The Re-Search Engine: Helping People Return to Information in Dynamic Information Environments

Jaime Teevan

Abstract

I investigate how people return to information in a dynamic information environment. For example, people often use search results not just to find new information, but also to return to information found via previous searches. However, search results change as search engines update their indices to reflect the current state of the Web. These changes can benefit users by providing new information, but they hinder returning to previously viewed information. I propose that the benefits can be gained without the drawbacks by giving people awareness and control over noticeable changes.

In this proposal, I present work I have already done to support this thesis and outline the further research in the area I plan complete as part of my dissertation. Thus far, I have conducted two observational studies, one a modified diary study of how people search and the other an analysis of instances found on the Web where people expressed difficulty returning to information that had moved. These studies highlight the importance of preserving the path originally taken to the information target. Consequently, to support returning to information in a dynamic environment, I will augment a standard search engine to create a "Re-Search Engine". The Re-Search Engine will preserve previously taken paths by remembering results to repeated queries. Based on a study presented here, it appears that people do not notice many changes, and this suggests it will be possible to preserve remembered search results while still presenting new ones. I will use implicit measures of attention to determine which aspects of the results page should and should not change.

Nothing endures but change.
Heraclitus (540 BCE - 480 BCE)

Illustration by Thomas Escher

Contents

1 Introduction **8**
- 1.1 The Problem of Re-finding in a Dynamic Environment . . . 8
- 1.2 The Proposed Solution . . . 9
- 1.3 Contributions . . . 11
- 1.4 Research Plan . . . 11
- 1.5 Proposal Outline . . . 12

2 Returning to Information **14**
- 2.1 Related Work . . . 14
- 2.2 Modified Diary Study . . . 15
  - 2.2.1 Related Studies . . . 15
  - 2.2.2 Study Methods . . . 16
- 2.3 Information Management . . . 16
  - 2.3.1 More Than Just Keyword Search . . . 17
  - 2.3.2 Information Target . . . 18
- 2.4 Finding the Information Source . . . 19
  - 2.4.1 Email . . . 20
  - 2.4.2 Files . . . 21
  - 2.4.3 Web . . . 22
- 2.5 Re-finding the Information . . . 22
  - 2.5.1 Defining Contextual Information . . . 22
  - 2.5.2 How Contextual Information Differs When Re-finding . . . 22
- 2.6 Individual Tactics . . . 22
- 2.7 Conclusions . . . 25

3 Problems Returning When the Information Environment is Dynamic **27**
- 3.1 Defining Dynamic Information . . . 27
- 3.2 On the Web . . . 28
  - 3.2.1 The Web is Dynamic . . . 28
  - 3.2.2 People Re-visit Web Information . . . 28
  - 3.2.3 Difficulties in Web Interaction . . . 29
- 3.3 "Where'd it go?" . . . 29
- 3.4 Methods . . . 30
- 3.5 Overview of the Data Collected . . . 31
  - 3.5.1 Understanding the Pages in Which the Phrase Occurred . . . 31
  - 3.5.2 What "It" Was . . . 32
  - 3.5.3 Where "it" went . . . 32

3.6 Describing the missing information . . . 33
3.6.1 Expressions of Frustration . . . 33
3.6.2 Shared Context . . . 34
3.6.3 The Importance of Path . . . 34
3.6.4 Time Not Important . . . 35
3.7 Answering "Where'd it go?" . . . 35
3.7.1 Explanations . . . 36
3.7.2 Work-Arounds . . . 37
3.7.3 Resolutions . . . 37
3.8 Multiple Users of the Same Information . . . 38
3.9 Design Implications . . . 39
3.10 Beyond the Web . . . 40
3.11 In the Future . . . 41
3.11.1 Helping Existing Applications . . . 41
3.11.2 Personal Information Becoming Dynamic . . . 42

4 Current Solutions to Dynamic Information Interaction 43
4.1 User Makes Changes/Current Information . . . 44
4.2 User Makes Changes/Interest in Returning . . . 45
4.2.1 Version Control Systems . . . 45
4.2.2 Undo . . . 46
4.3 Uncontrolled Change/Current Information . . . 46
4.3.1 Time Dependant Information . . . 46
4.3.2 Collaboration . . . 47
4.4 Uncontrolled Change/Interest in Returning . . . 47
4.4.1 Document Permanence and Digital Libraries . . . 47
4.4.2 Dynamic Menus . . . 48
4.4.3 Maintaining Context on the Web . . . 48
4.5 Generalizing Current Solutions . . . 48

5 Awareness and Control as a Solution 50
5.1 Related Work . . . 50
5.2 Building a Re-Search Engine . . . 50
5.2.1 Remembering What's Remembered . . . 51
5.2.2 Using What's Remembered . . . 53
5.3 Information Can Change While Maintaining Control . . . 54
5.3.1 Change Blindness Literature . . . 55
5.3.2 The Problem of Clustering . . . 55
5.3.3 Study Methods . . . 55
5.3.4 Clustering Implementation . . . 55
5.3.5 Understanding Conceptual Anchors . . . 56
5.4 Implicit Feedback for Inferring User Attention . . . 56
5.4.1 Classification of Implicit Feedback Techniques . . . 57
5.4.2 Examination of Key Papers . . . 60
5.4.3 Future Directions . . . 64
5.5 Evaluating the Re-Search Engine . . . 64
5.5.1 Deploying the Interface . . . 65
5.5.2 Laboratory Study . . . 65

6 Conclusions and Future Work **67**
6.1 Contributions 67
6.2 Future Work 68

List of Figures

1-1 An example of how returning to dynamic search results can be difficult. . . 10

2-1 For each user, the percent of email targets found in the user's Inbox. 23
2-2 The number of times participants used each search tactic in their files. . . . 24
2-3 The number of times participants used each tactic on the Web. Note that pilers appear to use specific search tools more often. 24

3-1 Three instances that contained the phrase "Where'd it go?". The first (a) contained a posting from a person looking for Web functionality. The second (b), titled "Where'd it go?", is a redirect page. The third (c) offers support in finding information that's moved as a result of a site reorganization. *Haven't yet pulled in the correct set of pictures ... Will do.* 31

5-1 The original search results returned (a), and the search results returned at a later date (b). These results can be merged together, as shown in Figure 5-2. 52
5-2 The results returned by the Re-Search Engine based on those shown in Figure 5-1. 52
5-3 Papers classified based on the observed implicit behaviors they discuss. The papers discussed in greater dept in Section 5.4.2 have been highlighted. . . 59

List of Tables

2.1 Information need by search strategy (19 unknowns removed). 18
2.2 The number of times participants used each search strategy for each corpus. 18

4.1 Examples of types of user interaction with dynamic information. Work is divided based on whether the user made the change and whether the user is interested in working only with current information versus returning to the information. 43
4.2 Related work in each of the user interactions with dynamic information. . 44

5.1 Classification of behaviors that can be used for implicit feedback from Oard and Kim [94]. Our additions are in italics. 58
5.2 Papers from the "Examine Object" cell in Table 5-3, classified by study type. The papers discussed in greater depth in Section 5.4.2 have been *italicised.* 60

That which is static and repetitive is boring. That which is dynamic and random is confusing. In between lies art.

John A. Locke (1632 - 1704)

Chapter 1

Introduction

For my dissertation, I will investigate how people return to information in a dynamic information environment. For example, people often use search results not just to find new information, but also to return to information found via previous searches. However, search results change as search engines update their indices to reflect the current state of the Web. These changes can benefit users by providing new information, but they hinder returning to previously viewed information. I propose that the benefits can be gained without the drawbacks by giving people awareness and control over noticeable changes.

In this proposal, I present work I have already done to support this thesis and outline the further research in the area I plan complete as part of my dissertation. Thus far, I have conducted two observational studies, one a modified diary study of how people search [4, 119] and the other an analysis of instances found on the Web where people expressed difficulty returning to information that had moved [118]. These studies highlight the importance of preserving the path originally taken to the information target. Consequently, to support returning to information in a dynamic environment, I will augment a standard search engine to preserve previously taken paths by remembering results to repeated queries. As I will show, it appears that people do not notice many changes, and this suggests it will be possible to preserve remembered search results while still presenting new ones without disrupting the user. I will use implicit measures of attention to determine which aspects of the results page are worthy of preserving.

In this chapter I introduce my thesis research. I begin by exploring in greater depth problems that can arise when returning to information in a dynamic information environment. I then summarize how I plan to address the problem. I highlight the contributions to the field that this research will make, and give an overview of my research plan. I conclude with an outline of the proposal.

1.1 The Problem of Re-finding in a Dynamic Environment

Electronic information, and in particular Web information, can be very dynamic. For example, online news sites change when new stories are written, personal Web sites change as their hosts edit them, and search results change as search engines update their indices to reflect updates on the Web. The growing ease of electronic communication and collaboration, the rising availability of time dependent information, and even the introduction of automated agents, suggest information will continue to become more dynamic in the future. Changes can benefit users by providing new information, but they hinder returning to previously viewed information. As stated by David Levy, "[P]art of the social and technical

work in the decades ahead will be to figure out how to provide the appropriate measure of fixity in the digital domain [75]."

Traditional human-computer interaction research with dynamic information has assumed that users are interested in the most recent state of dynamic information. For example, news stories are very dynamic, and, for the most part, when people want news they are interested in seeing only the most current stories. News delivery systems, therefor, focus primarily on providing new news. Similarly, with stock prices, a person generally wants just the latest quote, and providing this information has been at the heart of market information systems. Sometimes systems also make the user aware of recent changes. For example, recent movements in stock prices are often displayed, and some news delivery systems highlight new stories that might be of interest.

However, there are many other possible interactions a user might want to have with dynamic information, including those interactions they have with static information. One particularly important interaction with dynamic information that has not been well explored is that of returning to previously viewed information. For example, someone might chose to return to an interesting news story they read last week, in which case they don't care that there are new news stories that have since arrived. Similarly, when filing tax forms, a person needs to return to the stock prices at the time of purchase. On the Web, people regularly try to return to previously viewed information [25], despite the fact that the Web is very dynamic and the information might have changed.

Returning to information in a dynamic information environment is a particularly difficult problem because when people return to information, they rely heavily on the information environment in which they originally encountered the information. When the environment changes, often the cues that the user remembers to retrieve the information have changed. Consistency and control have long been tenets of user interface design [47, 110], but as the information that people work with becomes more and more dynamic, it will be an increasingly difficult task to provide consistency to the user as well as a sense of control over the information.

For example, when news is fairly static, returning to previously viewed articles is reasonably easy. When looking for a news article from last week's morning paper, a person can probably dig through their recycling bin, and return fairly easily to the context in which they originally encountered the article. In contrast, trying to return to a news story that was encountered via an online news service, where the information changes throughout the day, requires not only remembering the date of the story, but also the story's source and some uniquely identifying keywords. Furthermore, much of the remembered story context, such as the fact that it was originally advertised in the upper left hand corner of the Web site, would be entirely unavailable for its retrieval. Many examples of how changes to information like this can be a problem when returning to the information are given in Chapter 3.

1.2 The Proposed Solution

I propose that to support returning to information in a dynamic environment, users should be given awareness and control over noticeable changes to the information environment in which they originally encountered the information. Two observational studies presented in this proposal suggest the importance of the path taken to first get to the information target, and for this reason, I will look at providing users awareness and control over noticeable

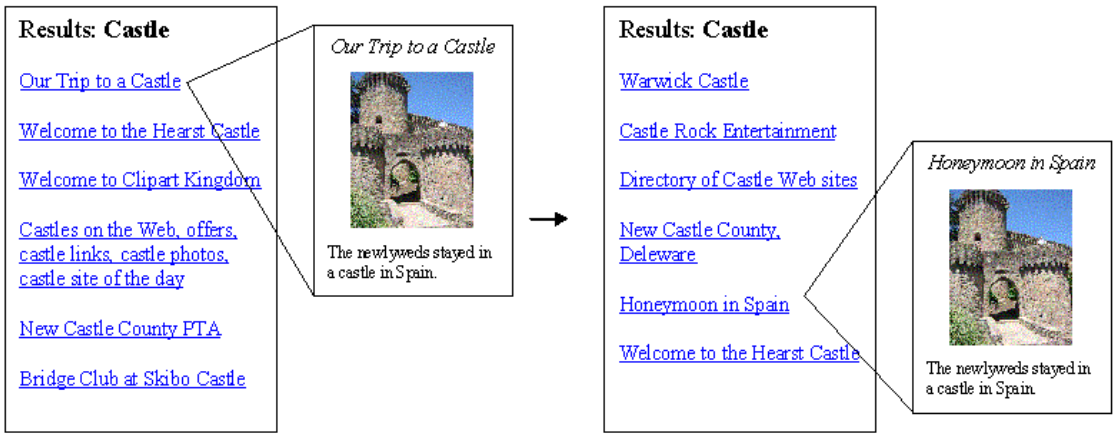

Figure 1-1: An example of how returning to dynamic search results can be difficult.

changes to a common and important step in many people's information access paths, search results.

People often use search results not just to find new information, but also to return to information they found via previous searches. But the Web is constantly changing, and online search engine results for the same query change over time. Results could change because the underlying information has changed. They could change because the search engine's algorithm has changed. Or they could change because the search engine discovers new information that is more relevant to the user's query. Such changes benefit the user who wants to find new information, but they hinder returning to old information.

Consider the example in Figure 1-1. Let us say we are interested in the page "Our Trip to a Castle", found during an initial search for the term "Castle". The same page is difficult to return to at a later date, even when using the same search term, because the search results for "Castle" have changed. New resources (e.g., "Directory of Castle Web sites") have been discovered, changing the page's position in the result list. And the name of the link to the page has changed to "Honeymoon in Spain", making it difficult to identify.

None the less, people commonly use search results as entry-ways into Web sites. People report considerable trouble returning to information on the Web [5], and it is conceivable that this is in part due to the fact that search results change. Further, preserving search engine related access paths is likely to become even more important in the future as search becomes a more integral part of file systems and email managers in addition to the Web. For example, the personal information management system, Haystack [54], intends to easily support personal information queries. Because such searches can be an expression of user interest, Haystack intends to have agents locate additional topical information to present to the user the next time they express that same interest. While this will help the user discover new relevant information, it will create problems very similar to that in Figure 1-1 when returning to information.

Stifling all change for the sake of consistency is not a viable solution to the problem of returning to information in a dynamic information environment. When information changes, it usually changes to benefit the user. In the case of news stories, it is good that a person has access to the latest news available. Similarly, in the case of search results, it is good when search engines return new, more relevant information. For example, in Figure 1-1 the

new search results for "Castle" are arguably better for a generic castle search than the old search results were. But change is not good when the dynamics of the information destroy the user's sense of control and understanding of that information. In my thesis I will look at continuing to allow information to change while supporting the user's interaction with that information.

Based on a study presented here, it appears that people do not notice many changes, and this suggests it will be possible to preserve remembered search results while still presenting new ones. I will use implicit measures of attention to determine which aspects of the results page should and should not change. I hypothesize that these solutions will make it easier for people to find old search results than current interfaces, while not interfering with their ability to find new relevant information.

1.3 Contributions

The most direct implications of this work will be for search engine user interfaces. I will implement and test a new approach to helping people find information through search with the specific aim of aiding searches intended to return to information. If successful, I will make it possible for information that has been found via past searches to be found in future searches, despite changes to the underlying search engine index. From this work, I hope to draw generalizations that can be applied to general solutions aimed at supporting information re-finding in dynamic information environments.

Additionally, my thesis research will give insight into how people re-find electronic information in a dynamic information environment, as well as broaden the current understanding of how people find, and especially re-find, electronic information, even when the information is not dynamic. The work I have done in this area is important because much research on people's electronic information behavior has been performed in a controlled setting. In contrast, This work is important because most research on people's electronic information seeking behavior has been performed in a controlled setting. In contrast, the modified diary study performed by Christine Alvarado, David Karger, Mark Ackerman and me [4] and the analysis I performed of people having trouble re-finding information that had moved, looked at information seeking behavior as it arose in the real world, not in a laboratory setting, and not with specific information seeking tasks.

1.4 Research Plan

This proposal already contains a considerable amount of background work and motivation for the importance of assisting people to return to dynamic information. It also contains several studies I have performed to understand the area better, including the results of a modified diary study performed by Christine Alvarado, Mark Ackerman, David Karger and me on how people search for electronic information, and an analysis of of instances found on the Web where people expressed difficulty returning to information that had moved. To further understand the problem, I plan to:

- Perform further analysis of the data collected from the modified diary study. We have already investigated the ways people look for information, and I will further investigate specifically how people look for information they have seen before. This work will be performed Summer 2004.

Based on the understanding of the problem I have developed, I hypothesize that giving people control and awareness over noticeable changes in search results will enable them to better find previously seen information, while still allowing them to find new information. The bulk of the remaining work to be done for my dissertation involves testing this hypothesis. Already included in this proposal are the results from a paper prototype study investigating what sort of changes people notice in information, as evidence that information need not be completely static for people to feel in control of it. For the completion of my thesis, I will also:

- Create, through iterative design, a search engine interface for search engines with indices that change over time. I will instrument this interface to give users awareness of and control over any noticeable changes to the results that occur. This implementation will be done during Summer, 2004.

- Deploy the interface to people in the lab and collect and analyze usage data to understand if the new features are used in realistic settings. I will deploy the interface at the end of Summer, 2004, and collect data during Fall, 2004.

- Run a laboratory study of the above mentioned interface to determine how the changes effect the user experience when re-finding information in search results that have changed, as well as when finding new information. This study will be performed during the Fall term, 2004.

I plan on completing additional analysis and writing my thesis during the Spring terms of 2005, for an expected June 2005 graduation.

1.5 Proposal Outline

- **Chapter 1** I begin the proposal with an introduction of my thesis topic.

- **Chapter 2** In Chapter 2, I discuss how people find, and particularly, re-find information. I present a modified diary study conducted by Christine Alvarado, Mark Ackerman, David Karger and me, in which we find that people rely heavily on the information environment surrounding their information target to actually find the target. We find that environment is often used to support a series of small steps to the target rather than one large jump.

- **Chapter 3** However, when the information is environment, returning to the information target in this manner can be difficult, and in Chapter 3 I show that this can be a considerable problem. I present numerous examples where people currently have trouble returning to information because the environment is dynamic, and give the results of a study I performed analyzing instances, collected via a Web search, of people expressing such troubles.

  In addition to discussing problems that already have arisen, I highlight potential future problems people could have re-finding in dynamic information environments. I give examples of applications that currently enforce static information environments to keep problems from arising, but that might benefit from allowing the environment to change in a way that didn't cause problems, and I show that many applications of the future will increase the amount of dynamic information available.

- **Chapter 4** In order to better support users in returning to dynamic information, in Chapter 4 I look at how current systems support any type of interaction with dynamic information. I find that for problems where a user is interested in returning to the information that has changed, time is an important element, and for interfaces where the information is dynamic due to changes beyond the user's control, maintaining context and awareness are central elements.

- **Chapter 5** While both time and maintaining context and awareness are interesting solutions to pursue, I explore the later. I present a paper-prototype study that suggests people do not notice many changes, and this suggests it will be possible to preserve remembered information, while still presenting new information. I will use implicit measures of attention to determine which aspects of the results page need to be preserved.

- **Chapter 6** In Chapter 6 I conclude, highlighting again the contributions of this work. I also discuss a number of interesting problems that arise in relation to this work for future consideration that I do not intend to pursue as part of my thesis.

You cannot step twice into the same river; for other waters are continually flowing in.

