"Where'd it Go?":
How People Ask After Lost Web Information

Jaime Teevan

Microsoft Research, One Microsoft Way, Redmond, WA, USA

teevan@microsoft.com

ABSTRACT This paper investigates the way people described the difficulties they encountered when returning to information on the Web. It presents analysis of Web pages, collected via a Web search, where the phrase, "Where'd it go?" was used. A number of interesting observations arose from this analysis, including that the path originally taken to get to the information target is commonly referred to during its re-retrieval, and that the temporal aspects of when the information was last seen were only important when considered relative to other events. While people expressed frustration when problems re-finding 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.

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 (1994), "[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."

This paper presents a naturalistic study of how people described the difficulties they encountered returning to information in a dynamic information environment, the Web. The study was conducted by analyzing instances, collected via a Web search, where the phrase, "Where'd it go?" was used to refer to a piece of information. The following quotation is an example from the data that illustrates a number of the observations discussed in greater detail in this paper:

> I remember when I first joined these forums! There were little "Did you know" facts about Star Wars at the front page, but they were replaced with movie quotes! 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 only important in a relative sense. 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 naturalistic 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.

---

[1] For readability, spelling and minor grammatical errors have been corrected in extracted quotations.

RELATED WORK

Studies investigating how people return to Web information (Capra and Pérez-Quiñones 2005; Jones et al. 2003; Maglio and Barret 1997; Teevan et al. 2004) and how Web information is kept for later access (Bruce et al. 2004; Jones et al. 2002), suggest that the information environment surrounding an information target is important when re-finding the target. For example, Teevan et al., (2004) found that people did not look directly for their target, but instead sought a broader information source that contained or pointed to it. Maglio and Barrett (1997) 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 say they returned when changes occurred to their information target and its environment.

Although Web browsers contain functionality designed to support re-finding, such as bookmarks and history, as do research systems like "Stuff I've Seen" (Dumais et al. 2003), these tools are not designed to handle situations where the information corpus may have changed. Systems that are designed with change in mind, such as version control systems (Østerbye 1992; Tichy 1985) and digital libraries (Hearst 1996; Reich and Rosenthal 2002) focus on preserving the original information, and not on helping people re-find it. Those systems designed for dynamic environments that do intend to provide an interface for re-finding (Freeman and Fertig 1995; Hayashi et al. 1998; Rekimoto 1999) deal primarily with the special case where changes are made by the person doing the re-finding. However, as the study presented in this paper suggests, re-finding problems in a dynamic environment often arise because the changes occur outside of the user's control.

Re-finding on the Web is particularly interesting because it is common and the environment is very dynamic. The percentage of Web page visits that are re-visits has been estimated as high as 80% (Cockburn et al. 2002). 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, during which time the page is likely to change. A study by Obendorf et al. (2007) found that 69% of all pages re-visited after a day had changed.

Although new content can benefit Web users, the negative ramifications of changes to Web content are numerous. A survey of problems using the Web (GVU 1998) 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 (2001) 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. Teevan et al. (2007) found that changes to search result pages slow re-finding.

Web tools are only just beginning to support returning to information that may have changed. Search engines and other tools for Web finding have traditionally reflected only the current state of the Web, although many now cache the pages in their indices. There is an effort underway to archive the Web (archive.org), as well as efforts to keep links from breaking (Ingham, Caughey and Little 1996). 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 instances of re-finding in a dynamic environment that are analyzed in this paper were found by collecting Web pages that contained the phrase "Where'd it go?" via a Web search. The phrase was selected because it is 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 Web search using the Google search engine. 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 MSN Live Search, and Yahoo. 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
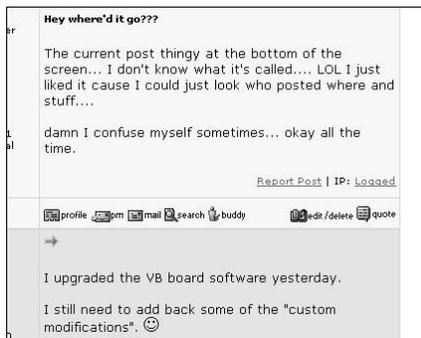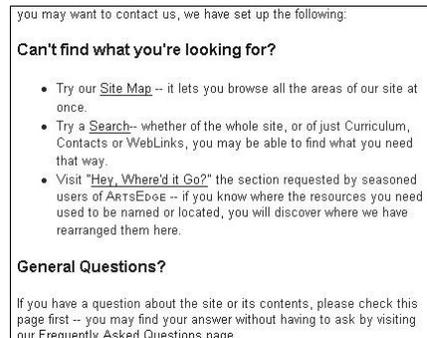
| (a) | (b) | (c) |

Figure 1: Three pages that contained the phrase "Where'd it go?". The first (a) is 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.

