HUMAN FACTORS
OF
PERSONAL WEB INFORMATION SPACES

by

David Abrams

A thesis submitted in conformity with the requirements for the degree of Masters of Science
Graduate Department of Computer Science
University of Toronto

© Copyright by David Abrams, 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-51586-9
HUMAN FACTORS
OF
PERSONAL WEB INFORMATION SPACES

Masters of Science, 1997
David Abrams
Department of Computer Science
University of Toronto

ABSTRACT

How do people use World Wide Web bookmarks? This in-depth empirical study of WWW human factors describes a user's personal Web information space. It is based on interviews of 12 users, a survey of 322 people, a mental maps survey of 27 users, and an analysis of the bookmark files of 56 users and the usage data of 23 users over 6 weeks.

The WWW is a complex information space with five defining properties: (1) users are overload with information, (2) the Web is polluted with redundant, erroneous and low quality information, (3) it progresses toward disorder according to the principle of entropy, (4) it has no aggregate structure which organizes distinct Web localities, and (5) users have no global view of the entire WWW from which to forage for relevant Web pages.

In dealing with this complex information space, users find localities of interest, create bookmarks to return to them, conceptualize WWW access through metaphors, and create a personal information space with the bookmarks they collect. Users employ this personal information space to counteract five problems of a complex information space in the following ways. (1) They prevent information overload by incrementally building a small archive. (2) They avoid pollution by selecting only useful items and creating a known source of high value. (3) They reduce entropy through maintenance, although they organize only when necessary. (4) They add structure by cost-tuning their information environment. (5) They compensate for the lack of a global view by creating their own personal view.

As a user's bookmark archive grows, it can begin to exhibit the properties of a complex information space. Thus, users are continually challenged with maintaining an effective personal Web information space.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Ron Baecker, for his encouragement, support, expertise and breadth of experience. Empirical studies can be tedious and lonely work, but Ron made this one fun and exciting.

Professor Mark Chignell was the second reader of this thesis. Many of the concepts about information spaces were based on his never-ending supply of creative ideas. Bruce Homer, University of Toronto, performed over 75 hours of statistical analysis and his cognitive science expertise was extremely valuable. Paul Lomic, University of Toronto, provided invaluable help in identifying relevant statistical tests and critiquing the data analysis. Linda Tauscher, University of Calgary, shared her usage data and generated URL vocabulary graphs which were used in the study.

Marshall McClintock, Microsoft, was the best devil’s advocate of my thesis and his in-depth critiques helped me identify and solve many problems. John Waterworth, UK, provided excellent feedback on the cognitive aspects of Web users. James Pitkow, Georgia Tech, helped us identify related work and his GVU survey was critical to the thesis. I would like to thank Stuart Card, Xerox PARC, for sharing related research. Jay Girotto, Harvard University, was a source of invaluable creative ideas and feedback. Colin Fumess, University of Toronto, played a critical role in the design of the surveys and provided expertise in the psychology of HCI. Paul Cooper, Northwestern University, was always a readily available source of scientific wisdom. William Hunt, Ian Graham, Ann-Barbara Graff, Kishore Jayabalan, David Modjeska, Nick Woolridge, Byron Long, Naomi Freelander, and Gale Moore, all from the University of Toronto, volunteered their
time to help proof-check and critique the surveys and the thesis. I would like to thank the
over 400 people who completed the various interviews and surveys.

I would like to thank my family for their never ending support and
couragement. Finally, my friends at the Newman Centre made life in Toronto an
adventure and extended their hospitality and home to me.
TABLE OF CONTENTS

1. INTRODUCTION .......................... 1-1
   1.1. MOTIVATION .......................... 1-1
   1.2. RESEARCH OVERVIEW .................. 1-3
   1.3. METHODOLOGY ......................... 1-4
   1.4. THESIS OUTLINE AND PREVIEW OF FINDINGS .......................... 1-10

2. INFORMATION SPACES AND THE WORLD WIDE WEB .......... 2-1
   2.1. COMPLEX INFORMATION SPACES ........... 2-2
   2.2. THE WWW AS A COMPLEX INFORMATION SPACE ..... 2-8
   2.3. USER BEHAVIOR IN THE WEB .............. 2-13
   2.4. PERSONAL WEB INFORMATION SPACES .......... 2-17

3. INFORMATION SPACE BEHAVIOR ON THE WWW ........ 3-1
   3.1. USERS FIND LOCALITIES OF INTEREST ........ 3-1
   3.2. USERS CREATE BOOKMARKS TO DISTINCT WEB LOCALITIES .......... 3-3
   3.3. USERS CONCEPTUALIZE WWW ACCESS THROUGH METAPHORS .......... 3-17
   3.4. USERS BUILD A PERSONAL WEB INFORMATION SPACE .......... 3-21

4. PWIS #1: START SMALL AND BUILD INCREMENTALLY .... 4-1
   4.1. SIZE AND GROWTH OF ARCHIVE ............. 4-1

5. PWIS #2: SELECT ONLY USEFUL ITEMS .......... 5-1
   5.1. WHAT USERS BOOKMARK .................. 5-1
   5.2. WHY USERS BOOKMARK ................... 5-10
   5.3. HOW BOOKMARKS ARE USED ............... 5-13

6. PWIS #3: ADD VALUE THROUGH ORGANIZATION .......... 6-1
   6.1. ORGANIZATIONAL METHODS ................. 6-2
   6.2. ORGANIZATIONAL FREQUENCY ............... 6-10
   6.3. INFLUENCE OF SCALE ON ORGANIZATION ........ 6-16

7. PWIS #4: ATTEMPT TO STRUCTURE FOR RETRIEVAL .......... 7-1
   7.1. FILING AND RETRIEVING BOOKMARKS .......... 7-1
   7.2. BOOKMARK AS REPRESENTATION ............. 7-7

8. PWIS #5: ESTABLISH A PERSONAL VIEW OF THE WWW .......... 8-1
   8.1. MENTAL MAPS OF THE WEB ................. 8-1
   8.2. HYPOTHESIS ABOUT PERSONALIZATION AND EVOLUTION OF GESTALT VIEWS OF THE WEB ........ 8-13

9. PERSONAL INFORMATION SPACES BECOME COMPLEX .......... 9-1
   9.1. COMPLEXITY PROBLEMS IN A PERSONAL WEB INFORMATION SPACE .......... 9-2
   9.2. CURRENT SOFTWARE SYSTEMS ............... 9-4

10. CONCLUSIONS ................................ 10-1
    10.1. CONTRIBUTIONS ......................... 10-1
    10.2. FUTURE WORK ......................... 10-2

REFERENCES ................................ R-1

APPENDICES ................................ A-1
   APPENDIX A ................................ A-1
   APPENDIX B ................................ A-5
   APPENDIX C ................................ A-8
# TABLE OF CONTENTS

## 1. INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1. MOTIVATION</td>
<td>1-1</td>
</tr>
<tr>
<td>1.2. RESEARCH OVERVIEW</td>
<td>1-3</td>
</tr>
<tr>
<td>1.3. METHODOLOGY</td>
<td>1-4</td>
</tr>
<tr>
<td>1.3.1. Review of Existing Systems</td>
<td>1-5</td>
</tr>
<tr>
<td>1.3.2. Informal Interviews</td>
<td>1-5</td>
</tr>
<tr>
<td>1.3.3. Initial Survey of 12 users</td>
<td>1-5</td>
</tr>
<tr>
<td>1.3.4. Formal Survey of 322 users</td>
<td>1-5</td>
</tr>
<tr>
<td>1.3.4.1. Population</td>
<td>1-6</td>
</tr>
<tr>
<td>1.3.4.2. Administration of Questionnaire</td>
<td>1-6</td>
</tr>
<tr>
<td>1.3.4.3. Respondents and Non-Respondents</td>
<td>1-6</td>
</tr>
<tr>
<td>1.3.4.4. Method of Analysis</td>
<td>1-6</td>
</tr>
<tr>
<td>1.3.5. Usage Data of 23 Users</td>
<td>1-7</td>
</tr>
<tr>
<td>1.3.6. Bookmark Files of 56 Users</td>
<td>1-8</td>
</tr>
<tr>
<td>1.3.6.1. Subjects</td>
<td>1-8</td>
</tr>
<tr>
<td>1.3.6.2. Data Collection</td>
<td>1-8</td>
</tr>
<tr>
<td>1.3.6.3. Method of Analysis</td>
<td>1-9</td>
</tr>
<tr>
<td>1.3.7. Survey of 27 Users</td>
<td>1-10</td>
</tr>
</tbody>
</table>

## 1.4. THESIS OUTLINE AND PREVIEW OF FINDINGS | 1-10

## 2. INFORMATION SPACES AND THE WORLD WIDE WEB | 2-1

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. COMPLEX INFORMATION SPACES</td>
<td>2-2</td>
</tr>
<tr>
<td>2.1.1. Information Overload</td>
<td>2-3</td>
</tr>
<tr>
<td>2.1.2. Pollution</td>
<td>2-4</td>
</tr>
<tr>
<td>2.1.3. Entropy</td>
<td>2-5</td>
</tr>
<tr>
<td>2.1.4. No Aggregate Structure</td>
<td>2-6</td>
</tr>
<tr>
<td>2.1.5. No Global View</td>
<td>2-7</td>
</tr>
<tr>
<td>2.2. THE WWW AS A COMPLEX INFORMATION SPACE</td>
<td>2-8</td>
</tr>
<tr>
<td>2.2.1. Overload on the Web</td>
<td>2-9</td>
</tr>
<tr>
<td>2.2.2. Pollution on the Web</td>
<td>2-10</td>
</tr>
<tr>
<td>2.2.3. Web Entropy</td>
<td>2-11</td>
</tr>
<tr>
<td>2.2.4. No Aggregate Structure in the Web</td>
<td>2-11</td>
</tr>
<tr>
<td>2.2.5. No Global View of the Web</td>
<td>2-12</td>
</tr>
<tr>
<td>2.3. USER BEHAVIOR IN THE WEB</td>
<td>2-13</td>
</tr>
<tr>
<td>2.3.1. Users Find Localities of Interest</td>
<td>2-13</td>
</tr>
<tr>
<td>2.3.2. Users Create References to Landmarks</td>
<td>2-14</td>
</tr>
<tr>
<td>2.3.3. Users Conceptualize Information Space Access through Metaphors</td>
<td>2-15</td>
</tr>
<tr>
<td>2.3.4. Users Build a Personal Information Space</td>
<td>2-16</td>
</tr>
<tr>
<td>2.4. PERSONAL WEB INFORMATION SPACES</td>
<td>2-17</td>
</tr>
<tr>
<td>2.4.1. Start Small and Build Incrementally (Prevent Information Overload)</td>
<td>2-17</td>
</tr>
<tr>
<td>2.4.2. Select Only Useful Items (Avoid Pollution)</td>
<td>2-18</td>
</tr>
<tr>
<td>2.4.3. Add Value Through Organization (Combat Entropy)</td>
<td>2-19</td>
</tr>
<tr>
<td>2.4.4. Attempt to Structure for Retrieval (Counteract the Lack of Structure)</td>
<td>2-19</td>
</tr>
<tr>
<td>2.4.5. Establish Personal View of Information Space (Compensate for No Global View)</td>
<td>2-20</td>
</tr>
</tbody>
</table>
3. INFORMATION SPACE BEHAVIOR ON THE WWW

3.1. USERS FIND LOCALITIES OF INTEREST 3-1
   3.1.1. Web Localities 3-1
   3.1.2. Revisitation via History Mechanism 3-2

3.2. USERS CREATE BOOKMARKS TO DISTINCT WEB LOCALITIES 3-3
   3.2.1. Paradox of Usage vs. Usefulness 3-4
   3.2.2. Bookmarks as a Launching Point for Web Exploration 3-6
      3.2.2.1. Navigation Patterns in WWW Browsing 3-7
      3.2.2.2. Bookmark Usage Functions 3-10
   3.2.3. Archival Use of Bookmarks 3-12
   3.2.4. Distinct Domain Names in Archive 3-15
   3.2.5. Delayed Value Judgment Problem: Forgetting to Bookmark a Page 3-17

3.3. USERS CONCEPTUALIZE WWW ACCESS THROUGH METAPHORS 3-17
   3.3.1. Marking/Tagging Information (Identification) 3-18
   3.3.2. Pulling Information into a Collection (Collection) 3-18
   3.3.3. Travelling Through Cyberspace (Movement) 3-19
   3.3.4. Remembering Episodes Related to Information (Episodes) 3-20

3.4. USERS BUILD A PERSONAL WEB INFORMATION SPACE 3-21
   3.4.1. Usage Statistics 3-21
      3.4.1.1. Experience Level 3-21
      3.4.1.2. Sessions per week 3-22
      3.4.1.3. Session Length 3-23
      3.4.1.4. Estimating Average Hours per Week 3-23
      3.4.1.5. Correlation of Usage Variables 3-24
   3.4.2. Use of Bookmark Functionality 3-24

4. PWIS #1: START SMALL AND BUILD INCREMENTALLY 4-1
   4.1. SIZE AND GROWTH OF ARCHIVE 4-1
      4.1.1. Small Bookmark Archive 4-1
      4.1.2. Incremental Growth of Archive 4-3
5. PWIS #2: SELECT ONLY USEFUL ITEMS 5-1

5.1. WHAT USERS BOOKMARK 5-1
5.1.1. Ontology of Bookmarked Sites 5-1  
5.1.2. Bookmarks to Organizational Home Pages 5-3  
5.1.3. Bookmarks to Search Engines 5-5  
5.1.4. Bookmarks to On-line News 5-8  

5.2. WHY USERS BOOKMARK 5-10  
5.2.1. Criteria for Bookmark Creation 5-10  
5.2.1.1. General Usefulness 5-11  
5.2.1.2. Quality 5-11  
5.2.1.3. Personal Interest 5-12  
5.2.1.4. Frequency of Use 5-12  
5.2.1.5. Future Use 5-13  

5.3. HOW BOOKMARKS ARE USED 5-13  
5.3.1. Reducing User Load 5-14  
5.3.1.1. Avoiding Managing URL Addresses 5-14  
5.3.1.2. Serving as a Memory Aid and History Mechanism 5-16  
5.3.2. Facilitating Navigation/Access 5-18  
5.3.2.1. Speeding Information Access 5-18  
5.3.2.2. Finding Web Information 5-18  
5.3.3. Collaborating/Publishing/Archiving 5-20  
5.3.3.1. Creating a Personal Web Information Space 5-20  
5.3.3.2. Authoring and Publishing Web Pages 5-20  
5.3.3.3. Collaboratively Using Web Information 5-21  

6. PWIS #3: ADD VALUE THROUGH ORGANIZATION 6-1

6.1. ORGANIZATIONAL METHODS 6-2  
6.1.1. Organizational Methods Defined 6-2  
6.1.1.1. No Organization 6-3  
6.1.1.2. Ordered List 6-4  
6.1.1.3. Set of folders 6-5  
6.1.1.4. Hierarchy of folders within folders 6-6  
6.1.1.5. Web Pages Recording Bookmarks 6-6  
6.1.1.6. External Applications 6-7  
6.1.1.7. Other Methods 6-8  
6.1.2. Experience Level of Organizational Methods 6-9  

6.2. ORGANIZATIONAL FREQUENCY 6-9  
6.2.1. Explaining Organizational Frequency 6-10  
6.2.1.1. Sporadic Filers 6-11  
6.2.1.2. Creation-Time Filers 6-11  
6.2.1.3. End-of-Session Filers 6-13  
6.2.2. Factors Influencing Organizational Frequency 6-14  
6.2.2.1. Cost Structure Matched to Current Usage Patterns 6-14  
6.2.2.2. Reduced Retrieval Time 6-15  
6.2.2.3. List Too Long to See 6-15  
6.2.2.4. Collaboration and Publishing 6-15  
6.2.2.5. Time to File Influences Frequency and Method 6-16
6.3. INFLUENCE OF SCALE ON ORGANIZATION 6-16
6.3.1. Aggregate Organizational Methods 6-17
6.3.2. User Characteristics by Scale 6-19
6.3.3. Factors of Scale Affecting Organization 6-21
  6.3.3.1. Users Find it Difficult to Organize a Large Number of Bookmarks 6-21
  6.3.3.2. Large Scale Requires More Organization 6-21
  6.3.3.3. Poor Scalability Forces Alternate Organizational Methods 6-22

7. PWIS #4: ATTEMPT TO STRUCTURE FOR RETRIEVAL 7-1
7.1. FILING AND RETRIEVING BOOKMARKS 7-1
  7.1.1. Selecting a Bookmark Retrieval Technique 7-1
  7.1.2. Retrieving from a Semantic Hierarchy 7-2
    7.1.2.1. "What folder did I put it in?" 7-2
    7.1.2.2. "What does this folder name stand for?" 7-3
  7.1.3. Creating a Semantic Hierarchy 7-3
    7.1.3.1. Naming Folders 7-3
    7.1.3.2. Forcing a Rigid Structure 7-4
  7.1.4. Managing a Stable Semantic Hierarchy 7-4
    7.1.4.1. Remembering the Classification System 7-5
    7.1.4.2. Significance Changes over Time 7-5
    7.1.4.3. "Why did I put it in this folder?" 7-6
  7.1.5. Filing Items in a Semantic Hierarchy 7-6
    7.1.5.1. Lack of an Immediate Filing Mechanism 7-6

7.2. BOOKMARK AS REPRESENTATION 7-7
  7.2.1. Problems with Bookmark Representation 7-8
    7.2.1.1. Non-Descriptive Titles 7-8
    7.2.1.2. Content Changes Without Warning 7-8
    7.2.1.3. Invalid URLs 7-9
  7.2.2. Customized Representation 7-9
    7.2.2.1. Changing the Name 7-9
    7.2.2.2. Annotating a Bookmark 7-10
    7.2.2.3. Assigning a Priority Rating 7-10
    7.2.2.4. Cropping a Semantic Icon 7-11

8. PWIS #5: ESTABLISH A PERSONAL VIEW OF THE WWW 8-1
8.1. MENTAL MAPS OF THE WEB 8-1
  8.1.1. Difficulties Drawing Mental Maps 8-1
  8.1.2. Respondents were Experienced Web Users 8-2
  8.1.3. No Two Mental Maps are the Same 8-3
  8.1.4. Abstract Landmarks 8-5
  8.1.5. Search Appears in the User's Personal View 8-8
  8.1.6. Me, My Computer and the Amorphous Cloud 8-9
  8.1.7. Bookmarks are Separate from the Web 8-11
8.2. HYPOTHESIS ABOUT PERSONALIZATION AND EVOLUTION OF GESTALT VIEWS OF THE WEB 8-12
9. PERSONAL INFORMATION SPACES BECOME COMPLEX

9.1. COMPLEXITY PROBLEMS IN A PERSONAL WEB

9.1.1. Too Many Bookmarks (Information Overload)
9.1.2. Polluted Bookmark Archive (Pollution)
9.1.3. Required Maintenance (Entropy)
9.1.4. Lack of Coherent Structure (No Aggregate Structure)
9.1.5. Difficulty in Visualizing Bookmarks (No Global View)

9.2. CURRENT SOFTWARE SYSTEMS

9.2.1. Folder System
9.2.1.1. Arranging Folders in the Hierarchy
9.2.1.2. Placing an Item in a Folder
9.2.1.3. Being Forced to use a Semantic Hierarchy
9.2.2. Browser-Bookmark Integration
9.2.2.1. Integrating Multiple Bookmark lists
9.2.2.2. Pull-Down Bookmark Menu

10. CONCLUSIONS

10.1. CONTRIBUTIONS
10.2. FUTURE WORK

REFERENCES

APPENDICES

APPENDIX A
APPENDIX B
APPENDIX C
1. INTRODUCTION

1.1. MOTIVATION

The World Wide Web (WWW) is a large hypertext system connecting vast resources of distributed, heterogeneous information and interactive services (Berners-Lee, et al. 1992). The Internet provides the underlying network infrastructure for this multimedia computing platform. The WWW is currently the definitive electronic medium for finding multimedia information and accessing interactive services.

In the huge, unstructured, polluted information space of the WWW (see 2.1., Complex Information Spaces), users need ways to keep track of valuable information once they have found it. Today’s Web browsers enable users to bookmark a Web page for later reference. This personal URL list, favorite’s list, hotlist, or bookmark file gives users a means of storing hypertext links to valuable Web sites. The screen shot below illustrates the bookmark pull-down menu in a Web browser:

Users routinely judge the relevancy of a Web page and decide whether to add it to their personal archive or not. As users collect bookmarks over time, they build up a small,
personal Web information space (see 2.4.). This serves as a known place of valuable sites with a familiar structure, tuned to their particular needs.

Bookmarks are essential for WWW browsing because they facilitate the locality of reference behavior inherent to users of a large information space and they reduce the cognitive and physical burdens of hypertext navigation. In two recent surveys totaling over 21,000 Web users, hotlists or bookmarks were identified as a WWW strategy for locating information by over 80% of respondents (Pitkow and Recker 1995), (Pitkow and Recker 1996). The use of bookmarks was cited slightly more than any other strategy including: entering queries in a search engine, referencing an index page, entering a known URL, and foraging for a specific page by traversing links (see 2.2.). In addition, over 92% of users surveyed have a bookmark archive and over 37% of users have more than 50 bookmarks (Pitkow and Recker 1995).

A user typically collects less than five bookmarks during each Web session (see 4.2.1.). As this user initiates more Web sessions and spends more time in this information space, he/she accumulates a larger collection of links to valuable sites. The size of a typical user’s bookmark archive increases over time (see 4.2.2.).

As existing Web users spend more time on the Web during the years to come their archives will most likely increase in size. New services and valuable resources appear on the Web daily. As more compelling content is available, users will need ways to manage a larger bookmark archive.

Users encounter significant problems as the size of their archive increases. For example, when the archive grows beyond approximately 35 items, they cannot be viewed with the pull-down menu because they run off the end of the screen (see 9.2.2.2.). In addition, in both GVU surveys, the process of organizing information found on the Web
was one of the top three problems reported with the Web (Pitkow and Recker 1995), (Pitkow and Recker 1996).

The demand for better tools is evident by the recent flurry in new commercial bookmark management tools (EastGate 1996), (Netscape 1996b), (DocuMagix 1996). Yet, there has been no detailed user study of WWW bookmarks to motivate the design of these systems. The WWW is in its infancy, and in order to build better systems which effectively meet user needs, we must understand the process, methods of use, and problems of browsing the Web and using bookmarks. This research is a detailed study and analysis of the human factors of bookmark use.

1.2. RESEARCH OVERVIEW

Web users collect bookmarks in order to create what we call a personal information space. This thesis is a broad empirical study of the creation, retrieval, and organization of a user’s bookmark archive, which becomes a personal information space. The hypothesis of this thesis is that users build a personal information space of bookmarks in order to mitigate the problems of finding, retrieving and managing information on the World Wide Web.

Personal information spaces have been studied in a variety of contexts (Burton 1987), (Barreau 1995), (Whitaker and Snider 1996). Malone (1983) studied how people manage paper work in the office. The psychology of personal information space management has been addressed by Lansdale (1983) and Jones (1986). The management of personal information on the computer desktops was analyzed by Fitzmaurice, Baecker and Moore (1994). Nardi and Barreau (1995) studied how people organize their computer desktops.
Although personal information spaces have been studied in a variety of contexts, there has been very little work on the use of bookmarks. In fact, there has been very little empirical research on the human factors of the Web. Catledge and Pitkow (1995) performed an initial quantitative study of Web use. Tauscher (1996) studied the use of history mechanisms. Berghel (1996) uses anecdotal evidence to describe user problems. Xerox PARC is developing a 3D personal information management system for the Web (Card, et al. 1996).

1.3. METHODOLOGY

We used a variety of empirical methods to understand how people manage a personal information space on the Web. We began with a review of the existing Web browsers and bookmark management tools because our initial goal was to develop a new Web browser with integrated bookmark management support. We generated such a large design space that we began to study the use of bookmarks. An initial set of informal interviews and a small survey motivated us to continue the user study. This helped us design a questionnaire and perform a formal survey of 322 users. The comments we collected from users formed the basis for our hypothesis about a personal Web information space (see 1.2).