Heraclitus (540 BCE - 480 BCE)

Chapter 2

Returning to Information

In order to support returning to dynamic information, I first look at the slightly better understood problem of how people return to static information. In this chapter I report on a modified diary study conducted by Christine Alvarado, Mark Ackerman, David Karger and me [4]. While the results I report on from our study currently only discuss how people search for information in general, for my thesis I will extend the research to look specifically at how people look for information they have seen before.

So far we have found that when locating a document or searching for a specific answer, people relied on their contextual knowledge of their information target to help them find it. For example, a participant might know she could find the phone number of a restaurant in a particular email from a colleague. As in this case, often people associated the information target with a specific document. People seemed to prefer to solve their information need by using contextual information as a guide in navigating locally in small steps to the desired document rather than directly jumping to their target. This navigation is similar to the Micronesian islanders' situated navigation in Suchman's work [114], and appears to be important but under-supported. We found that this behavior was especially true for people with unstructured information organization.

The structure of the chapter is as follows. I first discuss previous work related to re-finding information. I then describe the particulars of the modified diary study we performed and present what we learned from the study. I conclude by discussing the implications of our findings for the design of personal information management tools such as the one I plan to build for returning to dynamic information.

2.1 Related Work

Related work on re-finding information, to be fleshed out:

- Paul Maglio and Rob Barrett [8, 81, 82] have done work on how people return to information, and found that people return to information based on how they initially found it.
- William Jones, Susan Dumais and Harry Bruce [60, 61] have done a lot of work on "keeping found things found". They also have what seems to be a very relevant paper, based on the title, coming up at SIGIR [34].
- Rob Capra's Ph.D. [21, 55] work explores how people return to information remotely. For example, how people find information when they call in to their office looking for the location of an appointment.

2.2 Modified Diary Study

Christine Alvarado, Mark Ackerman, David Karger and myself [4] performed a qualitative study examining what people did when working with their email, their file system, and the World Wide Web. In performing the study, we did not set about to test specific hypotheses, but rather to understand our participants' information seeking behavior. We wanted to understand what people did with their electronic information, and we focused in particular on the situations in which people reported exerting effort in locating information. In this section I explain the set up for our study by first describing related user studies and then describing the methodology we used for our study.

2.2.1 Related Studies

Previous observational studies have focused on users' interaction with various different subsets of their personal information, such as their email, their files, and the Web. We use the term corpus to refer to any of these major subsets that are treated as a group by the user's tools. Early studies provided a theoretical motivation for electronic information management by examining the ways that people organized their paper documents. Lansdale [74] noted that the difficulty people had in retrieving information from this corpus arose because they were forced to group their information into categories that were not necessarily relevant for retrieval. Malone [83] also studied how people organized paper documents. More recently, Whittaker and Hirshberg [123] investigated personal paper archives to understand the value of paper over digital documents.

Similarly, a number of studies have investigated the different uses that people have for their email [79, 124]. Researchers found that in addition to using email as a communication tool, people also used it to keep track of upcoming appointments and often used their Inboxes as to-do lists. More recent studies on email, scheduling tools and instant messaging further explored these ideas [12, 92].

Another class of studies investigated how people organize their computer files. While file management systems rely largely on content hierarchy, two studies [91, 92] found that users preferred to use location-based search in storing and retrieving information, placing all of their currently relevant documents on the desktop and associating a location with each document. Fertig, Freeman and Gelernter [38] argued that people rely on temporal cues to locate files.

Finally, there is a large body of literature reporting on how people use the World Wide Web. These studies address both how people manage their bookmarks [1], the various tasks people perform using the Web [107], and how people keep track of information on the Web [60]. A body of literature on information seeking [22, 95, 100] discusses how people seek out information from large collections, often focusing on the Web.

Some observational studies have been conducted across different corpora. For example, [60] have recently looked at how Web information is stored by observing user interaction with the same three electronic corpora (email, files and the Web) that we investigated. However, they focused specifically on how people used files and email to support their Web interactions. We took a broader focus and aimed to uncover how people interacted with electronic information in general.

2.2.2 Study Methods

We conducted semi-structured interviews in which participants reported their information activities twice a day over the course of a week. The interviewer would interrupt participants' work and prompt them for their recent information activities in the three different corpora. We felt this method would enable us to properly understand how people interact with their electronic information, since participants worked with their own information. Our method was similar to the diary studies used in many information seeking studies, as well as the Experimental Sampling Method [98].

Our participants consisted of 15 graduate students (5 women, 10 men) in Computer Science at MIT. All participants were experts at computer use. The length of time the participants had been at the university varied from one year to seven. By looking at a range we were able to observe both those who were in the process of developing their information organization, and those who had a long standing structure.

While Computer Science graduate students were convenient to study, more importantly, they were experienced users with complex digital information spaces. As such, our participants could discuss the issues that arise in current information organization tools that occur despite users' considerable experience with those tools. In general, we believe that our participants are typical of many experienced users, but we will discuss the impact of MIT and Computer Science as cultures in Section 2.7.

We randomly interrupted each participant's work twice a day for five consecutive days. Each semi-structured interview lasted only five minutes in order not to unduly interrupt the participants' work. In the interviews, we asked the participants to describe what they most recently "looked at" and what they most recently "looked for" in each of the three corpora we studied (email, files and Web).

What precisely defined "looking for" versus merely "looking at", or *accessing*, was defined by the participants themselves based on what they considered effort. By allowing participants to self-categorize when they had to exert effort to find information (as in, for example, Bernard's work [13]), we were able to learn what types of information needs required effort and what techniques they relied on in those cases. We encouraged the participants to give as much detail as possible.

Overall, we obtained 151[1] interviews. In addition, we conducted longer semi-structured interviews (1 hour) with each participant about their information patterns and conducted some direct observations. The data were analyzed using standard qualitative techniques (e.g., Ackerman and Halverson [2]).

2.3 Information Management

This section describes the participants' reported information activities and focuses on specific types of behavior we found to be important. In general, participants in our study managed incredibly complex information spaces. For example, Alex[2] had two email accounts containing hundreds of email messages in his Inbox alone and many more in email directories (e.g., one email account was further divided into approximately 250 email directories, each containing anywhere from tens to hundreds of messages). His Web space and

[1] We inadvertently interviewed one participant 11 times. This participant is labeled "M" in future charts.

[2] All names and details reported here have been anonymized. Minor changes to the transcripts have been made for readability.

file system were similarly complex.

Despite their complex information spaces, participants were largely successful at managing and finding things. Most participants felt fairly in control of their information. For example, Alex, described above, when asked if he felt in control responded:

> That's an interesting question. I think my email is the worst, because I have so much of it. And there are people on the other end who expect me to reply to it. My file system is pretty well organized. I have to go through it every once in a while, every couple of months and just kind of push things into the right folders and delete the old stuff. The Web just works usually.

In fact, we find that, despite people's large and complex information repositories, most information activity involved simply accessing information and did not involve the user exerting effort to find information. When we asked about their most recent information activity, only in 50 of the 453 instances (151 interviews times 3 corpora) did participants report "looking for" something. However, when we specifically asked participants what they had looked for most recently (following up, in most cases, upon their initial response), they reported 200 instances (of 453-not everyone remembered looking for something in each corpus prior to each interview). As might be expected, a number of these activities occurred on the Web (n=83, 42%). However, a significant number also occurred within email (n=65, 33%) and files (n=52, 26%). The following subsection discusses a prevalent search strategy that relies on contextual information and that differs from the traditional concept of keyword search. In order to learn more about this under-explored area of search we then look at the different information targets people had when they searched.

2.3.1 More Than Just Keyword Search

In the interviews, people often reported looking for things without resorting to traditional keyword search, as the following example illustrates. Jim is looking for the office address of a professor named Connie Monroe:

> Interviewer: Have you looked for anything on the Web today?
> Jim: I had to look for the office number of the Harvard professor.
> I: So how did you go about doing that?
> J: I went to the home page of Math Department at Harvard.

This participant then goes on to explain that he knows there is a specific Web page with the address:

> I: Did you know it would be there [on a page] or you just hoped it would be there?
> J: I knew that she had a very small Web page saying, I'm here at Harvard. Here's my contact information.
> [...]
> I: So you went to the Math department, and then what did you do over there?
> J: It had a place where you can find people, a link to the page where you can find people and I went to that page and they had a dropdown list of visiting faculty, and so I went to that link and I looked for her name and there it was.

	Specific	General	Document	Total
Orient	47	19	41	120
Teleport	34	23	17	80
Total	81	42	58	200

Table 2.1: Information need by search strategy (19 unknowns removed).

	Orienteering	Teleporting
Email	59	6
Files	42	10
Web	19	64

Table 2.2: The number of times participants used each search strategy for each corpus.

While Jim only wants Connie's office number, he first goes to the Harvard Web page, and then navigates in toward the information he is after. This search by localized or situated navigation is an example of what we call *orienteering*. Orienteering involves using contextual information to narrow in on the actual information target, often in a series of steps. This definition is similar to that of O'Day and Jeffries [95], as we use orienteering to imply using information from the current location to decide where to go next. However, unlike their definition, our participants did not generally have an evolving information need. Orienteering, then, is an extension of situated activity reported in the Computer Supported Cooperative Work literature (e.g., Suchman [114]; Ackerman and Halverson [2]).

Participants also reported what we call *teleporting*. When a person teleports, they try to take themselves directly to the information they're looking for. For example, if instead of orienteering, Jim had tried to teleport, he could have typed "Connie Monroe office number" into a search engine hoping to find it directly.

It is important to note Bates' distinction between search strategies and search tactics here [9]. Orienteering and teleporting are strategies; participants can use the same search tactic, in some cases, to achieve either strategy. For example, participants sometimes reported using keyword search in orienteering as illustrated by Carla when she had to look for a page she lost when her network connection died. Although she performs a keyword search, her behavior is clearly orienteering because she takes small steps to narrow in on a goal.

> I did a re-search for it on Google and then I clicked into it. [...] I actually wanted not the main page for Bon Jour Quebec, but I kept clicking on the links, from the main page. It was very easy to get what I wanted to get.

In Tables 2.1 and 2.2 we use a very conservative measure of orienteering, classifying an instance as teleporting if it involved a keyword search at any stage in the process. Even with this conservative measure, 60% of the time people looked for something, they orienteered.

2.3.2 Information Target

People used additional information when orienteering to their information target. Because of the difference between the information people used in orienteering and the information they were actually looking for, it is fundamental in understanding orienteering to understand a

person's information target. We found that when people looked for information, they looked for qualitatively different things across instances.

We investigated people's information targets by first labeling the participant's particular goal in each seeking activity they reported and then clustering those goals into general categories. To help us do this, in our interviews we asked people to not only tell us what they were looking for, but also why they wanted it. We identified 21 different goals. Two of the authors independently categorized each seeking activity as one of these 21 goals. In the 3% of the cases where the researchers disagreed and in the 7% of the cases where the goal could not be determined, the activity was classified as "unknown".

When we further collapsed the 21 goals three main categories of information needs emerged: specific information, general information, and specific documents (emails, files or Web pages). A person has a specific information need when he is looking for a small fact, as typified by research on question answering (e.g., find the time of a meeting). When a user looks for general information, the user is interested in a broader set of information (e.g., determine a good pair of sneakers to buy). In the case when a specific document is the target, it is the actual document that is desired (e.g., a file to edit), and that document is not replaceable by the information contained within it.

The existence of these three categories revealed unexpected search strategy patterns. Table 2.1 shows how often people used the different search strategies for each of the different targets. The large variety in search strategy in relation to information target implies that the two are closely related.

We were not surprised to find that people orienteered to specific documents. We expected people to maintain a large amount of context about documents. However, we expected to find that people would try to go directly (with a keyword search) to a specific information target and were surprised that this was not the case. Instead, people were much more likely to orienteer when looking for specific information than general information.

2.4 Finding the Information Source

We observed that people not only used orienteering techniques to look for documents, but also when looking for specific information. This behavior implies that people maintained a large amount of related information to the specific piece of information they are looking for. In particular, part of this related information involved associating the piece of information with a source, such as an email or specific Web page containing the desired information. We observed documents and specific information were often conflated in participants' descriptions. Due to this tendency, in many instances, their strategies for finding a specific piece of information reflect the best way of locating the source of the information. Marchionini [85] observed the same information seeking behavior for large information collections, and Hearst [52] suggested this behavior holds on the Web.

To better understand how participants associated information with an information source, consider the following activity, in which Dan is searching within his email:

> Dan: Earlier today I was doing a search for a message that Kristi sent me that had this guy's address.
>
> Interviewer: So you needed the address and then you remembered Kristi had sent it?
>
> D: Yeah.

In this example, Dan tries to find someone's address, but instead of seeking the address directly (through a keyword search in his Inbox or on the Web), Dan associates the address with an email he received from a colleague and looks instead for that email.

Users also made similar associations in the other corpora. For example, one participant had difficulty finding the location of a city in Switzerland. He did not know exactly where to find that information, but he had four map sites bookmarked. Rather than relying on a keyword search directly to locate the city, he used the bookmarks to access the map sites and then clicked around to see if he could find a map with the information he was looking for.

Associating a specific information need with a source is one piece of contextual information that participants maintain about this information. In addition, they maintain a lot of contextual information about the information source itself that helps them in their orienteering process. In the Harvard office example (Section 2.3.1), Jim knew the approximate URL for the page that contained Connie's office number, as well as the size of the containing page and what additional content would appear there.

It should be noted that the contextual information used is not always definitive, such as in the following example of a participant looking to find a specific piece of information in her email. In this example, the participant has only the memory that a piece of email was in a specific directory:

> The last email I read was an email from Bill describing where to find the documentation on [a project]. I had searched for this email. It was an old email that I had. And I had to look for it. And I looked for it in the research directory which was where I put things that are sort of done for a research. [...] So anyway, so, in my research directory I found this. But I actually had to flip through all emails. I went and tried to look for the email that looked familiar for being the correct one. The only thing I had to go by was that it was probably from Bill. But I wasn't exactly positive on that. And I wasn't sure where it would be anyway. So it took a long time to find. But I found it.

Orienteering and teleporting were not used uniformly across the corpora. Table 2.2 shows the number of searches in each corpus by each search strategy. Recall that in the table below, any instance in which a keyword search was used was classified as teleporting, which is a conservative measure of orienteering. As we shall see below, many of those cases are actually instances of orienteering.

In general, participants preferred to orienteer toward their target. Although they occasionally used keyword search to help them narrow in on their target, they rarely teleported directly to what they were looking for. The next sections examine this behavior further for each of the three corpora.

2.4.1 Email

In email, almost none of the seeking activities involved teleporting, even though most participants' email programs facilitated it by supporting keyword search. Participants often looked in their email for specific information that they knew to be contained within a particular message. Most often people went directly to the correct folder (approximately half of the time, their Inbox) that contained the message. Then they would browse to the appropriate email, using either the date (the default ordering in most mail systems) or the

sender as a guide. In the following example, Carla looks in her email to find the location of a meeting:

> Carla: The last email I looked at was a email for Mischa's reading group I wanted to find out where we were meeting. [...] It was in my Inbox because I would have left it in my Inbox. I knew that it was in my Inbox and I searched for it knowing that it was by Mischa and I only had two emails from him.
>
> Interviewer: Did you sort?
>
> C: I'm always sorted by date.
>
> I: By date, okay, so you didn't sort by sender?
>
> C: I didn't sort by sender. I don't have so many people in my Inbox. I knew that it was rather recently. So, since I knew it was rather recently it was sort of easier just to visually sort. I just have to scroll down one or two pages to find the email.

In finding the location of the meeting, Carla relies on a large amount of information in addition to her information target: she knows the target is in an email sent to her by Mischa, that this email is in her Inbox because she left it there, that she only has two emails from Mischa, etc. Note that most of this additional information has to do with the email containing the information about the meeting and nothing to do with the meeting in question.

In three cases classified as teleporting, the same participant searched for the sender of the message in question and then browsed through the returned messages. Because this participant's email client did not support sorting the messages by sender, this participant likely relied on keyword search to fulfill the same role.

2.4.2 Files

Participants used orienteering techniques much more often than teleporting within their files. Most often their goal was actually to find a particular file to read or edit. Although participants did not often look for specific information within their files, when they did, they again tended to orienteer to the file containing that information, as in the following example:

> I actually was not looking for a file. I was looking for the contents of a particular file. ... I remembered that a Perl template that I had had those four lines of Perl in it. So I went to look at where my templates are.

In some cases the user did not appear to be able to associate the information with a particular source, and they used a keyword search to teleport to the information, as illustrated in this example:

> I needed to search through some Emacs source files in order to find a particular line of code that I needed. I went to the directory above all and then did a recursive grep [keyword search] down the tree looking for anything that had the appropriate set of characters in it.

2.4.3 Web

On the Web, participants' search techniques were more varied and they tended to rely on teleporting more often, possibly because keyword search tools are more sophisticated for the Web than for files or email, because of the larger search space, or because the information on the Web continually changes its structure.

Even on the Web, participants did sometimes associate their information needs with a specific information source and seemed to prefer to orienteer to that source when possible, as in the example in which Jim is looking for the Harvard professor's office number. Additionally, at least one fourth of the instances listed as teleporting in Table 2.2 were actually cases of orienteering where keyword search was used as a step, as in the example where Carla used a keyword search to get to a page on Quebec and then continued narrowing in on her target from this page. As well, 16 of the searches classified as teleporting consisted of searches where the user had already navigated to the neighborhood of their information target (e.g., a company's home page)-an orienteering activity-before performing the keyword search.

As with email and files, participants tended to teleport when they could not associate their information need with a specific source. In one case, a user spent a considerable amount of time looking for how much to tip hairdressers simply performing various keyword searches using the words "tip", "hairdresser", "percent", and "gratuities".

2.5 Re-finding the Information

We have yet to look at what distinguishes searches for information that has been seen before, and I will do this as part of my thesis research. I believe we will find that people use similar strategies whether finding information that they've seen before or finding new information. What will differ is the contextual information that they remember, and those differences I plan to discuss here.

2.5.1 Defining Contextual Information

The information source is just a type of contextual information. *Discuss exactly what is meant by contextual information, give other examples, etc.*

2.5.2 How Contextual Information Differs When Re-finding

Discuss how contextual information differs when re-finding information versus finding information for the first time.

2.6 Individual Tactics

While everyone orienteered as well as teleported as strategies, some people used keyword search more as a tactic. Surprisingly, these same individuals tended to put more effort into organization. These two groups emerged from our observations of people's email use. When people reported looking for information in their email, they either found that information in their Inbox a majority of the time, or they found it in a specific folder a majority of the time. Figure 2-1 shows the percentage of time the 13 participants who reported looked for

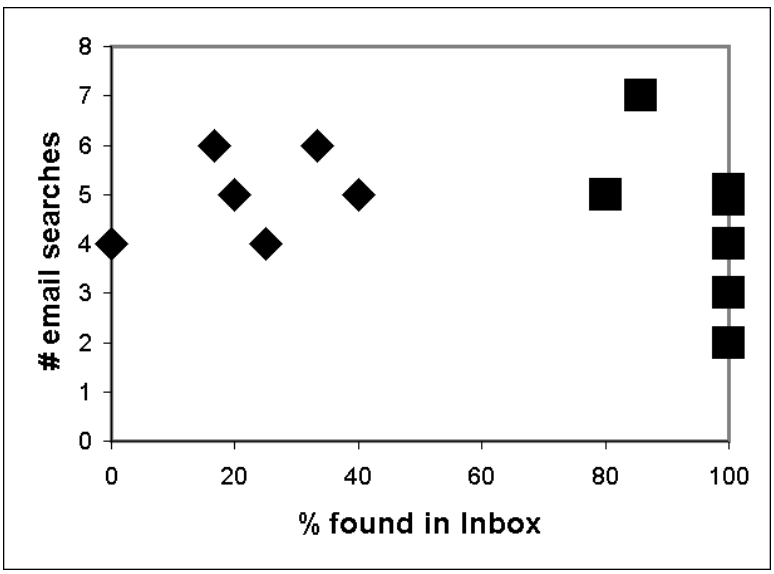

Figure 2-1: For each user, the percent of email targets found in the user's Inbox.

something in their email found it in their Inbox[3]. Notice that the difference was not because one group searched more, as the number of instances was similar for the two groups.

