“Where’d It Go?”: Re-Finding Information in the Changing Web

Jaime Teevan TEEVAN@CSAIL.MIT.EDU
MIT Computer Science and Artificial Intelligence Laboratory, 200 Technology Square, Cambridge MA, 02139 USA

I investigate the under-explored but significant interaction of returning to dynamic information. For example, one might encounter an interesting link on a Web page, only to find on a subsequent visit that the link has changed. While often changes benefit the user, they can hinder returning to old information. I performed a study of dynamic information re-finding by analyzing instances on the Web where the phrase “Where’d it go?” was used. I focused on how missing information was described and the answers given to the question. The study suggests that systems designed to support re-finding dynamic information should preserve the original information context (e.g., the path taken to get to it, or people associated with it) and provide awareness of changes to the user.

1. Introduction & Related Work

Electronic information is often dynamic. For example, online news changes because new news stories are written as events transpire. Search results change because search engines update their indices to reflect updates on the Web. Even personal documents might change as collaborators edit shared documents. The rising availability of time dependent information, the growing ease of electronic communication and collaboration, and even the introduction of automated agents, suggest information will continue to become even more dynamic in the future. 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 (Levy, 1994).”

Traditional human-computer interaction research with dynamic information has assumed that users are interested in the most recent state of dynamic information. For example, when people want news they are often interested in seeing only current stories. News delivery systems, therefore, focus on providing new news. Similarly, with stock prices, a person generally wants the latest quote and providing them is at the heart of market information systems. However, there are other possible interactions a user might have with dynamic information. A particularly important, but under-explored interaction is that of returning to previously viewed information. Someone might chose to return to a news story they read last week, in which case they don’t care that new stories have since arrived. Or when filing a tax return, a person needs to return to the stock prices at the time of purchase. Similarly, on the Web people regularly try to re-find information. Researchers have begun to investigate how people return to information (Capra, 2003; Jones et al., 2001), but not when it is dynamic.

Returning to dynamic information is a particularly difficult problem because in returning, people rely on a considerable amount of context (Alvarado et al., 2003). However, when information is dynamic, the retrieval cues remembered by the user might have changed. For example, when news is fairly static, returning to previously viewed articles is easy. To find an article from last week’s paper, a person can dig through their recycling bin, find the appropriate paper, and return to the context they originally encountered the article. In contrast, returning to a story found via an online news service, where the news 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 that it was originally advertised in the upper left hand corner of the Web site, is unavailable for its retrieval.

In an earlier study (Teevan, 2001) I investigated how people interacted with information that changed slightly, and found many changes went unnoticed. Here I explore the difficulties that arise when information changes are significant enough to cause problems. I conducted a naturalistic study on the Web, investigating instances where people expressed a belief that they couldn’t find information because it had moved. By looking at how they described their information need, and the solutions they were presented with, I suggest ways to support dynamic information re-finding.

2. “Where’d it Go?”

To find instances where people looked for information that had moved, I analyzed Web pages that contained the phrase “Where’d it go?”. In this preliminary study, I found these Web pages by running a Google search for the phrase and analyzed the top 100 of the 5,290 results. In future work I will look at all of the available Google results, as well as results from other search engines and for slightly modified queries. However, even among the top 100 Google results there were numerous examples that provide meaningful insight into types of problems encountered when re-finding dynamic information, and suggest unexpected solutions.

The data were analyzed using standard qualitative techniques, as, without a representative sample, quantitative statistics are misleading. For example, golf balls were the most commonly looked for item, but this does not imply they are the most commonly lost item. Nor does the number of results give any indication of the magnitude of the problem. However, the study does permit an otherwise unavailable perspective into the problem as it arises in the real world. For example, people joked about losing golf balls, implying that was part of the fun of the game, but expressed real frustration when they couldn't find a previously visited site, suggesting such loses were a serious problem.

2.1 Describing the Missing Information

I analyzed how people who used the phrase "Where'd it go?" described the missing information. The phrase was often used in a message board posting asking for help. In the postings, people provided very little surrounding context. However, while often I did not understand what was being sought, their intended audience clearly did. One example 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

Based on shared context, this confusing post was understood, although even the explanation still confuses me:

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.

A popular piece of context used in describing moved information was the source that originally provided that information. In fact, people regularly looked only for the source, and not their actual information target. In the following example, the poster, looking for obituaries, asks for a pointer to the containing newspaper:

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?

Searching for the source has been observed as an important information seeking behavior (Alvarado et al., 2003), and suggests that when information changes, maintaining the context of the source is particularly important.

Surprisingly, time did not appear to be an important piece of context. People rarely referred to exactly when they saw the moved information originally. However, time was often referred to vaguely (e.g., "recently", "earlier", not in "quite a while") or, occasionally, by relating the original access to an event (e.g., "when I first joined these forums"). This suggests that solutions that include time should allow for fuzziness and include personal events as reference.

2.2 Answering "Where'd it go?"

I also looked at how people answered "Where'd it go?" requests in order to understand what sort of solutions people were provided with. The most common solution was to suggest a work around to deal with not having the desired information. For example, in response to a posting asking where the program "gcore" had gone, someone replied,

...so where's gcore come into it? It doesn't sound as though getting a userland core would be of any use.

and then proceeded to suggest a solution without gcore.

Occasionally the "it" of "Where'd it go?" was found, or its absence was explained. In a large number of these cases, the person who found or explained the missing information was the one who changed the information. For example, missing posts were often explained by moderators who deleted them:

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.

Interestingly, explanations often served as acceptable substitutes to actually finding the information. The importance of understanding what had happened to the information on the part of the seeker, and of being involved in the change on the part of the finder, suggest that systems to support re-finding dynamic information should provide awareness and control over changes to the user.

3. Future Work

There are many more interesting observation as to how people deal with difficulties in re-finding information, than space permits discussion of here, and I will continue to explore the available data. Further, I plan to build a system that supports returning to dynamic information. The specific type of dynamic information I will look at is search results. Search engines are not just used to find new information, but also to return to information found before. However search results can and do change, and this can cause problems similar to those discussed above. I will extend a search system to allow users the benefit of new query results without hindering their ability to return to old query results. This work will influence the development of the information management system Haystack.

References

Alvarado, C., Teevan, J., Ackerman, M., & Karger, D. (2003). *Surviving the information explosion: How people find their electronic information* (AIM-2003-006). MIT AI Lab.

Capra, R. G. (2003). Mobile information re-finding as a continuing dialogue. *CHI '03*.

Jones, W., Bruce, H., & Dumais, S. (2001). Keeping found things found. *CIKM'01*.

Levy, D. M. (1994). Fixed or fluid? document stability and new media. *ECHT '94*.

Teevan, J. (2001). Displaying dynamic information. *CHI '01*.