We continued our user studies to test these five principles of a personal information space. A quantitative analysis of the survey was performed. We collected and studied the bookmark files of 56 users. We analyzed the usage data of 23 users over 6 weeks. Finally, we surveyed 27 users in a mental maps study.
1.3.1. Review of Existing Systems

The first step in this study was a review of the different Web browsers and bookmark management systems available. We identified design differences and evaluated the bookmark functionality provided. We brainstormed an extensive list of design ideas to prototype a new system. The result was a 60 page document of design ideas for a new bookmark management system. The number of possible avenues motivated us to change our focus to understanding the human factors of bookmarks in order to inform the design of potential functionality.

1.3.2. Informal Interviews

We then conducted a series of informal user interviews to identify a short list of problems and methods of use. This served as the foundation for an initial survey.

1.3.3. Initial Survey of 12 users

We surveyed 12 users on their use of bookmarks. The results showed serious flaws in our survey design and lead us to a re-design. The next survey was based on the user feedback and a better understanding of the process of using bookmarks.

1.3.4. Formal Survey of 322 users

We surveyed 322 participants at the Internet: Beyond the Year 2000 conference. This survey (see Appendix A) laid the foundation for our user study and directed our research. The structure of this thesis is based largely on our analysis of the qualitative and quantitative results. For example, the large number of comments users entered in the essay sections provided many surprising qualitative results. We used these comments to
select the specific problems to study in the thesis. Our five basic principles of a personal
information space emerged from this qualitative analysis.

1.3.4.1. Population

We distributed the questionnaire at The Internet: Beyond the Year 200
Conference. This population consisted of a mix of University of Toronto students,
faculty, and staff with a variety of professionals from industry. Most participants had an
expressed interest in the WWW and computer technology.

1.3.4.2. Administration of Questionnaire

Approximately 450 questionnaires were distributed at the beginning of a speaker
session. The audience was asked to fill out the questionnaire and drop it into a box at the
exit when they leave. Subjects were instructed to identify themselves by their ticket
number. All results would be held strictly confidential. A drawing took place at the end
of the second session and five Internet related books were presented to the winner.

1.3.4.3. Respondents and Non-Respondents

322 completed survey forms were collected from the 450 distributed. Of the
surveys completed, 200 people or 62%, were filled in correctly according to the
instructions on every question. 76.5% of the 322 surveys were male and 23.5% were
female, consistent across all sizes of bookmark archives. Age distribution was close to
normal and centered about 35 years old.

1.3.4.4. Method of Analysis

We analyzed both the quantitative and qualitative answers. Comments were
categorized based on the comment type and user characteristics. Repetitions of
comments were counted. Comments from users appear throughout the thesis. These are denoted by “an italic statement in quotes.”

We began our quantitative analysis by generating histograms of each question on the survey. We performed a correlation and ANOVA on the user’s experience, time per week on the Web, and session length (see 3.4.1.). A similar analysis was performed on the use of bookmark functionality (see 3.4.2.). Both correlations help verify the integrity of the survey. For example, users with more bookmarks tend to use the “Create New Bookmark” function more often as they should (see 3.4.2.).

We generated a size distribution of bookmark archives and compared this with the size of bookmark archives in the GVU survey (see 4.2.1.). Our survey results were similar to the GVU study.

When analyzing the organizational habits of users we identified two methods we did not anticipate when writing the survey. We generated a list of different organizational methods and created composite tables of user characteristics for each (see 6.2.1.). In 6.2.2 we correlated the experience level of organizational methods. In sections 6.3 and 6.4 we performed a similar analysis on filing habits and created composite user characteristic statistics.

1.3.5. Usage Data of 23 Users

We analyzed the usage data of 23 users over a 6 week period, provided by the University of Calgary (Tauscher 1996). This data of actual use of the Mosaic WWW Browser. Tauscher instrumented a Mosaic WWW browser and logged each user action during the six week period.
We analyzed the URL vocabulary graphs to see what hypertext navigation patterns emerged after the use of a bookmark. We identified a set of patterns, categorized them based on bookmark use, and created a composite distribution of emergent hypertext patterns following the use of a bookmark (see 3.2.4.). Appendix C has a table listing our analysis of each user in the Calgary study.

1.3.6. Bookmark Files of 56 Users

We collected 56 bookmark files through email. These files were analyzed to show the growth a user's archive over time, how often users retrieve bookmarks from their archive's, the types of Web pages they bookmark, and their use of folders in organizing a bookmark file.

1.3.6.1. Subjects

Subjects were all educated and had used the Web. 70% of subjects were from the Knowledge Media Design Institute (KMDI) at the University of Toronto. This consists of students and faculty members from a cross-disciplinary group, distributed in different departments throughout the university. 20% of subjects were from the graduate students in the Computer Science Department. The remaining 10% were professionals from industry.

1.3.6.2. Data Collection

We mailed an electronic message to users on the KMDI list server and the Computer Science Department. We collected email responses for two weeks and received 56 bookmark files in text format. The files were compressed and password encrypted for security.
A parser was written to extract each bookmark from the file for analysis. The table below lists the data for each bookmark:

<table>
<thead>
<tr>
<th>Field</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL</td>
<td>&lt;type&gt;://&lt;domain name&gt;/&lt;file name&gt;</td>
</tr>
<tr>
<td>Title</td>
<td>string</td>
</tr>
<tr>
<td>Date/Time of Creation</td>
<td>DD/MM/YY HR:MN:SS</td>
</tr>
<tr>
<td>Date/Time of Last Visit</td>
<td>DD/MM/YY HR:MN:SS</td>
</tr>
<tr>
<td>Date/Time of Last Modification</td>
<td>DD/MM/YY HR:MN:SS</td>
</tr>
<tr>
<td>Folder Location</td>
<td>&lt;leaf node in tree hierarchy&gt;</td>
</tr>
</tbody>
</table>

This data was used as the basis for our analysis.

1.3.6.3. Method of Analysis

A time-series analysis was performed on the bookmark files. We calculated the growth rate of a user's bookmark file over time (see 4.2.2.). In section 3.2.2. another time-series analysis shows how often users access items from their bookmark file. Both time series analyses were done on 50 of the 56 bookmark files because six files were missing date information.

We calculated how often users bookmark the same Web site or domain name more than once in section 3.2.3. We identified the set of distinct bookmarks and compared this the total number of bookmarks in each user’s archive.

In section 5.1 we created an ontology of bookmarked sites by categorizing the 30% most popular sites. A more detailed analysis of organizational home pages appears in 5.1.2. We compared bookmarks to search engines in section 5.1.3. A Chi-squared analysis was used to compare different types of search engine bookmarks.

Finally, in 6.4.1. we performed a regression analysis on the use of folders in bookmark archives.
1.3.7. Survey of 27 Users

Our final step was to perform another survey of users to identify how they conceptualize the WWW. The questionnaire was distributed at the beginning of a Toronto Web Society meeting. Respondents had 5 to 10 minutes to sketch their personal view of the Web (see Appendix D). The population had been using the Web for an average of 20.8 months (7.4 months standard deviation). We grouped users together based on their mental maps of the Web and extracted characteristics of each (see 8.1).

1.4. THESIS OUTLINE AND PREVIEW OF FINDINGS

Web users suffer from many problems because of its size and complexity. In chapter 2 we describe the concept of a complex information space which we developed through our study of the Web. Chapter 2 lays the theoretical foundation for the thesis. It defines both a complex and a personal information space based on a literature review and the key findings of our study. Chapter 2 also identifies some fundamental behaviors found in users of a complex information space.

In chapter 3 we describe four basic behaviors we identified in Web users. They find localities of interest, create bookmarks to return to those localities, conceptualize the use of bookmarks through metaphors, and create a personal information space out of their bookmarks.

Users create a personal information space in order to counteract the five inherent problems of a complex information space in the following ways:
1. They prevent information overload by incrementally building a small archive.
2. They avoid pollution by selecting only useful items and creating a known source of high value.
3. They reduce entropy through maintenance, although they organize only when necessary.
4. They attempt to add structure by cost-tuning their information environment.
5. They compensate for the lack of a global view by creating their own personal view.

Chapters 4 through 8 describe each of these five principles of a personal Web information space based on an extensive analysis of our empirical data of Web bookmark use.

In Chapter 9 we consider how successful users are in creating, organizing and maintaining a personal information space. Here we see that a personal information space can grow to exhibit the characteristics of a complex one, unless the user expends effort to maintain the integrity, structure, and organization of the bookmark archive.

Finally, in Chapter 10 we conclude by describing directions for future research and summarizing the most salient contributions of this thesis.
2. INFORMATION SPACES AND THE WORLD WIDE WEB

This chapter describes a complex information space as having five inherent properties: information overload, pollution, entropy, lack of structure, and no global view. We identified these five problems in the Web through our empirical study (see 2.2.). The Web is a complex information space and its users have four characteristic behaviors (see 2.3.): they find localities of interest, create references to landmarks, conceptualize information access with metaphors, and build a personal information space from their collection of references. The resulting personal information space (see 2.4.) has five basic properties paralleling those of a complex information space:

<table>
<thead>
<tr>
<th>Complex Information Space</th>
<th>Personal Information Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information Overload</td>
<td>1. Start Small and Build Incrementally</td>
</tr>
<tr>
<td>2. Pollution</td>
<td>2. Select Only Useful Items</td>
</tr>
<tr>
<td>3. Entropy</td>
<td>3. Add Value Through Organization</td>
</tr>
<tr>
<td>4. No Aggregate Structure</td>
<td>4. Attempt to Structure for Retrieval</td>
</tr>
<tr>
<td>5. No Global View</td>
<td>5. Establish Personal View of Information Space</td>
</tr>
</tbody>
</table>

Section 2.4. previews chapters 4 through 8 which take an in-depth look at each of these issues and explain Web user behavior based on our empirical study.
2.1. COMPLEX INFORMATION SPACES

A complex information space is a set of information units stored in a medium which provides people retrieval access. Below is a working definition:

<table>
<thead>
<tr>
<th>Definition of a Complex Information Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>A complex information space (C) stores a total number (N) of information units in a medium (M) of storage. A user (X) has relevant information units (R) in the information space according to the scope of X's information foraging goals. X suffers from information overload because C is saturated with a large quantity (N) of information units and C is polluted with low quality information units. C naturally progresses toward disorder according to the principle of entropy. C has no aggregate structure which organizes all the information units into a coherent whole. The user has no global view of the information space from which to forage for relevant information units.</td>
</tr>
</tbody>
</table>

We encounter complex information spaces in our everyday lives. It may be a physical place like the main office of a business which has memos, reports, periodicals and books. The office accumulates paperwork and becomes disordered as people misplace items. Individuals must expend effort on a regular basis to organize the information units into a useful structure. The people working in the office need to know where to look to find items.

A complex information space can readily emerge in the digital medium. For example, a user’s file system becomes overloaded with too many files. As data files become out of date or lose importance, the file system gets polluted with irrelevant files. The user has to deal with the resulting disorder by manually imposing an aggregate structure (e.g. directory tree, folders on the desktop). Items are grouped into folders and viewed in small collections with the user typically having no global view.

We assign people in organizations the special task of reducing the complexity of an information space. One example is the editorial review process of selecting papers for the proceedings of a conference. Papers are submitted from the scientific community. Reviewers fight pollution by selecting papers which meet publication criteria. The selected papers are organized into a bound volume with a table of contents and index. These are created to provide structure and enable the reader to find a given article.
Complex information spaces are characterized by five defining properties:

*information overload, pollution, entropy, no aggregate structure, and no global view.*

### 2.1.1. Information Overload

*Complex information spaces contain a large quantity of information units that exceeds the user’s information threshold and causes information overload.*

Complex information spaces contain a plethora of documents on a wide range of topics. In this environment the user is generally more concerned with the time and effort required to locate a piece of information rather than getting access to enough information. This *scarcity of attention* forces authors to create content that will vie for the user’s time and attention.

Humans have physiological tolerance levels to information. These threshold levels are described in biological systems at the cell, organ, organism, and group of organisms level by (Miller 1981). If the volume of information exceeds the capacity of our senses, then our comprehension declines.

"As information inputs impinge more and more rapidly on a system, they eventually overload its capacity to transmit information. Ultimately breakdown of efficient information processing occurs..."

(Miller 1981)

The human threshold for comprehending information is characterized by the graph below.
Keyes et al. (1989) showed that comprehension is positively correlated to the quantity of information only to a certain tolerance level. When the user is exposed to quantities of information beyond this point, comprehension declines rapidly.

Users tend to rely on resources that summarize and simplify information and direct them to more specific resources (e.g., the front page of a newspaper). Since a user cannot compare all of the documents which are available, the user must rely on a trusted information provider who filters out irrelevant information.

Even with such services, users often suffer from information overload. The quantity of available information is so large that it obscures the desired documents. Users are forced to select precise information foraging goals and narrow the scope of their interest area. They begin by sampling general resources which provide an overview (e.g., encyclopedias), and then search for specialized documents with more depth and less breadth (e.g., journal articles). This specialization process works in theory, but in practice users are still overloaded with too much information.

2.1.2. Pollution

*Complex information spaces are typically polluted with erroneous, redundant and low quality information units which obscure the information units of high value.*

A complex information space is usually polluted with information units that are not relevant to the user. There is much erroneous, redundant and low quality information which obscures the documents of high value (December 1995). Users rely on information providers to maintain current information and purge redundant items. The editorial and peer review process are used to ensure well defined levels of quality and correctness.

Users in a complex information space attempt to avoid pollution by seeking information of high value and avoiding superfluous items which would distract them. The scope of a user’s information foraging goals helps to define the fraction of available items which are useful. The huge quantity of documents (saturation) and the wide range of quality (pollution) in the complex information space force users to make precise
queries. Salton and McGill (1983) discusses this type of information retrieval. The number of information units which a user finds relevant is based on his/her task and related information foraging goals. Waterworth and Chignell (1989) used specificity as part of a model of user behavior while exploring information. Specificity measures how precise the user’s goals are for finding information.

2.1.3. Entropy

*A complex information space tends toward disorder unless effort is expended to organize the information units.*

Entropy measures the amount of disorder in a system. A system progresses toward disorder unless energy is expended to impose order on the individual elements of the system. The connection between entropy, originally the second law of thermodynamics, and information lies in the amount of order which organizes information units in a comprehensible or logical way. The entropy law states that

"when a system containing a large number of particles is left to itself, it spontaneously assumes a state of maximum entropy - that is, it becomes as disorderly as possible." (Young 1987)

Disorderly states are inherently more probable than orderly ones because there are significantly more possible disorderly states than possible orderly states. In a complex information space, users must expend effort to fight this natural progression toward chaos.

Returning to the example of an office, we see that knowledge workers expend significant effort to organize their information space. The books, periodicals and resources in the office must be catalogued, properly shelved, and organized on a continual basis. As the office grows, new items must be categorized into the proper place within the office’s ad-hoc or formal classification system. Without this continual effort, the office enters a state of disorder and this condition persists until sufficient energy is expended to fight entropy.
2.1.4. No Aggregate Structure

A complex information space often has no aggregate structure which organizes it into a coherent whole.

A complex information space must be carefully structured so that users can find documents and comprehend the contents of an item once it has been found. We define structure in an information space at three basic levels.

<table>
<thead>
<tr>
<th>Structure in a Complex Information Space</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit Structure</strong>: The structure within an individual document or information unit for the purposes of increasing the reader’s comprehension of the content of the document.</td>
</tr>
<tr>
<td><strong>Locality Structure</strong>: The structure connecting closely related documents (e.g. footnotes, hypertext links, order in a pile of paper) within the same information locality of reference.</td>
</tr>
<tr>
<td><strong>Aggregate Structure</strong>: The structure connecting different localities of information or groups of documents.</td>
</tr>
</tbody>
</table>

A document must have a well defined structure (*unit structure*) in order to facilitate the reader’s comprehension. From a cognitive science point-of-view, comprehension is the construction of a mental model that represents the objects and semantic relations described in the text (van Dijk and Kintsch 1983). The structure of a document enables readers to break the concepts down into manageable pieces as well as establish global coherence of the main ideas of the document (Mayer 1984).

An information space consists of localities of related information units. The *locality structure* is based on the links and substructures which connect related documents. The locality has a thematic organization which enables users to comprehend the main ideas.

Locality structure in hypertext can make navigation easier because users do not have to remember the location of nodes (Rouet, et al. 1996). Locality structure can also be found in a simple pile of papers (Mander, Salomon and Wong 1992). The pile is a locality of related documents within the information space of a person’s office. The order of the papers in this pile creates its locality structure. In *Information Anxiety*, Wurman (1989) identified five different methods of organizing information: (1) alphabetical, (2) by time, (3) by location, (4) along a continuum or magnitude, and (5) by category. These
methods are used to create structure for groups of related documents in a locality of information.

The aggregate structure of an information space connects distinct localities of information into a whole. All complex information spaces lack aggregate structure, but they may contain well structured localities and structured information units. A coherent aggregate structure links together separate localities of information units so that users have a systematic and logical method of finding a relevant document.

Aggregate structure generally connects the work of different authors. A document or group of documents can be structured by a small team of people into a coherent whole. Authors work to create a well organized structure through the editorial review process. But moving to the aggregate level, many people are involved. In addition, it usually requires moving from semantic structuring to a more systemic method based on computational means of automatically generating structure (e.g. Salton and McGill 1983, Everitt 1990). It is impractical for people to manually organize and manage a complex information space once its size increases beyond our threshold to process information.

2.1.5. No Global View

In a complex information space, there does not exist a complete map of the entire information space which provides a holistic, systemic view of all information units.

One defining characteristic of a complex information space is that there is no way to see everything in it. There is no high-level, global view of all its information units because of the immense quantity of information and the lack of an effective aggregate structure.

We use our perceptual system to build up mental models of where we are and where to go next (Levine 1981). Human information use involves cognitive integration of material beyond the level of just words, sentences and pictures (Dillon and Schaap 1996). Users of an information space have a limited view at any one time. Nielsen
(1990) found that people get disoriented when navigating large hypertext networks. Their restricted view causes them to "get lost in hyperspace" (Edwards and Hardman 1989).

A tool which helps people assimilate, process and understand information is known as a "cognitive prosthesis" (Wright 1991). For example, a map or a table of contents reduces the user's cognitive load because it abates the load on human long term and working memory, summarizing the information about the structure and organization that would otherwise have to be remembered (Chignell 1996). For example, Monk (1989) found that when users are burdened with finding their way about a hypertext structure, they often get distracted from their primary task. He showed that a user's performances was significantly improved when an overview map of the structure was provided. A hierarchical representation may help the reader build up a mental map of the hypertext (Dillon, et al. 1990).

2.2. THE WWW AS A COMPLEX INFORMATION SPACE

The WWW is a complex information space. In this section we show that the Web exhibits all five characteristics of a complex information space. The empirical data is based on our survey (see Appendix A) and related empirical studies of the Web.

Although the Web provides users with great access to information, users have many problems finding information and building knowledge from it. Two recent surveys by GVLU (Pitkow and Recker 1995, 1996) asked users to identify obstacles to using the Web. They identified access speed, finding information, and organizing retrieved information as the three most common problems among users. The table below orders the problems of WWW users:
### General User Problems with WWW

<table>
<thead>
<tr>
<th>Problem</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>Takes too long to view/download pages</td>
</tr>
<tr>
<td>Find</td>
<td>Not being able to find a page I know is out there</td>
</tr>
<tr>
<td>Organize</td>
<td>Not being able to organize well the pages and information I gather</td>
</tr>
<tr>
<td>Return</td>
<td>Not being able to find a page I once visited</td>
</tr>
<tr>
<td>Visualize</td>
<td>Not being able to visualize where I have been and where I can go</td>
</tr>
<tr>
<td>Lost in hypertext</td>
<td>Not being able to determine where I am</td>
</tr>
<tr>
<td>Software</td>
<td>My Web browser is poorly designed/does not work well</td>
</tr>
</tbody>
</table>

The graph below illustrates the results.

![Problems with WWW](chart.png)

In both studies, at least 69% of users complained about the slow speed of the Web, 34% have problems finding information, and 25% of all Web users surveyed have problems with organizing bookmarks and information gathered. 28% also complained about not being able to return to a page once visited. The last three problems, visualization, getting lost in hypertext, and software were noted by only 13%, 6%, 5% of respondents respectively. In the remaining parts of this section we will examine the underlying properties of the WWW which cause many of these problems.

### 2.2.1. Overload on the Web

The WWW continues to grow at a phenomenal rate. Over 30 million people are using the Internet today, and it is growing at 10 to 15 percent per month (Netscape...
The Web consists of over 230,000 sites run by a vast array of companies, universities and government institutions spanning every developed country in the world (Gray 1996). Over 9.5 million hosts were recorded in a recent domain name survey attempting to identify every uniquely reachable computer connected to the Internet. (Network Wizards 1996).

Users suffer from information overload when using the Web. One user in our survey wrote, "There are too many places that I visit to which I would like to potentially come back, but I don't know how to use existing bookmark support to do it conveniently." Another user described the process of locating relevant information on the Web at the intra-page level, as distinct from simply finding a specific page on the Web: "one page has information of interest to too many people and sometimes getting at only what is of interest to me becomes problematic." On the Web, where information is abundant, the user's time is in scarce supply. Another user described the role of time and the Web: "The biggest problem is just too little time to use all the interesting information I find."

2.2.2. Pollution on the Web

The Web is an extremely polluted information space. Users must wade through pages of wildly different types, topics, designs, and formats. Multimedia adds to the incredible variety of Web sites. Many sites do not incorporate graphic design and layout principles.

The Web consists of pages by a wide variety of authors and new people are writing Web pages every day. There is a tremendous variety of writing styles and quality of Web pages. For example, one user wrote: "I dislike so many different types of sites." In addition, pages vary in focus, quality and function (e.g., indexes to Web resources, interactive services, on-line publications, pages of specific content).
2.2.3. Web Entropy

The Web progresses toward chaos according to the general principle of entropy. As the Web grows, the amount of entropy in this complex information space increases toward states of more and more disorder. The Web’s growth is aided by the low barriers of entry to Web publishing (e.g., HTML authoring tools and inexpensive ISPs). More people have the ability to add pages to the WWW. When an author does this, it becomes part of the complex information space. These pages are often added without knowledge of all the related information which is available. It is difficult to even compare all the available Web pages on a single topic domain. Users struggle to find ways to deal with the chaotic nature of the Web.

2.2.4. No Aggregate Structure in the Web

The WWW lacks a coherent structure at the aggregate level. There does not exist a consistent, standard methodology for organizing distinct Web sites. Although two different Web site may be well structured according to design and layout principles, there is no structure between them. There does not exist a mechanism which clearly shows the interrelationship between localities on the Web.

Hypertext links connect two elements but do not provide structure by themselves. An author must carefully design a Web site and judiciously place links which connect related pages. A locality structure emerges which helps the user relate each subsection to the organization of the rest of the site (e.g., Kahn 1996 suggests visual cues for local and global coherence on the Web). This structure is consistent, supports a theme, helps the user abstract the key concepts, and enables the user to navigate between pages.

The authoring process is essential to creating a structure within an information unit and a locality within the complex information space. The Web has no aggregate structure because its size has grown to exceed our ability to categorize all the available
pages. Yahoo (Filo 1994) currently provides the most comprehensive ontology of Web sites, but this structured index cannot keep up with the growth of the Web.

Since the Web has no aggregate structure, users turn to search engines to locate Web pages. For example, Lycos receives an average of 4.5 million hits per day (Lycos 1996). Search engines index Web pages computationally and return the results of a search query on the index. The results of a search query often contain links to different Web sites which meet the user's search criteria. Although search engines are useful they do not provide a structured view of the Web beyond a single search query.

Efforts are underway to extract usable structure from the Web (Pirolli, Pitkow and Card 1996) at the site and Web locality level. These methods compute predictive selection based on text similarity, page metadata and prior usage patterns. The use of spreading activation to predict needed information is a means of providing structure within a Web locality. In addition, the artificial life research community is beginning to address the problem of searching in localities of Web documents (Menczer, et al. 1996). Yet, both of these projects focus on generating locality and not aggregate structure. Even if successful they will not create an aggregate structure for the Web.

