Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2004-041
AIM-2004-012

June 18, 2004

How People Re-find Information When the Web Changes

Jaime Teevan

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

ABSTRACT
This paper investigates how people return to information in a dynamic information environment. For example, a person might want to return to Web content via a link encountered earlier on a Web page, only to learn that the link has since been removed. Changes can benefit users by providing new information, but they hinder returning to previously viewed information. The observational study presented here analyzed instances, collected via a Web search, where people expressed difficulty re-finding information because of changes to the information or its environment. A number of interesting observations arose from this analysis, including that the path originally taken to get to the information target appeared important in its re-retrieval, whereas, surprisingly, the temporal aspects of when the information was seen before were not. While people expressed frustration when problems arose, an explanation of why the change had occurred was often sufficient to allay that frustration, even in the absence of a solution. The implications of these observations for systems that support re-finding in dynamic environments are discussed.

FUNDING
This research was supported by NTT, the Packard Foundation, MIT's Project Oxygen and the Arthur P. Sloan Foundation.

INTRODUCTION

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 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 [13]."

This paper presents 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 paper 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 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.

RELATED WORK

Observational studies investigating how people return to information [1, 2, 14] and how information is kept for later access [12], suggest that the information environment surrounding an information target is important when re-finding the target. For example, Alvarado, et al., [1] found that people did not look directly for their information target, but instead sought a broader information source that contained or pointed to it. Maglio and Barrett [14] observed that people returned to information using the path via which they initially found it. The study presented here expands on this work by investigating how people coped when changes occurred to their information target and its environment.

In addition to studies that have investigated re-finding behavior in natural settings, some systems, such as "Stuff I've Seen" [4, 20], have been developed to specifically support re-finding. Some of these systems, such as version control systems [16, 24] and digital libraries [10, 18], permit returning to information that might have changed. However, the focus of most systems that work within dynamic information environments is on preserving the original information, as opposed to helping people re-find it. Those systems that have intended to provide an interface for re-finding [5, 9, 19] deal primarily with the special case where any changes are made by the user. However, many of the problems with re-finding information in a dynamic environment arise because the changes occur outside of the user's control. As the study presented here gives insight into these problems, it has ramifications for the future development of re-finding systems. For example, many of the above systems emphasize time as a way to access old information, while time appears to be only one of a number of important aspects actually used.

Re-finding on the Web is particularly interesting because people commonly try to access Web information they have seen before, despite how often the Web changes. The percentage of Web page visits that are re-visits is estimated at between 58% [22] and 80% [3]. While many of these re-visitations occur shortly after the first visit (e.g., during the same session using the back button), a significant number occur after a considerable amount of time has elapsed. Thus it is not surprising that a survey of problems using the Web [8] found "Not being able to find a page I know is out there," and "Not being able to return to a page I once visited," accounted for 17% of the problems reported, and that the most common problem using bookmarks was, "Changed content." Whittaker and Hirschberg [25] found that people do not trust the Web as a repository for information, and instead keep paper copies of Web documents for archival purposes despite the costs incurred.

Web tools are only just beginning to support returning to information that has changed. Search engines and other tools for finding information on the Web tend to reflect the current state of the Web, although Google does cache the

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

pages in its index. There has been an effort to archive the Web (archive.org), as well as to keep links from breaking [11]. However, there is still much to be done as the problem becomes better understood. The implications of this study on the development of future solutions will be discussed later in this paper.

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 [6] to how they recover from injuries [17]. Observations have also been collected using search results. Good and Krekelberg [7] 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 [21]. 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.

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 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.

(a) **(b)** **(c)**

Figure 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.

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. However, 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.

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 [15]. 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.

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.

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.

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?"

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.

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. [1] 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.

Time Not Important

Despite the fact that time is often treated as a uniquely important feature in systems that support returning to information [5, 9, 19], 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.

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.

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.

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 WindowlAlign 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.

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 cancelled, 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.

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 [10].

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.

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 [12]. 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 [20]. 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 [4]. 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. Teevan [23] 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.

CONCLUSION

This paper presented an observational study of problems that arose during information re-finding in dynamic information environments. Analysis of Web pages that contained the phrase "Where'd it go?" provided insight into how people expressed the problems they encountered and the types of solutions they employed. The study has implications for the design of systems to support re-finding information in a dynamic environment, and they will be used to influence the development of a personalized information management system.