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 explored, 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 naturalistic studies. Several studies have analyzed postings collected from specific message boards to understand topics ranging from how people view robotic pets (Friedman et al. 2003) to how they recover from injuries (Preece 1998). Observations have also been collected using search results. Good and Krekelberg (2003) constructed KaZaA queries to see if people accidentally exposed personal data. 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 naturalistic observations of people having difficulties re-finding during personally motivated searches.

The data were analyzed by making an initial pass through the data 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 retrieved result 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 analyzed, several of which are shown in Figure 1, exclude duplicates and pages that did not involve the seeking of 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.

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, as is shown in Table 1. The data contained a number of 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.

Format	Total		Information Target						Rhetorical		Response	
			Web based information		Electronic information		Other information					
404/Redirect	17	7%	17	100%	0	0%	0	0%	17	100%	0	0%
Web log	21	9%	18	30%	3	52%	0	17%	2	74%	3	9%
Article	23	5%	7	58%	12	33%	4	8%	17	100%	2	0%
FAQ/Help	12	64%	7	64%	4	33%	1	3%	12	8%	0	68%
Message Board	165	8%	106	86%	54	14%	5	0%	14	10%	113	14%
Other	20	8%	19	95%	1	5%	0	0%	7	35%	1	5%
Total	258	100%	174	67%	74	29%	10	4%	69	27%	119	46%

Table 1: An overview of the type of data collected. A majority of the instances collected were message board or newsgroup postings. These instances are broken down by the type of information target "Where'd it go?" refers to, whether the question was used rhetorically, and whether a response to the question was provided by someone other than the searcher.

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 68% of them included responses to the query. The Web log pages also occasionally included responses in the form of comments. However, the phrase was not exclusively used by someone actively seeking information. In 27% 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 searches for information were analyzed. As seen in Table 1, the majority (67%) of those searches were for Web based information. The rest were for non-Web based electronic information (29%) and 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 (Marselas 2001). 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 (15%) where it had not, the seeker nonetheless had a strong expectation that the information used to exist in a particular place. This expectation most commonly 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 referring to? The post, "Technical and Plot Items 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 message board posting, it was often because a moderator had deleted it. 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 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 information was no longer available. However, the information target had not always moved, and in 23 instances (9%), it clearly had not. In these cases, the seeker appeared 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
The following section presents analysis of how people described their information target in the 165 instances collected where someone was actively searching for a piece of information (as opposed to, for example, where it was used rhetorically). People appeared to become frustrated when they could not find what they were looking for. They relied on shared context to express their difficulties, highlighting the path taken to the target and describing time in a relative manner. The analysis is summarized in Table 2.

Expressions of Frustration
People often 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 panicking 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

How People Described Missing Information	Example	Count	%
Expressions of frustration			
Frustration expressed	"Where'd it go.. gah.. I'm panicking now…"	41	25%
Seeker blamed self	"Maybe I'm doing something wrong?"	18	11%
Shared Context			
Other also missed target	"I noticed it too!"	22	13%
Importance of Path			
Path mentioned	"Okay, where's the link?"	52	31%
Asked for path to target	"Provide info on the demise of .. the .. newspaper."	17	10%
Time is Relative			
Time mentioned	"Recently…"	33	20%
Mention specific	"Half an hour ago…"	12	7%
Mention personal	"When I first joined these forums…"	5	3%

Table 2: How people described their information target in the 165 instances where someone was actively searching for a piece of information. People appeared to become frustrated when they could not find what they were looking for. They relied on shared context to express their difficulties, highlighting the path taken to the target and describing time in a relative manner.

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.

Shared Context

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?"

This communality of experience occurred despite the fact that 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 convenient! Where'd it go?

Nonetheless, the intended audience in both cases understood what was being referred to, and both queries 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 by its intended audience. Although several people expressed puzzlement, one person posted the following 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 likely 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 sections address the more commonly mentioned types.

The Importance of Path

The missing target was commonly described using path via which the target was originally found; 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 a piece of information, but rather to a step along the path used to get to the information. 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 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 [by] a competitor, and if so, and as importantly, where it's OBITs and other Personals may have be today?