2.2.5. No Global View of the Web

There does not exist a single, global view of the entire Web. The scale is so large that any such visualization would be so abstract that it may not be useful for navigation. Users need a high level visualization of the Web in order to relate their current location with their global understanding of the Web.

Web visualization falls into two broad categories: (1) Web site visualization, and (2) Web usage visualization. The topology of a Web site can be visualized through a variety of means. Cockburn and Jones (1996) demonstrate ways to visualize sub-sets of the entire information space. InContext (1996) has created a Web site visualization tool with concentric rings surrounding a page where distance is shown by each progressive
circle. The Internet publishing industry is filled with books which provide a structured view of the Web (e.g., Internet Yellow Pages by Hahn, 1994).

The local topology of the Web and a user's browsing history can be visualized by a graph structure. For example, the Hy+ system displays a user's history and the possible paths which lie ahead (Mendelzon 1996).

![Hy+ Web Visualization Tool (Mendelzon 1996)](image)

2.3. USER BEHAVIOR IN THE WEB

In this section we discuss four human behaviors which we have observed in Web users. (1) Users re-visit localities of relevant information. (2) They establish references back to those information patches. (3) Users have metaphors for accessing information which help them articulate their mental models of the information space. (4) They create a personal information space from the references they have collected.

2.3.1. Users Find Localities of Interest

*Users re-visit dense concentrations of related information units.*

As people forage for information in a complex information space they frequently discover useful information patches (Pirolli95), small areas of related and highly relevant information. Users seek out these dense concentrations of valuable information. The expected value from accessing an information patch is very high in comparison to the rest of the surrounding information space.
Locality of reference behavior is well known in the areas of computer desktop window systems (Card, Pavel and Farrell 1984) and paper-based office management (Malone 1983). Waterworth and Singh (1994) extends the idea of locality of reference to include a spatial metaphor of the information island.

In section 3.1 we show how Web users revisit localities of interest. Users revisit Web pages 58% of the time and a set of navigation patterns emerge over time (Tauscher 1996).

2.3.2. Users Create References to Landmarks

Users establish references to landmarks which identify distinct localities of information.

Users need a means of getting back to a locality of high-valued information units. They create references to return to groups of relevant documents. References tend to point to distinct landmarks which identify a locality.

People use landmarks in cities to guide them to a destination. Landmarks enable them to establish a link between their current location and their global coherence of the entire city. Thus, a landmark is the bridge between the streets a city dweller can see at any one moment and his/her cognitive map of the city. A landmark acts as a frame of reference in navigation. In Image of the City, Lynch found that people rely heavily on landmarks for their guides in cities (Lynch 1960). Landmarks have a clear form, contrast with their environment, and have some prominence of spatial location. The use of landmarks in hypermedia navigation has been studied by Valdez, Chignell and Glen (1988). They suggested several methods of identifying landmarks in hypertext and developed a method for assessing the quality of a landmark.

Bernstein first described the concept of a reference to a locality in hypertext. He called these "book marks" as shown below:

"Book marks let the reader modify a complex and potentially intimidating document in a way that makes the document simpler and more personal."(Bernstein 1988)
They allow users to quickly return to a familiar location. When users get disoriented navigating through a complex information space, they tend to return to a known starting point (Barret 1988).

Web users create references (often called bookmarks) which point to distinct Web localities. In section 3.2 we explain why bookmarks are only used 3% of the time, yet are "essential" to Web users. We shall see that users create bookmarks to localities of related Web pages and then use other navigational techniques to maneuver within that locality.

2.3.3. Users Conceptualize Information Space Access through Metaphors
Users have metaphors for accessing information which help them articulate their mental models of the information space.

Metaphors underlie our interaction with computer systems, our language and our everyday actions. Lakoff and Johnson (1980) argue the importance of metaphor

"metaphor is pervasive in everyday life, not just in language, but in thought and action." (Lakoff and Johnson 1980)

Information space users have conceptual models based on the metaphors suggested by the presentation media and the user interface.

Metaphors are used to convey abstract concepts in a more familiar and accessible form. We draw on prior knowledge in order to better understand a new domain.

"Metaphors function as natural models, allowing us to take our knowledge of familiar, concrete objects and experiences and use it to give structure to more abstract concepts." (Erickson 1995)

Our knowledge of a familiar domain, its elements and their relations to each other, is mapped onto the new domain. For example, we talk about understanding an idea as seeing an object. This thesis takes a closer look at bookmarks than previous research. The goal is to shed some light on how people use WWW bookmarks. Other researchers will undoubtedly take a different approach in order to show a different perspective. Our goal is to make apparent the idea of a personal Web information space.
Metaphors are essential for understanding user behavior because users describe their experiences in a complex information space through metaphors. They have incomplete mental models and multiple metaphors of the complex system (Norman 1993).

We extracted four basic metaphors from user comments about the Web in an empirical study (see 3.3.).

<table>
<thead>
<tr>
<th>Complex Information Space Metaphors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identification</strong>: users mark or tag information units so as to distinguish them from others and make them apparent for later recognition.</td>
</tr>
<tr>
<td><strong>Collection</strong>: users collect information units while using an information space and store them in a separate container.</td>
</tr>
<tr>
<td><strong>Movement</strong>: users move from location to location while navigating through the information space and revisiting localities of relevant information units.</td>
</tr>
<tr>
<td><strong>Episodes</strong>: users recall previous episodes associated with information units which they identified/collected/visited at some time in the past.</td>
</tr>
</tbody>
</table>

These concepts help users deal with the problems inherent in an information space. For example, the "inside-out problem" of hypertext navigation (Nielsen 1990) is generally described as disorientation according to the metaphor of user as traveler (i.e., movement). Section 3.3. will describe these four user metaphors for WWW bookmarks and browsing.

2.3.4. Users Build a Personal Information Space

*Users build a cost-tuned personal information space from the bookmarks they have collected.*

People collect information from their environment in order to create a structure which is cost-tuned for doing work. The goal is to expend the least amount of work building up a structure for their information that supports fast retrieval of relevant information. Card describes an information environment that is cost-tuned to doing information work,

"a small amount of information is organized to be available at very low cost, larger amounts are available at moderate costs, large amounts are kept at high cost. By so doing, they capitalize on the locality of reference and the activity is speeded considerably. A routine example would be a typical (ideal) office where a small amount of information is kept available on the desk; moderate amounts of information, moderately available in nearby files; and large amounts, slowly available are kept in a library down the hall. Users constantly rearrange the environments to tune the relative costs of the information, so as
to make them efficient. And if they don’t, they suffer accordingly." (Card, Robertson and York 1996)

People try to optimize their environment so they can spend the least amount of time and energy to get the greatest returns. The returns are based on the frequency of use and value of an information unit. Frequently used items should be kept close to enable fast retrieval, while less used items are stored further away. The result is a personal information space which is tuned to the individual needs of its users.

In section 3.4 we show how Web usage is correlated with bookmark use and organization, and how often available bookmark functionality is used.

2.4. PERSONAL WEB INFORMATION SPACES

A personal information space results as a user selects information units or references to them from the complex information space.

<table>
<thead>
<tr>
<th>Definition of a Personal Information Space</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User X incrementally builds a small-sized personal information space, P, through individual value judgments on information units. X judiciously selects only useful items to create a known source of valuable information. X adds value through organization and expends effort to manage P by optimizing the cost of managing items against the expected gains of each unit. X attempts to structure P so that it facilitates efficient retrieval based on expected use. P is a manifestation of X’s personal view on the complex information space, C.</strong></td>
</tr>
</tbody>
</table>

The next five sections describe each principle of a personal information space and preview our empirical study of Web users found in chapters 4 through 8. Each of these chapters provides empirical evidence to show that the principles of a personal information space are found in a Web user’s bookmark archive.

2.4.1. Start Small and Build Incrementally (Prevent Information Overload)

*Users incrementally build up a small sized personal information space through individual value judgments on information units.*
People judiciously select items from a complex information space in order to gradually build up a small, manageable personal information space. It is purposefully small so that the information space user can deal with the information overload inherent in the large source information space. In classical information theory this is known as the principle of selective omission of information (Resnikoff 1989). The modest size gives the person a basis for processing information units and retrieving items. It also enables the person to manage all the units cognitively.

The relatively small size of the personal information space is a direct result of the locality of reference behavior intrinsic to working within a complex information space. As an information consumer performs sense-making on a patch of information, he/she builds up a cognitive interpretation (Russell, et al. 1993). Items are judiciously added to the personal information space based on a relevancy rating of the information unit found within that locality.

Chapter 4 will describe how Web users incrementally build a personal Web information space through value judgments. Web users create 0 to 5 bookmarks each browsing session. We shall describe the growth rate of bookmark archives over time.

2.4.2. Select Only Useful Items (Avoid Pollution)

Users judiciously select useful items to create a known source of valuable information.

A personal information space typically has a large proportion of relevant information units. In relation to the percentage of relevant information units in the source information space, the personal information space has a much higher degree of relevant documents. This results directly from a host of individual value judgments which decide the admittance of each new item.

A personal information space is a known source of valuable information. The expected gain from referencing this locality of information units is also high. Bush (1945) introduced the idea of a Memex or “memory extender,” which serves as a
repository of valuable personal information to augment human memory (Fisher, et al. 1989).

In chapter 5 we show why a WWW bookmark archive is a known source of high value to users. Web users create bookmarks through individual value judgments according to five criteria for bookmark creation. They regularly reference their bookmark archive because it contains a high percentage of relevant Web sites. We consider two questions: what do users bookmark and why? We present a taxonomy of bookmark use and describe the uses of bookmarks while browsing the WWW: reducing user load, providing navigation and access to pages, helping in the collaboration and publishing of Web pages, and creating an archive of pages.

2.4.3. Add Value Through Organization (Combat Entropy)

Users expend effort to manage a personal information space by optimizing the cost of managing items against the expected gains of each unit.

People go through a process of adding new items, managing existing items and retrieving items from their personal information space. They expend energy to organize, tune, and manage the information space. This process iterates as the user continually adds new items and expends energy to organize them. Most users only organize what they have to, taking the time to organize if they believe the results will add value to the personal information space.

In chapter 6 we present detailed usage data of how Web users manage and organize their archives. Web users regularly iterate through a process of creating, managing and retrieving bookmarks from their archive. These tasks are performed within the context of a WWW browsing session. Users with large archives spend more time organizing and employ more advanced organizational techniques. Here we characterize
the influence of scale on the ways Web users organize bookmarks and their usage patterns.

2.4.4. Attempt to Structure for Retrieval (Counteract the Lack of Structure)
*Users attempt to structure their personal information space so that it facilitates efficient retrieval based on expected use.*

A personal information space is cost-tuned to minimized retrieval time with the least amount of effort. The structure which results from managing and organizing this personal information space is used to support efficient retrieval. Users tend to organize their personal information space with retrieval in mind (Lansdale 1983). The structure also facilitates adding new items and visualizing the personal information space.

In chapter 7 we describe how users retrieve items from their semantic hierarchy and maintain its organization, and consider the representation of Web pages and its effect on retrieval.

2.4.5. Establish Personal View of Information Space (Compensate for No Global View)
*The user's personal information space is a manifestation of the user's personal cognitive view of the complex information space.*

Users collect references and gradually build a personal cognitive prosthesis for the source information space. The result is a manifestation of the user's personal view on the complex information space. It is used to compensate for the lack of a coherent aggregate structure and the inability to see all the information units.

A user's cognitive view of the complex information space is based on experience. Since no global view exits, the user identifies landmarks in the information space which are fundamental to the user's mental map of the information space (Chignell 1996).

These landmarks are central to the formation of the personal information space.
In chapter 8 we describe how a personal Web information space is a manifestation of the user's personal cognitive view of the Web. We shall show mental maps of the Web collected in a survey of 27 users. These suggest that users identify abstract conceptual references and separate their bookmarks from the unmapped resources of the Web. Finally, we shall briefly consider the relationship between private and public space on the Web.
3. INFORMATION SPACE BEHAVIOR ON THE WWW

Web users exhibit at least four behaviors in a complex information space: (1) they frequently revisit localities of relevant documents, (2) they establish bookmarks so as to return to those localities, (3) they have metaphors for accessing information, and (4) they create a personal information space from the bookmarks they have collected.

3.1. USERS FIND LOCALITIES OF INTEREST

Web users re-visit groups of related Web pages. In this section we show how the locality of reference behavior indigenous to a complex information space is found in Web users. (Marchionini, et al. 1995) provides an overview of information access and shows how users access groups of documents. Rada and Murphy (1992) showed the difference between searching and browsing hypertext. In section 3.1.1. we define the concept of a Web locality which is the basis for our discussion of locality of reference behavior in Web navigation. Then in section 3.1.2. we summarize the revisitation properties of Web users.

3.1.1. Web Localities

A locality of information units on the Web must have a coherent, logical structure and be tightly coupled by hypertext links. The following definition will be used as we describe how bookmarks are used to navigate between different localities.
Our definition is different than Tauscher (1996) who employs the mathematical definition of a locality set to show revisitations in the use of history mechanisms (see 3.1.2.). Instead of investigating locality sets in user behavior, we focus on the structure of Web pages. We define a Web locality by the content and relationship between pages connected by hypertext links. This is not based on usage patterns. For example, my home page and its supporting pages are a Web locality. The Web site for my research group and my lab are also a Web localities because they have been organized thematically, have structural cues, and represent semantic relationships through hypertext links. In contrast, the results of a search engine query would not typically be a locality.

3.1.2. Revisitation via History Mechanism

Empirical studies by Tauscher (1996) showed that users frequently reference localities of related pages. Their analysis of the 6 weeks of usage data from 23 users showed the following revisitation patterns in Web navigation:

| Probability of Revisitation: 58% of an individual’s pages are revisits |
| New Pages: users continually add new pages to their repertoire of visited pages |
| Temporal Locality of Revisits: users tend to revisit pages just visited |
| Frequency of Use: users access only a few pages frequently |
| Related Clusters: users browse in small clusters of related pages |
| Repeated Paths: users generate short sequences of repeated URL paths |

The principle of locality of reference emerges in user behavior. There is a 58% probability that the next page visited will be one that the user has already encountered.

Users tend to browse in clusters of related Web pages.
3.2. USERS CREATE BOOKMARKS TO DISTINCT WEB LOCALITIES

Bookmarks are used in only 2.7% of all navigation actions, yet they are "essential" to users (see 3.2.1). Bookmarks are primarily used to jump between Web localities, after which users employ a variety of methods to navigate within the new locality.

Our usage data analysis identified navigational patterns emerging after the use of a bookmark (see 3.2.2). Our survey respondents identified the archival use of bookmarks significantly more than temporary navigational aid, publishing on the Web, or collaborating with others (see 3.2.3). Our analysis of user bookmark files found that 87% of bookmarks point to distinct domain names, independent of the size of the bookmark.
archive (see 3.2.4). In section 3.2.5, we see how a user loses access to a locality if he/she forgets to bookmark it.

### 3.2.1. Paradox of Usage vs. Usefulness

Tauscher (1996) found that a bookmark is used in only 2.7% of all navigation actions. Since bookmarks are used so infrequently, we might be inclined to conclude that they are simply not important to Web browsing. But this contrasts with the evidence from our survey. Users frequently wrote that bookmarks are "essential" to navigating the Web and returning to sites. In fact, a few users wrote that they "could not live without them."

Pitkow and Recker (1995) found that bookmarks were identified as a browsing strategy by over 80% of users surveyed. The table below summarizes the browsing strategies used:

<table>
<thead>
<tr>
<th>WWW Browsing Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hotlist: users revisiting pages they have added to their hotlist</td>
</tr>
<tr>
<td>2. Search: using search engines such as Lycos</td>
</tr>
<tr>
<td>3. Meta-index: using large indices such as Yahoo</td>
</tr>
<tr>
<td>4. Opportunistic: following links from page to page as they are encountered</td>
</tr>
<tr>
<td>5. URL: typing in known URLs</td>
</tr>
</tbody>
</table>

The use of a hotlist or bookmark was identified as a WWW browsing strategy more than any other strategy in both the surveys 4th (Pitkow and Recker 1995), and 5th (Pitkow and Recker 1996) by GVU. The graph below illustrates the results.
If bookmarks are so "essential" to users, then why are they used so infrequently (<3%)? The answer to this paradox lies in the special function which bookmarks provide. 

*Bookmarks are used to jump between localities.* They enable users to return to a Web locality. Then, other methods are used to navigate within that locality.
The distance traveled in a single action can be used to describe hypertext
navigation. The table below illustrates this concept.

<table>
<thead>
<tr>
<th>Walking vs. Jumping</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Walking</strong></td>
</tr>
<tr>
<td>Step (d=1)</td>
</tr>
<tr>
<td>Users walk (distance = 1) step by step down a hypertext path selecting links on the Web page. Back and Forward are used to traverse the path in each direction.</td>
</tr>
<tr>
<td><strong>Jumping</strong></td>
</tr>
<tr>
<td>Jump (d&gt;1)</td>
</tr>
<tr>
<td>Users jump (distance &gt; 1) between nodes that are not directly connected. Jumping commands include: bookmark, history list, entering URL, and pre-defined icons built into the browser (e.g., What's Cool).</td>
</tr>
</tbody>
</table>

Walking is the process of moving step by step down a hypertext path. In contrast, jumping is defined as transporting across any large number of nodes.

The history mechanism (e.g., Go pull-down menu in Netscape) can be used for both walking and jumping. Users can create multiple threads in hypertext by jumping between paths. They jump between Web localities and walk within a particular locality by traversing its paths.

Tauscher (1996) found that users tend to engage in a great deal of local navigation. She found that once users access a particular page, they tend to return to it frequently. Users employ a variety of methods to navigate within a locality of related pages (i.e., Open Anchor 41.6%, Back 30%, Forward 0.9%, History 0.8%). This most likely involves walking down a path. A guided-tour, hub-and-spoke, or depth-first-search pattern may emerge from traversing multiple threads in a locality (see 3.2.2.1.).

### 3.2.2. Bookmarks as a Launching Point for Web Exploration

Bookmarks are used as a launching point for exploring other Web pages. We analyzed the usage data of 23 users over a 6 week period and identified the patterns which emerge in hypertext navigation. Tauscher (1996) provided this data and generated URL
vocabulary graphs which illustrate the use of bookmarks in relation to other actions. In section 3.2.2.1., we review the hypertext patterns of Web users that appear in Tauscher's data. Then in section 3.2.2.2., we describe higher-level bookmark usage functions.

3.2.2.1. Navigation Patterns in WWW Browsing

Hypertext navigation patterns characterize users browsing within a particular Web locality (see Dillon, McKnight and Richardson (1990) for a review of hypertext navigation, and Paraneck (1989) for a discussion of patterns in hypertext use). Here we summarize the patterns found in a recent empirical study of the Web (Tauscher 1996).

<table>
<thead>
<tr>
<th>Hypertext Patterns in Web Navigation (Tauscher 1996)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guided-Tour:</strong> Users follow the structured links built into the Web page</td>
</tr>
<tr>
<td>(e.g. (Goto Next Page)) This is characterized by Open URL, Next and Previous actions by the user.</td>
</tr>
<tr>
<td><strong>Hub-and-Spoke:</strong> Users browse in one small area within a Web locality or Web site. They make frequent use of Back to return to a hub or index page. This serves as a launching-off point for exploration.</td>
</tr>
<tr>
<td><strong>Depth-First Search:</strong> Users step down a path with a series of Open URL actions followed by a series of Back actions.</td>
</tr>
<tr>
<td><strong>First Time Visit:</strong> The user selects a new URLs and follows a path of links via the repetitive use of Open URL.</td>
</tr>
<tr>
<td><strong>Revisit:</strong> The user re-visits a group of pages using the Back button or Bookmark.</td>
</tr>
<tr>
<td><strong>Submit a Form:</strong> The user enters data into a form and initiates (SUBMIT) from the page.</td>
</tr>
<tr>
<td><strong>Helper App:</strong> The user opens a helper application. For example, use an image viewer (e.g. ghostview), open a telnet session.</td>
</tr>
<tr>
<td><strong>Author a Page:</strong> The user repeatedly hits Reload while authoring a Web page.</td>
</tr>
</tbody>
</table>

The diagram on the next page illustrates five of the eight patterns in the usage data of one user. The X-axis denotes the number of URLs which the user has visited. The Y-axis shows the number of distinct URLs which form the user's URL vocabulary. As the diagram shows, revisits to pages appear as a horizontal line. The slope of the curve is steep at first-time visits.
A hub and spoke pattern emerges when the user uses Open URL intermittently with Back. Thus, the user makes short excursions ("spokes") from the hub and then returns in order to make another new excursion. Repeated use of the Reload action signifies that the user is authoring a page.

In a guided tour, the user follows embedded links. This appears as a step function in the usage data (see Tauscher 1996). In depth-first search, users step down a path, backtracking at certain points in order to start a new path.

We identified three additional navigation patterns in Tauscher's usage data. The table below summarizes these patterns.
Hypertext Navigation Patterns in Usage Data

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive Probes</td>
<td>Two or more bookmarks are selected one immediately after another.</td>
</tr>
<tr>
<td>Intermittent Probes</td>
<td>Two or more bookmarks are selected with short excursion between each use. The user jumps to a new location with a bookmark, then initiates less than four navigation actions before jumping to the next locality via a bookmark.</td>
</tr>
<tr>
<td>Dead End</td>
<td>The user makes a short series of actions after the use of a bookmark and hits a dead end.</td>
</tr>
</tbody>
</table>

The average size of Repetitive Probes is 2.63 bookmarks. Intermittent Probes have an average of 4.04 bookmarks used with one to three navigation actions between each (e.g., Bookmark, Open Anchor, Open Anchor, Back, Bookmark, Open Anchor, ... Bookmark).

The number of repetitions is shown below:

<table>
<thead>
<tr>
<th>Using Bookmarks to Probe Hypertext</th>
<th>Mean</th>
<th>St.Dev.</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive Probes</td>
<td>2.63</td>
<td>1.62</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Intermittent Probes</td>
<td>4.04</td>
<td>1.77</td>
<td>3.5</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

Bookmarks show up as clusters on the URL vocabulary graph resulting from the repetitive and intermittent probes into hypertext.

The graph below illustrates hypertext navigation patterns emerging after the use of a bookmark for another user. The user makes repetitive probes (URLs ~110 - 140) before a hub-and-spoke pattern emerges (URLs 140 - 165). In this instance of Repetitive Probes, the user may be searching for the correct bookmark by visiting a group of them. The user then begins exploration from the last bookmark in the cluster. The Intermittent Probes (URLs 325 - 335) shows up as a cluster of alternating uses of Open URL with a bookmark and Back.

In the next section we describe three usage functions which characterize a user's intentions for using bookmarks and the associated hypertext navigation patterns.
3.2.2.2. Bookmark Usage Functions

We categorized the patterns which emerged after the use of a bookmark into three groups: WhichBookmark?, LaunchingPoint, and DestinationPage:

<table>
<thead>
<tr>
<th>General Bookmark Usage Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WhichBookmark?: the user repeatedly uses bookmarks until he/she finds the right one; the user begins reading the page in detail once the correct bookmark has been selected.</td>
</tr>
<tr>
<td>2. LaunchingPoint: the bookmark is used as a starting point for exploration; a navigation pattern emerges after its use.</td>
</tr>
<tr>
<td>3. DestinationPage: the bookmark contains some useful content which the user studies; the page supplies content but does not have many links for further exploration.</td>
</tr>
</tbody>
</table>
These usage functions are an abstraction of the navigation patterns because they postulate users' intentions, that is, why each pattern emerges after the use of a bookmark:

<table>
<thead>
<tr>
<th>Patterns Emerging After the Use of a Bookmark</th>
<th>WhichBookmark?</th>
<th>LaunchingPoint</th>
<th>DestinationPage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetitive Probes</td>
<td>Hub &amp; Spoke</td>
<td>Submit a Form</td>
<td></td>
</tr>
<tr>
<td>Intermittent Probes</td>
<td>Depth First Search</td>
<td>Helper App</td>
<td></td>
</tr>
<tr>
<td>Dead End</td>
<td>Guided Tour</td>
<td>Author a Page</td>
<td></td>
</tr>
<tr>
<td>First Time Visit</td>
<td>Revisit</td>
<td>Probes/Dead End</td>
<td></td>
</tr>
</tbody>
</table>

In the first usage function, WhichBookmark?, the user is attempting to find the right Web page.

