscher, L. and Greenberg, S. (1997). How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *Int. Journal of Human Computer Studies*, 47(1), 97-138.
15. Whittaker, S., and Sidner, C. (1996). Email Overload: Exploring Personal Information Management of Email, *Proc. CHI '96*, 276-283.
16. Wright, P. (1991). Cognitive Overheads and Prostheses: Some Issues in Evaluating Hypertexts. *Proc. ACM Hypertext '91 Conference*, 1-12.
17. Wurman, Richard. (1989). *Information Anxiety*. Doubleday, New York, NY.
18. Yiu, K., Baecker, R.M., Silver, N., and Long, B., A Time-based Interface for Electronic Mail and Task Management, *Proc. HCI International '97*, 1997.