Those on the right-hand cluster of Figure 2-1 almost never spoke of interacting with emails that were not in their Inboxes and almost always expected to find messages in their Inboxes, implying they did not file their messages in general. Those in the left-hand cluster almost always went directly to folders and never expected to find messages in their Inboxes implying they regularly filed their messages. We can view the left-hand group in Figure 2-1 as *filers*, and the right-hand group as *pilers* [83]. A *filer* is a person who organizes information using a rigid structure, and a *piler* is someone who maintains a mostly unstructured information organization.

Filers and pilers tended to rely on different search tactics when looking for things within their files and on the Web. Figures 2-2 and 2-3 show the number of different search tactics used by participants in files and on the Web, respectively. The top 6 participants in each graph are those classified as filers based on their email search patterns. From Figure 2-2 it appears that filers reported having to look for files (or information within those files) more often than pilers. Furthermore, filers relied more on keyword search than pilers.

More work will be required to understand the nature of this association. It appears that both groups orienteered toward the information they were seeking, so we do not believe that filers were more likely to teleport. Rather, perhaps pilers associate a finer grain of contextual information with what they are looking for and then can take more local steps to get to their goal. Because they are more confident in their ability to rely on contextual information, they do not need to maintain complicated organizational structures to keep track of their information. Moreover, they do not need to rely on keyword search because they are able to take local steps using this contextual information.

[3]Two participants never reported finding anything in their email

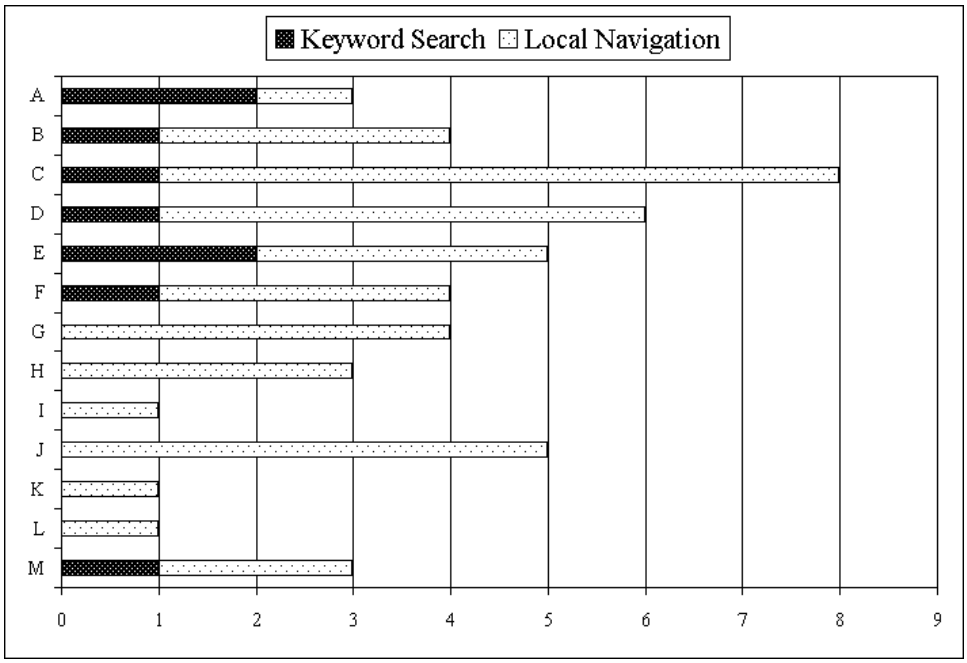

Figure 2-2: The number of times participants used each search tactic in their files.

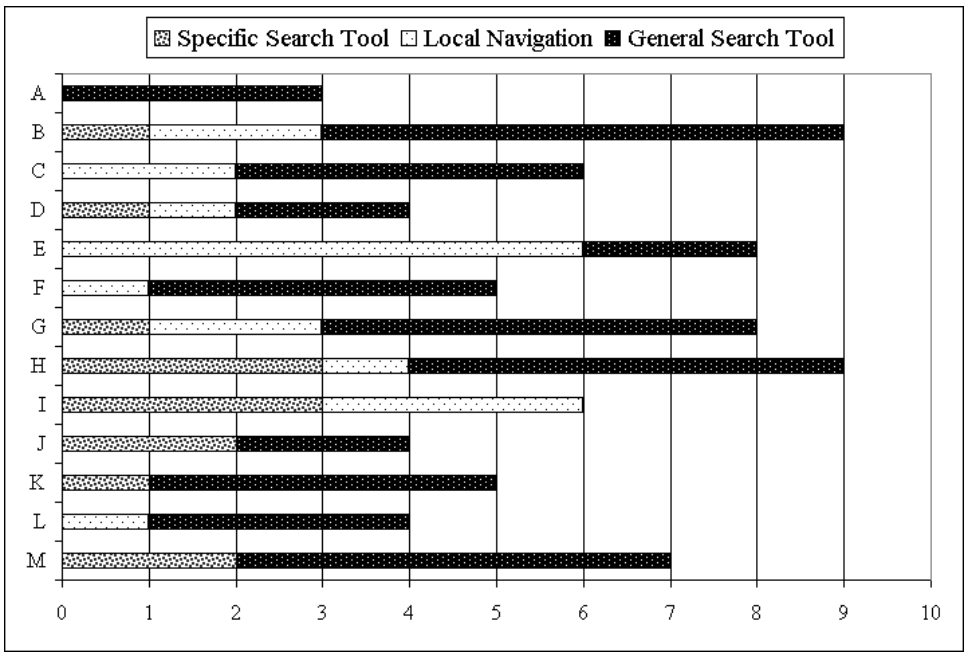

Figure 2-3: The number of times participants used each tactic on the Web. Note that pilers appear to use specific search tools more often.

2.7 Conclusions

In this chapter, I reported on a study of how people looked for electronic information in their email, their file system, and on the Web. We found that a majority of the times that people looked for things they used an orienteering strategy, taking advantage of the large amount of contextual information they had about their information target.

As with any qualitatively-based study, there are limitations to this study. We could investigate only to the existence of an interesting phenomenon, and discuss its potential associations and nuances. Further work will be required to assess how general and pervasive orienteering might be as well as to assess its causal roots and effects. Nonetheless, we believe we have described an important search strategy to support. Our participants were members of the MIT and Computer Science cultures, both of which place an interest in and a value on information handling. However, in our opinion, the observation that this population valued orienteering only shows orienteering's importance in dealing with large amounts of electronic information.

Several possibilities exist to explain our findings. It could be that people orienteer rather than teleport because current search tools do not work well enough to teleport to what the user requests. This problem may be alleviated by work in information retrieval. Or, current tools may not allow users to take advantage of their contextual information and fully specify the information they are looking for. New tools are attempting to address this by adding additional meta-data support, mostly within keyword search tools (e.g., Microsoft Windows XP).

However, as noted in this paper, we observed that even when people could use their contextual information to teleport directly to their information target, they often preferred to orienteer to the information instead. We believe this occurs because fully specifying an information need and all of its meta-data would require considerable cognitive effort. Often our participants were not entirely sure of the contextual information they could use to retrieve their target, or they were not even aware of it until reminded. Orienteering also has the added benefit of helping the user not to over-specify her target, allowing her to backtrack more easily. It also gives her information about the source, which may be important in determining the validity of the information. These possible explanations suggest that future systems should deeply consider orienteering or other approaches to help people use contextual data, perhaps by prompting them with contextual information instead of requiring them to fully specify all of this information at query time.

To support the use of contextual information, we must understand exactly what information people know about their information target. People often orienteered to targets they had seen before, and in these cases appeared to use different contextual information than in cases when they had not seen their target previously. As mentioned earlier, for my thesis, I will further examine the nature of the contextual information used in each case.

What people remember appears to be corpus dependent. What was sometimes considered "looking for" in one corpus was not the same for the others. The intricacies of these boundaries, and their use in search and retrieval, are other interesting areas to pursue. The tactics people prefer to use, while favoring orienteering, seem to vary by individual. It would be interesting to look at what different support the two different classes of people discussed in this chapter require when returning to dynamic information. Additionally, there are probably more subtle differences that could be learned from the user's behavior.

As the amount of information we interact with grows, electronic information management will increasingly become a problem we must deal with. Our study revealed behavioral

patterns we can examine further in order to build tools to make this interaction more manageable in the future.

Change is not made without inconvenience, even from worse to better.
Richard Hooker (1554- 1600)

Chapter 3

Problems Returning When the Information Environment is Dynamic

In Chapter 2 we saw that when returning to static information, people rely heavily on contextual cues. However, when information is dynamic and those contextual cues are destroyed, returning to information becomes even more difficult. While the problem of returning to uncontrolled dynamic information has not been a significant focus in human-computer interaction, there are numerous examples in the literature where a user's interaction with information fails because the user wants to return to previously viewed information that has since changed in some way. In this chapter I highlight those examples.

I begin by discussing what it means for information to be dynamic and precisely explain the problem of returning to uncontrolled dynamic information. I then give evidence of the problem in a well known and ephemeral domain, the Web, and discuss what I learned about the problem on the Web by studying the results to the query, "Where'd it go?" I then look at the existence of the problem in domains beyond the Web. I finish with a discussion of where we can expect to see this problem arise in the future, and highlight in particular how it will arise within an individual's own personal information interactions as that information begins to become dynamic.

3.1 Defining Dynamic Information

I use the term *dynamic information* to refer to any information that changes in any way. A piece of information could be dynamic because actual value of that piece of information has changed, as, for example, happens when the value of a company's stock changes over time. Or it could be dynamic because the organization of the information has changed, as when the organization of an email changes because a user files it into a folder. In the case of dynamic search results, the organization of the information changes when the ordering of the search results changes, and the value of the information changes when the underlying information being indexed is modified.

Changes to dynamic information could be made directly by the person using the information, as when an email is filed, or they could be made outside of the user's control, as when a stock price changes. When the changes are made outside of the user's control, they could occur synchronously with the user's own interaction with that information. For

example, someone watching a basketball game sees the score change synchronously with their viewing. Or they could occur asynchronously over multiple user interactions with the information, as is the case with dynamic search results that change between use.

The problem that I am addressing in my dissertation is that of returning to uncontrolled dynamic information. The information is *dynamic* because either its value or organization changes. It is *uncontrolled* because those changes occur outside of the user's control. And problem arises because the user wants to *return* to previously viewed information despite the fact that much of the user's context might have been destroyed by the information being both dynamic and uncontrolled.

3.2 On the Web

On the Web people regularly encounter trouble returning to some piece of information because it has changed outside of their control. In this section. I discuss how the Web is a very dynamic system, but that nonetheless people re-visit Web information regularly. I show that difficulties often arise when people try to re-visit Web information as a direct result of that information changing.

3.2.1 The Web is Dynamic

The Web is tremendously dynamic [71]. Examples of dynamic Web information include message boards, personal Web pages, stock quotes and Internet auction. One reason the information is dynamic is because it is controlled by other agents (in most cases those other agents are people). For example, a moderator could delete an inappropriate message from a message board, or someone could edit their personal Web page. Since any one individual only controls an extremely small portion of the Web, a majority of the Web is controlled by other agents and is likely to change outside of any one individual's control.

Additionally, the Web makes available a considerable amount of information that is dynamic because it is time dependent. For example, stock prices might go up throughout the course of a day, and an Internet auction might expire.

3.2.2 People Re-visit Web Information

Despite the fact that information changes so often on the Web, people commonly return to Web information they've seen before [20, 25, 115, 116]. Tauscher and Greenberg analyzed six weeks of detailed Web usage logs and found that 58% of all Web page visits were re-visits [115, 116]. In a more recent study by Cockburn, et al., users were found to revisit Web pages for as many as four out of every five page visits [25]. And while many of the Web page re-visitations occur shortly after a Web page is first visited (e.g., during the same session by using the back button), a significant number are visited after considerable time has elapsed. In their study, Tauscher and Greenberg report, "Users revisit pages that have not been accessed recently. For example, 15% of recurrences are not covered by a list of the last 10 URLs visited. Still, doubling or tripling the size of this list does not increase its coverage much [115]." Indeed, they found 6% of all Web page visits take place to pages that haven't been visited in over fifty visits. This is plenty of time for information on the Web to change, possibly causing disorientation for the user.

3.2.3 Difficulties in Web Interaction

There is evidence that people have trouble interacting with the Web because they are trying to return to information that has changed. A study of Web usage by the GVU Center at Georgia Tech [5] surveyed people on their biggest problems in using the Web, and found that "Not being able to find a page I know is out there" and "Not being able to return to a page I once visited" are significant problems. (*Talk more about this.*)

In another study of the Web, Cockburn, et al., found that 25% of all people's bookmarks no longer worked [25]. The broken URLs once pointed to information that the user indicated, through the process of bookmarking, as worth returning to. However, that information is no longer available where expected, making returning difficult. For this reason, there has been considerable effort in keeping links from breaking and fixing them when they do (e.g., [56, 99]), and in archiving the Web (e.g., *Alexa*).

Further evidence that the dynamic nature of the Web can cause problems with locating information, is that people do not trust the Web as a repository for information. Whittaker and Hirschberg found that people kept paper copies of documents they had found on the Web for archival purposes, even when keeping the documents incurred some cost to the keeper [123].

3.3 "Where'd it go?"

Both as additional evidence that people have trouble returning to information on the Web, and to understand better what sort of trouble people have in the real world, here I present an observational study of the difficulties people encountered returning to information in a dynamic information environment, the Web. The study was conducted by analyzing instances, collected via a Web search, where people expressed such difficulties. The following quotation is an example from the data that illustrates a number of the interesting observations that arose from the study:

> I remember when I first joined these forums! There was little "Did you know" facts about Star Wars at the front page, but they were replaced with movie qoutes! Why did they disappear?[1]

The description emphasizes that the Star Wars facts were originally encountered on the forum's front page, and there was a trend in the data to emphasize the importance of the original path used to encounter the information target. On the other hand, time is not mentioned directly in the quotation, but rather alluded to by relating earlier access to a personal event. The study suggests that the temporal aspects of when the information was seen before were surprisingly unimportant. Frustration, suggested in this example by the exclamation marks, was commonly observed, and it appeared that an explanation of why the change had occurred was often sufficient to allay frustration, even in the absence of a solution. In the example given above, instead of asking for a pointer to the missing information, the person asks for an explanation.

The discussion begins with a discussion of relevant re-finding research, highlighting research that has been done in the dynamic environment of the Web. After presenting the

[1] All quotations are reported exactly as they occurred. Spelling and grammar errors have not been corrected.

details of the observational study conducted and giving an overview of the data collected, the findings mentioned above, among others, are discussed in greater detail. The paper concludes with a discussion of the study's implications on systems that support re-finding information in dynamic information environments.

3.4 Methods

The study presented in this paper analyzed instances where people expressed difficulty returning to information that had changed. These instances were found by collecting Web pages that contained the phrase "Where'd it go?" The phrase was selected because it was general, yet implies that something cannot be found because it has moved. The term "move" is used here because change necessarily involves movement; move, remove and modify (i.e., remove-and-replace) all entail the movement of the originally presented information. When a person is only interested in returning to what has been observed before, any change can be viewed as a move. In the data collected, "Where'd it go?" was used in reference to all three types of change (move, remove and modify).

Web pages containing the phrase "Where'd it go?" were collected by performing a Google Web search. Because Google only returns the top 1000 results, the search yielded 1000 pages of 5,340 reported. This set of pages could have been supplemented by performing the same search on other search engines, such as AllTheWeb, AltaVista, and Lycos. However, there was considerable overlap among the result sets from different search engines, with 55% to 62% of the top 100 results already belonging to the Google set. Other phrases with similar meanings, such as "Where did it go?" and "I can't find it anymore," could also have been used to supplement the document set. "Where'd it go?" was selected because, of the phrases tried, it was found to be the one most commonly used in the appropriate context. Note that the additional instances found via other search engines and phrases appeared to merely enforce the phenomena observed in this paper. This suggests that little would have been gained by supplementing the data collected.

The Web is an emerging source of data for observational studies. Several studies have analyzed postings collected from specific message boards to understand topics ranging from how people view robotic pets [41] to how they recover from injuries [101]. Observations have also been collected using search results. Good and Krekelberg [45] constructed KaZaA queries to see if people accidentally exposed personal data. Clearly, data collected from the Web can be noisy, but the large quantity that can be cheaply gathered compensates for the noise. Further, data can be collected by mining the Web that might otherwise be unobtainable. For example, it would have been difficult to devise a study any other way that would have permitted observations of people having difficulties due to a dynamic environment during personally motivated searches.

The data were analyzed using standard qualitative techniques [113]. An initial pass through the data was made to develop a coding scheme and identify the 258 instances that contained expressions of difficulty finding information. A second pass was then made to code this subset. In the analysis, significantly more Web pages than instances were inspected, as each Web page's surrounding context was also explored. For example, if the page contained a message board posting, any responses were also analyzed. If the page no longer contained "Where'd it go?", the copy in Google's cache was analyzed.

Figure 3-1: Three instances that contained the phrase "Where'd it go?". The first (a) contained a posting from a person looking for Web functionality. The second (b), titled "Where'd it go?", is a redirect page. The third (c) offers support in finding information that's moved as a result of a site reorganization. *Haven't yet pulled in the correct set of pictures ... Will do.*

3.5 Overview of the Data Collected

This section gives an overview of the data collected. The 258 instances described, several of which are shown in Figure 3-1, exclude duplicates and pages that did not involve searches for information. The section begins with a brief overview of what the analyzed pages looked like. It then discusses the types of information people described seeking within them and the reasons the information being sought had moved. Subsequent sections discuss the patterns that emerged from this data.

3.5.1 Understanding the Pages in Which the Phrase Occurred

The topics of the Web pages collected ranged broadly, from technical software languages to teen sleeping disorders. The page format also varied. The data contained ten to twenty instances each of redirect pages (e.g., Figure 1(b)), Web logs (blogs), articles, and frequently asked question (FAQ)/help pages (e.g., Figure 1(c)). However, most of the pages in the collection (165 pages, 64%) were message board and newsgroups postings (e.g., Figure 1(a)). The popularity of this format could be due to "Where'd it go?" being an informal and conversational phrase, and thus commonly appearing in informal and conversational settings such as message boards.

When the phrase was used in a message board, it tended to be by someone who actually wanted help in locating a piece of information. Such postings were useful for analysis because 69% of them included responses to the query. The 21 Web log pages also occasionally (although less often, 14% of the time) included responses, in the form of comments. How-

ever, the phrase was not exclusively used by someone actively seeking information. In 68 instances, or 26% of the total instances, "Where'd it go?" was used rhetorically. Rhetorical use was particularly common when the phrase occurred in FAQs or on redirect and help pages. The instances where the phrase was used rhetorically provided insight into how information re-finding in dynamic information environments is currently supported.