Once the user right finds the right bookmark, it serves as a launching point for exploring the Web. We identified five of the hypertext navigation patterns from section 3.2.2.1. which emerge after a LaunchingPoint bookmark. For example, a bookmarked page serves as the hub in a hub and spoke pattern. A bookmark begins a depth first search. A bookmark might also point to the first page in a guided-tour.

A bookmark is a DestinationPage when the page contains the information the user needs. For example, the bookmark points to a page where the user can enter data into a form and press [SUBMIT]. Note that repetitive and intermittent probes can also be part of a DestinationPage. This occurs when the user needs information on more than one page and uses bookmarks to quickly jump between them. A dead end emerges after a DestinationPage when the user simply uses the content on the page without making any deep hypertext explorations from the page.

We compiled an aggregate distribution of emergent bookmark patterns from the table in Appendix C which lists each user's bookmark use. When a bookmark is used, what is the probability of each pattern emerging? The graph below provides an estimate.
of this distribution based on our data analysis. Each of our three basic categories and their composite patterns are shown:

![Graph showing patterns emerging after use of a bookmark]

Bookmarks are used as a launching point for Web exploration 44% of the time.
Bookmarks are used as probes into hypertext (repetitive + intermittent) 42% of the time.

3.2.3. Archival Use of Bookmarks

Bookmarks are used as a reference to store for use in future WWW sessions. It is very rare that a bookmark is used as a temporary navigational landmark to come back to later during the same session. This supports our notion of bookmark as a tool to jump between localities instead of navigating within a locality.

The graph below shows the mean self-reported importance of four different types of bookmarks based on question 2.3 of our survey (see Appendix A). All error bars in the thesis denote a single standard deviation from the mean.
Archival bookmarks were significantly different than temporary, publishing or collaborative ($F_{3,1064} = 2.61, P <= 0.05$). A Tukey test showed that archival elicited a significantly higher response ($P < 0.05$) than all others.

Our analysis of bookmark files substantiates the archival use of bookmarks. Only 50 of the 56 files had date information. We calculated the number of days since the last time the user visited each bookmark. This was calculated against the date at which the user submitted the bookmark file to the study. The average percent of bookmarks retrieved over the past year was calculated.

The next figure plots the percent of the bookmark file visited over a one year period. This function is best described by the following quadratic equation:

$$R = 0.202 + 0.004*T + (-5.0x10^{-5})*T^2$$

where $R$ is the percentage of bookmarks retrieved from the archive in the last $T$ number of days. This explains most of the variance ($r^2 = 0.996$). Thus, 67% of bookmarks have been visited in the last 6 months, 96% in the last year.
In the next graph we take a closer look at one user.

This graph illustrates the sporadic use of bookmarks over a period of approximately 19 months. We calculated that this user has only visited 30% of the archive in the last 3 months and only 6% in the last month.
Bookmarks are created and stored for archival purposes, and often not visited for months. They are a means of returning to a Web site or cluster of related pages, and as references for future exploration. In the next section, we describe a defining property, distinct domain names.

3.2.4. Distinct Domain Names in Archive

WWW bookmarks point to distinct domain names. Users very rarely create more than one bookmark to the same domain name (in most cases a domain name is equivalent with a single Web site). We used the bookmark files of 56 Web users and analyzed the URLs in each file, calculating the percentage of distinct domain names as shown below:

<table>
<thead>
<tr>
<th>Percentage Distinct Domain Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U(1..N)$ is a set of URLs. Let each URL in bookmark file, $F$, be $u$, where $U(1..N) = u_1, \ldots, u_k, \ldots, u_s$ possibly contains non-distinct elements.</td>
</tr>
<tr>
<td>Let $D(u_j)$ be the domain name of URL $u_j$. Then, let the set of distinct domain names, $D(U)<em>{distinct}$, be $D(U)</em>{distinct} = {D(u_1)YD(u_2)Y\ldots YD(u_i)Y\ldots Y(u_s)}$</td>
</tr>
<tr>
<td>Let the percentage of distinct domain names, $PercentDistinct$, in $U$ be $PercentDistinct = \frac{</td>
</tr>
</tbody>
</table>

First we construct a set of distinct domain names, $D(U)_{distinct}$, from the set of URLs in a bookmark file. Then we calculate the percentage of distinct domain names, $PercentDistinct$, found in the set of URLs. This measures the amount of overlap in domain names. A high percentage indicates that the user very rarely creates more than one bookmark to the same domain name.

3-15
We found a high percentage, 86%, of distinct domain names in bookmark files:

<table>
<thead>
<tr>
<th>Percent Distinct Domain Names</th>
<th>Summary Statistics of 56 users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>St. Dev.</td>
</tr>
<tr>
<td>85.95%</td>
<td>9.35%</td>
</tr>
</tbody>
</table>

Bookmark files were widely distributed between 1 and 325 bookmarks. Users create bookmarks to distinct domain names, irrespective of the size of the archive.

We divided the above data set into four equal groups and tested PercentDistinct for significance ($F_{3, 52} = 0.115, P > 0.05$). We used an arcsine transform on PercentDistinct and our ANOVA showed no significant difference in the four groups.

We list the mean and standard deviation of each group in the table below:

<table>
<thead>
<tr>
<th>Number Bookmarks</th>
<th>Mean PercentDistinct</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 29</td>
<td>93%</td>
<td>10%</td>
</tr>
<tr>
<td>30 - 64</td>
<td>85%</td>
<td>8%</td>
</tr>
<tr>
<td>65 - 125</td>
<td>84%</td>
<td>6%</td>
</tr>
<tr>
<td>126 - 325</td>
<td>81%</td>
<td>9%</td>
</tr>
</tbody>
</table>

The number of bookmarks explains very little of the variance ($r^2 = 0.14$) in PercentDistinct domain names. Although PercentDistinct decreases slightly as the
number of bookmarks increases, the group with the largest size bookmark files has over 80% distinct domain names.

3.2.5. Delayed Value Judgment Problem: Forgetting to Bookmark a Page

Users forget to bookmark pages on the Web and have difficulty returning to the desired Web site. This is indicative of the experience of traveling through a complex information space with a limited view and no global-map to guide the way. When the user forgets to bookmark a page, the user loses quick access to the respective Web locality.

One user wrote, "Bookmarks are often an afterthought." The user does not know that the page is worth bookmarking until later in the session or during a different session altogether. Many times, a user can only make a value judgment until after many pages have been examined and the user has compared their relative merits. The process of identifying a page to bookmark is critically tied to the local browsing context.

When users forget to bookmark a page, they often lose access to its respective Web locality. For example, one user complained "I forget to add a bookmark so the page is lost. Perhaps browsers could ask if you want a bookmark when you stay on the page for a long time." When the value judgment is delayed, it becomes difficult to re-locate the page.

3.3. USERS CONCEPTUALIZE WWW ACCESS THROUGH METAPHORS

We observed Web users with at least four metaphors for accessing pages: identification, collection, movement, episodes. In this section, we examine each user metaphor identified in our survey.
Mental models of a user manifest themselves in the ways users describe experiences with Web browsers and related tools, since metaphoric expression is part-and-parcel of language (Lakoff and Johnson 1980).

We identified metaphors for using bookmarks and accessing Web pages by analyzing the comments provided in the survey. Users described their experiences and problems with the Web in essay format.

3.3.1. Marking/Tagging Information (Identification)

_Users rank the value of a Web page by actively assigning distinctive labels via bookmarks._

The marking metaphor for bookmarks is based on the concept of placing a small tag or distinctive label on a piece of information. The label represents the relevance or value of the associated 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." These comments demonstrate the concept of actively assigning distinctive labels to a Web site for future reference.

Current bookmark tools only provide a Boolean ranking. When marking a Web site, users can either mark it or not. One user described his wishes for a more detailed labeling scheme: "It would be nice to be able to assign a star rating or number." Thus, even within a user's own personal Web information space, there are degrees of relevance; some bookmarks are more important than others.

3.3.2. Pulling Information into a Collection (Collection)

_Users pull Web pages out of the hypertext into their collection of bookmarks._

Bookmarks can be used to retrieve sites from the Web. This metaphor is based on the notion that the user is stationary and he/she is pulling specific information from the
vast information space of the WWW. One user likes “the ability to quickly retrieve sites” with bookmarks. Bookmarks are considered a local storage of links to information which the user has selected or pulled down from the Web. The principle of collection is central to this metaphor. Bates termed this metaphor “berry-picking” (Bates 1989). The Web is not a broadcast medium pushing a limited set of choices on users. Consequently, they pull valuable bits and pieces into their collection (Negroponte 1995).

3.3.3. Traveling Through Cyberspace (Movement)

Users move from site to site on the Web and return to valuable ones via bookmarks.

Many users employ notions of location and movement in describing the process of browsing the Web. They are traveling through a vast space of information. Bookmarks are an effective means of “returning to a site,” or “getting back to a site.” Another user believes that bookmarks are “extremely useful for going back to certain pages,” and “easy to navigate back to favorite sites.” The notion of traveling is illustrated by a user who wrote, “many time I run across a useful page by chance.” Upon stumbling across this valuable page, a bookmark is created to provide a direct link to return to that location.

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.” These comments demonstrate the relationship between locations in space and time. Users associate being at a particular place in Cyberspace with the time of their visit. “Bookmarks are essential to find good sites again and remember previous sessions.” Bookmarked pages stand out as landmarks, as temporal and navigational guides.

A subset of the traveling metaphor is information foraging. This metaphor is based on the concept of actively seeking information. Foraging encompasses an active
process of finding or searching for a piece of information. It involves "relocating" a piece of information based on a goal-directed search.

3.3.4. Remembering Episodes Related to Information (Episodes)

Users recall previous episodes associated with the Web pages they have stored in their bookmark archive.

A list of bookmarks can function as a chronological list of episodes. For example, 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. In this sense bookmarks are used to make cognitive inferences about information foraging tasks. Reviewing a list of bookmarks can remind a user through episodic memory (Murdock 1974) of the bookmark's (i) utilitarian characteristics, (ii) associated tasks, and (iii) represented information content.

People make mental inferences based on short textual phrases. A headline reminds us of what the article was used for and the embedded task environment. Reviewing a set of bookmarks is essentially a process of making inferences based on a textual representation of the content of each Web page (see 7.2.).

Creating a bookmark requires a value judgment based on personal criteria (see 4.1.2.). An individual title can spawn a mental inference of the particular episode which the user was doing at the time, the value judgment determining why he/she created the bookmark, and what the page content is used for. The degree to which a user makes mental inferences based on the title of the page depends on how well it describes the content, the user's familiarity with the page title, and how much time and attention the user expends using the bookmark. For example, Craik and Lockhart (1972) showed that recall and recognition are improved through elaboration. As new associations are created
the levels of cognitive processing increases and so does the subject's memory recall. In section 7.2, we show how users encounter problems with remembering the Web page associated with a bookmark.

3.4. USERS BUILD A PERSONAL WEB INFORMATION SPACE

In this section we show how users attempt to cost-tune their environment through the functionality provided. As they gain more experience, they utilize more of the available functionality. Users with more bookmarks are more likely to cost-tune their archives. This section does not address why or how successful users are in optimizing their information environment. Chapters 4 through 8 address these question in more detail by explaining how each of the five principles of a personal information space apply to WWW bookmark archives.

3.4.1. Usage Statistics

Here we show the relationship between the user's self-reported experience level and amount of Web usage.

3.4.1.1. Experience Level

In general, the survey participants were moderate to highly experienced Web users. This is expected since the sample group consisted of Internet: Beyond the Year 2000 conference participants, a group likely to have Web expertise.
The experience level of respondents is shown in the graph above. The mean self-reported experience level was 4.95 with a standard deviation of 1.45 on a standard 7 point scale (see question 1.1 in Appendix A).

3.4.1.2. Sessions per week

We asked users "On average, how many times do you make use of the WWW per week?" The histogram below shows the distribution.
Answers ranged from 0 to well over 50 times per week. Most users, 84.3%, use the Web less than 20 times per week. 64.0% of users have less than 10 WWW browsing sessions per week. Very few users, 11.6%, make use of the Web more than 20 times per week.

3.4.1.3. Session Length

Users were asked about the length of a typical WWW browsing session. The chart below shows the results of all users:

![Session Length Chart](image)

Hence, 75.9% of users spend less than 1 hour in a typical WWW browsing session.

3.4.1.4. Estimating Average Hours per Week

We calculated the average hours per week browsing the Web based on the product of frequency and session length. The result for each user is a range (minimum vs. maximum hours per week). Based on the survey results obtained here, half of the respondents spend five or less hours per week. Note that this estimate is calculated by multiplying estimates of frequency and session length together and that it must be interpreted cautiously because of likely errors in making these estimates. The chart below shows the composite of the average of each users range in terms of hours per week.
3.4.1.5. Correlation of Usage Variables

Users acquire more experience as they collect bookmarks during their browsing sessions. The number of bookmarks is significantly correlated with self-reported expertise \( (r = 0.55, p < 0.01) \), and with time spent per session \( (r = 0.17, p < 0.01) \). Note that this correlation explains only a small part of the variance and other factors are likely to be involved. The table below summarizes our correlation of usage variables from the survey.

<table>
<thead>
<tr>
<th></th>
<th># Bookmarks</th>
<th>Experience</th>
<th>Sessions/week</th>
<th>Time/session</th>
</tr>
</thead>
<tbody>
<tr>
<td># Bookmarks</td>
<td>1.0000</td>
<td>0.5522**</td>
<td>0.1102</td>
<td>0.1715**</td>
</tr>
<tr>
<td>Experience</td>
<td>0.5522**</td>
<td>1.0000</td>
<td>0.2030**</td>
<td>0.1054</td>
</tr>
<tr>
<td>Sessions/week</td>
<td>0.1102</td>
<td>0.2030**</td>
<td>1.0000</td>
<td>-0.0113</td>
</tr>
<tr>
<td>Time/session</td>
<td>0.1715**</td>
<td>0.1054</td>
<td>-0.0113</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Note that the number of sessions per week spent browsing did not correlate with the length of each session.

3.4.2. Use of Bookmark Functionality

We asked users how often they use the functionality available by current tools in question 2.1 of the survey (see Appendix A).
Bookmark functionality is used to tune the cost structure of the user's archive. The infrequent use of functionality presented in the graph above suggest that these functions are not predominant during a browsing session. They are only used when necessary.

We identified trends in the use of functionality in relation to experience level and number of bookmarks in a user's archive. The table below summarizes the correlation between use and functionality.

<table>
<thead>
<tr>
<th>Correlation of Use with Functionality (r values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Experience</td>
</tr>
<tr>
<td># Bookmarks</td>
</tr>
<tr>
<td>Sessions/week</td>
</tr>
<tr>
<td>Significance</td>
</tr>
</tbody>
</table>

As users acquire more experience, so does their use of the available functionality for managing bookmarks. The number of bookmarks significantly correlates with all functionality except annotation.
4. PWIS #1: START SMALL AND BUILD INCREMENTALLY

In this chapter we present empirical data which shows how a user incrementally builds-up a small bookmark archive through individual value judgments on Web pages. The resulting personal Web information space is much smaller than the source information space (i.e., the WWW).

4.1. SIZE AND GROWTH OF ARCHIVE

In this section, we present data from our survey and collection of bookmark files to illustrate how personal Web information spaces grow.

4.1.1. Small Bookmark Archive

Users create a small personal information space so that they can cognitively manage all the items. The largest group of users in our survey, 44% or 141 respondents, have 26 to 100 bookmarks in their archive. The groups shown below were selected through our initial pilot study of 12 interviews and an informal survey.
We performed an analysis of skewness to identify its fit with a normal distribution curve to satisfy the pre-conditions for performing statistical tests on this data set. The negative skewness (skew=-0.4, observations=321) indicates that the more extreme values are less than the mean. Our distribution is also slightly more heavy in the center (26-100 bookmarks) than a normal distribution curve.

The bookmark files we collected were within the same order of magnitude as survey respondents. Most bookmark files varied from 1 to 200 bookmarks.

The graph below shows the number of bookmarks reported by users in the GVU study (Pitkow and Recker 1995, 1996).
We generated composite categories to compare our data with the GVU study. The chart below shows that our results are similar to the GVU results. The studies are shown in chronological order.

![Comparing Number Bookmarks](image)

In all three studies, over 55% of users have 11-100 bookmarks. The remaining users seem to be evenly split between 0-10 bookmarks and more than 100 bookmarks. Approximately 40% of the 56 bookmark files we collected via email had more than 100 items. This differed from the other three surveys and may be due to self-selectivity of this data set.

### 4.1.2. Incremental Growth of Archive

Bookmark archives grow incrementally. 83% of users create 1 to 5 bookmarks during each browsing session. The chart below illustrates the results of our survey.
We analyzed the growth of each bookmark file collected. The graph below shows the aggregate growth of bookmarks files.

To examine the rate of growth in bookmarks over time, the bookmark files of 50 participants were used to make an aggregate file. The file traced one year’s growth,
starting from the creation of the first bookmark. In the combined data, number of bookmarks was highly correlated with time ($R = .9958$, $P < 0.0001$). The regression equation for number of bookmarks is as follows:

$$\text{NumberBookmarks} = 7.45 + .18 \times \text{NumberDays}$$

We analyzed individual users to identify patterns in the growth of bookmark archives. The graph below shows how one user adds bookmarks sporadically over many browsing sessions in small clusters.

The vertical line segments in the graph above denote a group of bookmarks added on the same day.

Bookmarks are created sporadically in small clusters. The average number of bookmarks added per session was examined for the 50 bookmark files. We identified days where at least one bookmark was added and created four categories based on the size of the cluster: (1) Only 1 added bookmark, (2) from 2-5 bookmarks, (3) 5-10 added
bookmarks and (4) more than 10 bookmarks. There were significant differences in the number of sessions that matched to the 4 categories ($F_{3,196} = 107.07, p<.0001$). Most common were sessions with only 1 added bookmark, 19.8, which differed significantly from all the other categories. The 2-5 bookmarks added category also differed from the others.

![Bar chart showing distribution of bookmark additions per session.]

We analyzed clusters with less than 5 bookmarks in more detail. There were differences, however, between the number of session in which 1, 2, 3, and 4 or more bookmarks were added. An analysis of variance indicated that the differences were significant ($F_{3,16} = 55.65, P < 0.0001$) and post hoc analysis (Tukey) indicated that sessions where only 1 bookmark was created (19.8) was higher than all other groups. Clusters with 2 to 5 bookmarks was significantly higher than those of size 6-10 and 11 or more.
5. PWIS #2: SELECT ONLY USEFUL ITEMS

A user's bookmark archive is a known source of valuable Web information. In this chapter we describe the value of bookmarks from three perspectives. First, what do users bookmark? Second, why do users create bookmarks? Third, how are bookmarks used? In this way we describe some of the critical factors which underlie the value of a personal Web information space.

5.1. WHAT USERS BOOKMARK

What kinds of sites do users bookmark? We analyzed the bookmark files of 56 users to provide some initial answers to this difficult question. We generated a simple ontology to categorize the most popular bookmarks among users. Then we looked closely at the use of bookmarks in three specific domains: (i) organizational home pages, (ii) search engines, and (iii) on-line news. In all three domains, users create bookmarks to pages which are starting points for exploring the Web.

5.1.1. Ontology of Bookmarked Sites

We generated a simple ontology based on the contents of bookmarked pages. Our 56 bookmark files contained a total of 4902 bookmarks (represented as URLs). We built our ontology from all domain names with 4 or more references in order to make the data set manageable. This provided us with a total of 1459 bookmarks or 29.76% of all bookmarks in the study. The table below shows the simple ontology we generated from the top 29% most popular bookmarks.
Appendix B lists the actual domain names and frequency of each reference. In the graph below we illustrate the distribution of bookmark content according to this ontology.

The prevalence of university-related sites in the bookmark archives of these users is not surprising since 90% of our user population was from University of Toronto students, faculty and staff.
We calculated the percentage of users that have collected at least one bookmark to each category in the ontology.

In the chart above, we see that the domain names listed in Appendix B are widely distributed among the 56 bookmark files we collected.

Further research is needed to relate bookmarks to work context, compare diverse user groups, identify common patterns in the contents of bookmarks, consider the embedded task environment, and investigate the utilitarian properties of bookmarks (e.g., do they point to content pages or index pages?).

5.1.2. Bookmarks to Organizational Home Pages

We calculated the percentage of bookmarks with URLs to domain name prefixes only. The example below illustrates this concept:

<table>
<thead>
<tr>
<th>Examples of Prefix Only vs. Prefix+Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Name Prefix Only</td>
</tr>
<tr>
<td><a href="HTTP://WWW.CS.UTORONTO.CA/">HTTP://WWW.CS.UTORONTO.CA/</a></td>
</tr>
<tr>
<td>Domain Name Prefix + File Path Suffix</td>
</tr>
<tr>
<td><a href="HTTP://WWW.CS.UTORONTO.CA/~abrams/thesis.html">HTTP://WWW.CS.UTORONTO.CA/~abrams/thesis.html</a></td>
</tr>
</tbody>
</table>
In the above example, the file path "~abrams/thesis.html" is the suffix. This URL has a domain name prefix and file path suffix. In order to calculate an empirical lower bound on the number of organizational home pages bookmarked, we calculated the percentage of bookmark URLs that have only a domain name. On average, 23% of bookmarks in each user file has no file path suffix; they contain a domain name prefix only.

<table>
<thead>
<tr>
<th>% BK URLs to Domain Name Prefix Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>23.24%</td>
</tr>
</tbody>
</table>

The graph below illustrates the clustering in more detail.

Changes in the number of bookmarks had very little effect \( (r^2 = 0.009) \) on the percentage URLs with a domain name prefix only. Subjects were divided into four equal groups based on the size of the archive. The percentage data was adjusted with an arcsine transformation and an ANOVA found no significant difference \( (F_{3, 52} = 0.115, P > 0.05) \) in the percentage of URLs to a domain name prefix only. The number of bookmarks explains very little of the variance \( (r^2 = 0.009) \) with a linear regression equation:

\[
D = 0.219 + 1.58 \times 10^{-4} * N
\]
where D is the percentage of URLs in each file with a domain name prefix only and N is the number of bookmarks in each file. Future research should investigate how the percentage of home pages in a user's archive compares to the actual percentage of home pages in the entire Web.

Do bookmarks tend to point to organizational home pages? This calculation does not provide a complete answer but does suggest a lower bound. That is, on average 23% of bookmarks do point to organizational home pages. More may be possible because some home pages have a path name suffix of the form: "home.html", "index.html".

In order to calculate the exact percentage times a bookmark points to a head page (i.e. the natural starting point for a set of related pages) we would need to explicitly visit each of the 4902 bookmarks in our study. This is left for future research.

5.1.3. Bookmarks to Search Engines

We studied bookmarks to search engines in detail and found that they contrast from the general bookmark population. Over 95% of users with 20 or more bookmarks have at least one search engine in their archive; either a query engine (e.g., Lycos) or a meta index (e.g. Yahoo). The graph below shows the prevalence of search engines in users’ bookmark archives.
We analyzed the URLs of each search engine bookmark and found that on average 65% of them point to the search engine's home page. We divided search engine bookmarks into three categories.

<table>
<thead>
<tr>
<th>Types of Bookmarks to Search Engines</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sub-Index Page:</strong> bookmark points to a part of the hierarchy of manually indexed pages which have been organized into a subject category.</td>
</tr>
<tr>
<td><strong>Search Results Pages:</strong> bookmark points to a page generated dynamically by the search engine in response to a user query.</td>
</tr>
<tr>
<td><strong>Query Page:</strong> bookmark points to a page with an entry form and [SUBMIT] button. This is usually the search engine's home page.</td>
</tr>
</tbody>
</table>

These groups categorize search engine bookmarks according to their utility. An average of 84.27% of all query engine (e.g. OpenText, Excite, Lycos, AltaVista, WebCrawler, Infoseek) bookmarks point to the *query page*. Query engines enable users to enter search queries into a computationally generated index. In contrast, a *meta-index* like Yahoo provides users a manually authored hierarchy of sites organized into subject categories. Query results point to locations in the Yahoo hierarchy with related sites. Only 34.0% of bookmarks to Yahoo, a *meta-index*, point to the *query page*. Most bookmarks to Yahoo, 58.5% of them, point to a *sub-index page* within the hierarchy.

