An Examination of How Dynamic Content Impacts Refinding and Keeping Behavior

Barbara Burton
Rutgers University
School of Communication and Information
New Brunswick, NJ 08901
908-672-2208

burtonb@rci.rutgers.edu

ABSTRACT
Web content has become increasingly dynamic with greater use of real time updating in many popular sites. Adar, Teevan and Dumais [1] found different user behavior and intent in web sites characterized as fast, medium, slow and hybrid in updating. The pace of updating also seems likely to impact the "keeping" decision described by Bruce, Jones and Dumais [2]. Proposed here is a research study examining refinding behaviors and keeping decisions with three web sites that have varying degrees of dynamism in content. Due to its relevance in refinding and keeping behaviors, this paper reviews the Morrison, Pirolli and Card taxonomy of Web use in and examines the purpose, method and content of Web use with emphasis on content attributes, including dynamism and quality.

General Terms
Measurement, Experimentation, Human Factors

Keywords
Refinding, Web Use, External Memory, Personal Information Collection.

1. INTRODUCTION
Plato suggested that writing would "create forgetfulness in the minds of those who learn to use it" and the debate about the impact of recorded knowledge on memory continues to today. Psychologists Sparrow, Liu and Wegner [9] have only recently concluded that "the Internet has become a primary form of external or transactive memory" although this has been a tenet of the study of Personal Information Management for many years. Bruce, Jones and Dumais [2] identified the Web as place where people go to find and refind information in 2004. They also described the keeping and leaving decisions focusing on how users make the decision whether or not to move content to a Personal Information Collection from the Web. Web content has grown increasingly dynamic over the past few years as evidenced by Google's 2010 change to its indexing system [5] and 2011 changes to its search engine algorithm to increase freshness [8] and this is clearly having an impact on user refinding and keeping behavior.

2. WEB USE TAXONOMY AND REFINDING
A broad overview of Web use may help us gain a fuller understanding of finding and refinding. Morrison, Pirolli and Card [6] present three taxonomic classification schemes based on Web activity which include the purpose of people's search, the method they use to search and the content of the information they seek. Refinding is an aspect of all three schemes.

2.1 Purpose Taxonomy
Morrison, Pirolli and Card [6] state that the purpose taxonomy includes finding, comparing/choosing and understanding. The ultimate goal in seeking information on the Web may be decision making, an increase in knowledge or simply entertainment. They do not include refinding, or remembering, as a discrete purpose of Web. The importance of refinding is emphasized by Dumais, Cutrell, Cadiz, Jancke, Sarin and Robbins [3] in their citing of studies showing that 58 to 81% of web pages accessed were re-visits to pages previously seen.

2.2 Method Taxonomy
The method taxonomy is a particular focus of refinding research and it includes exploring, monitoring, finding and collecting. Exploring may be related to browsing behavior, monitoring to alerts and finding to traditional search. Collecting is an interesting construct and refers to users making the decision to pull information from the public Web into their Personal Information Collections (PICs). Bruce, Jones and Dumais [2] identified and discussed the important distinction between keeping and leaving in refinding information, noting that the leaving decision occurs when an individual makes a conscious decision to leave useful information *in situ*. They identified many different methods that users may employ for keeping including bookmarks, emailing urls and doing nothing. The doing nothing decision depends on the user's ability to refind the web page either by remembering the url, finding it from a known access point, or searching for it.

There have also been several detailed studies evaluating users' refinding and remembering skills including a doctoral dissertation by Capra [4] titled "An Investigation of Finding and Refinding Information on the Web" which examined the ability of people to refind information and the techniques they used to do so. Capra found six major ways that users refound information which confirm the Bruce, Jones and Dumais findings about bookmarks, remembering URLs and searching. These methods seem to bring a large degree of success.

2.3 Content Taxonomy

The content taxonomy of Morrison, Pirolli and Card is comprised of topical areas that people often search. They list business, education, finance, job search, medical, news, people, product information and purchase, travel and miscellaneous. A May 2011 Pew Internet survey [7] finds that 92% of online adults use search engines to find information on the Web making it the most popular online activity, tied with email. A closer look at the Pew list of activities supports the continued importance of the areas named in the Morrison, Pirolli and Card content taxonomy although recent developments in the area of the social aspects of Web appear. Internet activities such as visiting a social networking site, using Twitter and updating and commenting on blogs and photos appear throughout the Pew survey results and represent a large amount of Web activity where refinding has not yet been studied.