3.5.2 What "It" Was

While the phrase "Where'd it go?" was used to refer to everything from physical objects (e.g., "Where'd the spider go?") to abstract concepts (e.g., "Where'd the day go?"), only the 258 searches for information were analyzed. Of these searches, 174 were for Web based information (67%), 74 were for non-Web based electronic information (29%), and 10 were for non-electronic information (4%).

The most common piece of Web-based information referenced was general Web content, as illustrated both by Figure 1(c) and the following FAQ:

> You used to have a "Nekkid People" section on your Web Site. Where'd it go?

Web sites (e.g., Figure 1(b)) and message board postings were also frequent targets. Slightly less common targets included pictures, message board threads, information to download, and Web functionality (e.g., Figure 1(a)). Non-Web based information searches were similarly varied. For example, one page contained a paper describing the problem of losing shared files in a collaborative work environment [86]. However, a particularly large percent (46%) of the non-Web based electronic information searches involved re-locating features in applications or operating systems, primarily after an upgrade.

The information target had been seen before by the seeker in a large majority of the cases. In the 38 cases where it had not (15%), the seeker nonetheless had a strong expectation that the information used to exist in a particular place. This expectation often came from others. For example, people sometimes wondered where information pointed to by a link that had been made by someone else had gone. In the following instance, the seeker wanted a message board posting others had said was interesting:

> Where'd the post go that you are refering to? The post "Technical and Plot Itmes of Importance" seems to have been deleted. What did it say?

The expectation that the target existed also came from related experience. As mentioned earlier, people often asked where functionality went after upgrading software. Although the functionality could not be found in the new version of the software, the seeker had an expectation that it would be there based on their experience with the old software. In general, these cases where the seeker did not have immediate experience with the information appeared very similar to the cases where the seeker did.

3.5.3 Where "it" went

The most common reason the information target being sought had moved was that another person had changed it or its information environment. Fifty instances contained explicit mentions of another person moving the target, and many others implied it. For example, when someone could not find a posting on a message board, it was often because a moderator had deleted the message. Similarly, missing Web content tended to have been moved by

the site administrator. However, there were instances where changes occurred for other reasons. In 24 cases information moved because a site had gone down or a piece of software had failed. There were no instances where people wanted to find information that had changed because it was time dependent (e.g., last week's weather or old stock prices). This could be because people had strong expectations that time dependent information might change, and thus did not expect to be able to relocate it. Regardless of why the change occurred, in 95 of the cases (37%) the missing information became unavailable.

The information target had not always moved, and in 23 instances (9%), it clearly had not. Instead, the seeker was simply unable to locate what they were looking for. Consider the following posting, titled "Where'd it go???":

> I must be blind! I posted my intro and first time weigh in - I saw it posted - honest! and now its gone...unless I'm blind! lol Help?????

The posting being sought had not moved, but instead had been posted on a different list than the seeker remembered. Still, as the phrase "Where'd it go?" implies, the seeker believed it had moved, and this belief of change, even when inaccurate, was present in virtually all of the examples.

3.6 Describing the missing information

This section discusses how people expressed the problems they encountered that led them to use the phrase, "Where'd it go?" As such, it addresses the 165 instances where the phrase was used by someone actually looking for a piece of information (rather than, for example, rhetorically). The percentages reported in this section are out of 165.

3.6.1 Expressions of Frustration

People expressed frustration when they could not locate information they believed had moved. In 41 of the 165 instances where someone was trying to locate a piece of information (25%), there was a clear statement of frustration, such as "Ahhh *pulls out masses of hair* Where'd it go?!?!" or "where'd it go.. gah.. i'm panicing now.. ahhhh.. ok.. ok.. settle..". There are many reasons why people might have felt such frustration. For example, when information moves, it challenges the control a person has over their information space and destroys their sense of continuity of the information. One explanation that appeared in the data was that losing information made people feel bad about themselves. In 18 of the cases (11%), people who could not find information called themselves stupid or crazy (e.g., "I thought I was going crazy on this one") or assumed blame for their difficulties (e.g., "maybe i'm doing something wrong?"). As will be discussed in a later section, an explanation of why information had moved was often a satisfactory answer. This could be because while explanations do not solve the problem, they remove the stress of losing things and allay the fear of being stupid.

Of course, the large amount of frustration observed could be in part due to the fact that people only went through the effort expressing their difficulties on the Web when a problem was particularly frustrating. Most people do not announce to the Web every difficulty they have re-finding information. This is supported by the fact that in 13% of the instance (22 times), people who had not originally mentioned having trouble re-finding something agreed when someone else did, saying, "I noticed it too!" or, "I was wondering the exact thing. Where DID it go?"

3.6.2 Shared Context

The phrase "Where'd it go?" often appeared with very little explicit surrounding context. An illustrative example of this can be found in Figure 1(a), where the information target is described only as a "thingy". Similarly, the person who posted the following could not name their target:

> I miss that little tab thingy on my profile that took me straight to my groups...that was convienient! Where'd it go?

Nonetheless, the intended audience in both cases understood what was being referred to, and both received responses. An instance of a particularly cryptic posting was posted under the title "ALRIGHT WHERE'D IT GO!":

> HEY! who thieved the guids to dotb solo'n, and neriad shall solo'n-i knowfaint poitns not the detailed particulars-so uh someone post the url, or email me or somthin

Even this confusing post was understood. Although several expressed puzzlement, one person posted an explanation:

> I do believe she/he is referring to the drums of the beast, and neriad shawl guides, mainly how to obtain each of them solo, most likey either a thread or a link on the old site would be my guess.

Relying on shared context relieved some of the burden from the seeker of expressing their information need. The types of context that were explicitly stated suggest what the seeker considered necessary to specify their target, and the following addressed the more commonly mentioned types.

3.6.3 The Importance of Path

The path via which the target was originally found appeared to be very important, and in 52 of the instances (31%) the path was explicitly mentioned. As an example, 17 times (10%) the query "Where'd it go?" clearly referred not to the information target, but rather to a step along the path. This is illustrated in the following quotation, where the target was a recipe, but the seeker asked for help getting to the containing Web site:

> Okay, where's the link? I wanted to try this quick and delicious recipe everyone raved about

Similarly, someone else asked for a pointer to a newspaper, despite their target being the obituaries it contained:

> Can anyone please provide info on the demise of the Jersey City Observer newspaper? In particular, whether or not it was bought a a competitor, and if so,and as importantly, where it's OBITs and other Personals may have be today?

Alvarado, et al. [4] observed this same behavior for search in general, and suggested several advantages to searching this way, such as that the source is often easier to locate than the target, and that the source can provide valuable information about the target, such as its trustworthiness.

3.6.4 Time Not Important

Despite the fact that time is often treated as a uniquely important feature in systems that support returning to information [39, 48, 104], the instances analyzed in this study did not contain many references to exactly when target was seen before. The temporal aspects of previous interactions with the information target were mentioned in 33 instances (20%), but less than half of those instances made specific references to time in terms of dates or time intervals. When they did, the event usually occurred that same day (e.g., "this morning", "earlire today", "half an hour ago"), although twice the event had occurred exceptionally long ago (e.g., "for over twelve years now").

There were few specific references to time in the interval between the recent past and long ago. Instead, the references were vague (e.g., "recently", "earlier", "way back when", not in "quite a while", and not "for some time"). Consider as an illustrative example five different people's postings looking for an online intelligent agent that could be talked to via instant messaging. Only two of the postings made any reference to time at all:

> i) OH MY GOD, where is SmarterChild, he's been ofline for a LONG time, and...WHERE DID HE GO?
>
> ii) Smarter Child has been offline for some time. What's going on?

However, based on these references, it is impossible to tell how long the agent had been missing.

Time was sometimes referred to in a personal manner. In five cases, previous interaction with the information was related to a personal event. This can be seen in the quotation in the introduction ("when I first joined these forums"). Regularity of access was also mentioned eight times. One person, looking for a Web site that had disappeared, said, "I check it almost every day". Another poster looked for an advertisement seen many times before:

> For awhile now, ive been seeing an advertisement ... Now I cant find the Inside Sun advertisment ... So, the question is, what happened to it?

Regularity of access appeared to be used as proof that missing information once existed, and that the seeker once knew how to find it.

3.7 Answering "Where'd it go?"

In addition to looking at how people described missing information, the answers people received to "Where'd it go?" requests were analyzed in order to understand how the problems encountered were solved. Solutions ranged from explanations of what had happened, to work-arounds so the seeker could deal with not having the information, to actual resolutions. The three types of solutions (explanations, work around, and resolutions) were not mutually exclusive, and sometimes all three occurred in a single instance.

The question "Where'd it go?" was sometimes anticipated, used rhetorically by information providers trying to ease the re-finding of information they had changed. For example, "Where'd it go?" occurred twelve times in frequently asked questions (FAQs) (e.g., "Retrieving the Office Toolbar - Where'd it go?") and on help pages (e.g., Figure 1(c)). Other pages referenced a Macintosh manual's appendix titled "Where'd it go?" The appendix

linked common tasks in other operating systems, such as Windows or older Macintosh versions, with the new operating system:

> "Where'd it Go?" is a cleverly conceived reference for OS 9 users. It isn't just some skimpy table that tells you which menu now contains a given command. It's a reasonably good translation dictionary for work habits that includes explanations of the new way to think about the task.

Clearly the problem of re-finding information that has changed is a significant enough problem for people to invest considerable effort helping others deal with it. As such, these instances provide insight into how information re-finding in dynamic environments is currently supported. For example, the fact that people remember the path that they originally encountered information was sometimes taken advantage. The data set contained twelve redirect pages (e.g., Figure 1(b)), and five "404: file not found" pages. These pages provided information about where and why the target had moved at the site it used to be located. Thus, while the previous analysis focused solely on those instances where information was actually being looked for, the analysis in the rest of this paper includes all of the 258 cases where "Where'd it go?" referred to information.

3.7.1 Explanations

The question "Where'd it go?" was often answered with an explanation of where "it" had gone. Even in the absence of an actual pointer to the sought after information, it appears explanations were important in allaying some of the frustration people felt at not being able to re-find information that had moved. Explanations were the most common solution observed, occurring in 33% of the instances (85 times). Explanations were particularly common when "Where'd it go?" was used rhetorically in reference to information that had became unavailable, occurring in 19 out of 23 such cases (83%). For example, all five of the "404: file not found" pages provided an explanation of what had happened to the information, as exemplified by the following:

> I haven't been able to maintain these pages the way I would like to. I've removed the pages I used to have here. If you need a link from one of my old pages, I may be able to retrieve the page from my archives. I'd be happy to send you, via e-mail, any information that was on those pages.

It appeared that explanations were so important that they were often made up. In 38 instances, "Where'd it go?" was asked with a hypothesis of where it had gone. In an illustrative example, someone noted a missing message board with a suggestion for why it might have disappeared:

> Nothing posted after December 6 went onto the board, then today it disappeared completely! Maybe Eric didn't pay his web page hosting fee.

Replies also often guessed at what might have happened (22 times). While the following is not an explanation of why someone's post had moved, it is a hypothesis:

> Well cindi.....in my experience, if Spike doesn't like how a post is going, or if it is too off topic or controversial, he'll take it out. Which post was it? Sorry!

Explanations often seemed to be sufficient to allay the frustration of the searcher, and people who provided explanations were often thanked, but rarely followed up with. In fact, explanations were sometimes the sole target of the query. This was the case for the quotation in the introduction, and the following is a more extreme instance; here the person created a thread titled "Where'd it go?" despite having already found the target:

> Knox [a server] just seemed to disapear for a couple of minutes and then came back again

These cases where the target was already found highlight the importance of explanations when information moved.

3.7.2 Work-Arounds

Another solution, observed in 22 of the pages analyzed (9%), was to suggest a work around to deal with not having the desired information. For example, someone looking for functionality that had changed asked:

> Where'd it go to? I know I can use guides to manually center elements, but I kinda miss the Center command from FW4.

The respondant pointed the seeker to a worthy substitution, saying, "I found it, or something better, under Window—Align menu." Similarly, a "404: file not found" page suggested alternatives that might be of interest. The page, which once provided satirical content, recommended another Web site with comic information:

> For the time being, I (Pete) reccomend you go here and read some comics, as we all need our daily dose of funny, don't we.

Work-arounds were not always satisfactory, however. This is illustrated in the following instance where the seeker was provided with a successful work-around:

> whatever modules ARE working right now seem to be what i need... but–where'd it go off to? if i do need it in the future, how can i restore it?

In this case, the person still wanted an explanation, and perhaps even a resolution to the problem.

3.7.3 Resolutions

The information being looked for was successfully located in 82 of the cases (32%). An analysis of these instances where the problem was resolved suggests the importance of being involved with the change; when a definitive solution was provided, it was often provided by the person who had made the change. While this obviously occurred regularly when "Where'd it go?" was used rhetorically, it was also common when "Where'd it go?" was used by people actually trying to locate a piece of missing information. Of the 47 instances where people trying to locate information were told where it had gone, ten of the responses were clearly from the person who made the change. In the following instance, the person looking for a posting they had made was pointed to its new location by its mover:

> I moved it to the bug reports forum since it seems to be a bug that is effecting all premium stores.

The person who changed the information also was often the one to provide an explanation of why the information had moved. People trying to locate information received 52 explanations, and 22 of those were obviously from a person involved in the change. As an example, when people asked where a message board posting had gone, it was almost always the moderator who explained that it had been deleted. In another example, someone asked:

> I won the "Jr. Wolfer, 75 posts" contest, but, where did the "Contests and Stuff" section go? And I think the contests idea is pretty good, too. I'm wondering if you got rid of it?

The seeker received an explanation from the person who organized, and subsequently canceled, the contest:

> Well, it's like that: Being a global moderator needs tons of posts, but the contest only required 75 posts, wich is a very litle number, so i cancelled, and maybe i'll put a new contest soon.

While it was often difficult for people not involved in the change process to locate missing information, people who were involved appeared to maintain a good understanding of the information and what had happened.

3.8 Multiple Users of the Same Information

People often had different intentions with the same information, as illustrated by the fact that the most common reason for information to move was another person. As a result, several interesting problems worthy of further investigation arose. For example, sometimes information was removed because people in general were not interested in it, despite the information being of interest to the seeker:

> I think they got removed because there were only about three of them, and they got old fast

Information was also sometimes removed because the information provider actively did not want the information to continue to be available. For example, the author of the following quotation references a previous posting he did not want others to be able to read:

> I was hoping nobody saw it, oops. I got taken in by that Metallica spoof going around the net. I found out it was a parody site so I deleted [the posting].

This same conflict was also evident in the seven instances when information was removed for copyright reasons:

> [T]he French site Elostirion.com was asked to take down the image of the Ringwraiths. You can still read the news on this story from this morning which ended with the confirmation of these characters in fact being uncloaked ringwraiths.

The conflict of interest between information users, who want the information they interact with to be persistent, and information providers, who want control over their information, is related to copyright issues that have arisen in making digital library documents persistent [50].

Another interesting conflict that arose was highlighted by the large number of message board postings that went missing because they were deleted by moderators:

> The web site you list is commercial & is the reason your post was removed. I have now edited out the site so you will understand. Please read the goals & rules of posting on sleepnet.com forums.

In these cases, the people looking for their past postings were not interested in finding the information for themselves, but rather in ensuring that others could see it. This was in direct conflict with the information providers, who had removed the posting because they explicitly did not want the content to be available.

3.9 Design Implications

The previous sections have discussed the patterns that emerged from the data in re-finding behavior as it occurs in dynamic information environments. This section discusses the implications of these observations on the development of future solutions. People currently employ many tools to return to information, from search engines to bookmarks to email messages [61]. While the information environments that these tools work in can be dynamic, the tools do not explicitly support such interactions.

Systems that provide information access to a number of users, such as Internet search engines, have a particularly difficult task because while whether information is being found or re-found is inherently user dependent, the systems do not know about individual users. Nonetheless, this study suggests several ways such systems could better support information re-finding. Because it was common for information that moved to become unavailable, systems should cache as much information as possible. However, if time is used to access this cache, it should be in a relative sense, much as has been explored by Ringel, et. al [105]. Furthermore, time should not be a uniquely important access point into the cache.

Systems could take advantage of the importance of the path taken to originally locate information by not just supporting search for old information, but also preserving the original navigation. For example, a news Web site should not just archive past articles, but also archive the news digest page that originally presented the news. The number of people who said, "Me too," when a change was observed by someone suggests that perhaps people tend want to return to the same information and notice the same changes. This could mean that information that is returned to by a number of people should be made easy to find for others. Desktop systems and other systems that can track individual users are clearly at an advantage. Personalized systems need not cache all information, but rather only the information the user has seen before, much as in the "Stuff I've Seen" system [34]. Access into personalized caches can be improved to include the personally relevant information, such as the path the user used to access the information, the regularity with which the information was accessed, and temporal aspects related to personally relevant events.

The large number of times that "Where'd it go?" was answered by the person who moved "it" demonstrates the importance of being involved in the process of change to retrieving old information. While it is not necessarily the case that information must remain static for

users to feel comfortable with it, this study suggests that users should have an understanding of what happens to it. One way a personalized system could support returning to dynamic information would be to provide awareness and control over any changes. For example, when a person clicks on a bookmark, if a copy of the bookmark from their previous visit has been cached, any changes could be highlighted, and accepted or rejected by the user. By including the user in the change process, chances are greater that the information can be found again, and, at the very least, the user will have an explanation of what has happened, alleviating much of the potential frustration. Note that it is not necessary to provide awareness of every single change, only significant changes. I [117] investigated people's interactions with information that changed slightly, and found many changes went unnoticed.

The study also raises several problems that might arise for systems that individualize what a person sees. It was sometimes important for a person to know what others see. For example, when a person looked for a past posting they had made, they were not interested in finding the information for themselves, but rather to ensure that others could see it. If the user could still find their old posting because, for example, it was cached for them, the user might not even know that it was not accessible to others. Also, people often removed content because they did not want it to be found again, either because they deemed the content inappropriate or because of copyright issues. These examples suggest potential pitfalls for personalized systems supporting dynamic information re-finding.

3.10 Beyond the Web

There is also evidence that people have trouble returning to uncontrolled dynamic information outside of the Web domain. For example, White, Ruthven and Jose [122] proposed a new user interface for a search system that tried to dynamically help the user find their information target. They did this by providing the most relevant sentences from relevant documents in addition to the relevant documents as a search aid. These sentences were dynamically reordered based on passive relevance feedback from the user's search behavior.

White, Ruthven and Jose were surprised to find that this dynamic reordering actually hurt search performance rather than help it, and blamed this on the way way they laid out the information. However, when the fact that users like consistency and control is taken into account, their findings are not at all surprising. Each time the relevant sentences were reordered, the user had to spend time looking at the whole list again. Even if the new sentence ordering was objectively "better", it could actually be worse because it did not allow the user to use any of the context already built around the sentences.

Dynamic menus are also a good example of the problem in a non-Web domain [89, 106, 111, 121]. The speed at which people access items in a menu is a function of that item's location in the menu [106]. For this reason, for long menus attempts have been made to dynamically pull up the more frequently accessed items to the top of the menu. However, even though this might seem like it should help, Mitchell and Shneiderman [89] conducted an experiment to see whether organizing information in menus dynamically helped users find that information faster. They found that users not only started out faster with static menus, but also that they preferred static menus to dynamic menus even when their performance with both was the same.

In another study of dynamic menus, Somberg [111] found that while people perform better initially with some sort of rule based menu organization, after a moderate amount of practice, they prefer a positionally constant arrangement. This can be explained by Sears

and Shneiderman's [106] observation that people can find frequently accessed items more quickly than less frequently accessed items. This implies that people have some sort of memory about the item's location and use that memory during access, and changing the information the user remembers to present the information target sooner does not actually help the user find the items faster.