The chart below shows the difference between the use of a query engine and a meta-index.
The compiled data are shown in the contingency table below:

<table>
<thead>
<tr>
<th>Meta-Index vs. Query Engine</th>
<th>Query Page</th>
<th>Results Page</th>
<th>Sub-Index Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-Index (Yahoo)</td>
<td>18</td>
<td>4</td>
<td>31</td>
</tr>
<tr>
<td>Query Engine (OpenText, Excite, ...)</td>
<td>75</td>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>

Data from Yahoo was compared to the combined data of the other search engines (see table above). Chi-squared analysis indicates that bookmarks pointing to a Meta-Index differ significantly from the Query Page, Results Page and Sub-Index Page of a Query Engine ($\chi^2(2)=55.72$, $P < 0.001$). We calculated the standard Z scores to see how much each cell differed from the expected value. An examination of each cell in the table indicates that for Meta-Index, subjects bookmarked the Query Page significantly less than would be expected ($z = -2.8$, $P < 0.01$) and bookmarked Sub-Index Page significantly more ($z = 5.1$, $P < 0.01$). For the combined Query Engine group, however, we find exactly the opposite; Query Pages were bookmarked significantly more than was expected ($z = 2.2$, $P < 0.01$) while Sub-Indices were bookmarked significantly less ($z = -$-
4.0, \( P < 0.01 \). There was no difference for either group in number of bookmarks to search Results Page.

In summary, users generally bookmark the home page of a query engine and selected parts of the meta-index hierarchy. They rarely bookmark the search results page. Yahoo differed significantly from all other query engines. In addition, none of the query engines differed significantly from one another.

5.1.4. Bookmarks to On-line News

In the on-line news domain, users tend to bookmark the cover page with links to more detailed information, instead of specific articles. Many users had bookmarks to the NEWS sites listed in Appendix B. The chart below shows the distribution based on four equal sized groups of users.

We separated bookmarks to news sites into two basic (1) categories: those used to facilitate navigation and (2) those used to convey information. The complete ontology is shown below:
We categorized bookmarks according to the definitions shown below.

### Ontology of Bookmarks to On-line News

<table>
<thead>
<tr>
<th>Navigation</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Entrance/Login</strong></td>
<td>The primary purpose of the page is to facilitate browsing and provide links to other pages.</td>
</tr>
<tr>
<td>The page is a login prompt to the news resource or a cover page welcoming the reader.</td>
<td></td>
</tr>
<tr>
<td><strong>Site Index</strong></td>
<td>The primary purpose is to convey information through the content of the page, not to support navigation.</td>
</tr>
<tr>
<td>The page is organized into subject categories with links. Examples:</td>
<td></td>
</tr>
<tr>
<td>1. sports, business, arts, etc.</td>
<td></td>
</tr>
<tr>
<td>2. <a href="http://www.pathfinder.com">www.pathfinder.com</a> has links to magazines: Time, Fortune, etc.</td>
<td></td>
</tr>
<tr>
<td><strong>Newspaper Front Page</strong></td>
<td>The page combines both navigation and content. Subject category links point to general topic areas. Headlines and short articles appear in a newspaper layout. Selecting a headline loads the respective article.</td>
</tr>
<tr>
<td><strong>Article(s)</strong></td>
<td>The page contains one or more news articles. It has very few links to other pages, if any.</td>
</tr>
<tr>
<td><strong>Download Software</strong></td>
<td>The page has a list of files which can be downloaded.</td>
</tr>
<tr>
<td><strong>Stock Quote</strong></td>
<td>The page presents a query to stock quotes via ticker symbol.</td>
</tr>
</tbody>
</table>

An **Entrance/Login** page is used to give users access to the site. The **Site Index** organizes links to related material by subject. The **Newspaper Front Page** is used for both navigation and content because it combines headlines, short captions and subject categories. Bookmarks to content fell under three categories: (1) one or more news articles, (2) a list of free software to download, (3) a query to a stock quote service.
As shown in the chart above, most users either create bookmarks to a page which is the natural starting point for locating specific news articles on the site.

5.2. WHY USERS BOOKMARK

We extracted comments from the answers to our questionnaire in order to understand why users create bookmarks. We have organized their comments into five criteria for bookmark creation.

5.2.1. Criteria for Bookmark Creation

Bookmarks take very little physical interaction to create and they are "easy to make." Users must continually judge the value of pages according to their own criteria for bookmark creation. A user makes a conscious decision to select a page out of all other available pages to add to his/her personal archive. This decision is based on a user's personal criteria which ranks the relevance of the page content. The value of information is critically tied to an associated task (Repo 1986). The relevance of a piece
of information is not an intrinsic property of the document itself, but it can "only be assessed in relation to the embedding task environment" (Pirolli and Card 1995).

Survey respondents volunteered descriptions of the sites they bookmark in the essay section of the survey (see Appendix A). Their comments provide evidence about a user's criteria for selecting pages out of the WWW to bookmark. Web users employ five general 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.

5.2.1.1. General Usefulness

Many users explained that the usefulness of the Web page is an essential factor in their selection criteria. For example, bookmarks "allow me to remember useful sites, and refer students to those sites." In addition, bookmarks are an "easy way to get back to useful sites on the web."

Usefulness is not only an important factor in bookmark creation, but in organizing and pruning a personal bookmark archive. One user described that usefulness plays a role in his organization process: "I typically only organize useful information." Another user deletes bookmarks which he no longer finds useful. He complains that "weeding ones [which are] no longer useful" takes to much time and cognitive effort. Consequently, a user's criteria for the usefulness of information changes over time (Schamber, et al. 1990). They must regularly re-evaluate the usefulness of pages stored in their personal archive.

5.2.1.2. Quality

The quality of a Web page is an important criteria for creating bookmarks. For example, bookmarks provide "quick access to key sites" and make it "easy to return to
key pages." Another user wrote, "I use [bookmarks] to get at and find important sites." They like having "direct access to important pages." They are "essential to finding good sites." Here, the relevance is a function of the perceived quality.

5.2.1.3. Personal Interest

Bookmarks are created so that users have fast access to sites of personal interest. Users bookmark sites which they find interesting. They like being "able to return to useful/interesting sites" and "the ability to quickly get to the URL of a site of personal interest." For many users, their bookmark archive represents their personal interest view of the Web. One survey respondent described his bookmarks as a "customized interest list." Users like the ability to "index [sites] which [are] interesting." The result of bookmarking a set of Web pages is a composite of the user's interests.

Personal interest also plays a role while exploring lengthy pages. For example, "They point to information that is too rigid. For example, one page has information of interest to too many people and sometimes getting at only what is of interest to me becomes problematic." The level of hypertext granularity currently available on the Web prevents this user from creating a bookmark to any particular location of interest on the page; unless the page has anchors, the user will return to the beginning of the page instead of the specific location of interest within that page.

5.2.1.4. Frequency of Use

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." "WWW bookmarks provide me easy access to my most frequently visited sites."
When users bookmark frequently used sites, they reduce their information foraging time; this is part of the process of optimizing the cost-structure of their information environment (see 2.1.3.4.).

5.2.1.5. Future Use

Users take advantage of the asynchronous aspect of bookmarks and mark pages for future reference during a browsing session. Instead of stopping to analyze the page in detail, a bookmark is created so that the page can be retrieved in a future session. For example, one respondent uses them to "store all the useful sites that I would use for future use." Another wrote that bookmarks "useful for retrieving sites for future sessions." A bookmark is created specifically to remind the user to come back to that page at a later date and complete the task associated with that page.

5.3. 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) create a personal information space for themselves and others. We have organized our taxonomy based on these three primary reasons for using WWW bookmarks.
The remaining sections of this chapter will describe each node in the taxonomy with user comments.

5.3.1. Reducing User Load

Bookmarks reduce the cognitive load and physical load of browsing hypertext (Wright 1991). Bookmark users avoid long URLs and their archives serve as a memory aid.

5.3.1.1. Avoiding Managing URL Addresses

Bookmarks insulate users from the tedious task of typing, managing, storing and interpreting URL addresses.
Facilitating Browsing: Avoiding entering URLs

Users like bookmarks because they do not have to re-type the URL address in order to return to the page. They like the "convenience of not having to retype a long Web site address." This speeds up access to sites and simplifies the browsing process because users do not have to pause to enter a long URL address. They are a "keystroke savings" and enable users to "ease of access a site without typing in the address." In general, users "do not have to type in the URLs again and again" when using bookmarks. The result is a fluid movement from one page to the next.

Facilitating Archiving: Avoiding Managing Detailed Site Information

Bookmark users do not have to write down URL addresses in order to keep track of important pages. For example, one user likes the fact that "writing down terse addresses is not necessary" with bookmarks. This enables users to quickly mark information for later use and continue on with their current information foraging task. Thus, bookmarks are a "convenience" to users because they "don't have to type any detailed information about the site." One user likes "the fact that they permit me to easily store an address without having to write it down." As a result, users avoid the time and cognitive effort needed to write down URLs when keeping track of information found on the Web.

Reducing Cognitive Load: Using Page Titles instead of URLs

Bookmarks enable users to abstract from the computer readable URL address and simply manage the title of each page. This reduces cognitive load and enables users to focus on the semantic representation of the page based on the author's textual title. URLs
are "cumbersome", "long", "tedious" and "terse." Bookmarks "make life easier; instead of remembering the address." They "save having to keep a logbook of Web addresses."

5.3.1.2. Serving as a Memory Aid and History Mechanism

Users remember previous tasks associated with bookmarks and browsing (see 3.3.). They use them as a cognitive prothesis to supplement memory and access the browsing history of previous sessions.

Serving as a Generic Memory Aid

Bookmarks serve as a "mnemonic device" for many users. They are a "memory jogger" because they remind users of important information they have found on the Web. For example, one user wrote "they [bookmarks] free me from the exhaustive task of remembering everything." Thus, bookmarks are a "memory replacement" and they "function as a memory list." This concept is similar to Bush's Memex or memory-extender (Bush 1945). Norman (1993) discusses a variety of ways that people use external artifacts to mediate cognition.

Facilitating the Remembering of Previous Tasks

Users bookmark pages to remind themselves of the associated tasks. For example, a bookmark "allows me to remember where I've been." They associate bookmarks with the temporal nature of a browsing session. One user wrote "I can keep track of what I was doing lately and a few weeks earlier" with bookmarks. In this light, a single bookmark represents an individual (or discrete group of) browsing session to the user. When reviewing a list of bookmarks, the user sees a time-ordered view of separate
information foraging tasks over a period of days or weeks. One user wrote that he views bookmarked "pages as a history of using the Web."

Providing an Inter-Session History Mechanism

Bookmarks provide the only currently available means of keeping browsing history information between Web sessions. One user wrote that bookmarks "take me to the last site where I was not finished during the last session." Not only does this user associate a task with individual browsing sessions, but he also uses bookmarks as an inter-session history mechanism. Remembering previous sessions is important to this users who wrote, "bookmarks are essential to find good sites again and remember previous sessions." Another user likes bookmarks because "they are useful for between sessions to store or return to for future research."

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. For example, "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. Often, the page is still in my Netscape cache, but I have no way to find it." This demonstrates that 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 so as to access previous browsing sessions.
5.3.2. Facilitating Navigation/Access

Bookmarks are used to jump between Web localities (see 3.2.1.). Users create bookmarks to help them quickly find Web pages, preserve their time spent foraging, and keep track of serendipitous sites found along the way.

5.3.2.1. Speeding Information Access

Bookmarks reduce the time spent returning to a site. In particular, the use of bookmarks “is a time-saver and a shortcut to information sources.” Another user wrote that bookmarks are “really handy” because they provide “quick retrieval” of information.

Serving as a Readily Available Information Resource

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.” In chapter 7 we discuss bookmark retrieval in more detail.

5.3.2.2. Finding Web Information

Bookmarks are used to relocate information found on the Web. For example, "They provide about the only way to easily re-locate web pages." Bookmarks are “extremely useful for going back to certain pages” because they make it “easy to return to key pages.”

Avoiding Search Engines

Some users bookmark pages to avoid using a search engine. For example, a user queries the Web with AltaVista and browses to a site of interest. The user bookmarks the
site so that he/she does not have to perform the query over again to re-locate the page. One user wrote that "bookmarks are essential" because "I do not have to 'search'" when using bookmarks.

There exists a tradeoff between the amount of time spent using a search engine and keeping an organized set of bookmarks. This comparison can be characterized by:

<table>
<thead>
<tr>
<th>Idealized Page Retrieval Hueristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$ = desired page</td>
</tr>
<tr>
<td>$A$ = user's personal archive (size and degree of organization)</td>
</tr>
<tr>
<td>$T_1$ = retrieval time of locating a search engine, initiating the correct query and browsing the results to find page $P$</td>
</tr>
<tr>
<td>$T_2$ = retrieval time of finding the bookmark for page $P$ in the user's personal archive $A$. Cost of managing the organization of the archive $A$ is amortized over all retrieval times $T_2$,</td>
</tr>
</tbody>
</table>

$$\begin{align*}
(T_1 < T_2) \text{ THEN } & \text{ <retrieve via search engine>} \\
\text{if} & \\
(T_1 \geq T_2) \text{ THEN } & \text{ <retrieve via bookmark>} 
\end{align*}$$

This simplified analysis demonstrates that the optimal method depends on the estimated cost of information access.

Preserving Information Foraging Investment

Bookmarks preserve the time invested locating a particular site. "Without them a great deal of collective time would be wasted." In short, users like "the ability to retrieve quickly sites that you've spent [time] finding." This caching mechanism can be used to optimize a user's information foraging by keeping an index to important results.

Taking Advantage of Serendipitous Discoveries

A personal archive enables users to keep track of serendipitous information found on the Web. One user likes "the ability to organize and return to useful locations"
because very often you can 'stumble' upon something useful which you may not have been looking for.” This illustrates the serendipitous nature of browsing the Web. For example, “many times I run across a useful page by chance - like finding an interesting book in a library when looking for another volume - without on-line-bookmarks I might not find my way back to that page.” The personal archive enables a user to keep track of these serendipitous sites and study them at a later time.

5.3.3. Collaborating/Publishing/Archiving

Users collect bookmarks so that they can create their own personal information space and share it with others.

5.3.3.1. Creating a Personal Web Information Space

Some users like the ability to create a personal archive of Web pages. 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.”

5.3.3.2. Authoring and Publishing Web Pages

Users create Web pages out of bookmarks by authoring an HTML file with bookmarks as hypertext links. Users are able to easily add annotations and create customized views of their bookmarks with an HTML file. They are able to manage bookmark organization in a text file instead of using the folders interface (see 6.2.). An HTML file of bookmarks enables users to view and scroll through their bookmarks using
the browser's main window. Other users can access their bookmark archive and the user can author professional Web pages based on his/her bookmarked URLs.

**Providing an Editable HTML Bookmark File**

Bookmarks are usually stored in HTML format. One user likes the fact that "the hot-list provides a user editable HTML file." Users tend to bookmark pages of interest and move these to a personal Web page with annotations (see 6.2.1.5.).

**Publishing Web Information**

The resulting Web page can be published on the Web. This gives others easy access to the links and the supplementary annotations authored by the user. Bookmarks are often used in the literature search or collection phase of projects. Then a Web page is created out of the results. One user describes "*Project Bookmarks: a number of references needed to work on a particular project and deleted afterwards.*" For example, "*If I'm working on an Olympic web site I will surf the web and bookmark all Olympic sites I can find, and delete them when the job is completed.*" Here bookmarks are used as a temporary storage space before authoring a Web page with the selected URLs. Then users share this page with others: "*I download URLs and other details about the site; I load the into a database on my PC and then create a list which I mount on the UofT Astronomy department home page.*"

**5.3.3.3. Collaboratively Using Web Information**

Bookmarks are used to share Web resources with third parties. A group of users working on the same project will share bookmarks in order to collaborate on joint tasks. For example, "*I just mail collaborative bookmarks to other people.*" 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 user wrote that sharing bookmarks "is very valuable for the teams I work on. They often say thanks for doing this. I keep 5 bookmarks for common/free accessed links (search AltaVista) phone, directory, and stock quotes." Another user explained he uses bookmarks "my course to provide students with initial navigational markers." A librarian wrote that she "collects [bookmarks] for clients."

Enabling Presentations with Bookmarks

Bookmarks are used in presentations and lectures. One user wrote "I mostly use them while giving lectures." He spends time locating valuable resources on the Web. A set of bookmarks are created and he uses them during his lecture to move quickly from one site to the next. No browsing is required to find a particular page during the presentation.
6. PWIS #3: ADD VALUE THROUGH ORGANIZATION

Users *economize* their management of information. They try not to waste their time and energy organizing information from which they will not benefit. In a personal Web information space, we find that users rarely organize until they are forced to do so. In this chapter we look at the methods and frequency with which users organize bookmarks, and how the size of the archive influences bookmark organization and management.

Users organize their archive’s in order to fight entropy. This maintenance process creates many problems. For example, “There’s not an easy way to maintain them. I want to add and delete easily. I want to file them under multiple categories, rank their importance, and store comments about them. I’d also like to be able to create subsets that I could send to others.” Users explained that there is “no quick and easy way to organize them.”

This “labor intensive” process “requires extensive time to manage effectively.” A few survey respondents wanted to avoid the task altogether: “I would like someone else to organize for me.” A steep learning curve often impedes organization. Users complained that bookmarks are “not intuitive for reorganizing, and I can’t be bothered with learning how to do it right.”
6.1. ORGANIZATIONAL METHODS

Users organize bookmarks through a variety of methods. In this section we look in detail at the various ways users organize their bookmarks and how this relates to browsing the Web.

6.1.1. Organizational Methods Defined

Users have different habits for organize bookmarks. The method and frequency depend on the user’s particular habits and work context. Users tend to fall into seven different categories which describe how they organize bookmarks.

<table>
<thead>
<tr>
<th>Methods Used to Organize Bookmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Organization: users who do not organize bookmarks. The bookmarks stay in the order in which they were created.</td>
</tr>
<tr>
<td>Ordered List: users who manually re-arrange bookmarks in a list.</td>
</tr>
<tr>
<td>Set: users who create folders to group together related bookmarks.</td>
</tr>
<tr>
<td>Hierarchy: users who create folders within folders to create a hierarchy of bookmarks.</td>
</tr>
<tr>
<td>External: users who export their bookmarks to a separate system (e.g. database, bookmark organization program).</td>
</tr>
<tr>
<td>Web Page: users who create Web pages out of their bookmarks in order to organize them.</td>
</tr>
<tr>
<td>Other: users who search, sort and use unusual organizational schemes.</td>
</tr>
</tbody>
</table>

The chart below shows organizational strategies of users surveyed.
6.1.1.1. No Organization

36.9% of respondents checked "I don't organize bookmarks - they stay in the order which I created them." (see Appendix A) The chart below shows the percentage of users in each category of bookmarks who do not organize bookmarks.

![No Organization Chart](chart.png)

This graph illustrates the steady decrease in percentage of users who do not organize as the number of bookmarks increases.

Why do 36.9% of all users not organize their bookmarks? There are three primary factors: expertise, time/effort, and number of bookmarks. Users must weigh the costs of organizing bookmarks against the expected gains. This behavior in a personal information space is one form of economizing the management of Web information.

Expertise Required is too High

Users do not organize bookmarks when they do not have the expertise needed to use the tools. This may result from lack of experience. For example, one user wrote "I don't know how to organize them." Users encounter problems learning to use the existing tools.
**Time/Effort Dissuades Organization**

The time and effort needed to organize bookmarks can dissuade users from even attempting. Organizing bookmarks is labor intensive, requires extensive time and is difficult to do. As a result, many users avoid the chore of organizing.

**Scale Does Not Warrant Organization**

Users who do not have many bookmarks do not need to organize them (i.e. economy of management does not support organization at this point). Beginners have not collected enough bookmarks to require any organization. "I'm just getting started, I didn't even consider this [organizing bookmarks] yet." Our analysis shows than until a user collects 35 bookmarks he/she is not required to organize them. That is, the user can visualize and retrieve bookmarks in a simple list, in the order in which they were created.

6.1.1.2. Ordered List

Very few users manage a list without folders. The graph below shows how few users rearrange a list. The largest percentage group is 14.7% of users with 11-25 bookmarks manually rearrange bookmarks in a list.

The first two groups can see all of their bookmarks on the screen at the same time.
6.1.1.3. Set of folders

A set of folders is equivalent to a single-tiered hierarchy. These are users who do not have a deep hierarchy like the folders within folders group, but instead employ a simple use of folders.

An analysis of variance indicated that the self-reported frequency of creating folders differed with the reported total number of bookmarks ($F_{4,257} = 6.24, P < 0.001$). Post-hoc analysis (Tukey) indicated that the group with the most reported bookmarks (300+) differed significantly from all other groups ($P < 0.05$).

The chart shows that a higher percentage of users in the 1-10 and 11-25 bookmark groups organize them with a single tier of folders instead of a hierarchy of folders within folders. The use of a set of folders peaks with the 101-300 user group and drops dramatically when users collect 300 or more bookmarks. The 300+ bookmark users seem to prefer to use a hierarchy in order to manage all their bookmarks.
6.1.1.4. Hierarchy of folders within folders

Users with 26-100 and 101-300 create folders within folders 28.7% and 24.1% respectively. A very significant percentage, 44.4% of users with 300+ bookmarks create a hierarchy.

The reported frequency of use of sub-folders (a hierarchy) differed significantly with reported total number of bookmarks ($F_{4,249} = 3.63, P < 0.01$). Post-hoc analysis indicate that group with the most bookmarks (300+) differed significantly from the 3 groups with the lowest number of bookmarks (none, 1-10, and 11-25) ($P < 0.05$).

6.1.1.5. Web Pages Recording Bookmarks

2.2% of users accumulate bookmarks and then author a Web page from the collection of URLs. "I make web pages out of them." "I put links on my bookmarks page." One user described how he creates a Web page once his collection reaches 10 bookmarks: "I create web pages once I get more than 10 bookmarks."

11.8% of 1-10 bookmark users create HTML Web pages from their bookmark list. The reason these users only have 1-10 bookmarks is because all of their bookmarks are being off loaded to separate HTML files where they are independent of the browser. These users avoid the functionality provided by the browser and have specifically chosen to create Web pages out of their bookmarks instead of using the existing tools. The table below lists comments provided by users regarding the authoring of HTML pages based on bookmarks.
**Bookmarks Integrated with Descriptions and Organizational Structure**

Web pages enable users to include comments and links in the organizational structure. This provides a powerful and expressive medium for managing bookmarks. "I organize certain categories into home pages." I organize "by section in sub-pages of my home page." "I create my own web page to manage bookmarks." These users are very experienced (see 6.1.2.).

**The Need for Seamless Migration of Bookmarks to the Web**

Bookmarks "should have browser support to automatically put bookmarks in a full hypertext page and move them easily so I can incorporate them in my notes." The ability to easily migrate bookmarks to an HTML page is essential. The current tools create an HTML file of bookmarks. More innovations will definitely be needed in this area in the future (e.g. integration with HTML authoring, automatically updating database, etc.).

6.1.1.6. **External Applications**

2.9% of users in our sample create bookmarks and then off-load them to a separate application. There are a variety of motivations behind this organizational strategy. Existing tools are difficult to use, they are not powerful enough to manage a large number of bookmarks, and they have a steep learning curve. One user wrote, "I didn’t like something the first time I’d used them so I decided to organize them myself in a text-file." "I store them externally because they slowed down my browser." 44.4% of users with 300+ bookmarks offload them to separate programs.
Database

A database can be used to organize bookmarks. One user organizes with a hierarchy in the browser and “Lotus Notes database connected to HTML via InstantNotes WebPublisher” to offload bookmarks.

Text File

A simple text file provides one means of organizing bookmarks. This enables users to organize with their favorite editor and incrementally build up a structure. “I manually arrange in a file with heading index.” Annotations can be integrated into the file: “I keep them in a browser-independent file with my own comments.” Lists can be used to categorize bookmarks: I use “multiple bookmark lists by topic or work/personal/public.” Another uses “several separate files” to organize bookmarks.