Teevan et al. (2004) observed the importance of path during 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 is Relative

Time is often treated as a uniquely important feature in systems that support returning to information (Freeman and Fertig 1995; Hayashi et al. 1998; Rekimoto 1999). However, 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 fewer 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", "earlier 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 made any reference to time at all:

i) OH MY GOD, where is SmarterChild, he's been offline for a LONG time, and...WHERE DID HE GO?

ii) Smarter Child has been offline for some time. What's going on?

It is impossible to tell from the comments how long the agent had been missing.

Five times time was referred to in a personal manner, related, for example, to a personal event, as in the quotation in the introduction ("when I first joined these forums"). Regularity of access was 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, I've been seeing an advertisement … Now I can't find the Inside Sun advertisement … So, the question is, what happened to it?

Such mentions offer proof that the missing target 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 to understand the different ways the question could be answered. These instances provide insight into how information re-finding in dynamic environments is currently supported, and suggest ways tools might better deal with the problems that arise. For example, support for re-finding sometimes took advantage of the path by which information was originally encountered. The data set contained 12 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 where it used to be located.

In these cases, the question "Where'd it go?" was anticipated and used rhetorically by information providers trying to ease the re-finding of information they had changed. Rhetorical use also occurred 12 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 instances referenced a Macintosh manual's appendix titled "Where'd it go?" The appendix of the manual 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.

Because of the value of rhetorical instances, while the previous analysis focused solely on those cases where "Where'd it go?" was used by someone who actually did not know where "it" had gone, the following analysis includes all of the 258 cases where "Where'd it go?" referred to information, including rhetorical use.

Answers 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 answers (explanations, work around, and resolutions), summarized in Table 3, were not mutually exclusive, and sometimes all three occurred in a single instance.

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. 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 become

How "Where'd it go?" Was Answered	Example	Count	%
Explanations			
Explanation provided	"I've removed the pages I used to have here."	85	33%
Seeker hypothesizes	"Maybe Eric didn't pay his web page hosting fee?"	38	15%
Respondent hypothesizes	"If Spike doesn't like .. a post, he'll take it out."	22	9%
Work-Arounds			
Alternative suggested	"I found it, or something better…"	22	9%
Resolutions			
Target found	"I moved it to the bug reports forum."	82	32%

Table 3: The answers people received to "Where'd it go?" requests were analyzed to understand the ways the question could be answered. The analysis includes the 258 cases where "Where'd it go?" referred to information. Answers ranged from explanations of what had happened, to work-arounds so the seeker could deal with not having the information, to actual resolutions.

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 sometimes even made up. In 38 instances, the seeker asked "Where'd it go?" 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.

Respondents 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, as well as the following instance; here the person created a thread titled "Where'd it go?" despite having already found the target:

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

These cases where the target was already found highlight the importance of explanations during change.

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 respondent 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 content:

> For the time being, I (Pete) recommend 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 most 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 affecting 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, the answer was almost always given by 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?

An explanation was given by 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, which is a very little 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 what had happened.

MULTIPLE USERS OF THE SAME INFORMATION

That the most common reason for information to move was another person illustrates that people often had different intentions with the same information. For example, sometimes information was removed because people in general were not interested in it, despite it 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 conflict was also evident in the seven instances where information was removed for copyright reasons:

> The 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 (Hearst 1996).

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 discussed patterns that emerged from the data collected about re-finding behavior as it occurred in the dynamic information environment of the Web. The data suggested that people felt a lot of frustration when they could not find something they had seen before. One reason for this frustration appeared to be that the inability to re-find hurt the seeker's self-image, evidenced by the assumptions of blame and self-degrading comments made by frustrated seekers. The data also revealed that people relied on shared context to describe their information target, highlighting in particular the path originally taken to the target and describing the time the target was last seen in a relative, rather than absolute, manner.

This understanding of how people reacted to being unable to re-find and how they expressed the experience can aid in the design of future re-finding solutions. This section builds on the answers to "Where'd it go?" observed in the data, such as explanations, work-arounds, and resolutions, to suggest future solutions. People currently employ many tools to return to information, from search engines to bookmarks to email messages (Jones et al. 2002). While the information environments that these tools work in can be dynamic, most tools do not explicitly support such interactions.