For this reason, we see a "hot-list" solution suggested [106, 121], where the frequently accessed menu items are put in a hot list, while the overall menu remains the same. I will discuss this solution further when we look at current work in returning to dynamic information in Chapter 4.

3.11 In the Future

I have described several examples of the problem of returning to dynamic information that currently exist. However, as suggested by Hearst [50], "documents containing .. dynamic elements may soon no longer be considered an aberrant form," and it is likely that the problem of returning to dynamic information will become even more important to consider in the future than it currently is.

In this section, I look at two places the problem could arise in the future. First, there are many applications that currently have restrictions placed on them to keep users from interacting with dynamic information. Should we have a good way to support such interactions, these restrictions would not be necessary. Second, in addition to the increasing prevalence of the Internet and other such externally dynamic information sources, I believe we will soon find that even the information a person considers directly under their control will soon become dynamic.

3.11.1 Helping Existing Applications

There are a number of applications that don't currently display dynamic information, but could provide some benefit to their users, such as quicker response times, if we had a good way to help users receive newer and better information.

As an example, meta-search engines collect search results from numerous different search engines and combine the results. Because of this, they are dependent on a variety of resources with a varying response time before returning anything to the user. The user must either wait until the slowest resource responds or the meta-search engine must place a time limit on response. For example, Baldonado and Winograd impose a time limit on their meta-search system [7]. Because they are aware of the trade-offs between wait time and quality of results, they give some control of the time limit to the user. But if the user instead could begin interacting with the first available results and then receive new results in a way that didn't destroy their context, there would be no need for such limits.

Another possible example: I know that Kai Shih, in his work, argues that clustering has to be fast to provide quick results to the user, and for this reason a lot of work has been done of fast clustering algorithms. However, if clustering algorithms could provide incremental results, and it were possible to allow people to interact well with dynamic information, the clustering algorithms wouldn't have to be so fast.

3.11.2 Personal Information Becoming Dynamic

As electronic information becomes cheaper and easier to create, collect and store, we will soon find individuals owning a phenomenal amount of information. This could include every Web page the individual has ever seen, every email that was ever at all possibly of interest, every photograph a friend ever wanted to share, and more. In fact, it is likely that it will be too much information for the individual to successfully manage alone.

For this reason we are beginning to see personal information systems such as Haystack [54] and Chandler [63]. These systems are intended to help users deal with the overwhelming amount of data they own, and one of the ways that they will do this is by dynamically guessing what the user is interested and providing that to the user. For example, a person's *haystack* is a repository of all of the information that that individual comes into contact with. The system is constantly looking over this information, and drawing new conclusions about the data it contains. For example, it may decide that a certain email I wrote my advisor is related to a paper I'm writing, and link the two. Or it may decide that I am interested in cooking, and run a search on the Web to collect new recipes for me. In this way, the information within a haystack, information that I consider directly under my control, is constantly changing, both in what is available, and in the relationships between objects.

More about agents (e.g., [6]).

The problems that people have when returning to dynamic information that is not under their direct control will only be exacerbated when the information a person considers directly under their control becomes dynamic. With personal information a user has an even greater expectation of consistency and control than they do with external information, and so any system that introduces dynamics into personal information will have to be very aware of this. But by effectively integrating the display of dynamic information into systems like Haystack, I believe it will be possible to create an interactive and dynamic desktop where the system and the user work together to organize the information.

Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information on it.

Samuel Johnson (1709 - 1784)

Chapter 4

Current Solutions to Dynamic Information Interaction

In order to better support users returning to uncontrolled dynamic information, in this chapter I look at how current systems support any type of interaction with dynamic information. Relatively little research has been done on the specific problem I am investigating, but by looking at how related research suggests supporting user interaction with dynamic information where either the user doesn't have an interest in returning to the information or the user is in control of the changes, we can get an idea of how to address this problem.

It is useful to think of the current work in dynamic information interaction as divided along two different axes. One axis involves the user's interest in the dynamics of the information. Traditional research in human-computer interaction with dynamic information has assumed that the user is interested either in the dynamic of information that changes or in the most recent state of the dynamic information. For example, a user watching the stock market is typically interested in either the current stock prices or how the stock prices have changed recently. However, the user could also be interested in re-finding information that's been seen before, as is the case when a person wants to know, for example, what the stock price was when a particular stock was purchased.

The other axis we look at is whether the information is dynamic because the user was the one to make the changes to the information or whether those changes were made outside of the user's control. This is important because whether or not a user is involved in the changes can affect the cues the user can use in retrieval. As discussed in Chapter 2, people use context when information seeking, and when the user performs the changes those contextual cues are still available. This is supported by Lansdale's work [74], that suggests the more involved in storage one is the easier retrieval is.

		Interest in current information versus returning	
		Current	Returning
User makes changes	Yes	Drawing a picture, writing a letter, changing queries into a databases	Returning to an old version of a file, finding an email in a folder
	No	Collaborating on a project, watching stock market or weather	Examples from Chapter 3, using search results as bookmarks

Table 4.1: Examples of types of user interaction with dynamic information. Work is divided based on whether the user made the change and whether the user is interested in working only with current information versus returning to the information.

		Interest in current information versus returning	
		Yes	No
User makes changes	Yes	Dynamic queries	Version control, undo
	No	Collaborative systems information visualization	*???*

Table 4.2: Related work in each of the user interactions with dynamic information.

Table 4.1 shows examples of types of interactions a user might have with dynamic information broken down along these two axes. The specific problem I will address, of returning to search results, exists in the lower right-hand quadrant of the table. Table 4.2 shows the related work in each of these quadrants, and it is in the lower right-hand quadrant that the least amount of research has been done. In this chapter I discuss each of the quadrants individually, and then look at how I can use the work in the other quadrants to generalize solutions for the quadrant I am interested in. Note that the boundaries are somewhat fuzzy. For example, version control systems are often used by many people, in which case it is not only the user who makes the changes. I will point out cases that cross the boundaries as it applies.

The discussions are currently somewhat rough ... Sorry.

4.1 User Makes Changes/Current Information

The upper left-hand quadrant of the table contains the cases most distant from our problem. These are the cases where the user makes the changes to the information and is interested in viewing only the information that is most current. A good example of this type of interaction is when a user repeatedly queries a large database. The user looks at the large amount of data by creating complex queries. The information that is changing is the query and the query result set. For example, the user could be interested in purchasing a house, and create queries to select available houses based on cost, size, location, etc. By changing the queries, the user is changing the view into the data, but typically does not want to return to previous views (although one could imagine one might want to).

To support this interaction, there has been work with dynamic queries, which allow the user to create complex and easily changed queries. Shneiderman and Ahlberg [3, 109] have done work on creating dynamic queries into large, static databases, handling cases such as the house purchasing example. With dynamic queries, it is the query, or view into the information that's dynamic, not the underlying information. This is somewhat similar to the organization of the information being dynamic. More recent work has been done with dynamic queries [76]. Also very similar to the dynamic queries is work by Hearst, et al., [52, 49, 125] to support search using faceted meta-data. The various facets are essentially used to create dynamic queries into a database. An important aspect of this work is that the user feels in control of the information. Control over the information makes the users happy: "The enthusiasm users have for dynamic queries emanates from the sense of control they gain over the database [109]."

None of this work, however, has addressed what could be a potential issue with these dynamic views into information, namely returning to information that has been seen before. A user could find something of interest, but, after changing the query considerably, forget the attributes necessary to run a query to return to that information. Instead, the interest

has always focused on displaying the current information. Although the faceted meta-data work by Hearst, et al., does help in returning to old queries somewhat in that it provides a lot of context and prompting. We can't really know if returning to the old queries was a problem with any of the systems, because the user studies with dynamic queries have not looked for this problem but instead have been confined to short periods of time.

Information filtering also belongs somewhat in this quadrant of Table 4.2. In information filtering, user profiles are used to describe what a user is interested in when accessing information. These profiles can then be used to filter the information that the user sees, providing the user only with relevant information. Information filtering also belongs somewhat in the quadrant where the user does not make the change and is interested in only most current information (lower left-hand), because the information that is being viewed is also often dynamic as well, beyond the user control. Here, however, I consider the case where the user changes the profile attributes.

Generally the profile attributes are specified by the user, but can also be learned from user behavior [10]. Regardless of how they are specified, the point is that the profiles can change, and thus the information that the user sees also changes. Like with dynamic queries, this could cause problems if the user wants to return to information that's been seen before with a different profile.

Baldonado and Winograd [7] work with SenseMaker is very similar to dynamic queries and information filtering. They create a profile of user interest that evolves, and allow the user to iterate during the search procedure, essentially creating dynamic queries. They do address the issue of returning to information, by allowing the users to store old contexts and return to them. Here, the problem could be not with being unable to return to the information because the query has changed, but with the fact that the underlying information (either the search results to the query or the Web pages that the search results point to) may change. This case is dealt with more in Section 4.3.

4.2 User Makes Changes/Interest in Returning

The upper right-hand quadrant of the table represents cases where the user is still the one making the changes, but in this case is also interested in returning to information despite any changes. For example, the user might be editing a document, and want to replace a paragraph she deleted earlier. To support this type of interaction there are version control systems, that allow a user to return to an earlier version of a document, and support for undo/redo.

4.2.1 Version Control Systems

Version control systems [86, 120] use complex time lines, or version trees, to manage the different versions of a document. However, beyond the complex notion of time in version control systems, there is little we can take from these systems as cues for what a system that helps users return to uncontrolled dynamic information, as the field of version control has placed very little focus on user interfaces. (*I'm still looking for more work here, have some leads to follow still: Osterbye[96], Haake and Haake [46], Fuller, Mujica and Pino [43], Davison and Zdonik [29], Hicks, et al. [53] and Whitehead [35].*) Version control interfaces also tend to focus on a single document, and not on how to deal with changes in the organization of documents.

In his work on version control systems, Tichy also offers this advice in dealing with dynamic information: "Indiscriminately storing every change produces too many revisions, and programmers have difficulties distinguishing them" [120]. Keeping this in mind, the system I build to help users return to information that is dynamic should not overwhelm the user with every version of everything they have seen.

It should be noted that sometimes the problems attempted to be solved by version control systems can cross over into the lower right-hand quadrant (the area I am interested in) because they are sometimes designed to be used by large numbers of people. For example, Marselas [86] discusses the problem of losing files because different people worked with the same information, resulting in considerable effort spent re-finding or re-creating what is missing. As a solution, he investigates version control systems.

4.2.2 Undo

Another common way to allow users to return to information that they themselves have changed is through undo [58, 23, 18]. Undo tends to involve a sophisticated notion of what a user understands to be change, and can, like version control systems, involve instances where the user wants to undo changes that have been made by other users [23]. *More about what we learn from this solution...*

4.3 Uncontrolled Change/Current Information

The bottom half of the table deals with the situation where the user is not the one making the changes to the information. This could happen because the information is time dependent, like stock quotes and news stories are. The information might also be changing because it is controlled by other people. For example, a shared document might be edited by co-authors. There has been considerable research as to how people should interact with this information when the user is interested in only the most current information and not in returning to information they've already seen. Note that the work in this quadrant primarily focuses on data that are changing, but sometimes also covers the case where the organization of those data is changing [42].

4.3.1 Time Dependant Information

A number of systems have been developed to help users interact with rapidly changing information, such as stock quotes, air traffic, sports scores or weather information. In these cases, the dynamics of the information are well understood by the user, and the most recent information is what's important. For example, a person keeping track of a stock price is primarily interested in the current stock price, and perhaps its recent change. While quite complex systems have been built to display stock prices (get MarketMapp/Wattenberg citation), they don't focus on returning to the information.

McCrickard, Stasko and Zhao [88] developed an interface to display dynamic information in the background, so a person can keep apprised of the current information. For example, they allow for basketball scores to kept in the background, but ideally only distract the user when something very interesting has happened.

They focus on the importance of awareness with dynamic information. They argue that "often with static information, a person seeks to answer a question or make a decision. Consequently, they examine an interface, come to some conclusion, and move on. But with

dynamic information, people's tasks more closely align to *awareness*, a constant knowledge about the state of and changes to a body of information [88]." However, I argue that there also exist similar tasks for dynamic information as for static, that people will also see to answer a question or make a decision based on the dynamic information.

None the less, they do provide some guidelines for maintaining awareness of dynamic information that could be useful in the development of a system to allow users to returning to dynamic information. *Give examples?*

Another area of research that focuses on supporting user interaction with information that is changing beyond their control is information visualization. Information visualization deals with looking at large complex information spaces. Often the information being visualized is static, but it can be dynamic.

Fry [42] (also more recent gene paper?) proposes visualizing the information using simulated organic properties, and uses this to show emergent Web traffic patterns, relationships between words in documents that have new words being added, and Web site structure, as well as human genes. He finds this is good to get a qualitative overview of the data.

Bederson, Shneiderman and Wattenberg [11] also look at information visualization when the information is dynamic. They note that many information visualization techniques could cause the display to change drastically, and note that changes in the data that cause abrupt changes in the display bad. They suggest that systems try to help users maintain locational context for the information that doesn't change. "Even occasional abrupt changes mean that it is hard to find items .. by memory, decreasing efficacy for long-term users." However, while they do focus on conserving some context for the information, the context of old information is not preserved, and they do not address the issue of how a user can find the information that has disappeared.

4.3.2 Collaboration

Often information that is changing beyond the user's control changes because it is being acted on by other agents (primarily people). *Talk about collaborative interfaces.* [30, 37, 36] In collaborations, people are generally creating something, and the focus has been on displaying the latest version of this something.

Systems like Haystack [54] could also have agents change the information. Currently has focused on displaying the latest information. I hope to allow users to return to information that has changed within this system.

4.4 Uncontrolled Change/Interest in Returning

The area of the problem I am addressing in this thesis falls into the lower right-hand quadrant of Table 4.2. Here the user is not the one making the changes to the information, but is interested in returning to information that has already been viewed. Examples of this type of interaction are plentiful, as discussed in Chapter 3. None the less, there has been very little work in this quadrant. The work that has been done falls into two categories: work with dynamic menus, and work with maintaining context on the Web.

4.4.1 Document Permanence and Digital Libraries

Some discussion of the permanence of documents. We have version control systems and digital libraries [103, 64] as an attempt to make electronic documents permanence.

Levy [75] looks at whether documents are fixed or fluid. "Indeed, part of the social and technical work in the decades ahead will be to figure out how to provide the appropriate measure of fixity in the digital domain." Also discussed by Hearst [50, 51].

4.4.2 Dynamic Menus

Dynamic menus are a good example of where a user has interest in returning to information that is changing beyond their control [89, 111, 121, 106]. Dynamic menus are menus try to place what the user is most likely to select at the top. The purpose of doing this is to save the user the effort of scrolling down long lists to find what they're looking for. Ideally, when the user returns to the menu to find the desired menu item they will be able to find it more quickly because it is at the top.

However, as discussed in Chapter 3, studies have found that people both perform better with static menus and prefer static menus to dynamic menus [106, 111]. For this reason, split menus have been used [106, 121], which provide several suggested, commonly accessed menu items at the top of the list, and then the menu below, unchanged. Sears and Shneiderman [106] found that these split menus were more helpful than either menus ordered alphabetically or ordered by frequency of access.

This suggests that a possible solution to the problem of returning to dynamic information could be to keep almost everything constant, but have a small window with the changes available in it. Maybe this is something to look into ...

4.4.3 Maintaining Context on the Web

There has also been some work with maintaining context on Web. Hayashi, et al. [48], looks at changes where external Web sites change independent of the progress of a user's activities. They deal with this problem by asserting that because the information is dynamic, time is a special and unique access point to the information. They cache the version that the user has seen, and give the user access to that. *More needs to be said about this paper, as it's relevant.*

Another source: Alexa was built to archive all of the pages on the Web. Perhaps there's a white paper on this.

4.5 Generalizing Current Solutions

In this section I look at commonalities along the two axes of Table 4.2 to understand what is an appropriate solution to the problem of using dynamic information that is both uncontrolled and something that the user wants to return to. It is by understanding what makes systems where the user doesn't control the changes different from systems where the user does, and what makes systems where the user wants to return to the information different from systems where the user doesn't, that I approach the problem of the lower right-hand quadrant, returning to uncontrolled dynamic information.

The work along the "Interest in current information only - Yes" axis of Table 4.2 is different from the work along the "No" axis in that all systems here contain a notion of time as a special dimension of the data. The work along the "User makes changes - Yes" axis is different from the work along the "No" axis in that they focus on awareness and giving user control, probably because of their lack of it. In Chapter 5 I look at awareness

and control as a solution to the problem of re-finding information in a dynamic information environment.

Chapter 5

One always begins to forget a place as soon as it's left behind.

Charles Dickens

Awareness and Control as a Solution

User control and awareness were important aspects discussed in Chapter 4 for helping people interact in dynamic information environments. In this chapter, I discuss the system I plan to build that will take advantage of user control and awareness to support returning to information in dynamic information environments. I begin with a discussion of related work, and then present the system I plan to implement. This system will rely on the fact that people do not remember everything about the information they interact with, and thus do not need to be in control of every change to the information in order to feel in control of it. I explore this through a paper-prototype study. Because I will use implicit measures of user attention to determine what the user will remember, I discuss related work using implicit feedback for inferring user preference and argue that similar measures can be used for inferring user attention. I conclude with a discussion of how I plan to evaluate the system I implement.

5.1 Related Work

Context is related to control and awareness because to give the user control, it is necessary to preserve the context, and to give the user awareness, it is necessary to let the user know when the context is changed. This section will contain a survey of related work on context: Its importance is discussed Chapter 2. People use a lot of context [7]. When retrieving an old document, people want its context [26, 31]. Perils of losing context. Context is particularly important for browsing [4, 33]. People navigate by landmarks [52, 82]. Benefit of browsing [52, 49, 4]. Context dynamic, a function of task [32]. [73] People remember gists, recall depends on our context [74, 87].

Awareness: Discuss interfaces that make users explicitly aware of changes [88].

5.2 Building a Re-Search Engine

I plan modify a search engine to create the "Re-Search Engine". The Re-Search Engine will echo a normal search engine for most tasks, but will assists a user in re-finding information by 1) remembering what the user remembers about previous search results and 2) presenting that information to the user in an appropriate manner.

5.2.1 Remembering What's Remembered

Based on related literature on implicit measures of user interest for search and a preliminary understanding of psychology literature, it appears the following information might be remembered about previously performed searches:

- Search results that were clicked on.
- Anomalous search results.
- The first and last search result in the list.

Consider as an example the search for "neon signs" shown in Figure 5-1(a). The user has looked at the search results shown in maroon, and it is likely that if the user were to perform the same search again, these results would be expected to appear in the results. The Re-Search Engine will remember clicked on search results. A more sophisticated version of the Re-Search Engine might use more sophisticated measures of implicit feedback to determine interest, such as the dwell time or the number of times a URL is followed.

The result "History of Neon Signs", even though it has not been clicked on, might be remembered because it is anomalous (most other results appear to pertain to selling neon signs). It might be possible for the Re-Search Engine to determine and remember very anomalous results using the text of the search results.

That the first search result appeared first is likely to be remembered. However, the ordering of the other two maroon search results probably won't be remembered. Thus the Re-Search Engine will note if a remembered search results appeared first or last, but it will not remember positional information for intermediate positions. The first and last results are also more likely to be memorable, and if a result appears in the first or last position, it should be more likely to be remembered. For example, the "History of Neon Signs" result might not be anomalous enough to warrant remembering on its own, but as it is both anomalous and last, it perhaps should be remembered.