Bookmark Management Tool

Special bookmark management tools supplement the existing tools. For example, one user does not organize “a few key bookmarks,” but uses a “separate organizer program, Grabnet” to manage bookmarks. Another uses "Smartmarks and folders" to organize them.

6.1.1.7. Other Methods

1.3% of users apply other organizational methods. Below is a short list of the various methods used.

Searching: “I use the find command when I want something.”
Sorting: "Alphabetized automatically."
Divided List: Manually rearrange in a list "with separation lines."
Thematic Lists: Organize bookmarks "thematically, into a list of links."
6.1.2. Experience Level of Organizational Methods

The experience level of users increases with the sophistication of the organizational method used. The graph below shows this relationship with mean and one std error bars.

![Experience Level of Organizational Methods](image)

A Spearman correlation coefficient was calculated and experience group was found to significantly correlate with organization method ($r_{291} = 0.3333$, $P < 0.001$). Users who author Web pages out of bookmarks have a mean experience level of 6.50, while those who do not organize have mean of 4.36.

6.2. ORGANIZATIONAL FREQUENCY

Users file bookmarks at different times and rates. The table below summarizes the four basic user groups.

<table>
<thead>
<tr>
<th>Frequency of Organizing Bookmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Filers:</strong> users who never organize bookmarks. The bookmarks stay in the order in which they were created.</td>
</tr>
<tr>
<td><strong>Creation-time Filers:</strong> users who store a new bookmark in the appropriate category when the bookmark is first created.</td>
</tr>
<tr>
<td><strong>End-of-session Filers:</strong> users who organize all their new bookmarks at the end of the session.</td>
</tr>
<tr>
<td><strong>Sporadic Filers:</strong> users who organize bookmarks occasionally or sporadically. Sporadic Filers may go many sessions without organizing bookmarks.</td>
</tr>
</tbody>
</table>
The graph below shows the percentage of users surveyed for each frequency of use category.

Approximately half of all users surveyed organize their bookmarks sporadically. 26% never organize bookmarks, while 23% store each new bookmark in its proper place at the time it is created. Very few users, 7%, organize bookmarks at the end of every session.

6.2.1. Explaining Organizational Frequency

Users spend the time and energy to organize bookmarks because of the expected benefits which they will get back in return (i.e., the economy of information management). The rate at which they organize is based on a series of conscious or unconscious cost-justifications. Whether to organize or not to organize lies at the heart of this process. Users must mitigate the costs of organizing with the beneficial results. In sections 6.2.1.1. through 6.2.1.3. we will describe the habits of users who organize bookmarks.
6.2.1.1. Sporadic Filers

Almost half, 49%, of users surveyed file sporadically in spring cleaning sessions.

<table>
<thead>
<tr>
<th>Times per week</th>
<th>Experience Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>under 10 min.</td>
<td>5.495</td>
<td>1.191</td>
</tr>
<tr>
<td>11-20</td>
<td>10-30 min.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>30-60 min.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>1-2 hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td>2-3 hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+</td>
<td>3+ hr</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sporadic filers choose when to optimize their work environment. They start a special spring-cleaning session to organize their bookmarks.

As a result of these decisions to clean up their bookmark, they tend to “organize 2-3 times/week” or “organize once a week.” Sometimes this is a Friday afternoon ritual: “I organize at the end of the week (a Friday project).”

6.2.1.2. Creation-Time Filers

Creation-time filers categorize a new bookmark and store it in the appropriate folder at the time when the bookmark is created. These users avoid a pile-up of unorganized bookmarks because they “organize often at the time of creation.”

They take advantage of situational awareness by filing immediately after reading the Web page. This helps reduce representation failure problems (see 7.2.).
23% of all users surveyed file at creation time. 67% of users with over 300 bookmarks file at creation time. This is in contrast to the 15%-24% in each of the other groups. The graph below illustrates this relationship.

Why do so many users with 300+ bookmarks tend to file at creation-time? 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 far outweighs the distraction from browsing that is required to file a bookmark. Below is a list of plausible motivations:

- Bookmark users must keep a stable ontology in order to manage them in a semantic hierarchy. The scale of the folder system forces them to keep it up-to-date.
- Retrieval is incredibly difficult in an archive of over 300 bookmarks. In order to warrant finding a bookmark in the archive instead of using a search engine, the user must keep items filed in a structured environment. This way the archive will facilitate efficient retrieval.
- The archive's structure may be important for reasons other than simply bookmark retrieval. For example, users with a large detailed hierarchy may share this resource with others. Collaboration drives organization.
6.2.1.3. End-of-Session Filers

End-of-Session filers organize bookmarks at the end of every browsing session.

<table>
<thead>
<tr>
<th>Times per week</th>
<th>Experience Level</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>under 10 min.</td>
<td>4.906</td>
<td>1.573</td>
</tr>
<tr>
<td>11-20</td>
<td>10-30 min.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td>30-60 min.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td>1-2 hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td>2-3 hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+</td>
<td>3+ hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td>6.25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81.25%</td>
<td>12.50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td>6.25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td>6.25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.25%</td>
<td>18.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.00%</td>
<td>31.25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organization Method</td>
<td>Number of Bookmarks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does organize</td>
<td>93.33% Some use folders</td>
<td>80.00%</td>
<td>1-10 BK: 25.00%</td>
</tr>
<tr>
<td>Not Organize</td>
<td>6.67% Folders</td>
<td>40.00%</td>
<td>11-25 BK: 12.50%</td>
</tr>
<tr>
<td>Rearrange list</td>
<td>13.33% Nested folders</td>
<td>40.00%</td>
<td>26-100 BK: 50.00%</td>
</tr>
<tr>
<td>Web pages</td>
<td>0.00% Offload</td>
<td>0.00%</td>
<td>101-300 BK: 6.25%</td>
</tr>
</tbody>
</table>

Very few users (7%) organize bookmarks at the end of every browsing session. There are many reasons why so few users file at the end of each browsing session. In general, there are very few benefits to this method. Three primary factors describe this behavior:

- End-of-session filers must set aside a special period of time to organize at the end of each session. As a result, they do not benefit from contextual awareness like creation-time filers.
- Most users, 94% of all respondents, create 0 to 5 bookmarks per WW browsing session. There is no pile-up of unorganized bookmarks at the end of each session which forces them to file.
- They must incur the costs of placing items in folders before ending the session. Without any real factors forcing them to file (e.g. cannot retrieve an item), the benefits do not outweigh the costs.

End-of-session filers tend to have the following characteristics:

- **Longer sessions**: An analysis of variance indicated that the average time spent per session varied significantly with when bookmarks were organized ($F_{3,278} = 3.5237$, $P < 0.0155$). Post hoc analysis indicates that subjects who organize their bookmarks at the end of the session spend significantly more time per session than do subjects who either sporadically organize or never organize ($P < 0.05$).
- **Heavy use of folders**: 80% these end-of-session filers use folders in some manner to organize bookmarks.
- **No difference in number bookmarks created per session**: An analysis of variance indicated that the average number of bookmarks created per session did not vary significantly with when subjects organized their bookmarks ($F_{3,272} = 2.4249$, $P=0.0660$).
6.2.2. Factors Influencing Organizational Frequency

Users must justify the cost of filing bookmarks with the expected returns of faster retrieval time, better visualization, a cost-tuned information workspace, and a structured index to share with others. These factors which influence the frequency with which users organize bookmarks are essentially reasons why users organize their bookmarks. We can then explain user behavior in terms of the expected gains they hope to receive as a result of filing a single bookmark. In all cases, the expected rate of return should exceed the predicted cost of filing.

6.2.2.1. Cost Structure Matched to Current Usage Patterns

Users organize their archive when the cost-structure of this information workspace needs tuning. This is in an attempt to optimize the cost of retrieving items based upon expected frequency of use. Since the cost structure is based on the relevancy of information units or the frequency with which users re-visit bookmarked Web pages, it can change as the user's tasks and interests change. Thus, the user expends the energy to organize his/her personal Web information space to meet current needs and create an up-to-date, optimal environment for doing work. Usage patterns are matched against the environment.
6.2.2.2. Reduced Retrieval Time

When the cost of retrieving a bookmark from the archive becomes too high, it is time to organize them. Users organize their bookmarks in order to reduce the retrieval time. In order for the user's personal Web information space to be a viable entity, the time to retrieve a bookmark must be less than other browsing strategies (e.g. search engine) for re-locating Web sites.

6.2.2.3. List Too Long to See

One user said that he organizes bookmarks "when they no longer fit in my drop down menu." Other users complain about not being able to see all their bookmarks on the screen at one time because of "long lists" of bookmarks. Thus, users can live with a pile-up of unorganized bookmarks until they can no longer see them because they run off the screen. When this happens, retrieval is impossible unless the user switches to the special bookmark archive window or separate tool.

Users create folders and file bookmarks into them during a special organizational session in order to split-up the pull-down menu into tiers. In short, the inability to view all the items within the integrated pull-down menu system may force users to organize their bookmarks.

6.2.2.4. Collaboration and Publishing

Users sometimes organize bookmarks when they expect to share them with others or publish them on the Web. Collaboration is a motivating factor because the archive must be organized such that another people can interpret its structure. Users publish Web pages for a variety of reasons. Regardless of the exact driving force behind Web page authoring, they usually take the time to organize bookmarks into a coherent structure.
within the Web page. This demonstrates an alternative organizational method to filing in folders. In summary, both collaboration and publishing are means of distributing a user's personal Web information space. Before distributing it, user's generally create at least some structure.

6.2.2.5. Time to File Influences Frequency and Method

Users do not have an immediate filing mechanism for storing a bookmark in their archive. The time it takes to file a bookmark not only affects the organizational frequency but the method as well. For example, “I do both [organize at creation-time and sporadically] depending on my surfing objective at the time.” The time it takes to file a new bookmark affects his choice of when to file. If a user does not care about being distracted, then he/she may file at creation-time. In contrast, if the user is working on a very directed search and serendipitously finds an interesting page, he/she may bookmark it and file it later. The lack of an immediate filing mechanism keeps some users from filing at creation time. “I would like to store each new bookmark in its place when I create it but it is too slow and awkward with Netscape.” One user wrote that he files bookmarks at creation-time when creating “via LotusNotes”, but files sporadically when using the browser.

6.3. INFLUENCE OF SCALE ON ORGANIZATION

Users take the time to organize their bookmarks so as to add value to the structure of their archive. They economize their energy and tend to only organize when they have to. In this section we look at how the size of a user’s archive affects the organizational methods used.
6.3.1. Aggregate Organizational Methods

The pull-down menu for bookmarks holds approximately 35 items. Most users with less than 35 bookmarks have 0 folders in their archive. They do not need to create a folder because their bookmarks will fit in the list. Once the user's archive grows beyond the size which fits on the screen, they are forced to use folders so that they can get access to them all with the pull-down menu.

![Use of Folders Graph](image)

After the 35 bookmark threshold, users create folders in direct relation to the number of bookmarks in their archive. The total number of bookmarks and the total number of folders was correlated \( (r^2 = 0.60) \). The following linear regression equation was calculated to predict the number of folders by the number of bookmarks:

\[
\text{NumberFolders} = 1.14 + 0.14 \times \text{NumberBookmarks}
\]
This equation is significant, \(F_{1.26} = 38.52, P < 0.0001\). This shows a strong linear relationship between folder use and the number of items.

Users organize with different methods depending on the number of bookmarks they have. The chart below shows a steady increase in the use of organizational methods corresponding to the number of bookmarks.

![Chart showing aggregate use of organization methods](chart.png)

This chart shows the aggregate use of organizational methods in relation to the number of bookmarks a user has. The category "some organizational method" is a composite of all organizational methods presented on the survey. For example, 70.2% of users with 26-100 bookmarks use some sort of organizational method to manage bookmarks.

The use of folders is one instance of an organization method. The chart shows the steady increase in the percent of users who create and manage folders as the number of bookmarks increases. Obviously, from each user group the percent using folders is less
than the percent organizing in any way. For example, 54.9% of users with 11-25 bookmarks organize them in some way. 34.3% use folders to organize bookmarks.

6.3.2. User Characteristics by Scale

We generated summary statistics from the survey respondents to provide snapshots views of different users grouped together by the number of bookmarks as in the survey. The first group has very few bookmarks:

<table>
<thead>
<tr>
<th>Number of Bookmarks</th>
<th>Experience Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Times per week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td></td>
<td>4.529</td>
<td>1.375</td>
</tr>
<tr>
<td>11-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>under 10 min.</td>
<td></td>
<td>11.76%</td>
<td></td>
</tr>
<tr>
<td>10-30 min.</td>
<td></td>
<td>44.06%</td>
<td></td>
</tr>
<tr>
<td>30-60 min.</td>
<td></td>
<td>23.53%</td>
<td></td>
</tr>
<tr>
<td>1-2 hr</td>
<td></td>
<td>5.88%</td>
<td></td>
</tr>
<tr>
<td>2-3 hr</td>
<td></td>
<td>5.88%</td>
<td></td>
</tr>
<tr>
<td>3+ hr</td>
<td></td>
<td>5.88%</td>
<td></td>
</tr>
</tbody>
</table>

The next group of users can still see all of their bookmarks in the pull-down menu.

<table>
<thead>
<tr>
<th>Number of Bookmarks</th>
<th>Experience Level</th>
<th>Mean</th>
<th>St. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-25</td>
<td></td>
<td>4.647</td>
<td>1.246</td>
</tr>
<tr>
<td></td>
<td>Times per week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41-50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>under 10 min.</td>
<td></td>
<td>3.92%</td>
<td></td>
</tr>
<tr>
<td>10-30 min.</td>
<td></td>
<td>33.33%</td>
<td></td>
</tr>
<tr>
<td>30-60 min.</td>
<td></td>
<td>13.73%</td>
<td></td>
</tr>
<tr>
<td>1-2 hr</td>
<td></td>
<td>1.96%</td>
<td></td>
</tr>
<tr>
<td>2-3 hr</td>
<td></td>
<td>3.92%</td>
<td></td>
</tr>
<tr>
<td>3+ hr</td>
<td></td>
<td>3.92%</td>
<td></td>
</tr>
</tbody>
</table>

The largest group of users is shown below:
Heavy bookmark users are forced to organize more than those with less bookmarks.

The users shown in the group below offload their bookmarks to a separate organizer 44% of the time.
6.3.3. Factors of Scale Affecting Organization

In this section we describe some of the factors underlying the way users organize as the size of their archive increases.

6.3.3.1. Users Find it Difficult to Organize a Large Number of Bookmarks

Users complained about "the inability to organize large quantities of bookmarks." One user wrote that there is "no good organized way to control an increasing-sized bookmark list." Not only is difficulty an issue but the time required limits a user's effectiveness. "My problem is that [there is no] time to organize masses of information."

6.3.3.2. Large Scale Requires More Organization

User's with a large number of bookmarks need a way to organize them. One user wrote that "once you're over 100 [bookmarks], [you] need a means to index [your bookmarks]." Organization becomes more important as the size of the archive grows.

For example, "I would like to be more organized to cope with the increase." Another user wrote "I've never bothered to use folders, but I don't have an enormous number of bookmarks... yet!"
6.3.3.3. Poor Scalability Forces Alternate Organizational Methods

Some users with a large number of bookmarks have begun to use alternate methods of organization because the current tools are not effective when managing large numbers of bookmarks. "I've browsed for 2 years and the biggest problem with bookmarks is the list gets too big too fast and the ability to organize this is poor (plus I can't share these with anyone, etc.) So I have my own web page to manage, organize, add/remove and shuffle." Off-loading bookmarks to Web pages was most common in users with 300+ bookmarks.
7. PWIS #4: ATTEMPT TO STRUCTURE FOR RETRIEVAL

Users attempt to structure their personal Web information space so that it is tuned for fast retrieval. In this chapter we discuss problems encountered when structuring a personal Web information space with a semantic hierarchy. Section 7.1 illustrates how some of the traditional problems of filing and retrieval appear when users manage bookmarks. Section 7.2 discusses representation issues which limit retrieval. We shall see that, although users may attempt to structure their archives for fast retrieval, they encounter many problems.

7.1. FILING AND RETRIEVEING BOOKMARKS

In this section we describe the process of retrieving a bookmark from the archive. We identify significant factors which affect retrieval and relate bookmark retrieval to other techniques for locating a Web page.

7.1.1. Selecting a Bookmark Retrieval Technique

There are a variety of techniques which can be use to retrieve a bookmark. Below is a general list of retrieval methods.

<table>
<thead>
<tr>
<th>Bookmark Retrieval Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pull-Down Menu</strong></td>
</tr>
<tr>
<td>Select an item from the Bookmark pull-down menu.</td>
</tr>
<tr>
<td><strong>Bookmark Window</strong></td>
</tr>
<tr>
<td>Open the separate bookmark window and select an item from it.</td>
</tr>
<tr>
<td><strong>Web Information Management Program</strong></td>
</tr>
<tr>
<td>Use a separate Web information management program to retrieve the bookmark.</td>
</tr>
<tr>
<td><strong>Personal Web Page</strong></td>
</tr>
<tr>
<td>Access a Web page authored with bookmarks.</td>
</tr>
<tr>
<td><strong>Desktop Alias</strong></td>
</tr>
<tr>
<td>Locate and select an alias (e.g. MS Win95 &quot;shortcut&quot;) on the desktop file system.</td>
</tr>
</tbody>
</table>
The user must select the optimal technique. This is based on where the bookmark is stored and the user's effectiveness with the available tools. If the user believes the pull-down menu is too awkward for wading through a large number of bookmarks, then opening the separate Bookmark Window may be the best alternative. In contrast, if the user remembers creating and annotating a Web page based on bookmarks, then selecting one of the URL links on his/her page would be the optimal method.

7.1.2. Retrieving from a Semantic Hierarchy

Finding a bookmark in a semantic hierarchy is based on an interactive process between browsing a hierarchy of folders and making cognitive inferences to direct the search. The heart of this process is remembering which folder a particular item was stored in. Research shows that users encounter many problems when retrieving items from folders (Malone 1983).

7.1.2.1. “What folder did I put it in?”

Users must remember which folder a bookmark is stored in. Research shows that it is very challenging to find an item deeply nested inside a hierarchy (Lansdale 1988). Accessing long-term memory is based on initiating and making inferences. Many times the top level of the semantic hierarchy does not present enough distinct mental cues in order to spawn a chain of inferences successfully. The user must browse through the top tier of folders until the user stumbles upon a distinctive cue, key chunk of information, or sense-making attribute which reminds the user of where the item can be found. This process iterates until the search has been refined enough so that the correct folder can be located within a branch of the semantic hierarchy's tree-like structure.
7.1.2.2. "What does this folder name stand for?"

While maneuvering through the semantic hierarchy, the user must interpret the meaning of each folder name in order to associate it with the bookmarks encapsulated within. This is a problem because the user's interpretation of a textual title can change over time.

The meaning of a categorical name is tied to the information it encapsulates. When a folder name is originally created, it is based on a finite set of items. As the user adds bookmarks to the semantic hierarchy, the meaning of the folder name can change dramatically. In addition, the content of the pages change without the user's control since Web resources are distributed throughout the Internet. The user may not know the contents have changed until he/she re-visits the page.

7.1.3. Creating a Semantic Hierarchy

Users encounter many problems with a personal filing system (Jones 1988). In this section we review two problems: naming folders and forcing a rigid structure.

7.1.3.1. Naming Folders

Naming folders is very difficult for many users (Carroll 1982). It requires cognitive effort and time. This is because the user must analyze the list of bookmarks, and then decide on a baseline set of metrics for dividing the items into groups. Next the user must evaluate all items in a particular group and generate a succinct textual title or representative label for the information.

Domain expertise is needed to generate a categorical name. Representing a set of Web pages with a textual title requires an understanding of the information in the Web pages in question. A more complicated, fine-grain semantic hierarchy forces users to articulate a more detailed and precise taxonomy. Users with a semantic hierarchy of 300 or more bookmarks must have a keen understanding of the information they are creating.
categories for, while users with a hierarchy of small bookmarks can avoid such domain expertise.

7.1.3.2. Forcing a Rigid Structure

When semantic hierarchies are the only effective organizational tool available to users, it forces them to apply a rigid semantic structure to disparate information. It can be distracting for a user to take the time to build up this structure. For example, a user has collected 50 bookmarks. If this user wants to organize them in some fashion to make retrieval efficient, then the user must create folders and build up a semantic hierarchy. This process diverts the user from the principle task of information foraging or finding Web pages. Users who like a rigid semantic structure should be able to impose one, but this should not be the only method with which users can organize Web information.

Applying structure to Web information is particularly difficult because of the disparate nature of Web pages and the rapid evolution of the Web itself. Web page authors vary widely in intent, skill, writing style, domain expertise, and graphic design expertise. The HTML standard is rapidly evolving and the number of new entrants into the Web publishing market is increasing. The embryonic nature of the Web as a medium of authoring, presenting and interpreting information makes it difficult to categorize Web pages. Therefore, creating a stable categorical structure for this type of information is very challenging.

7.1.4. Managing a Stable Semantic Hierarchy

Categorizing information requires a stable hierarchical structure to support incremental growth. As users find more Web pages to bookmark, they add items to their hierarchy. The naming structure must be stable enough to facilitate the filing process and support efficient retrieval. A stable ontology is a set of categories that does not change over time. The structure must be consistent.
7.1.4.1. Remembering the Classification System

A stability of the ontology is based, in part, on the level of understanding the user maintains about the exact nature of the classification system. For example, one user wrote that bookmarks are "difficult to organize." He explained that "part of this problem is software related, also it is related to creating stable categories in one’s mind." When he refers to "stable categories" he is talking about the fact that he must create the ontology by naming categories and consistently grouping bookmarks according to their semantic relationship. This entire process increases cognitive load because the user must remember the classification system which forms the basis of the semantic hierarchy. A breakdown may lead to incorrect similarity measurements and items being placed in the wrong category or an inconsistent organizational structure resulting from a constantly changing mental model of the classification system.

7.1.4.2. Significance Changes over Time

Managing a stable ontology is very difficult for many users because the significance of categories changes over time. An area of research may be important and useful for a short period of time, but as the user moves on to different projects its relevance fades (Schamber, et al. 1990). Users collect Web information based on personal criteria or value judgments. When enough information in one area is collected, a folder is created to store the information. This category is based on the user's interests at the time. When this interest level fades, the significance of the category diminishes as well. Thus, keeping a stable categorical structure for Web information is difficult.
7.1.4.3. “Why did I put it in this folder?”

This problem surfaces when the user cannot remember the criteria he/she used to file the bookmark into a folder. After stumbling upon an item in a part of the semantic hierarchy, the user identifies an inconsistency with his current conceptualization of the folder’s semantic representation. This interpretation is based on the location of the folder within the entire hierarchy, the textual label assigned to the folder, the items stored in the folder, and the user’s current mental model.

“Why did I put it in this folder?” is a manifestation of two underlying causes (1) recalling the decision used to file an item is difficult and forces the user to access long-term memory (2) the significance of parts of the semantic classification scheme changes over time as the user’s interests change.

7.1.5. Filing Items in a Semantic Hierarchy

Sometimes the user needs to store an item under multiple categories. The user needs a way to quickly select multiple categories to store the bookmark. With the current browsers, a user must go through all the steps to create a bookmark and store it in the correct location. For example, to store a new bookmark under two different headings would require the user to create two bookmarks separately and store one at a time.

7.1.5.1. Lack of an Immediate Filing Mechanism

Filing a bookmark requires the cognitive tasks of (i) selecting a bookmark to file, (ii) reviewing the semantic hierarchy, and (iii) selecting a destination folder. In addition, the user is burdened with the interactive tasks of opening the archive window, browsing the hierarchy, and moving the bookmark into its folder. Users complained about the tools available to file bookmarks. For example, one user wrote that interacting with the interface is “too slow to immediately file the bookmark in the appropriate place.” The
user must always switch context to the bookmark archive window, and perform "multiple clicks and drags" to file the bookmark.

7.2. BOOKMARK AS REPRESENTATION

A bookmark represents the URL address of a Web page. In this section we look at the issues involved in creating, managing and interpreting these representations. Good representations are needed for fast retrieval from the user’s archive.

A bookmark is, first and foremost, something that stands for a URL. Usually the bookmark is a textual title or name which represents the content material stored in the Web page. The diagram on the next page illustrates this concept.