ACKNOWLEDGEMENTS

Mark S. Ackerman, Susan T. Dumais, David R. Karger and Paul Moody provided valuable feedback on this work. This research was supported by NTT, the Packard Foundation, MIT's Project Oxygen and the Arthur P. Sloan Foundation.

REFERENCES

1. Alvarado, C., Teevan, J., Ackerman, M. S. and Karger, D. R. (2003). Surviving the information explosion: How people find their electronic information. Technical Report AIM-2003-006, MIT AI Lab.
2. Capra, R. G. and Pérez-Quiñones, M. A. (2003). Re-finding found things: An exploratory study of how users re-find information. 2003. Technical Report, Virginia Tech.
3. Cockburn, A., Greenberg, S., Jones, S. McKenzie, B. and Moyle, M. (2002). Improving Web page revisitation: analysis, design and evaluation. *IT & Society*, 3(1): 159-183.

4. Dumais, S. T., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff I've seen: A system for personal information retrieval and re-use. In Proceedings of SIGIR '03.
5. Freeman, E. and Fertig, S. (1995). Lifestreams: Organizing your electronic life. In AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval.
6. Friedman, B., Kahn, P. H. and Hagman, J. (2003). Hardware companions? – What online AIBO discussion forums reveal about the human-robotic relationship. In Proceedings of CHI '03.
7. Good, N. S. and Krekelberg, A. (2003). Usability and privacy: a study of KaZaA P2P file-sharing. In Proceedings of CHI '03.
8. GVU Center at Georgia Tech (1998). CVU's 10th WWW user survey. http://www.gvu.gatech.edu/user_surveys
9. Hayashi, K., Nomura, T, Hazama, T., Takeoka, M., Hashimoto, S. and Gudmundson, S. (1998). Temporally-threaded workspace: A model for providing activity-based perspectives on document spaces. In Proceeding of HyperText '98.
10. Hearst, M. (1996). Research in support of digital libraries at Xerox PARC, part I: The changing social roles of documents. *D-Lib Magazine*, May 1996.
11. Ingham, D., Caughey, S. and Little, M. (1996). Fixing the "broken-link" problem: The W3Objects approach. *Computer Networks and ISDN Systems*, 28(7-11): 1255-1268.
12. Jones, W., Dumais, S., and Bruce, H. (2002). Once found, what then?: A study of "keeping" behaviors in the personal use of Web information. In Proceedings of ASIST '02.
13. Levy, D. (1994). Fixed or fluid? Document stability and new media. In Proceedings of European Conference on Hypertext.
14. Maglio, P. P. and Barret, R. (1997). On the trail of information searchers. In Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society.
15. Marselas, H. (2001). Managing 40,000 assets per game ("Hey, where'd it go? It was just here!"). Game Developers Conference.
16. Osterbye, K. (1992). Structural and cognitive problems in providing version control for hypertext. In Proceedings of European Conference on Hyptertext.
17. Preece, J. (1998). Reaching out across the Web. *Interactions*, March-April 1998.
18. Reich, V. and Rosenthal, D. S. H. (2002). LOCKSS: a permanent Web publishing and access system. *D-Lib Magazine*, June 2002.
19. Rekimoto, J. (1999). Time-machine computing: A time-centric approach for the information environment. In Proceedings of UIST '99.
20. Ringel, M. Cutrell, E., Dumais, S.T. and Horvitz, E. (2003). Milestones in time: The value of landmarks in retrieving information from personal stores. In Proceedings of Interact '03.
21. Strauss, A. and Corbin, J. (1990). Basics of Qualitative Research Grounded Theory Procedures and Techniques. Newbury Park, CA: Sage Publications.
22. Tauscher, L. and Greenberg, S. (1997). How people revisit Web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47:97-137.
23. Teevan, J. (2001). Displaying dynamic information. In Proceedings of CHI '01 (Extended Abstracts).
24. Tichy, W. F. (1985). RCS – a system for version control. *Software – Practice and Experience*, 15(7): 637-654.
25. Whittaker, S. and Hirshberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions of Computer-Human Interaction*, 8(2): 150-170.