The role of content in finding, refinding and keeping decisions has been primarily focused on topics such as news, shopping and sports, rather than attributes such as timeliness or quality. The quality of Web information has been extensively studied as a standalone topic and there is much good work available in that area. The actual size and growth rate of real time information on the Web is difficult to calculate but the recent Google changes to their indexing system and algorithm and the continued growth of Facebook and Twitter emphasize its importance. The Pew survey notes that 82% of Web users list "get news" as an online activity without developing the question about how the dynamic nature of news affects this popular activity. Undoubtedly most users are seeking the news of the moment but the continually changing nature of many news sites may make it difficult to refind something only a few hours old. The same applies to financial information. Pew finds that 37 percent of online users pursue "financial information such as stock quotes or mortgage interest rates" and also that 11 % list "buy or sell stocks, bonds or mutual funds" as an important Web activity. Current stock prices are generally quite easy to find, you need only put a ticker symbol into the Google search box to find those, but the current instability of the stock markets may mean historical prices are also important to understand properly a company's performance. Finding the information that was at one time so available is more difficult once it is archived. The greater dynamism of these sites must also be having an impact on the keeping decision as users ability to refind changes. If the Web is going to continue to serve a source of external memory, then it must excel at providing easy access not only to current information but to archived information as well. On that level, dead links and the inability to refind what was once found can be seen to represent lapses in collective memory and there is ample evidence that this continues to increase. The only way for the individual to maintain the information needed is to move it into a PIC.

3. Research Proposal

How does the increasing dynamism of the content on the Web affect refinding and keeping? It is proposed here that this area could be studied by focusing on how users refind and make the keeping decision with information in a three selected web sites with varying degrees of dynamism in content. A reference site such as Wikipedia is subject to change at any time but would seem to have a relatively low level of change. Users might have little reason to do more than enter a url to find it or choose it from a list of Web results to refind information once left there. A news site such as The New York Times has less stability in terms of content and may require different refinding and keeping strategies. Upon finding information that users would like to include in their PIC, it would seem that they would need to employ different strategies such as downloading or emailing the desired information rather than to simply return to the url and use the search function. Financial sites such as Yahoo Finance have perhaps the highest degree of change with financial instrument pricing updated continuously. While users may feel little need to refind a stock specific price (other than the one at market close) at any time, they may need to come up with specific refinding activities to handle such dynamic data once it is archived.

There is also another potential benefit to this proposed research with its findings having the potential to add to the ongoing debate and discussion about whether refinding personal information is equal to re-finding information on the Web. Teevan and Belkin [10] in their report revealed complete agreement that "re-finding in one's personal information space is the same as re-finding on the Web" although it was noted that there is not general agreement in the PIM community about this.

4. CONCLUSION

Examining how the increasingly dynamic content of many Web sites is impacting refinding and keeping behavior would seem to have practical implications for understanding and guiding users as well as shedding light on the differences between refinding in PICs and on the Web. Information in a PIC is generally static and only changed by the owner who would then have knowledge of any changes. Changes, updates and deletions on the Web are more collective and likely to have an impact on refinding and keeping issues in a different manner. If the trend towards more social and collaborative activity continues, the issue could grow in importance.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Adar, E., Teevan, J., Dumais, S.T. and Elsas, J. (2009). The Web Changes Everything: understanding the dynamics of web content. In Proceedings of the WSDM 2009 Workshop on Exploiting Semantic Annotations in Information Retrieval (WSDM 2009), 282-91.

[2] Bruce, H, Jones, W. and Dumais, S. (2004). Information behaviour that keeps found things found Information Research 10 (1). Retrieved from http://InformatinR.net/ir/10-1/paper207.html

[3] Dumais, S., Cutrell, E., Cadiz, J.J. Jancke, F., Sarin, R. and Robbins, D. (2003). Stuff I've Seen: A system for personal information retrieval and re-use. In Proceedings of SIGIR, 2003. Retrieved from http://research.microsoft.com/en-us/um/people/cutrell/SISCore-SIGIR2003.pdf

[4] Capra, R. (2006). *An Investigation of Finding and Refinding Information on the Web.* Doctoral dissertation. Department of Computer Science, Virginia Polytechnic Institute and State University. Blacksburg, VA, February 2006.

[5] Grimes, C. (2010). Our new search index: caffeine. Retrieved from http://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html

[6] Morrison, J., Pirolli, P., and Card, S. (2001). A Taxonomic Analysis of What World Wide WebActivities Significantly Impact People's Decisions and Actions. In Proceedings of the 2001 Conference on Human Factors in Computing Systems (CHI '01) Retrieved from Extended Abstracts. ACM Press, New York, 163-164.

[7] Pew Research Center (2011). Search and email still top the list of most popular online activities. Retrieved from http://www.pewinternet.org/Reports/2011/Search-and-email/Report.aspx

[8] Singhal , A. (2011). Giving you fresher, more recent search results. Retrieved from http://googleblog.blogspot.com/2011/11/giving-you-fresher-more-recent-search.html

[9] Sparrow, Liu and Wegner (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. Science Express, Published Online July 14 2011, p. 1-4.

[10] Teevan, J. and Belkin, N.J. (2005). Don't Drop the Ball: Re-finding Personal Information The PIM Workshop, An NSF Sponsored Invitational Workshop of Personal Information Management (NSF PIM Workshop), Seattle, WA, January 2005. Breakout Group Report.