The bookmark name should reflect some aspects of the page being represented. It models a set of relations among conceptual objects of the page.
7.2.1. Problems with Bookmark Representation

Users encounter problems because Web information is dynamic and distributed. The titles do not always effectively describe the contents of the page. When the representation breaks down, the user must revisit the page to remember its content.

7.2.1.1. Non-Descriptive Titles

"Bookmarks are not descriptive enough" because they "aren't great describers of actual content" of the Web page. The "descriptions are typically vague" and this makes it "difficult to understand page contents" from the "misleading titles." 276 respondents reported (mean = 4.2, std. = 2.1) on a scale of 1 (not a problem) to 7 (very significant problem) that titles are "non-descriptive."

The title of the Web page is used as the bookmark's representation by default. The user must interpret the author's title for the content material. For example, one user complained about "people that don't appropriately title their pages." The result is a bookmark title that is "poorly constructed" and "not informative" to the user.

7.2.1.2. Content Changes Without Warning

The user is not notified when the author of a bookmarked page updates the its content. In fact, unless the user explicitly initiates a special time-stamp comparison procedure, the user has no way of determining changes in page content without specifically re-visiting the page. One user explained this problem: "The web has no back-pointers. Consequently, there is no way I can be notified if something to which I have a link changes." Another wrote, "I hate bookmarks that change if I loose track of them."

How much has the content of the Web page changed? One user explained that there is "no way to determine automatically if a bookmarked page has changed (on a
scale between small to significant change).” This user wants the capability to distinguish between a significant percentage change in content as opposed to a tiny modification, which will help to determine the necessity of re-visiting a page.

7.2.1.3. Invalid URLs

Bookmarks pointing to URLs that no longer exist is a problem for users. “It’s only a link - no guarantee that the destination is still there.” Users complained about “canceled URLs” and “stale links” in their bookmark archive. In general, “becoming out of date is problem” with bookmarks because “they can reference a non-existent site.”

When the URL is no longer available, users lose access to information which they spent time locating. “I hate it when the URL is no longer accessible. It was one I liked a lot.” The bookmark archive may contain links to non-existent URLs. This creates problems because “I don’t know when they are no longer valid until I try them.” Thus, “knowing when a site has changed” addresses could be useful. For example, “I cannot easily tell when bookmarks have become obsolete because URL has changed or page has disappeared (would like a link checker).”

7.2.2. Customized Representation

Users rarely customize their representation by changing the title of the page (see 3.4.2.). In this section we describe why it is difficult, the benefits that can result, and two alternative ways to customize a bookmark’s representation.

7.2.2.1. Changing the Name

Users change the name of a bookmark in order to make it a more meaningful and effective representation of the Web page content. Users complained that it is “difficult to
change the name of a bookmark to something more descriptive." This is a combination of the system design and the cognitive load needed to generate "something more descriptive." Some users find the current tools difficult to learn. For example, one user said that "I don't know how to change the bookmark name."

7.2.2.2. Annotating a Bookmark

An annotation is a short paragraph of text describing an item. A bookmark's annotation does not show up in the normal representation. The annotation is used when the user initiates a search or wants a more detailed view of a particular bookmark in the archive. This enables the user to view the representation at different levels of detail: (1) a simple, short title (2) a more detailed description.

7.2.2.3. Assigning a Priority Rating

A few users in our study wanted to articulate their value judgments through priority ratings. Existing browsers and bookmark management tools do not provide the ability to rate the importance of a bookmark. Assigning a value rating is different than classifying information into a semantic hierarchy because the information is being evaluated on a limited, well-defined scale. For example, one user stated, "It would be nice to be able to assign a star rating or number, like *** extremely useful."
7.2.2.4. Cropping a Semantic Icon

One bookmark management tool enables users to crop a screen-shot image from the browser window so as to represent the page iconically (ForeFront 1996). The user selects a portion of the screen and generates a copy of that image. Bookmarks are represented as snap-shot images of the contents of the associated Web pages.

This is another form of expressing the user's initial value judgment. The user specifically indicates the portion of the Web page that will be used to succinctly represent the entire page as a graphic icon.
8. PWIS #5: ESTABLISH A PERSONAL VIEW OF THE WWW

8.1. MENTAL MAPS OF THE WEB

This chapter is based on a survey of 27 users who sketched their personal views of the Web. We employed an extremely simple methodology (see 8.1.1.) to acquire drawings which describes a user's personal, cognitive view of the Web. We shall use illustrative examples from these drawings to characterize a user's mental map of this complex information space. Our data suggests that a user's gestalt view of the Web is closely related to personal experience and information which the user has consciously judged as valuable. In the absence of a global view of the Web, users identify abstract landmarks. Search engines and meta-indices like Yahoo appear because they provide context to information. As user's personalize information, they create a personal Web information space which they conceive of as separate from the unmapped regions of the Web.

8.1.1. Difficulties Drawing Mental Maps

The survey had a 37.5% response rate from a self-selected group of 72 users believed to have previous Web experience. We received 20 responses from a group of 56 industry professionals and academics at a Toronto Web Society meeting. In addition, 16 Web users at the University of Toronto (10 students, 6 staff members) were asked to fill out the survey and 7 responded.

In addition to the low response rate, 40% of the respondents did not draw diagrams. 4 respondents described their personal view of the Web in essay form.
Another 4 respondents had limited sketches with a very strong textual component.

Finally, 2 of the sketches were incomprehensible scribbles.

The survey asked users to draw their personal view of the Web, as shown in the box below:

![SKETCH YOUR PERSONAL VIEW OF THE WORLD WIDE WEB](image)

Some users had difficulty understanding the directions. Many subjects asked us to clarify what we meant by this question. We responded by saying "There is no right answer. Just simply draw the Web from your personal view."

8.1.2. Respondents were Experienced Web Users

Our sample was small and biased toward experienced users. Respondents reported an average of 20.8 months of Web experience (7.4 standard deviation). Only one user reported less than 10 months of Web experience and this user simply sketched a mass of scribbles. The 19 respondents that actually sketched pictures averaged 22.8 months (6.4 standard deviation). Future research should sample a broader population, compare user characteristics from diverse experience levels, and explore the relationship between the population’s experience level and the diagrams of respondents.
8.1.3. No Two Mental Maps are the Same

The 27 drawings varied greatly. The diagrams below show examples of personal views that are very different.

Abstract Conceptual View

WEB ACCESS

Heyday

Propaganda

Communal

Knowledge

This depicts the user surrounded by Web content shown as paintings or artwork.

Artistic View

The link between human knowledge and the content on the Web is still under construction.

Users tend to apply their own theories to explain information spaces in their mental maps. For example, in the "Artistic View" the user draws the Web as artwork and sketched herself surrounded by paintings which contain Web content. Another user did not draw a picture but wrote "Northrop Frye used to say that the centre of reality is where the individual is at the moment, and reality's circumference is what that individual can make sense of. This is my non-technical perspective."

Another user described the process by which an image travels from a server to his Web browser, as shown below:
This denotes the process by which data flows from a Web server to the user's Web browser. The network infrastructure is shown as pipes where the size of the pipe is proportional to the relative speed of the network connection.
The user who sketched the "Browser Interface View" has depicted the main window of a Web browser.

In the "History List View" the Web is drawn as a simple pull-down menu of browsing history items.

```
History List View

This user views the Web as his pull down history menu. "Until a browser/search engine truly 'maps' the results of a surf session in some way other than a text list, how can it be otherwise?"
```

Because a complex information space has no global view, or a map as a guide, users know the Web primarily by their experience in it. Because every user visits different pages and has different browsing episodes, it is not surprising that our data showed wide variations in user's cognitive views of the Web. Despite this, there were some similarities in mental maps of the Web, which we shall describe in the next four sections.

8.1.4. Abstract Landmarks

Landmarks were prevalent in 70.4% of the 27 drawings. 81.5% of these landmarks appeared as abstract entities and only 18.5% were specific Web sites, even though the instructions specifically asked the users to "denote any Web sites which are particularly important to your mental map." We identified landmarks in the mental maps by noting any small, self-contained picture, entity, or node which was accompanied by a short textual title. The table below summarizes the landmarks we found:
Notice that we did not explicitly classify the 95 landmarks to general topic domains.

Search engines were singled out as landmarks in the sketches. The only Web sites to appear were Netscape, Microsoft, Interlog and University of Toronto. Users tend to denote content on the Web by abstract landmarks, but specifically point out sites which give them access to information or have some personal relevance. The “Central Server View” illustrates the use of landmarks centered about the user’s access point to the Internet:

Landmarks are the fundamental building blocks of a person’s cognitive map in the physical world. For example, Thorndyke (1982) demonstrated the difference between
people who learn a terrain through experience and those who learn it through an overview map. People who learned a shopping mall with a map would continue to use the map metaphor as their mental guide. In contrast, those who learned by experience would derive their spatial judgments by simulating a walk through the environment and relating it to landmarks they identified. In the next example, we see a user who has identified a variety of abstract landmarks while exploring the Web.

Abstract Landmark View

The Web is shown as a series of abstract landmarks connected by unidirectional arrows. The user's computer links directly to search engines and "My bookmarks." Search engines access "Dead Sites" and the unmapped regions of the Web. The user's bookmarks point to more specific sites landmarks. In this view, bookmarks are separated from the unmapped regions of the Web.
The next example, "Dial-Up Access View," shows the user's computer accessing the Internet through any one of three other machines. The Web is made up of a network of abstract landmarks.

8.15. Search Appears in the User's Personal View

Our data suggests that many users' experience with the Web is mediated through search engines. 51.9% of the mental maps included a reference to a search engine. For example, Yahoo provides context for information by categorizing it in an index. The "Process View" shows how one user conceptualizes the process of accessing information on the Web.
The user is shown in the middle of this diagram surrounded by a cloud of unknown resources. This view depicts the process of accessing information. Search is used to access unmapped resources in the cloud. Clocks and "time lag" show how it takes time to get access to information on the Web. Just beyond the brick wall is a pot of gold containing "good medical info."

8.1.6. Me, My Computer and the Amorphous Cloud

8 sketches denoted "Me" in center with access to the rest of the Web. Similarly, "My Computer" appeared in 4 drawings. "My Home Page" came up 5 times. In all these drawings personalization or taking ownership over information was prevalent. In the next example, "Relevancy View," the user describes Web content in terms of its personal relevance. At each step away from the user, information become less valuable and more abundant.
The user is depicted as a spider in the middle of the Web. The most relevant sites to the user appear in the center of the Web and represent only 1% of Internet sites. At each progression the sites decrease in relevance but increase in the percentage of Internet sites. The outside circumscribing line represents 44% of Internet sites and contains "sites interesting to none."

In section 5.2 we saw that users bookmark Web pages through value judgments which are based (i) general usefulness, (ii) quality, (iii) personal interest, (iv) frequency of use, and (v) future use. This criteria is used to filter information on the Web. In the “My Computer View” shown below we see an example of an amorphous cloud denoting “The rest of the Internet.”
The user's computer has access to various abstract landmarks as well as a cloud denoting "The rest of the Internet."

The user has access from "My computer" to sites which have already been identified as relevant as well as the cloud of unmapped resources. Users' mental maps of the Web separate information which has been personalized from that which is not relevant or has not been visited.

8.1.7. Bookmarks are Separate from the Web

As we have seen throughout this thesis, bookmarks are an important means of personalizing information. The user's bookmark archive appears separate from the rest of the Web in a location proximate to "Me" or "My Computer." 4 users sketched their bookmarks in their mental maps. Every time bookmarks were drawn, they appeared separate from, but with access to the "cloud." In the "Separate Bookmarks View," the Web is shown as a cloud of "unmapped resources, diversions, etc." with both "accurate
and inaccurate information.” The user visits her bookmark archive or her home page of frequently used links in order to access the Web.

![Separate Bookmarks View](image)

The user has direct access to his bookmarks which are divided into classified and unclassified. The browser also provides access to the user’s Web page which contains frequently used links. Search engines access the unmapped resources of the Web.

Information which has been filtered by the user is denoted in a separate location.

8.2. HYPOTHESIS ABOUT PERSONALIZATION AND EVOLUTION OF GESTALT VIEWS OF THE WEB

We have shown that users create a personal Web information space through the use of bookmarks, and that users separate “My Bookmarks”, “My Computer,” “Me,” and “My Home Page” from the cloud of unmapped Web resources. How does a user’s gestalt view of a complex information space change over time as the user personalizes more information and gradually creates a personal information space?
Waterworth (1994), for example, describes the difference between public and private places in an information space. The metaphor of the user a traveling in a vehicle to information islands which are public spaces is used to show how a user creates a personal spaces while exploring the public information space.

Although we do not have enough data to describe the evolution of a user's gestalt view over time, two mental maps from our survey are suggestive. The user with the least experience of any respondent, 3 months, sketched the "Surrounded by Total Chaos View," as shown below.

```
Surrounded by Total Chaos View
```

![Surrounded by Total Chaos View](image)

A user with 3 months of Web experience sketched "me" surrounded by total chaos.

In contrast to this mental map, one user with 800 bookmarks sketched a picture of the semantic hierarchy in his bookmark archive as his personal view of the Web (see "Bookmarks View" on the next page).

Thus, users may develop a detailed personal view of the Web through experience. Future research should investigate how users personalize information and thereby develop incomplete or multiple mental maps of the Web.
This user sketched his view of the Web as his semantic hierarchy of 800 bookmarks. "I usually type in URL's for personal surfing - they are memorized (internalized!). For business surfing, I use a structured set of over 800 marks, with focus on software, hardware, media, and network infrastructure categories. I enter each bookmark title by hand, not using the pre-programmed title of the page, so that it is in my own style of expression, and easier/more effective to search."

As a user maps more resources on the Web, does his/her gestalt view mature into a personal space center around those sites? For example, a user's gestalt view might start with a "Surrounded by Total Chaos View" with no sites mapped, and then advance to a "My Computer View" with only a few charted landmarks and a cloud of unmapped resources. As the user creates bookmarks, he/she may progress to a "Separate Bookmarks View" denoting a personal space, and finally move to a "Bookmarks View" in which one conceptualizes the Web through the extensive bookmark archive which has been created over time.
9. PERSONAL INFORMATION SPACES BECOME COMPLEX

In chapters 4 through 8 we saw how users create personal information spaces in order to deal with five inherent problems of complex information spaces. In this chapter we look briefly at how successful they are reducing the complexity. We found that although a personal information space is a valuable resource it can grow to exhibit the properties of a complex information space (see 9.1.). A few of the users in our study identified the problems of a complex information space in their bookmark archive.

Complex information space users reduce the complexity by creating a small, personal archive. As this archive grows incrementally, users must continually manage it. If the user fails to optimize this information environment and tune it accordingly, then entropy will create enough disorder that the personal information space will start to become a complex one.

We can idealize this relationship as a continuum where users expend energy to keep their bookmark archive personal and not allow it to become complex.

In section 9.2 we look briefly at problems with current software tools for managing bookmarks. Tools aid the user in the battle to keep a well organized, high-value, personal information space. Comments from users in our survey describe
problems with current pull-down menus, folder systems, and the integration of bookmark management with Web browsers.

9.1. COMPLEXITY PROBLEMS IN A PERSONAL WEB INFORMATION SPACE

A personal information space can exhibit the properties of a complex information space. In this section we look at each property in the context of a user's bookmark archive.

9.1.1. Too Many Bookmarks (Information Overload)

Users collect valuable information on the Web through the use of bookmarks. Although the decision to bookmark a page is based on a value judgment, some users still collect too many bookmarks. For example, one user explained the problem of information overload in a personal Web information space: "They [bookmarks] pile up too fast and become unmanageable." Over time, users find more and more valuable sites, after enough bookmarks are collected, their personal Web information space becomes saturated with too much information.

9.1.2. Polluted Bookmark Archive (Pollution)

Pollution can occur in a personal Web information space. Pollution often becomes a more prevalent problem as the size of the archive increases. For example, one user wrote "I end up with many bookmarks that I don't use; it clutters my list." This extra clutter increases cognitive load, reduces the overall value of the archive, increases retrieval time, and tasks the user to delete superfluous information. The relevancy of Web pages, and information in general, changes over time and is critically tied to the user's tasks (Schamber 1986). A page may have been valuable at one time, but is now considered clutter and pollutes the information space. The value of information units
tends to diminish over time (Chignell 1996). One user complained about his problems dealing with pollution: “I don’t delete often enough.” The net result is that the personal Web information space exhibits saturation and pollution, and the percentage of relevant information units decreases.

9.1.3. Required Maintenance (Entropy)

A personal Web information space becomes more disordered as items are added to it. Users must expend time and energy to manage bookmarks so that the archive retains the benefits of a small, structured, personal information space. One user wrote that he least likes “maintaining” bookmarks and “weeding ones [which are] no longer useful.”

Without expending time and energy to manage a bookmark archive its usefulness degrades. For example, one user wrote that bookmarks are “difficult to use without manually organizing.” A variety of organizational methods arise (see 6.2.). These methods describe how users try to reduce entropy in a personal information space by expending energy.

9.1.4. Lack of Coherent Structure (No Aggregate Structure)

Large archives lack a coherent structure which supports visualization and retrieval. For example, “bookmarks are unstable unless they are kept in meaningful categories, which takes a lot of time.” Users are forced by the current tools to organize bookmarks into a semantic hierarchy so that a structure can be developed within the archive (see 9.2.1.3.). Users need tools which are scalable and can help the user add structure to an archive even with a large number of items.

9.1.5. Difficulty in Visualizing Bookmarks (No Global View)

A personal Web information space suffers visual clutter. When a user collects more than approximately 35 bookmarks (see 9.2.2.2.), visualization problems occur.
One user explained the no global view problem: "I can't see them all at once."
Visualizing a personal Web information space is difficult because of the limitations of
screen space. For example, "My only problem is that I have to get all the bookmarks on
the screen." Another user wrote that it is "hard to get an 'overall' view of bookmarks in
long lists."

9.2. CURRENT SOFTWARE SYSTEMS

Software tools help mitigate the problems of a complex information space
creeping into a user's personal information space. User's need tools which scale up and
support the management and visualization of bookmarks. In this section we look at the
current folder systems (e.g., Netscape 2.0) and the integration of bookmarks with the
browser.

9.2.1. Folder System

Users have trouble using a folder system to organize bookmarks. One user
complained that there are "poor organizational and sorting facilities" for bookmarks.
Users reported problems with arranging folders into a structured hierarchy and placing a
single item into a folder.

9.2.1.1. Arranging Folders in the Hierarchy

A few users complained about arranging folders while managing the
organizational structure of the semantic hierarchy. Arranging folders within the hierarchy
is difficult to do with current tools. For example, one user wrote that "It is not easy to
rearrange folders." Another user dislikes "the organization of bookmarks." He says "the
directory structure when you have nested bookmarks is a little bit irritating."
9.2.1.2. Placing an Item in a Folder

Users encounter problems when placing bookmarks into folders. This results from having to highlight an item, select a destination folder, and drag the item to the destination folder. The user’s in our survey described general problems with folder systems used to manage bookmarks.

Not Intuitive, Awkward Mechanism

Users complained that “moving stuff around in the bookmark list is not intuitive and is very awkward.” One user explained that the system “does not make it easy to move bookmarks around or group a new one into an existing group.” Placing a bookmark into the destination folder is difficult to do, requires time, is hard to learn, and requires physical and cognitive effort to do. One user wrote, “there must be an easier way to place bookmarks in their appropriate section.”

Intricate Manipulation Required

The interaction process required to place a bookmark into a folder requires detailed manipulation. One user wrote, “It always takes multiple clicks and drags to store a bookmark.” Although it may be easy to create a bookmark, the process of moving that bookmark into its destination folder is laborious.

No One Step Placement

The current tools for managing bookmarks do not enable the user to place a bookmark into a folder in a single step. It requires intricate manipulation of the interface. For example, one user complained that he “cannot place it in the prescribed folder easily at the mouse click.” This can be further described as the “inability to store bookmarks into the appropriate folder in one step.”
9.2.1.3. Being Forced to use a Semantic Hierarchy

Semantic hierarchies are the primary organizational facility in most of today's Web browsers and bookmark management tools. Users complained that no other method exists for organizing information. One user wrote that he hates "having to sort them into those folders." Without an alternate method for organizing information, bookmarks become unmanageable. This results in a polluted, unstructured archive which inhibits fast retrieval.

The value of a bookmark diminishes and alternate information foraging methods are used to re-locate pages. For example, one user wrote that "Bookmarks are unstable if they are not kept in meaningful categories, which takes a lot of time." Another complained that "having to sort is not simple." Thus, existing tools force users to manage bookmarks in a semantic hierarchy, as implemented via folders. Users are forced to categorize information in order to keep up with information overload and make efficient retrieval possible. For example, bookmarks are "difficult to use without manually organizing."

One user explained the lack of any other metaphor for managing bookmarks. "I'd like to have a better metaphor for managing bookmarks than folders." Folders create many problems for users. The lack of any other effective means of organizing information forces users to deal with the problems associated with semantic hierarchies.

All of today's Web browsers and bookmark management tools are based on semantic hierarchies in some form. Folders are the most common implementation. WebForager has extended the state-of-the-art and introduced 3D spatial alignment and the book metaphor as a categorical artifact (Card, Robertson and York 1996). Users are
forced by the existing tools to categorize bookmarks in order to structure their personal
Web information space.

If we can invent computational artifacts which automatically generate automatic
structure and provide context to an information space, then manual filing and
categorization would not be required. In the absence of such technology, a well
structured semantic hierarchy is required to make efficient retrieval of bookmarks feasible
(Fertig, et al. 1996).

9.2.2. Browser-Bookmark Integration

Users complained that none of today’s browsers have adequate integrated
bookmark management support. For example, there is “no really good bookmark
organizer that doesn’t demand you exit the browser to delete or edit.” The archive
structure can only be edited from a separate window. Another user discussed the
relationship between search engines and bookmarks: “I tend to collect and organize
resources in batches. An interface between the search queries and my bookmarks would
be nice.” In this section we look at problems with the integration of bookmarks and
browsers.

9.2.2.1. Integrating Multiple Bookmark lists

Users have no way to integrate two or more bookmark lists from different
machines. “I have two computers and two web browsers. I need a simple way to
synchronize both sites.” Another user says he has “too many separate bookmark files”
because he has a “couple of different accounts with different bookmark files.” There
currently is no way to integrate separate bookmark files into a single composite bookmark
file.
One user suggested "networked bookmarks" as a possible solution. "I can't reference a single bookmark file across multiple computing platforms." He believes we "need NFS-like networkable bookmarks."

9.2.2.2. Pull-Down Bookmark Menu

Users have difficulty managing the pull-down menus when they collect a large number of bookmarks. They complained that the "menu system is awkward" and the "folders are in the way when browsing."

Long Lists

Users quickly generate long lists of bookmarks and the pull-down menu becomes unmanageable. "I've browsed for two years and the biggest problem with bookmarks is that the list gets too big too fast and the ability to organize this is poor (plus I can't share
these with anyone, etc.) So I have my own web page to manage organize, add/remove
and shuffle.” Another user wrote that it is difficult to scroll long lists of bookmarks “even
when I created nested folders to minimize the depth of the overview list.” It takes
“multiple clicks” to find a bookmark because users “must scroll through long lists.”

**Viewing all Bookmarks**

Users do not have tools to visualize their bookmarks. It is “hard to get an
‘overall’ view of bookmarks in long lists.” “My only problem is that I have to get all the
bookmarks on the screen.” Another user wrote: “I can’t see them all at once.”

**Multi-Tiered Pull Down**

Folders are integrated into the pull-down menu. One user complained that there is
“limited information in the top level listing of bookmarks” when it contains folders
instead of actual bookmark URLs. The width of the pull-down is a function of the length
of the longest title. One user suggested that he needs “to make a short descriptive title
available at the top level.”