It is likely that more is remembered about searches that were performed recently or that are performed regularly than searches that were performed a long time ago or rarely. Thus for recent or common searches, a sophisticated version of the Re-Search Engine might maintain:

- Every search result that was clicked on.
- The first and last search result in the list, whether or not it was clicked on.
- Many anomalous search results.

On the other hand, for uncommon or long past searches, it is probably only worth remembering:

- Search results that were dwelt on.
- The first search result in the list, if it was clicked on or is anomalous.
- Very anomalous search results.

By "forgetting" information as time passes, much as the user does, the Re-Search Engine can both present more new information to the user (because it is not encumbered with so much old information) and have a lower overhead in remembering and recalling every Web search result a user has ever seen.

(a) Original results

(b) New results

Figure 5-1: The original search results returned (a), and the search results returned at a later date (b). These results can be merged together, as shown in Figure 5-2.

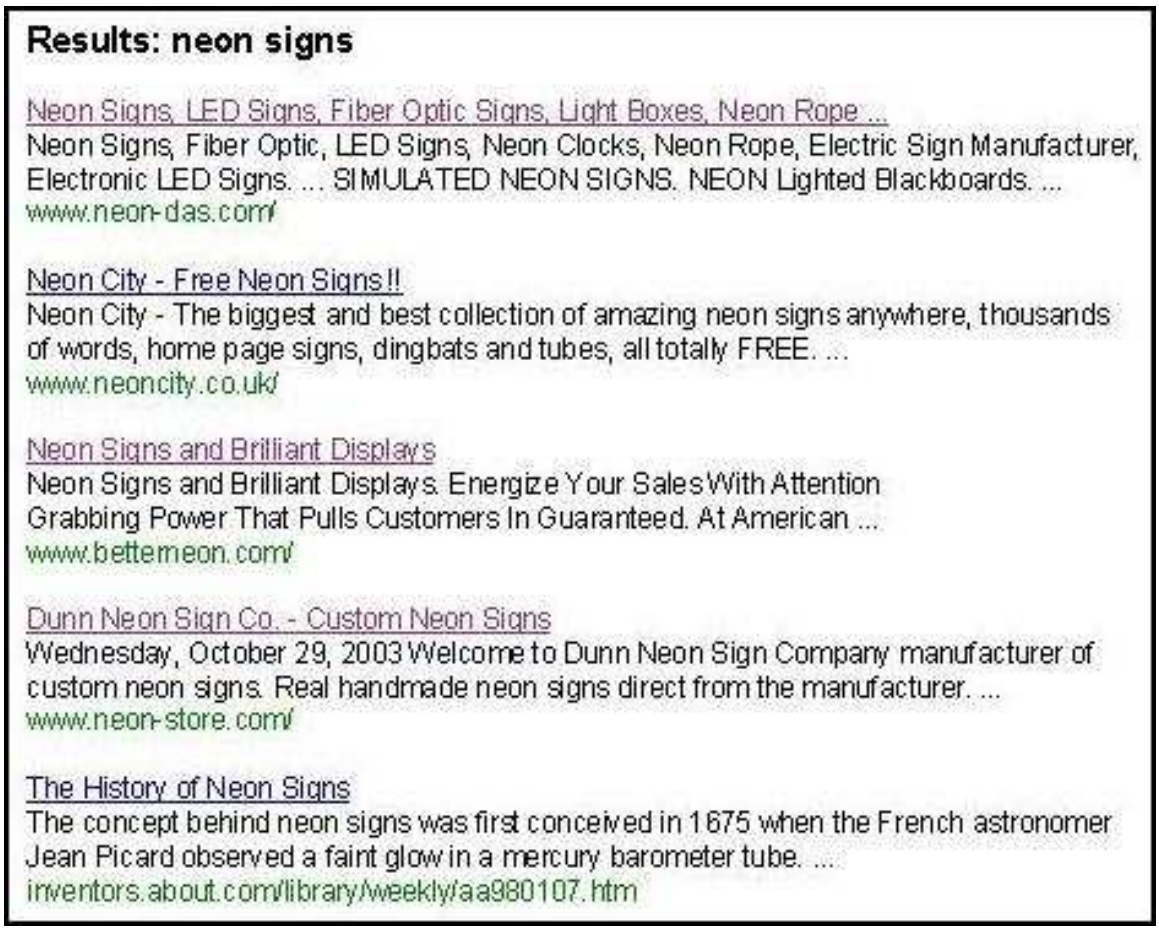

Figure 5-2: The results returned by the Re-Search Engine based on those shown in Figure 5-1.

5.2.2 Using What's Remembered

While for most searches the Re-Search Engine will perform exactly as a regular search engine, when a user performs a search that has been performed before, if the search results have changed, the Re-Search Engine will use the remembered information to assist the user in re-finding the original search results while still permitting access to new information.

In this section, I first discuss how the Re-Search Engine will guess that the user is trying to return to old search results. I then discuss how the aspects of the original search results the user remembers are displayed when the search results have changed. Because this display could block new information the user might be interested in, the Re-Search Engine will highlight areas of important change, and present them to the user for approval.

It is interesting to note that encouraging direct user involvement with any changes to their information also involves the user more intimately in the processes that change the data. Because the user must actively approve changes, in some cases the user may be able to easily incorporate his personal opinions into the changes, and can even stop mistakes as they are happening. Relevance feedback from the user becomes both a natural and integrated part of the interface.

Determining if the User's Returning

When a user repeats a previously performed query exactly, the Re-Search engine guesses they might be trying to return to previously viewed information, and will try to assist the user by displaying the remembered aspects of the previous search, as discussed in the following section.

However, a person might not always use exactly the same search query when trying to return to information. A more sophisticated Re-Search Engine might use a more sophisticated matching algorithm (e.g., significant overlap in the query results) to guess when the user is trying to return to previously viewed information. In these cases, the change in results would not be because the search engine index had actually changed, but because the query results changed slightly in response to the changed query. If the Re-Search Engine were to match in this manner, it would be important to ignore searches happening in quick succession, as in these cases it is likely the user is trying to use different but similar queries as part of their search session to find some information, and preserving their previous results would probably be counter-productive.

Presenting the Remembered Information

When a user performs a search that might be for previously viewed information and the results change, the Re-Search Engine will present the user with the aspects of the original search results the user remembers. For example, although the search for "neon signs" originally returned the results shown in Figure 5-1(a), at a later date it might return the results shown in Figure 5-1(b). The Re-search Engine will merge what is memorable from Figure 5-1(a) with Figure 5-1(b) to produce, for example, the result set in Figure 5-2. The types of changes that I expect the Re-Search Engine to deal with are discussed below:

Ordering changes As mentioned earlier, ordering is important for the first and last result, but less important for intermediate results. Thus I propose that if the Re-Search Engine remembers a first or last result, the ordering for that result be maintained, but the results it remembers that falls in the middle of the list, its ordering is free to change.

Figure 5-2 preserves the ordering of first and last search result from the original result set because both results were memorable. However, the ordering for the other two search results that were remembered because they were clicked on is not important, and they appear in a different order in Figure 5-2 than they did in Figure 5-1(a).

Additions and subtractions Since the user probably doesn't remember all of the results seen in the original search result list, it is ok for those results that are not remembered to be dropped from the new search result list. For example, the user probably does not remember the result "Neon Signs by Neon Central" from the first search result page, and since that result no longer occurs in the new result set, it is removed. This creates room to display new results such as "Neon Cit - Free Neon Signs!!" However, "The History of Neon Signs", because it is anomalous and last in the list, might be remembered, and for this reason, it is not dropped even though it also no longer appears in the new result set.

Title and description A sophisticated Re-Search Engine might also remember and maintain the title and description of previously viewed search results, but I do not plan to implement this in the first version.

Giving Access to New Information

Because this display could block new information the user might be interested in, the Re-Search Engine will highlight areas of important change, and present them to the user for approval. For example, in Figure 5-2, the system might suggest to the user that there is an additional search result of interest ("Outdoor Neon Signs, ...") or that another search result ("The History of Neon Signs") might no longer be of interest.

5.3 Information Can Change While Maintaining Control

I have discussed how contextual information is important in retrieving, and especially in re-retrieving, information. And I hypothesize that in the study described above I will find that giving the user control and awareness over changes to context is important for user interaction with dynamic information. However, while people remember a lot of information, not everything they see is internalized or remembered exactly. For example, people often only remember the gist of what they've seen, and cannot recall the exact words used [74]. If I told you that the name of this chapter were "User Control and Awareness", you'd probably believe me, despite the fact that you just read something different as the title.

The ability to change something without the observer noticing is called *change blindness*, and we can exploit it when updating information. Thus it is not important for a user to have control over every piece of dynamicism for them to feel in control. In this section, I first elaborate on the phenomena and give examples from the literature, and then discuss a paper prototype [117] study I performed to exploit it. In the study, I explored what sort of information people remember and use to return to information. I found that I was able to change large amounts of information without people noticing, and, more importantly, with people still feeling in complete control of the information.

5.3.1 Change Blindness Literature

Here I will talk more about change blindness [57]. People only remember the gist of what they've seen [74], etc. Change blindness might be too specific a term, since it's really more a question of memory in general. I need to review the psychology literature further.

Mayes, et al., found that people don't remember the words used in menus [87]. This can be related to the dynamic menu example I discussed earlier. Although numerous studies have shown that information in menus should be positionally constant, it is likely that the words could change without effecting performance.

Levy [75]: "Depending on our point of view, the granularity of our looking, we may or may not see change at all – or will see certain changes but not others."

5.3.2 The Problem of Clustering

I performed a paper prototype study on a sample problem to understand what elements of context are important. The sample problem I used was text clustering. Many clustering algorithms are able to roughly group documents into an initial clustering, but may take more time than the user is willing to wait to create a good clustering. The goal in presenting the clusters was to allow a user to begin working with the initial clustering immediately, while still providing them with the benefit of later clustering improvements.

5.3.3 Study Methods

Here I will talk about the study methods. I will describe what a paper prototype study is, give the number of participants, discuss how the feedback was qualitative, etc.

5.3.4 Clustering Implementation

In this section, I will describe the paper prototype that I implemented (give a picture). The interface is similar to the interface described in Section 5.2, in that it gave the user awareness of noticeable changes and required user approval for them to be made. The difference was that un-noticeable changes were allowed to occur without the user's permission.

Clearly a user does not have an understanding of information that she has not seen; so, unseen information is free to change as needed. Since many of the instances where the underlying information is changing involve very large collections of data, it is quite likely that the user will never see most of it, making the issue of how to maintain the user's context trivial. With my implementation of clustering, the information that the user has not seen includes the documents in the clusters that she has not visited and the documents in the clusters she has visited that are not ranked highly enough to be displayed.

On the other hand, the information that the user has seen contains both noticeable and un-noticeable changes. I call the information that the user has noticed *conceptual anchors*. The user develops some conceptual anchor into information, and when she wants the details of the information again, she will use that same anchor into the information to retrieve it. I propose that a good interface for interacting with dynamic information allows as much information as possible to change, while ensuring that those anchors the user has developed remain constant unless she has explicitly understood them to have changed.

5.3.5 Understanding Conceptual Anchors

Conceptual anchors are a function of what a user expects from the information she is working with. When you watch news in the evening, you may expect the news presented to you on the television to be changing, so in addition to remembering pieces of the story, you may also relate it to a specific time to put it in a context. But unless you have recorded it, you do not expect to be able to return to the story, so you may not develop anchors to aid in returning to the story. Note that this is different from when you read a newspaper article. In this case, you expect to be able to return to the news story. Instead of remembering the time you read the article, you may remember the section of the paper you saw it in. I found that if I preserved several conceptual anchors in the clustering problem the other information could change as needed. For example, a cluster is described by a set of keywords, found based on common word occurrences within the documents contained in the cluster. From initial tests, it seems that the user generates a general theme for the cluster from the keywords, and does not notice small word changes within the keyword list. This is especially true because I represent each cluster with a unique color. The user quickly associates the color with the cluster, and uses this mapping to navigate between clusters, rather than using the keywords.

Within each cluster a short preview of its documents is displayed, including a title and short summary for each document. Documents are ordered by their relevance to the cluster. From the tests I have performed, the user does not seem to notice the order in which the documents are displayed, as long as the first document remains first, and all of the visible documents remain visible. I believe that the first document's position is important for several reasons. It is likely the first document looked at, and may be remembered for that reason. Additionally, being first is more distinctive than being in the middle. Its position is obvious, where as a document located half way down the list could be perhaps fourth or perhaps fifth. The user does care about which cluster a document was located in when he first saw it, but does not seem to mind if it later also shows up in a cluster where it is related by content.

As long as the anchors are kept constant, each of the test subjects expressed a feeling of total control over the information, and often articulated surprise when I informed them that they had been working with information that was changing. "You say information was changing," one woman said to me, "but I did not feel like it was changing."

5.4 Implicit Feedback for Inferring User Attention

(From Diane Kelly and my paper [67]. Needs to be made to fit in better.)

Relevance feedback has a history in information retrieval that dates back well over thirty years (c.f. [112]). Relevance feedback is typically used for query expansion during short-term modeling of a user's immediate information need and for user profiling during long-term modeling of a user's persistent interests and preferences. Traditional relevance feedback methods require that users explicitly give feedback by, for example, specifying keywords, selecting and marking documents, or answering questions about their interests. Such relevance feedback methods force users to engage in additional activities beyond their normal searching behavior. Since the cost to the user is high and the benefits are not always apparent, it can be difficult to collect the necessary data and the effectiveness of explicit techniques can be limited.

In this paper we consider the use of implicit feedback techniques for query expansion

and user profiling in information retrieval tasks. These techniques unobtrusively obtain information about users by watching their natural interactions with the system. Some of the user behaviors that have been most extensively investigated as sources of implicit feedback include reading time, saving, printing and selecting. The primary advantage to using implicit techniques is that such techniques remove the cost to the user of providing feedback. Implicit measures are generally thought to be less accurate than explicit measures [93], but as large quantities of implicit data can be gathered at no extra cost to the user, they are attractive alternatives. Moreover, implicit measures can be combined with explicit ratings to obtain a more accurate representation of user interests.

Implicit feedback techniques have been used to retrieve, filter and recommend a variety of items: hyperlinks, Web documents, academic and professional journal articles, email messages, Internet news articles, movies, books, television programs, jobs and stocks. There is a growing body of literature on implicit feedback techniques for information retrieval tasks, and the purpose of this article is to provide a brief overview of this work. Our intention is not to be exhaustive, but rather to be selective, in that we present key papers that cover a range of approaches. We begin by presenting and extending a classification of behaviors for implicit feedback that was previously presented by Oard and Kim [94], and classifying the selected papers accordingly. A preponderance of the existing work clusters into one area of this classification, and we further examine those papers. We then provide a brief overview of several key papers, and conclude with a discussion of future research directions suggested by our analysis.

5.4.1 Classification of Implicit Feedback Techniques

Implicit feedback techniques take advantage of user behavior to understand user interests and preferences. Oard and Kim [94] classified observable feedback behaviors according to two axes, Behavior Category and Minimum Scope. The Behavior Category (Examine, Retain, Reference and Annotate), refers to the underlying purpose of the observed behavior. Minimum Scope (Segment, Object and Class), refers to the smallest possible scope of the item being acted upon. This classification scheme is displayed, with example behaviors, in Table 5.1.

Based on our examination of the literature, we added a fifth Behavior Category, "Create", to Oard and Kim's [94] original four. The "Create" behavior category describes those behaviors the user engages in when creating original information. An example of a "Create" behavior is the writing of a paper. We also added some additional commonly investigated observable behaviors, and they have been highlighted. Like Oard and Kim [94], we make no claim that this table of behaviors is exhaustive. Rather, we suggest that Table 5.1 be viewed as a sample of the possible behaviors that users might exhibit. It should be noted that Table 5.1 includes both implicit and explicit observable behaviors. In our discussion of implicit measures, we do not consider explicit observable behavior, such as "rate".

Categorizing an observable behavior into the appropriate cell in Table 5.1 can be difficult, because both the intent of the behavior and its scope can be ambiguous. Thus, while the Behavior Category for saving a newly created document could appear to be "Retain", the behavior is probably more appropriately considered "Create". Similarly, while find and query behaviors involve the creation of text, they are primarily used to locate information for examination, and thus are classified in the "Examine" category. For example, a person might use find to locate a term or passage to examine in a document. Similarly, they might perform a query to locate a document for examination. While querying traditionally applies

	Segment	**Object**	**Class**
Examine	View Listen *Scroll* *Find* *Query*	Select	*Browse*
Retain	Print	Bookmark Save Delete Purchase *Email*	Subscribe
Reference	Copy-and-paste Quote	Forward Reply Link Cite	
Annotate	Mark up	Rate Publish	Organize
Create	*Type* *Edit*	*Author*	

Table 5.1: Classification of behaviors that can be used for implicit feedback from Oard and Kim [94]. Our additions are in italics.

to documents, the behavior is classified with a Minimum Scope of "Segment" because some systems return best passages rather than documents.

It is also difficult to assign the Minimum Scope of a behavior, as the scope can be ambiguous. For example, a behavior such as bookmark acts on a Web page, which is traditionally considered an "Object". However, when a Web page is considered in the context of its containing Web site, it can be understood as a "Segment" instead. Note, too, that observable behaviors are classified according to the minimum scope for which the behavior could be observed. For example, the minimum scope we might observe for the behavior type is a "Segment" although it is also common for typing to occur during the creation of an object. Similarly, view is identified in the "Examine Segment" category. However, most research has investigated viewing as it relates to objects, and thus that research belongs in the "Examine Object" category.

We classified the thirty papers we selected to include in this article according to the Behavior Category and Minimum Scope of the implicit measures addressed by the paper. The classification is shown in Table 5-3. Some of the papers, such as [17], [90] and [102], overlap a number of categories and are shown in overlapping gray boxes. Those papers discussed in greater depth in Section 5.4.2 are highlighted.