A simple way to ensure that previously viewed information does not become unavailable is to cache earlier versions of changing information. This can be done on a global scale, such as is done by the Internet Archives, archive.org. However, doing so requires massive amounts of storage and infrastructure. Furthermore, systems that provide access to a group of users cannot differentiate between re-finding and finding because while whether information is being found or re-found is inherently user dependent, the systems do not know about individual users. Desktop systems and other systems that can track individual users are clearly at an advantage for supporting re-finding. 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 (Dumais et al. 2003).

Caching can ensure that the information a person is trying to re-find still exists, but it does nothing to guarantee that the person can still re-find it. The study presented in this paper suggests several ways to support access to previous versions of information. The use of shared context and personally relevant information to describe missing information that was commonly observed suggests an understanding of a person's previous information interactions can help re-finding. For example, allowing an individual to express the time the target was last seen would be helpful, particularly when expressed in a personally relevant way such as has been explored by Ringel et al. (2003).

While time was important in re-finding, other attributes of the original information encounter were commonly described during re-finding as well. The path taken appeared particularly important. Systems could take advantage of the importance of the by not just supporting the ability to search for old information, but by also preserving the original navigation. For example, a news Web site could archive the news digest page that originally presented a news story in addition to archiving the actual story. And when the content of a bookmarked page changes, the bookmark could still provide access to the originally viewed content in addition to pointing to the newly available information.

Caching previously viewed content individualizes what a person sees. However, there are challenges to this approach. As was observed, 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.

Users may also be interested in knowing that they previously viewed interesting content on a Web page even if that content is no longer available for, for example, copyright reasons. Actively exposing changes to the information environment could help keep the searcher from assuming the responsibility and self-blame for re-finding problems. It may also be useful to help the seeker understand why the sought after

information is no longer available. While specific explanations may not always be possible, it may be useful to suggest common reasons.

There are also benefits to be gained by the fact that different people view the same information. The number of people who said they had noticed changes observed by other people suggests that people may tend to be interested in returning to the same pieces of information and notice the same types of changes. This could mean that information that is returned to by a number of people should be made easy to find for others.

CONCLUSION

This paper presented a naturalistic 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 to answer the question. Looking forward, the effective management of changing information will be essential to successfully supporting complex re-finding behavior. The study presented here has implications for the design of systems to support re-finding information in dynamic environments. Research is currently underway to make use of these findings in the development of a personalized information management system.

REFERENCES

Bruce, H., Jones, W. and Dumais, S. (2004). Keeping and re-finding information on the Web: What do people do and what do they need? In Proceedings of ASIST '04.

Capra, R. G. and Pérez-Quiñones, M. A. (2005). Using Web Search Engines to Find and Refind Information. *IEEE Computer*, 38(10): 36-42.

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.

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.

Freeman, E. and Fertig, S. (1995). Lifestreams: Organizing your electronic life. In AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval.

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.

Good, N. S. and Krekelberg, A. (2003). Usability and privacy: a study of KaZaA P2P file-sharing. In Proceedings of CHI '03.

GVU Center at Georgia Tech (1998). CVU's 10th WWW user survey. http://www.gvu.gatech.edu/user_surveys

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.

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.

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.

Jones, W., Bruce, H. and Dumais, S. (2003). How do people get back to information on the Web? How can they do it better? In Proceedings of INTERACT '03.

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.

Levy, D. (1994). Fixed or fluid? Document stability and new media. In Proceedings of European Conference on Hypertext.

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.

Marselas, H. (2001). Managing 40,000 assets per game ("Hey, where'd it go? It was just here!"). Game Developers Conference.

Obendorf, H., Weinreich, H., Herder, E. and Mayer, M. (2007). Web page revisitation revisited: Implications of a long-term click-stream study of browser usage. In Proceedings of CHI '07.

Østerbye, K. (1992). Structural and cognitive problems in providing version control for hypertext. In Proceedings of European Conference on Hyptertext.

Preece, J. (1998). Reaching out across the Web. *Interactions*, March-April 1998.

Reich, V. and Rosenthal, D. S. H. (2002). LOCKSS: a permanent Web publishing and access system. *D-Lib Magazine*, June 2002.

Rekimoto, J. (1999). Time-machine computing: A time-centric approach for the information environment. In Proceedings of UIST '99.

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.

Teevan, J., Adar, E., Jones, R. and Potts, M. A. S. (2007). Information re-retrieval: Repeat queries in Yahoo's logs. In Proceedings of SIGIR '07.

Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In Proceedings of CHI '04.

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

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.