**Limited Screen Space**

Long lists of bookmarks and scaling tiers of the menu cover-up the browser
window. In addition, our study showed that the pull-down menu holds approximately 35
items before it runs off the bottom of the screen. Once the 36 bookmark is added, the
user cannot see it on the screen with the pull-down menu.
10. CONCLUSIONS

10.1. CONTRIBUTIONS

The Human Factors of Personal Web Information Spaces has four salient contributions:

1. This thesis is the first in-depth analysis of how users mitigate the problems of the WWW through bookmarks. It is based on interviews of 12 users, a general survey of 322 people, a mental maps survey of 27 users, and an analysis of the bookmark files of 56 users and the usage data of 23 users over 6 weeks. This rich data set provides evidence which describes how users behave in a complex information space.

2. This thesis has analyzed Web users' cognitive and physical behavior from many different perspectives including navigational patterns, archival use, metaphors, growth rate of the archive, criteria for bookmarking, organizational methods and habits, and mental maps. The breadth of this thesis provides a rich source for future Web human factors research and grounds the theoretical principles of complex and personal information spaces in an empirical study.

3. This thesis defines complex information spaces and discusses how the Web exhibits each of the five inherent properties: information overload, pollution, entropy, lack of aggregate structure, and no global view. We also showed how Web users find localities of interest, create bookmarks to return to them, conceptualize WWW access through metaphors, and create a personal information space from the bookmarks they have collected.

4. Finally, this thesis shows how users create a personal Web information space so as to mitigate the five inherent problems of a complex information space. (1) They prevent
information overload by incrementally building a small archive. (2) They avoid pollution by selecting only useful items and creating a known source of high value. (3) They reduce entropy through maintenance and they organize only when necessary. (4) They add structure by cost-tuning their information environment. (5) They compensate for the lack of a global view by creating their own personal view.

This thesis addresses a broad range of issues that are particularly timely in view of the recent rapid growth in size and usage of the Web. Many of the questions addressed here are new. Future studies should confirm these results by exploring each area in more detail. In the next section, we describe important future work which could extend upon this thesis.

10.2. FUTURE WORK

This thesis suggests many new opportunities for Web human factors research.

We have shown five properties of complex information spaces in this thesis. Future work should develop these principles in more detail through mathematical modeling. Any general theory of information spaces must hold true in both the electronic and paper domains. In pursuit of this goal, future work should test the feasibility of our definition of a complex information space in a variety of domains and note similarities and differences between user behavior in electronic and paper environments. Furthermore, the notion of bookmarks should be extended to a more general notion of "information tags."

The structure of complex and personal information spaces was an important issue in this thesis. Future work should define more precisely the notion of structure in information spaces and how users create a structured environment to minimize the cost of
accessing information units. In addition, a precise definition of "localities" in information spaces would bridge the gap between Tauscher's (1996) mathematical definition and our definition based on the semantic relationship of information units (see 3.1.1.). Empirical studies of Web user behavior should use this definition to relate a user's information foraging goals with the hypertext navigation patterns which emerge through actual use.

We identified four user metaphors for using bookmarks on the Web (see 3.3.). Future empirical studies should identify the circumstances in which each metaphor is used, and identify the frequency with which each is used. These metaphors should be tested by designing bookmark management functionality with each and performing usability tests on each. The notion of relating episodes related to information is particularly important. Future research should investigate time-based Web information management systems which provide users with cues to remind them of previous browsing sessions through novel visualization techniques.

In section 4.1.2, we showed how the growth of a bookmark archive is linear throughout the first year. Once Web users have more experience, future empirical studies should identify the saturation point of a user's bookmark archive and relate this to the archival use of bookmarks. At what point will users stop creating bookmarks and why?

In section 5.2 we presented our simple ontology of the types of sites users bookmark. Future research should relate the content of a user's bookmark archive to work context and demographics. Diverse user groups should be considered. An analysis of this study should identify characteristics of bookmarked sites which make them different from sites not bookmarked. For example, what can a designer do to increase the
likelihood that users will bookmark his/her page? In addition, factor analysis could be used to estimate the probability that a site will be bookmarked.

In sections 5.1.3. and 5.3.2.2. we discussed bookmarks and search engines. Future research should investigate the relationship between bookmarks, history mechanisms and search engines in more detail. For example, describe the tradeoffs between using a search engine and finding a bookmark in the archive. This suggests a tool which integrates bookmarks and history with an interface to query engines and meta-indices on the Web.

In chapter 6 describe the organizational behavior of Web users. Future research should track changes to a user’s bookmark archive over time. In addition, the semantic structure of a user’s bookmark archive should be analyzed. For example, how do users name folders and what are their basic level categories for Web information?

In chapter 8 we described users’ mental maps of the Web. Future research should employ a more sophisticated methodology in order to identify the relationship between private and public spaces on the Web.

In chapter 9 we showed how a personal information space can become complex. Empirical studies are needed to show how often this occurs and in what context. Through such research we will continue to explore the relationship between personal and complex information spaces.

Finally, this empirical study of user behavior on the Web should be applied to building systems which help users generate and maintain personal Web information spaces.
REFERENCES


APPENDICES

APPENDIX A

This survey was originally distributed on both sides of a single sheet of legal sized paper.

World Wide Web Survey

The Computer Science Department at the University of Toronto is exploring ways in which people use the World Wide Web (WWW). Please assist us by filling out the following survey. Your answers will be kept strictly confidential. You may contact David Abrams (abrams@cs.toronto.edu) if you have any questions. Thank you for your participation.

Win A New Computer Book

Your completed survey will be part of a drawing for five new computer books complements of Prentice Hall Canada, PTR Division Marketing Department. Enter your ticket number in the space provided below to win.

Ticket Number: ______________________

1) General Usage Characteristics

1.1) How do you describe your WWW experience?

No Experience

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

[Circle a number]

1.2) On average, how many times do you make use of the WWW per week? _____________

1.3) On average, how much time do you spend during each WWW session?

☐ under 10 min  ☐ 10-30 min  ☐ 30 min - 1 hr  ☐ 1-2 hrs  ☐ 2-3 hrs  ☐ 3+ hrs

1.4) How many WWW bookmarks (hotlist items) do you have?

☐ none  ☐ 1-10  ☐ 11-25  ☐ 26-100  ☐ 101-300  ☐ 300+

1.5) How do you organize your bookmarks?

☐ I don’t organize - they stay in the order which I created them.

☐ I manually re-arrange bookmarks in a list.

☐ I create folders to group together related bookmarks.

☐ I create folders within folders to create a hierarchy of bookmarks.

☐ Other (describe) ________________________________
1.6) When do you usually organize your bookmarks?

☐ I store each new bookmark in its place when I create it.
☐ I re-organize my bookmarks at the end of each browsing session.
☐ I organize bookmarks occasionally/sporadically.
☐ I never organize bookmarks.
☐ Other (describe)__________________________________________________________

2) Your Typical WWW Browsing Patterns
Think back to a typical WWW browsing session, for example, the last time when you browsed the WWW for 20 minutes or more.

2.1) Indicate how often you use the following techniques to organize bookmarks on a scale from 1 to 7.

<table>
<thead>
<tr>
<th>Never</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Very Often</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

_____ create a new bookmark
_____ manually re-arrange bookmarks in a list
_____ create folders to group together related bookmarks
_____ create sub-folders within folders to manage a hierarchy
_____ change the title of a bookmark
_____ annotate a bookmark
_____ delete a bookmark

=> OVER =>
2.2) Approximately how many new bookmarks do you create during a typical session?

- [ ] none
- [ ] 1-5
- [ ] 6-10
- [ ] 11-20
- [ ] 21 or more

2.3) Rank the importance of each reason for creating bookmarks on a scale of 1 to 7.

Not Important: 1  2  3  4  5  6  7  Very Important

[ ] Temporary Bookmark: a temporary navigational landmark to come back to later during this session
[ ] Archival Bookmark: a reference to store for use in future WWW sessions
[ ] Publishing Bookmark: a hypertext link to add to my own WWW site
[ ] Collaborative Bookmark: an item to give to a friend or third person
[ ] Other (describe)

3) Rate Problems You Find with Existing Bookmarks

How significant/bothersome are each of the following possible bookmark problems?

Not a Problem: 1  2  3  4  5  6  7  Very Significant Problem

[ ] I have trouble finding a bookmark stored somewhere in my history/archive.
[ ] I open a bookmarked page to remember what's in it because the title doesn't describe the content.
[ ] There's no good way to organize my bookmarks.
[ ] I cannot see all my bookmarks on the screen at one time.
[ ] I cannot easily tell when the content of a bookmarked page changes on the WWW.
[ ] There's no quick and easy way to store a new bookmark in the correct place and continue browsing the WWW.
[ ] Other (describe)

4) What do you like most and/or least about WWW bookmarks?

______________________________________________________________________________
5) Personal Attributes

Gender:  □ Male      □ Female

Age:  □ under 20 □ 20-29 □ 30-39 □ 40-49 □ 50-59 □ 60+

THANK YOU FOR YOUR SUPPORT
APPENDIX B

Ontology of Bookmarked Sites
Each table lists the number of repetitions of each domain name in the top 30% most bookmarked sites. All domain names with 4 or more repetitions are included. Domain names were grouped together if they represent the same Web site.

<table>
<thead>
<tr>
<th>159 SEARCH ENGINES</th>
<th>152 NEWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query Engine</td>
<td>General News</td>
</tr>
<tr>
<td>31 ALTAVISTA.DIGITAL.COM</td>
<td>13 <a href="http://WWW.CNN.COM">WWW.CNN.COM</a></td>
</tr>
<tr>
<td>19 <a href="http://WWW.LYCOS.COM">WWW.LYCOS.COM</a></td>
<td>11 PATHFINDER.COM</td>
</tr>
<tr>
<td>12 <a href="http://WWW.INFOSEEK.COM">WWW.INFOSEEK.COM</a></td>
<td>6 <a href="http://WWW.NEWswire.CA">WWW.NEWswire.CA</a></td>
</tr>
<tr>
<td>11 <a href="http://WWW.OPENTEXT.COM">WWW.OPENTEXT.COM</a></td>
<td>4 <a href="http://WWW.RADIO.CBC.CA">WWW.RADIO.CBC.CA</a></td>
</tr>
<tr>
<td>9 WEBCRWLER.COM</td>
<td>4 <a href="http://WWW.CBC.CA">WWW.CBC.CA</a></td>
</tr>
<tr>
<td>7 <a href="http://WWW.EXCITE.COM">WWW.EXCITE.COM</a></td>
<td>4 <a href="http://WWW.NBC.COM">WWW.NBC.COM</a></td>
</tr>
<tr>
<td>Meta-Index</td>
<td>4 <a href="http://WWW.PBS.ORG">WWW.PBS.ORG</a></td>
</tr>
<tr>
<td>53 <a href="http://WWW.YAHOO.COM">WWW.YAHOO.COM</a></td>
<td></td>
</tr>
<tr>
<td>News Groups</td>
<td>Newspaper</td>
</tr>
<tr>
<td>7 <a href="http://WWW.DEJANews.COM">WWW.DEJANews.COM</a></td>
<td>15 <a href="http://WWW.USA">WWW.USA</a> Today.COM</td>
</tr>
<tr>
<td>Find People</td>
<td>11 <a href="http://WWW.GLOBE">WWW.GLOBE</a> AND MAIL.CA</td>
</tr>
<tr>
<td>5 <a href="http://WWW.FOUR11.COM">WWW.FOUR11.COM</a></td>
<td>8 <a href="http://WWW.NYTimes.COM">WWW.NYTimes.COM</a></td>
</tr>
<tr>
<td>5 <a href="http://WWW.SWITCHBOARD.COM">WWW.SWITCHBOARD.COM</a></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>189 CORPORATE</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>47 HOME.Netscape.COM</td>
<td>15 <a href="http://WWW.ZDNet.COM">WWW.ZDNet.COM</a></td>
</tr>
<tr>
<td>36 <a href="http://WWW.MICROSOFT.COM">WWW.MICROSOFT.COM</a></td>
<td>14 <a href="http://WWW.HOTWired.COM">WWW.HOTWired.COM</a></td>
</tr>
<tr>
<td>21 <a href="http://WWW.SGI.COM">WWW.SGI.COM</a></td>
<td>9 <a href="http://WWW.ZIFF.COM">WWW.ZIFF.COM</a></td>
</tr>
<tr>
<td>16 <a href="http://WWW.APPLE.COM">WWW.APPLE.COM</a></td>
<td>5 <a href="http://WWW.MACWorld.COM">WWW.MACWorld.COM</a></td>
</tr>
<tr>
<td>7 <a href="http://WWW.CISCO.COM">WWW.CISCO.COM</a></td>
<td>4 TECHWEB.CMP.COM</td>
</tr>
<tr>
<td>7 <a href="http://WWW.REALAUDIO.COM">WWW.REALAUDIO.COM</a></td>
<td></td>
</tr>
<tr>
<td>7 <a href="http://WWW.SUN.COM">WWW.SUN.COM</a></td>
<td></td>
</tr>
<tr>
<td>6 <a href="http://WWW.ADOBE.COM">WWW.ADOBE.COM</a></td>
<td></td>
</tr>
<tr>
<td>5 <a href="http://WWW.MACROMedia.COM">WWW.MACROMedia.COM</a></td>
<td></td>
</tr>
<tr>
<td>5 <a href="http://WWW.WILTEL.COM">WWW.WILTEL.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.ALIAS.COM">WWW.ALIAS.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.IBM.COM">WWW.IBM.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.BORLAND.COM">WWW.BORLAND.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.DISNEY.COM">WWW.DISNEY.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.EPICURIOUS.COM">WWW.EPICURIOUS.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.EXTRACorp.COM">WWW.EXTRACorp.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.INTEL.COM">WWW.INTEL.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.MASTERCorp.COM">WWW.MASTERCorp.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.MASTERCARD.COM">WWW.MASTERCARD.COM</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.PCFunds.BOMIL.CA">WWW.PCFunds.BOMIL.CA</a></td>
<td></td>
</tr>
</tbody>
</table>

Financial
- 6 WWW.PCQUOTE.COM
- 5 WWW.SECAPL.COM
<table>
<thead>
<tr>
<th>493 UNIVERSITY</th>
<th>132 MISC/SPECIAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>493 UNIVERSITY</strong></td>
<td><strong>132 MISC/SPECIAL</strong></td>
</tr>
<tr>
<td><strong>MIT</strong></td>
<td><strong>Medicine</strong></td>
</tr>
<tr>
<td>22 <a href="http://WWW.AI.MIT.EDU">WWW.AI.MIT.EDU</a></td>
<td>17 MED.TORONTO.EDU</td>
</tr>
<tr>
<td>11 <a href="http://WWW.MIT.EDU">WWW.MIT.EDU</a></td>
<td>4 TIE.TELEMED.ORG</td>
</tr>
<tr>
<td>10 WEB.MIT.EDU</td>
<td>4 <a href="http://WWW.NLM.NIH.GOV">WWW.NLM.NIH.GOV</a></td>
</tr>
<tr>
<td>8 THE-TECH.MIT.EDU</td>
<td>4 <a href="http://WWW.PSLGROUP.COM">WWW.PSLGROUP.COM</a></td>
</tr>
<tr>
<td>8 <a href="http://WWW.MEDIA.MIT.EDU">WWW.MEDIA.MIT.EDU</a></td>
<td><strong>Cars</strong></td>
</tr>
<tr>
<td>4 <a href="http://WWW.MITPRESS.MIT.EDU">WWW.MITPRESS.MIT.EDU</a></td>
<td>5 <a href="http://WWW.EDMUNDS.COM">WWW.EDMUNDS.COM</a></td>
</tr>
<tr>
<td><strong>Universities</strong></td>
<td>4 <a href="http://WWW.Z28.COM">WWW.Z28.COM</a></td>
</tr>
<tr>
<td>27 <a href="http://WWW.CS.CMU.EDU">WWW.CS.CMU.EDU</a></td>
<td><strong>Government</strong></td>
</tr>
<tr>
<td>23 CSZX.OPEN.AC.UK</td>
<td>4 <a href="http://WWW.MAINE.COM">WWW.MAINE.COM</a></td>
</tr>
<tr>
<td>20 SUNSITE.UNC.EDU</td>
<td>6 <a href="http://WWW.DSIEGEL.COM">WWW.DSIEGEL.COM</a></td>
</tr>
<tr>
<td>20 <a href="http://WWW.NCSA.UIUC.EDU">WWW.NCSA.UIUC.EDU</a></td>
<td><strong>Entertainment</strong></td>
</tr>
<tr>
<td>14 <a href="http://WWW.CIS.OHIO-STATE.EDU">WWW.CIS.OHIO-STATE.EDU</a></td>
<td>14 <a href="http://WWW.UNITEDMEDIA.COM">WWW.UNITEDMEDIA.COM</a></td>
</tr>
<tr>
<td>10 <a href="http://WWW.LYSATOR.LIULU.SE">WWW.LYSATOR.LIULU.SE</a></td>
<td>6 CYCLERY.COM</td>
</tr>
<tr>
<td>9 <a href="http://WWW.CC.GATECH.EDU">WWW.CC.GATECH.EDU</a></td>
<td><strong>Environment</strong></td>
</tr>
<tr>
<td>8 <a href="http://WWW.CS.UBC.CA">WWW.CS.UBC.CA</a></td>
<td>6 <a href="http://WWW.ON.DOE.CA">WWW.ON.DOE.CA</a></td>
</tr>
<tr>
<td>7 WINGS.BUFFALO.EDU</td>
<td><strong>Maps</strong></td>
</tr>
<tr>
<td>6 INFO.MED.YALE.EDU</td>
<td>5 <a href="http://WWW.MAPQUEST.COM">WWW.MAPQUEST.COM</a></td>
</tr>
<tr>
<td>6 <a href="http://WWW.CS.INDIANA.EDU">WWW.CS.INDIANA.EDU</a></td>
<td><strong>Local Hard Drive</strong></td>
</tr>
<tr>
<td>5 HIRU.MCMASTER.CA</td>
<td>53 File on local hard drive</td>
</tr>
<tr>
<td>5 PUBWEB.NEXOR.CO.UK</td>
<td></td>
</tr>
<tr>
<td>5 UNION.NCSA.UIUC.EDU</td>
<td></td>
</tr>
<tr>
<td>5 WEBER.U.WASHINGTON.EDU</td>
<td></td>
</tr>
<tr>
<td>5 WHEAT.U.WATERLOO.CA</td>
<td></td>
</tr>
<tr>
<td>5 WWW-LELAND.STANFORD.EDU</td>
<td></td>
</tr>
<tr>
<td>5 <a href="http://WWW.DNA.LTH.SE">WWW.DNA.LTH.SE</a></td>
<td></td>
</tr>
<tr>
<td>5 <a href="http://WWW.IOP.BPMF.AC.UK">WWW.IOP.BPMF.AC.UK</a></td>
<td></td>
</tr>
<tr>
<td>5 <a href="http://WWW.NWU.EDU">WWW.NWU.EDU</a></td>
<td></td>
</tr>
<tr>
<td>5 <a href="http://WWW.OISE.ON.CA">WWW.OISE.ON.CA</a></td>
<td></td>
</tr>
<tr>
<td>4 AKEBONO.STANFORD.EDU</td>
<td></td>
</tr>
<tr>
<td>4 ARCHIVE.CIS.OHIO-STATE.EDU</td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.ECS.NWU.EDU">WWW.ECS.NWU.EDU</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.INDIANA.EDU">WWW.INDIANA.EDU</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.PUBLIC.IASTATE.EDU">WWW.PUBLIC.IASTATE.EDU</a></td>
<td></td>
</tr>
<tr>
<td>4 <a href="http://WWW.SDSC.EDU">WWW.SDSC.EDU</a></td>
<td></td>
</tr>
</tbody>
</table>
### 122 SERVICE PROVIDERS

<table>
<thead>
<tr>
<th>On-Line</th>
<th>ISP</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 AOL.COM</td>
<td>13 <a href="http://WWW.IO.ORG">WWW.IO.ORG</a></td>
</tr>
<tr>
<td>4 COMPUSERVE.COM</td>
<td>11 <a href="http://WWW.WEBCOM.COM">WWW.WEBCOM.COM</a></td>
</tr>
</tbody>
</table>

### 197 COMPUTER TECHNOLOGY

<table>
<thead>
<tr>
<th>Professional Societies</th>
<th>Web Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>27 <a href="http://WWW.W3.ORG">WWW.W3.ORG</a></td>
<td>7 <a href="http://WWW.HYPE.COM">WWW.HYPE.COM</a></td>
</tr>
<tr>
<td>14 <a href="http://WWW.ACM.ORG">WWW.ACM.ORG</a></td>
<td>6 <a href="http://WWW.WSMITH.COM">WWW.WSMITH.COM</a></td>
</tr>
<tr>
<td>4 SIGGRAPH.ORG</td>
<td>4 WEB.IDIRECT.COM</td>
</tr>
<tr>
<td>4 <a href="http://WWW.IEEE.ORG">WWW.IEEE.ORG</a></td>
<td>4 <a href="http://WWW.ACTWIN.COM">WWW.ACTWIN.COM</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Java</th>
<th>Software Archives</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 JAVA.SUN.COM</td>
<td>8 <a href="http://WWW.GNN.COM">WWW.GNN.COM</a></td>
</tr>
<tr>
<td>7 <a href="http://WWW.GAMELAN.COM">WWW.GAMELAN.COM</a></td>
<td>7 <a href="http://WWW.SHAREWARE.COM">WWW.SHAREWARE.COM</a></td>
</tr>
<tr>
<td>5 <a href="http://WWW.JAVASOFT.COM">WWW.JAVASOFT.COM</a></td>
<td>6 <a href="http://WWW.WINDOWS95.COM">WWW.WINDOWS95.COM</a></td>
</tr>
<tr>
<td>4 <a href="http://WWW.JAVAWORLD.COM">WWW.JAVAWORLD.COM</a></td>
<td>5 <a href="http://WWW.CNET.COM">WWW.CNET.COM</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unix</th>
<th>Graphics</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 <a href="http://WWW.SCO.COM">WWW.SCO.COM</a></td>
<td>6 <a href="http://WWW.3DSITE.COM">WWW.3DSITE.COM</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eng Labs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24 COS.GDB.ORG</td>
<td>6 <a href="http://WWW.SANDIA.GOV">WWW.SANDIA.GOV</a></td>
</tr>
<tr>
<td>4 LCWEB.LOC.GOV</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

Analysis of [Tauscher96] usage data. Bookmarks were classified into each column based on the pattern which emerged after the bookmark's use [see 3.2.2. for analysis].

<table>
<thead>
<tr>
<th>User #</th>
<th>Hub-Spoke</th>
<th>DPS</th>
<th>Guided-Tour</th>
<th>First</th>
<th>Review</th>
<th>Repetitive Probes</th>
<th>Intermitent Probes</th>
<th>Dead End</th>
<th>Submit</th>
<th>Helper App</th>
<th>Author</th>
<th>Unexplained</th>
<th>TOTAL</th>
<th>NUM URLS</th>
<th>% BK USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>57</td>
<td>370</td>
<td>1.89%</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>27</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>57</td>
<td>1570</td>
<td>4.16%</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>3</td>
<td></td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td>35</td>
<td>725</td>
<td>4.83%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1080</td>
<td>0.19%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>475</td>
<td>2.32%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>6</td>
<td></td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>24</td>
<td>800</td>
<td>3.00%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
<td>1</td>
<td>6</td>
<td></td>
<td>8</td>
<td>1</td>
<td>20</td>
<td></td>
<td>20</td>
<td>375</td>
<td>5.33%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td>23</td>
<td></td>
<td>33</td>
<td>600</td>
<td>5.50%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>570</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
<td>345</td>
<td>3.19%</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>625</td>
<td>1.44%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>670</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td>950</td>
<td>2.53%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>500</td>
<td>10.00%</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td></td>
<td>375</td>
<td>3.73%</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>250</td>
<td>3.20%</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td>4</td>
<td>1950</td>
<td>0.21%</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>700</td>
<td>2.57%</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>550</td>
<td>1.27%</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>1180</td>
<td>0.76%</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>540</td>
<td>2.41%</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>2900</td>
<td>0.69%</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>1050</td>
<td>1.33%</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>72</td>
<td>32</td>
<td>7</td>
<td>43</td>
<td>16</td>
<td>78</td>
<td>84</td>
<td>21</td>
<td>12</td>
<td>8</td>
<td>2</td>
<td>15</td>
<td>390</td>
<td>18750</td>
<td>2.08%</td>
</tr>
</tbody>
</table>