A preponderance of the literature falls into the "Examine Object" category. This is not surprising, as document selection and viewing time, both measures included in "Examine Object", are relatively easy to obtain and are available for every object with which a user interacts. Other areas of Table 5-3 contain little or no work, suggesting possible categories of observable behavior to explore. One likely reason for the dearth of literature across the Minimum Scope categories of "Segment" and "Class" is that for many systems, the unit with which the user interacts is the object. An exception to this is that many annotation

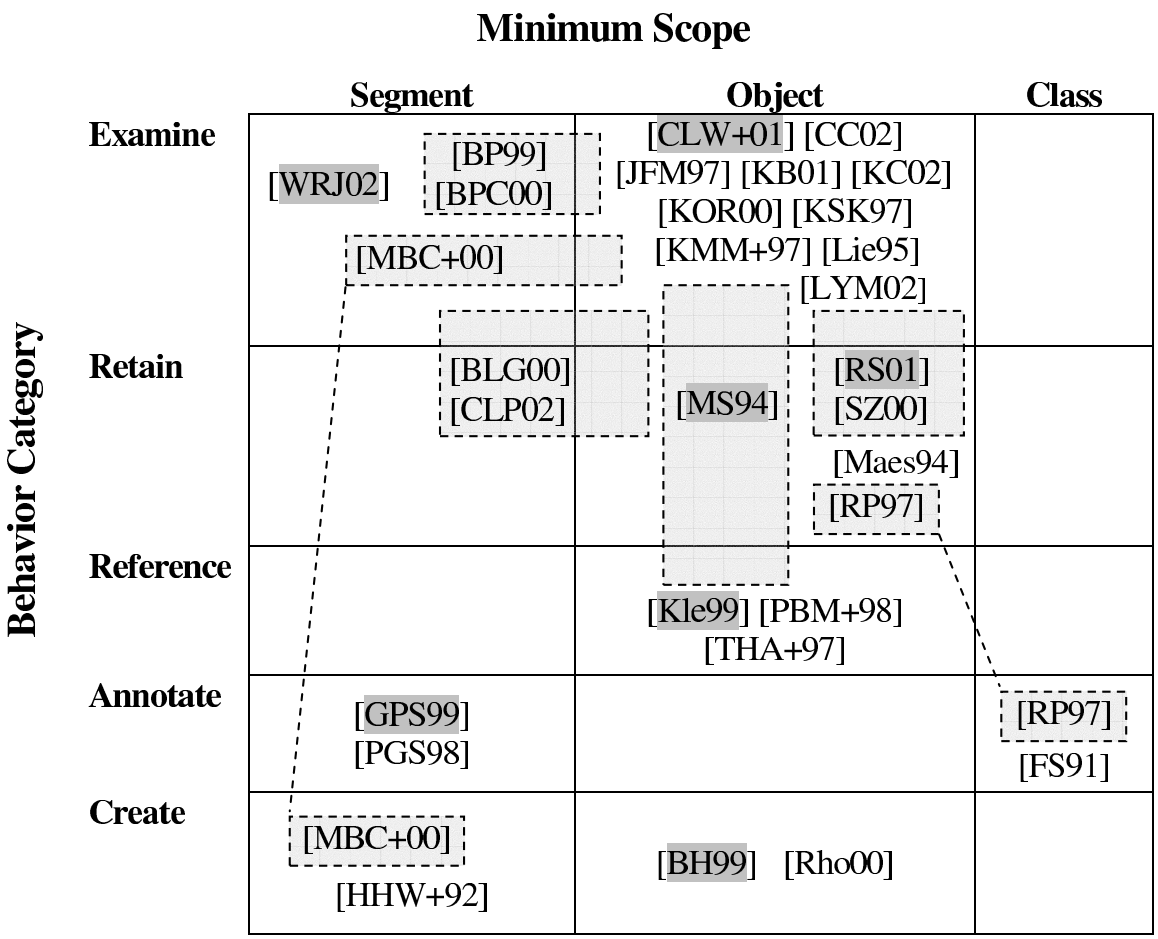

Figure 5-3: Papers classified based on the observed implicit behaviors they discuss. The papers discussed in greater dept in Section 5.4.2 have been highlighted.

	Design	Implementation	Evaluation
Individual	[65][66][80] [62] [77] [78]	[14] [17] [15]	
Group	[27] *[24]* [68] *[90]* *[102]* [108]	[59] [72]	[28]

Table 5.2: Papers from the "Examine Object" cell in Table 5-3, classified by study type. The papers discussed in greater depth in Section 5.4.2 have been *italicised.*

systems consider segments, and this could suggest why much of the annotation literature falls into this category.

We further examined the 18 papers that fell into the "Examine Object" category, classifying them into Table 3 along two additional axes. One axis represents the standard software lifecycle based on the spiral model of software development (c.f. [16]): design, implementation, evaluation. Papers in the "Design" category address the issue of what are good implicit measures to use. The "Implementation" category contains papers about implementing systems that use implicit feedback, and those in the "Evaluation" category focus on frameworks for evaluation. Of course, there is overlap among all three of these categories, particularly because the work with implicit measures is still in its infancy. For example, because there do not yet exist many test beds for system evaluation, most system implementation research has necessitated the development of an evaluation scheme. We classify the papers according to the stage they primarily address, but encourage the reader to explore papers from other categories as well.

The other axis in Table 5.2 focuses on whether the research deals with user preferences on an individual or group level. For example, in the understanding of implicit measures, the amount of time an individual spends reading a document can be compared to that individual's explicit relevance judgment to understand if reading time is a good implicit measure for relevance, or reading times can be averaged across many users, and compared to a global relevance judgment for that document. Similarly, systems that use implicit measures can use them to help retrieve, filter and recommend items for individual users, or they can provide feedback on an aggregate level by, for example, clustering the documents or highlighting popular articles. Note that many implicit feedback systems built to support individuals do so based on analysis performed over groups. For example, a system that infers an individual's relevance judgments based on his or her reading time may base the judgment on a threshold derived from averaging the reading time over a group of users. None the less, because such work focuses on supporting the individual, we classify it as "Individual".

While the papers from the "Examine Object" category of Table 5-3 spread evenly across several of the categories of Table 5.2, it is evident that little work has focused primarily on the "Evaluation" category. This is probably because the field is still young, and until now it has been difficult to determine what sort of evaluation test beds would be appropriate.

5.4.2 Examination of Key Papers

In this section, we provide a more in depth analysis of several papers that we believe are good representatives of the various different areas of Table 5-3. Our purpose in examining these papers in more detail is to present the reader with a better idea of how studies of implicit feedback are conducted, how this feedback is typically used and what the key issues

and problems in this area are. These papers are necessarily biased toward the work that has examined reading time, as a majority of research has focused on this behavior.

[24] Claypool, M., Le, P., Waseda, M., and Brown, D. (2001). Implicit interest indicators. In Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI '01), USA, 33-40.

Claypool, Le, Waseda, and Brown provide a categorization of different interest indicator categories, both explicit and implicit, and address the fundamental question of which observable behaviors can be used as implicit measures of interest. The authors create a customized browser and record the online behavior of seventy-five students, who were instructed to use the browser for 20 to 30 minutes of unstructured browsing. Several behaviors were examined: mouse clicks, scrolling, and time on page. Mouse clicks and scrolling were measured both as a frequency number (i.e. number of mouse clicks) and as total time spent. Scrolling was further measured both at the keyboard and with the mouse. Users were asked to explicitly rate each page that they viewed just before the page closed and these ratings were used to evaluate the implicit measures. Users looked at a total of 2,267 Web pages and made ratings on 1,823 (80%) of these. The authors found that time spent on a page, the amount of scrolling on a page (all scrolling measures combined) and the combination of time and scrolling had a strong positive correlation with the explicit ratings. However, the number of mouse clicks and the individual scrolling measures were found to be ineffective in predicting the explicit ratings.

[90] Morita, M., and Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '94), Ireland, 272-281.

Morita and Shinoda explored how behaviors exhibited by users while reading articles from newsgroups could be used as implicit feedback for profile acquisition and filtering. For six weeks, eight users were required to read all articles that were posted to the newsgroups of which they were members and to explicitly rate their interest in the articles. The authors measured the reading time, the saving or following-up of a story and copy for each of the 8,000 articles read by their users. They further examined the relationship of three variables on reading time: the length of the document, the readability of the document and the number of news items waiting to be read in the user's news queue. Very low correlations (not significant) were found between the length of the article and reading time, the readability of an article and reading time and the size of the user's news queue and reading time. Although no statistics are presented, the reading time for articles rated as interesting was longer than for articles rated as uninteresting. Saving, following-up and copying of an article were not found to be related to interests. Based on these results, the authors examined several reading time thresholds for identifying interesting documents. When applied to their data set, they found that the most effective threshold was 20 seconds, resulting in 30% of interesting articles being identified at 70% precision.

[102] Rafter, R., and Smyth, B. (2001). Passive profiling from server logs in an online recruitment environment. In Proceedings of the IJCAI Workshop on Intelligent Techniques for Web Personalization (ITWP 2001), USA, 35-41.

While the two preceding works found reading time to be related to interests, it is often difficult to effectively deal with reading time distributions because the curves are not normal. Instead, the curves have long tails, with a majority of points at the low end (toward zero). When collected in natural settings, there are often numerous outliers. These distributions often make statistical analysis challenging and may require some transformations. Rafter and Smyth perform a two-step process to prevent spurious reading times in data collected from the log records of users accessing job postings. In the first step, the median of median reading time values per individual job access for both users and jobs were used to calculate a normal reading time for the collection. Spurious reading times were then identified using this normal reading time, and outliers were replaced by this value. Graded reading times per job were then produced by calculating in each user's profile the number of standard deviations each job's newly adjusted reading time was above or below the user's mean reading time. In addition to the reading time, the authors also used raw visits to a job, incorporating a threshold on revisits, as implicit feedback and used the behaviors of applying for a job or emailing the job to oneself to evaluate the implicit measures. Users who had a profile of at least fifteen jobs were included in the analysis (412 total users). Using the adjusted revisit data and adjusted reading time data was found to result in better prediction performance than using their unadjusted counterparts, suggesting that collection, task and user specific transformations and normalizations on the raw behavioral data can produce more effective predictions of usefulness.

[122] White, R. W., Ruthven, I., and Jose, J. M. (2002). Finding relevant documents using top ranking sentences: An evaluation of two alternative schemes. In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02), Finland, 57-64.

In a controlled laboratory study, White, Ruthven and Jose, examined reading time as a technique for automatically re-ranking sentence-based summaries for retrieved documents. Users completed simulated tasks using three types of systems, one of which automatically re-ranked the top sentences in the summaries based on the user's reading time of each summary. They normalized the reading times for individuals by requiring users to perform a timing task before each search, where they were presented with a search description and the text of thirty summaries, and asked to read all documents and mark the relevant ones. To derive baseline reading times for each user, reading times for each summary were normalized by the length of the summary and divided by the number of characters to arrive at a character based, user-specific reading time for both relevant and non-relevant summaries. Performance results regarding the implicit system were inconclusive.

[44] Golovchinsky, G., Price, M. N., and Schilit, B. N. (1999). From reading to retrieval: Freeform ink annotations as queries. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99), USA, 19-25.

Research which uses the text that a user generates, be it an annotation or text from a word processing application, has shown promising results with regard to using this text as implicit evidence of user interests. Golovchinsky, Price and Schilit constructed full text queries based on users' annotated passages of documents and compared these to queries

constructed using standard relevance feedback techniques. The motivation for this work was that the words and passages that users mark can provide the system with a more refined, user-specific unit with which to perform relevance feedback and that these passages can help in establishing a context that is better than using just a list of terms. Results from an experiment with ten users annotating and evaluating documents for six topics found that queries derived from users' annotations produced retrieval performance that was better than standard relevance feedback techniques.

[19] Budzik, J., and Hammond, K. (1999). Watson: Anticipating and contextualizing information needs. In Proceedings of the 62nd Meeting of the American Society for Information Science, USA, 727-740.

Budzik and Hammond proposed a system that automatically retrieved documents and recommended URLs to the user based on what the user was typing. This work was motivated in part by the observation that users typically pose short queries that are highly ambiguous and often lack context. Like Golovchinsky, Price and Schilit [44], Budzik and Hammond suggest that evidence of context can be found in numerous other applications with which the user interacts. To initially and informally provide some support for their hypothesis, the authors asked ten researchers to submit an electronic version of a paper that they wrote and then asked these users to evaluate the documents that their experimental system had retrieved based on these texts. The results were encouraging, with at least eight of the ten users indicating that at least one of the retrieved results would have been useful. While Budzik and Hammond also provide results from several other informal evaluations, a full-scale, formal evaluation has yet to be performed.

[70] Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), 604-632.

Perhaps the most impressive large-scale use of implicit feedback comes in the form of Web link analysis. An example of this is Kleinberg's work with hubs and authorities. Authorities are authoritative information sources on a topic, and hubs are collections of authorities. Kleinberg suggested that good hubs could be recognized because they point to many good authorities, and similarly, good authorities could be recognized because they are pointed to by many hubs. Thus the links that people make in the course of Web page authoring are interpreted as endorsement. Link analysis, in the form of PageRank [97], is used to great success in practice by Google.

From the work reported in this section, it is clear that numerous problems arise when trying to infer information from observable behaviors, because what can be observed does not necessarily reflect the user's underlying intention. For instance, the amount of time that an object is displayed does not necessarily correspond to the amount of time that the object is examined, yet display time is traditionally treated as an equivalent to reading time. Further, the amount of time an object is actively examined does not necessarily correspond to the user's interest in that object. As is evidenced by the work described above, it appears that while implicit measures can be useful, they are not necessarily inherently so. Implicit feedback is often difficult to measure and interpret, and should be understood within the larger context of the user's goals and the system's functionalities.

5.4.3 Future Directions

We have looked at some of the relevant work on implicit feedback, and classified and highlighted a diverse set of papers that lay a foundation for the field. We believe that using implicit feedback is an exciting and promising approach to identifying user preference, and in this paper we have called attention to the areas where research in the field has focused, as well as illustrated several areas where there does not exist much work. We have had to be brief in our examination of key papers, and regret the exclusion of many interesting papers from this discussion. We did not consider some types of behaviors that could also be useful, such as those not covered by Table 5.1 and feedback from outside the digital domain (e.g., eye movements and gesture). For instance, Maglio, et al., [80] suggested using eye movements to infer user interests and there is a large body of research in the HCI community using eye movements to infer attention. We encourage the interested reader to explore the references provided in this paper further and assure that a longer review of implicit feedback is under construction.

To allow for the effective use of implicit feedback, more research needs to be conducted on understanding what observable behaviors mean and how they change with respect to contextual factors. Along with the papers discussed in Section 3, there is additional evidence that individual, task, topic and collection differences have some effect on the use of reading time as an effective measure of implicit feedback [65, 66]. While some work has limited the particular type of task under investigation, a more systematic investigation of the relationship between various contextual factors and potential behavioral indicators of interests needs to be undertaken.

Not all implicit measures are equally useful and some may only be useful in combination with others. For instance, the selection of an object is different, and perhaps weaker, evidence of interests than the printing or saving of an object, and a document with a low reading time might be printed or saved. It is likely, also, that how implicit measures are collected influence their effectiveness. More tools that allow for the accurate and reliable collection of data, such as the browser developed by Claypool, et al. [24], need to be developed, tested and shared, and further research should be done into how the collection process can encourage implicit feedback to closely match the user's underlying intent.

An in-depth investigation into the research that looked at object examination as a type of implicit feedback (Table 5.2) revealed that implicit feedback is used to recommend, retrieve and filter objects on both an individual and group level. Our examination further highlighted the lack of literature on developing test beds and evaluation metrics for implicit measures. We hypothesize that this is due to the novelty of the field; it is difficult to develop a good testing framework while all of the assumptions underlying implicit measures are still being explored. Perhaps now is a good time to look at developing test beds to encourage the further development of implicit measures systems.

5.5 Evaluating the Re-Search Engine

I will test this interface in two ways:

- By deploying the interface to people in the lab and collect and analyze usage data to understand if the new features are used in realistic settings.
- By running a laboratory study of the above mentioned interface to determine how the

changes effect the user experience when re-finding information in search results that have changed, as well as when finding new information.

5.5.1 Deploying the Interface

I will get people to use the interface, and record how often they repeat searches, as well as how often they take advantage of the new functionality I will provide.

5.5.2 Laboratory Study

The laboratory study will involve giving users a list of documents very similar to a search results page. The users will be asked to interact with this list by performing some retrieval tasks. Then, after a period of time, they will be asked to perform three tasks: 1) re-find the information they previously found, 2) find information they previously encountered that was not directly related to their primary task, and 3) find new information. These tasks will be performed on one of three possible search result lists:

- A list where nothing has changed at all since the user first saw it. In this case, the users should have an easy time re-finding information, but difficulty finding new information.
- A list that has changed. In this case, the user should have an easy time performing new tasks, but I expect to see that they have trouble re-finding information that they saw before.
- A list where the changed information is presented as discussed in Section 5.2. I hypothesize that this interface will give the users some of the benefit of the above two cases.

There are a number of details that I will need to consider before performing these experiments. These include:

- The exact period of time between the first interaction and the next interaction. In his experiments with re-finding, Capra [21, 55] waits a week, and that seems like a reasonable period of time.
- The population I will study. One population I am considering is business school students. This is because I have close connections to Yale's School of Management and MIT's Sloane School. Thus I could probably easily get involvement from both MIT and Yale students, giving me diversity of institution. In addition to being easily available, business school students seem like a fairly representative group of computer literate people business people in general, coming from a large variety of industries and companies. Most are only students briefly. The average student has been in the business world five years before attending, and as the program is only two years, a vast majority of the students return to the business world quickly.
- The size of the population I will study. I think thirty participants is a reasonable goal (ten for each type of interface tested), although the specific number will depend on the details of the study.
- Compensation for participants. I will likely compensate participants by holding a raffle at each school for a trendy gadget.

- The topic of the documents in the search results list. It is important that the participants be interested in the study topic to get good results [69]. This is likely to be a function of what segment of the population I use in testing. For example, if I do use business school students, it could be business or job related.
- The exact tasks. The exact tasks I will ask participants to perform will be a function of the information contained in the list, but will cover re-finding information from previous tasks, re-finding information incidentally encountered on previous tasks, and finding new information.
- Exactly how I will measure success for the various tasks. I will definitely look at time to task completion and success of task completion, and elicit feedback from the users for structured information regarding their satisfaction with the interface. It may or may not include more unstructured feedback from the user.

Change has a considerable psychological impact on the human mind. To the fearful it is threatening because it means that things may get worse. To the hopeful it is encouraging because things may get better. To the confident it is inspiring because the challenge exists to make things better.

King Whitney Jr.

Chapter 6

Conclusions and Future Work

In this thesis proposal, I have introduced the under-explored but increasingly important problem that people have returning to uncontrolled dynamic information. I have supported both the existence of this problem and its importance, and discussed the two solutions, time and user control, that I will explore as part of my dissertation work. In this chapter, I highlight the contributions this research will make, and discuss future directions for this work that are not within the scope of this thesis.

6.1 Contributions

As I have discussed, there are several important contributions of this thesis work. This work will have obvious implications for search engine user interfaces. I have proposed to implement and test giving users awareness and control over noticeable changes to search results, to helping people find information through search with the specific aim of aiding searches across time. If successful, I will make it possible for information that has been found via past searches to be found in future searches, despite changes to the underlying search engine index.

In investigating the problem of returning to dynamic search results, I have also show how I will highlight and support importance of the under-explored problem of returning to dynamic information. Because the problem has received relatively little focus, I propose to invest considerable effort into supporting and understanding it, and have conducted an observational study of the problems that can arise. I hope to draw generalizations from what I learn about re-finding via dynamic search results that apply to the broader problem in general.

In addition to directly dealing with the problem of returning to uncontrolled dynamic information, my thesis will also provide insight into the more general problem of how people find, and especially re-find, electronic information. While there currently exist some studies that try to understand people's electronic information seeking behavior, there is very little understanding as to how people return to information they have already seen [21]. Additionally, previous information seeking research has focused on specific corpora of information, such as email messages or Web pages, and has been performed primarily under tightly controlled conditions. The modified diary study I presented in this thesis gives a previously unavailable insight into people's search behavior across all of their electronic information, when not in a laboratory setting and not performing a specific information seeking task.

6.2 Future Work

There are many interesting areas for further exploration into the problem of returning to uncontrolled dynamic information that are not within the scope of this thesis. As part of my dissertation work, I will only explore one way that the information environment can change (that search results change). There are clearly many other ways that changes can affect users. It is by exploring multiple problems within this same space that a general solution can be found.

Explore time as a solution ([36, 38, 39, 40, 48, 84, 104].).

Solutions may vary by individual. As I discussed in Chapter 2, individual patterns of returning to information vary. The variations appear to be predictable, a function of whether the person is a piler or filer. I am interested in pursuing the different support these two different classes of people require when returning to dynamic information. Further, there are probably more subtle differences between individuals than filer or piler that could be learned from the user's behavior, allowing greater individualized support. Haystack, in focusing on an individuals information management, provides a good framework to explore this problem.

If we knew what it was we were doing, it would not be called research, would it?

Albert Einstein (1879-1955)

Bibliography

[1] David Abrams, Ron Baecker, and Mark Chingell. Information archiving with bookmarks: Personal web space construction and organization. In *ACM Computer Human Interaction Conference Proceedings*, pages 41–48, April 1998.

[2] M. S. Ackerman and C. Halverson. Considering an organization's memory. In *Proceedings of CSCW-98*, 1998.

[3] Christopher Ahlberg and Ben Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Human Factors in Computing Systems (CHI '94)*, pages 313–317, 1994.

[4] Christine Alvarado, Jaime Teevan, Mark Ackerman, and David Karger. Surviving the information explosion: How people find their electronic information. Technical Report AIM-2003-006, MIT AI Lab, April 2003.

[5] GVU Center at Georgia Tech. GVU's 10th WWW user survey, 1998.

[6] Tamara Babaian, Barbara J. Grosz, and Stuart M. Shieber. A writer's collaborative assistant. In *Interational Conference on Intelligent User Interfaces (IUI '02)*, 2002.

[7] Michelle Q. Wang Baldonado and Terry Winograd. Sensemaker: An information-exploration interface supporting the contextual evolution of a user's interests. In *Human Factors in Computing Systems (CHI '97)*, 1997.

[8] Rob Barrett, Paul P. Maglio, and Daniel C. Kellem. How to personalize the web. In *Human Factors in Computing Systems (CHI '97)*, 1997.

[9] M. Bates. Information search tactics. *Journal of the American Society for Information Science*, pages 205–214, 1979.

[10] Patrick Baudisch. *Dynamic Information Filtering.* PhD thesis, ??, 2001.

[11] Benjamin B. Bederson, Ben Shneiderman, and Martin Wattenberg. Ordered and quantum treemaps: making effective use of 2d space to display hierarchies. *ACM Transactions on Graphics (TOG)*, 21(4):833–854, 2002.

[12] V. Bellotti and I. Smith. Informing the design of an information management system with iterative fieldwork. In *Proceedings of the ACM SIGCHI Conference of Desiging Interactiv Systems: Processes, Practices, Methods and Techniques (DIS-00)*, pages 227–238, 2000.

[13] H. R. Bernard. *Research methods in anthropology: Qualitative and quantitative approaches.* Altamira Press, Landham, MD, 1994.

[14] D. Billsus and M. J. Pazzani. A personal news agent that talks, learns and explains. In *Proceedings of the 3rd International Conference on Autonomous Agents (AGENTS '99)*, pages 268–275, 1999.

[15] D. Billsus, M. J. Pazzani, and J. Chen. A learning agent for wireless news access. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI '00)*, pages 33–36, 2000.

[16] B. Boehm. A sprial model of software development and enhancement. *IEEE Computer*, 21(5):61–72, 1988.

[17] K. D. Bollacker, S. Lawrence, and C. L. Giles. Discovering relevant scientific literature on the web. *IEEE Intelligent Systems*, 15(2):42–47, 2000.

[18] Aaron B. Brown and David A. Patterson. Undo for operators: Building an undoable e-mail store. In *2003 USENIX Annual Technical Conference*, June 2003.

[19] J. Budzik and K. Hammond. Watson: Anticipating and contextualizing information needs. In *Proceedings of the 6th international conference on Intelligent User Interfaces (IUI '01)*, pages 727–740, 1999.

[20] Michael D. Byrne, Bonnie E. John, Neil S. Wehrle, and David C. Crow. The tangled web we wove: A taskonomy of www use. In *Human Factors in Computing Systems (CHI '99)*, pages 544–551. Addison Wesley, 1999.

[21] Robert G. Capra. Mobile information re-finding as a continuing dialogue. In *Human Factors in Computing Systems (CHI '03)*, pages 664–665, 2003.

[22] E. H. Chi, P. Pirolli, K. Chen, and J. Pitkoww. Using information scent to model user information needs and actions on the web. In *CHI '01*, pages 490–497, 2001.

[23] Rajiv Choudhary and Prasun Dewan. Multi-user undo/redo. Technical Report SERC-TR-125-P, Perdue University, Software Engineering Research Center, August 1992.

[24] M. Claypool, P. Le, M. Waseda, and D. Brown. Implicit interest indicators. In *Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI '01)*, pages 33–40, 2001.

[25] A. Cockburn, S. Greenberg, S. Jones, B. McKenzie, and M. Moyle. Improving WEB page revisitation: analysis, design and evaluation. *IT and Society Journal*, 3, December 2002.

[26] Terry Cook. Do you know where your data are? *Technology Review*, January 1995.

[27] M. D. Cooper and H. M. Chen. Predicting the relevance of a library catalog search. *Journal of the American Society for Information Science*, 52(10):813–827, 2002.

[28] D. Cosley, S. Lawrence, and D. M. Pennock. REFEREE: An open framework for practicle testing of recommender systesm using ResearchIndex. In *Proceedings of thte 28th Very Large Database (VLDB '02) Conference*, pages 35–46, 2002.

[29] Jay W. Davison and Stanley B. Zdonik. A visual interface for a database with version management. *AC Transaction on Office Information Systems*, 4(3):226–256, July 1995.

[30] Prasun Dewan. Designing and implementing multi-user applications: a case study. *Software - Practice and Experience*, 23(1):75–93, 1993.

[31] Alan J. Dix. Design of user interfaces for the Web. In *User Interfaces to Data Intensive Systems*, pages 2–11, 1999.

[32] Paul Dourish. What we talk about when we talk about context (draft). 2002.

[33] Paul Dourish, Victoria Bellotti, Wendy Mackay, and Chao-Ying Ma. Information and context: Lessons from a study of two shared information systems. In *ACM Conference on Organizational Computing Systems*, 1993.

[34] Susan Dumais, Edward Cutrell, J.J. Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. Stuff i've seen: A system for personal information retrieval and re-use. In *SIGIR '03*, 2003.

[35] Jr. E. James Whitehead. Design spaces for link and structure versioning. In *HT '01*, 2001.

[36] W. Keith Edwards and Elizabeth D. Mynatt. Timewarp: Techniques for autonomous collaboration. In *CHI '97*, pages 218–225, 1997.

[37] C. A. Ellis, S. J. Gibbs, and G. L. Rein. Groupware: some issues and experiences. *Communications of the ACM*, 34(1):38–58, 1991.

[38] Scott Fertig, Eric Freeman, and David Gelernter. Lifestreams: An alternative to the desktop. In *ACM SIGCHI Conference on Human Factors in Computing Systems Conference Companion (CHI '96)*, pages 410 – 411. ACM Press, 1996.

[39] Eric Freeman and Scott Fertig. Lifestreams: Organizing your electronic life. In *AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval*, Cambridge, MA, November 1995.

[40] Eric Freeman and David Gelernter. Lifestreams: A storage model for personal data. *SIGMOD Record (ACM Special Interest Group on Management of Data*, 25(1):80–??, 1996.

[41] B. Friedman, P. H. Kahn, and J. Hagman. Hardware companions? - what online AIBO discussion forums reveal about the human-robotic relationship. In *Proceedings of CHI '03*, 2003.

[42] Benjamine Jotham Fry. Organic information design. Master's thesis, Carnegie Mellon University, May 1997.

[43] David A. Fuller, Sergio T. Mujica, and Jose A. Pino. The design of an object-oriented collaborative spreadsheet with version control and history management. In *Symposium on Applied Computing (SAC '93)*, 1993.

[44] G. Golovchinsky, M. N. Price, and B. N. Schilit. From reading to retrieval: Freeform ink annotations as queries. In *Proceedings of the 24th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '99)*, pages 19–25, 1999.

[45] N. S. Good and A. Krekelberg. Usability and privacy: a study of KaZaA P2p file-sharing. In *Proceedings of CHI '03*, 2003.

[46] Anja Haake and Jorg M. Haake. Take CoVer: Exploiting version support in cooperative systems. In *Proceedings of INTERCHI '93*, 1993.

[47] W. J. Hansen. User engineering principles for interactive systems. In *Proceedings of the Fall Joint Computer Conference*, volume 39, pages 523–532, Montvale, New Jersey, November 1971. AFIPS Press.

[48] Koichi Hayashi, Takahiko Nomura, Tan Hazama, Makoto Takeoka, Sunao Hashimoto, and Stephan Gudmundson. Temporally-threaded workspace: A model for providing activity-based perspectives on document spaces. In *HyperText '98*, 1998.

[49] Marti Hearst, Ame Elliott, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ka-Ping Yee. Finding the flow in web site search. *Communications of the ACM*, 45(9):42–49, September 2002.

[50] Marti A. Hearst. Research in support of digital libraries at xerox PARC, part I: The changing social roles of documents. *D-Lib Magazine*, May 1996.

[51] Marti A. Hearst. Research in support of digital libraries at xerox PARC, part II: Paper and digital documents. *D-Lib Magazine*, June 1996.

[52] Marti A. Hearst. Next generation web search: Setting our sites. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2000.

[53] David L. Hicks, John J. Leggett, Peter J. Nurnberg, and John L. Schnase. A hypermedia version control framework. *ACM Transactions on Information Systems*, 16(2):127–160, April 1998.

[54] David Huynh, David R. Karger, Dennis Quan, and Vineet Sinha. Haystack: A platform for creating, organizing and visualizing semistructured information. In *Proceedings of IUI 2003*, 2003.

[55] Robert G. Capra III and Manuel A. Perèz-Qui nones. Conversational information re-finding using a voice interface, 2004.

[56] David Ingham, Steve Caughey, and Mark Little. Fixing the "broken-link" problem: The W3Objects approach. *Computer Networks and ISDN Systems*, 28(7–11):1255–1268, 1996.

[57] Stephen S. Intille. Change blind information display for ubiquitous computing environments. In *Proceedings of the Fourth International Conference Ubquitous Computing*, pages 91–106, 2000.

[58] JR. James E. Archer, Richard Conway, and Fred B. Schneider. User recovery and reversal in interactive systems. *ACM Transactions on Programming Languages and Systems*, 6(1):1–19, 1984.

[59] T. Joachims, D. Freitag, and T. Mitchell. Webwatcher: A tour guide for the world wide web. In *Proceedings of the International Joint Conference on Artificial Intelligence (ICML '97)*, pages 770–777, 1997.

[60] William Jones, Harry Bruce, and Susan Dumais. Keeping found things found. In *Proceedings of CIKM'01*, pages 119–126, November 2001.

[61] William Jones, Susan Dumais, and Harry Bruce. Once found, what then?: A study of "keeping" behaviors in the personal use of web information. In *Proceedings of ASIST'02*, 2002.

[62] T. Kamba, H. Sakagami, and Y. Koseki. Anatagonomy: A personalized newspaper on the world wide web. *International Journal of Human-Computer Studies*, 46:789–803, 1997.

[63] Mitch Kapor. Chandler (need specific citation), 2003.

[64] Michael A. Keller, Victoria A. Reich, and Andrew C. Herkovic. What is a library anymore, anyway? *First Monday*, 8(5), May 2003.

[65] D. Kelly and N. J. Belkin. Reading time, scrolling and interaction: Exploring implicit sources of user preferences for relevance feedback during interactive information retrieval. In *Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval (SIGIR '01)*, pages 408–409, 2001.

[66] D. Kelly and C. Cool. The effects of topic familiarity on information search behavior. In *Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL '02)*, pages 74–75, 2002.

[67] Diane Kelly and Jaime Teevan. Implicit feedback for inferring user preference: A bibliography. *SIGIR Forum*, 32(2), 2003.

[68] J. Kim, D. W. Oard, and K. Romanik. Using implicit feedback for user modeling in internet and intranet searching. Technical Report 00-01, University of Maryland CLIS, 2000.

[69] Kyung-Sun Kim and Bryce Allen. Cognitive and task influences on web searching behavior. *Journal of the American Society for Information System Technology*, 52(2):109–119, 2002.

[70] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[71] Wallace Koehler. Web page change and persistence – a four-year longitudinal study. *Journal of the American Society for Information System Technology*, 52(2):162–171, 2002.

[72] J. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(7):77–87, 1997.

[73] Mik Lamming and Mike Flynn. Forget-me-not: intimate computing in support of human memory. In *Proceedings of FRIEND21 Symposium on Next Generation Human Interfaces*, February 1994.

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

[75] David M. Levy. Fixed or fluid? document stability and new media. In *ECHT '94*, pages 24–31, 1994.

[76] Qing Li, Xiaofeng Bao, Chen Song, Jinefei Zhang, and Chris North. Dynamic query sliders vs. brushing histograms. In *CHI '03*, 2003.

[77] H. Lieberman. Letizia: An agent that assists web browsing. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 924–929, 1995.

[78] F. Liu, C. Yu, and W. Meng. Personalized web search by mapping user queries to categories. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM '02)*, pages 558–565, 2002.

[79] W. E. Mackay. More than just a communication system: Diversity in the use of electronic mail. In *CSCW 88*, pages 344–353, 1988.

[80] P. P. Maglio, R. Barrett, C. S. Campbell, and T. Selker. Suitor: An attentive information system. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI '00)*, pages 169–176, 2000.

[81] Paul P. Maglio and Rob Barrett. How to build modeling agents to support web searchers. In *Proceedings of the Sixth International Conference on User Modeling (UM97)*, pages 5–16, 1997.

[82] Paul P. Maglio and Rob Barrett. On the trail of information searchers. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, Mahwah, NJ, 1997. LEA.

[83] Thomas W. Malone. How do people organize their desks? *ACM Transactions on Office Information Systems*, 1(1):99–112, 1983.

[84] Richard Mander, Gritta Salomon, and Yin Yin Wong. A 'pile' metaphor for supporting casual organization of information. In *Proceedings of ACM CHI '92 Conference on Human Factors in Computing Systems*, pages 627–634, 1992.

[85] Gary Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.

[86] Herb Marselas. Where'd it go? It was just here! Managing 40,000 game assets, 2001.

[87] J. Terry Mayes, Stephen W. Draper, Alison M. McGregor, and Keith Oatley. Information flow in a user interface: the effect of experience and context on the recall of MacWrite screens. *People and Computers IV*, 1988.

[88] D. Scott McCrickard, John T. Stasko, and Q. Alex Zhao. Exploring animation as a presentation technique for dynamic information sources. Technical Report GIT-GVU-99-47, GVU Center, Georgia Institute of Technology, 1999.

[89] J. Mitchell and B. Shneiderman. Dynamic versus static menus: an exploratory comparison. *ACM SIGCHI Bulletin*, 20(4):33–37, 1989.

[90] M. Morita and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '94)*, pages 272–281, 1994.

[91] Bonnie Nardi and Deborah Barreau. "finding and reminding" revisited: Appropriate metaphors for file organization at the desktop. *ACM SIGCHI Bulletin*, 29(1):76–78, 1997.

[92] Bonnie Nardi, Steve Whittaker, and Erin Bradner. Interaction and outeraction: Instant messaging in action. In *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work (CSCW-00)*, pages 79–88, New York, 2000. ACM Press.

[93] D. M. Nichols. Implicit ratings and filtering. In *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, pages 31–36, 1997.

[94] D. W. Oard and J. Kim. Modeling information content using observable behavior. In *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, pages 38–45, 2001.

[95] Vicki L. O'Day and Robin Jeffries. Orienteering in an information landscape: How information seekers get from here to there. In *ACM Conference on Computer Human Interaction Proceedings*, pages 438–445, April 1993.

[96] Kasper Osterbye. Structural and cognitive problems in providing version control for hypertext. In *Preoceedings of European Conference on Hypertext (ECHT '92)*, 1992.

[97] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1998.

[98] L. Palen and M. Salzman. Voic-mail diary studies for naturalistic data capture under mobile conditions. In *CSCW '02*, pages 87–95, 2002.

[99] Seung-Taek Park and David M. Pennock. Analysis of lexical signatures for finding lost or related documents. In *SIGIR '02*, 2002.

[100] P. Pirolli and S. Card. Information foragin. *Psychological Review*, 106(4):643–675, 1999.

[101] J. Preece. Reaching out acress the web. *interactions*, March–April 1998.

[102] R. Rafter and B. Smyth. Passive profiling from server logs in an online recruitment environment. In *Proceedings of the IJCAI Workshop on Intelligent Techniques for Web Personalization (ITWP 2001)*, pages 35–41, 2001.

[103] Vicky Reich and David S. H. Rosenthal. LOCKSS: A permanent web publishing and access system. *D-Lib Magazine*, June 2001.

[104] Jun Rekimoto. Time-machine computing: A time-centric approach for the information environment. In *ACM Symposium on User Interface Software and Technology*, pages 45–54, 1999.

[105] M. Ringel, E. Cutrell, S. T. Dumais, and E. Horvitz. Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proceedings of Interact '03*, 2003.

[106] Andrew Sears and Ben Shneiderman. Split menus: effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction*, 1(1):27–51, 1993.

[107] A. J. Sellen, R. Murphy, and K. Shaw. How knowledge workers use the Web. In *Proceedings of CHI-02*, pages 227–234, 2002.

[108] Y. W. Seo and B. T. Zhang. A reinforcement learning agent for personalized information filtering. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI '00)*, pages 248–251, 2000.

[109] Ben Shneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11(6):70–77, 1994.

[110] Ben Shneiderman, Don Byrd, and W. Bruce Croft. Clarifying search: a user-interface framework for text searches. *D-Lib Magazine*, January 1997.

[111] Benjamine L. Somberg. A comparison of rule-based and positionally constant arrangements of computer menu items. In *SIGCHI/GI conference on Human factors in computing systems and graphics interface*, 1987.

[112] A. Spink and R. M. Losee. Feedback in information retrieval. *Annual Review of Information Science and Technology*, 31:33–78, 1996.

[113] A. Strauss and J. Corbin. *Basics of Qualitative Research Grounded Theory Procedures and Techniques*. Sage Publications, 1990.

[114] L. A. Suchman. *Plans and Situated Actions*. Cambridge University Press, 1987.

[115] Linda Tauscher and Saul Greenberg. How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47:97–137, 1997.

[116] Linda Tauscher and Saul Greenberg. Revisitation patterns in world wide web navigation. In *Human Factors in Computing Systems (CHI '97)*. ACM Press, 1997.

[117] Jaime Teevan. Displaying dynamic information. In *CHI '01*, 2001.

[118] Jaime Teevan. "where'd it go?": An observational study of how people re-find information when the Web changes. In *Submitted to CHI '04*, 2004.

[119] Jaime Teevan, Christine Alvarado, Mark Ackerman, and David Karger. The perfect search engine is not enough: An observational study of orienteering behavior in directed search. In *Submitted to CHI '04*, 2004.

[120] Walter F. Tichy. RCS – a system for version control. *Software – Practice and Experience*, 15(7):637–654, 1985.

[121] James R. Warren and Patrick Bolton. Intelligent split menus for data entry: A simulation study in general practice medicine. In *??*, 1999.

[122] Ryen W. White, Ian Ruthven, and Joemon M. Jose. Finding relevant documents using top ranking sentences: An evaluation of two alternative schemes. In *SIGIR 2002*, 2002.

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

[124] Steve Whittaker and Candace Sidner. Email overload: Exploring personal information management of email. In *Proceedings of the Conference on Human Factors in Computing Systems: Commun Ground*, pages 276–283. ACM Press, 1996.

[125] Ka-Ping Yee, Kristen Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. In *CHI '03*, 2003.