

Factors and Evaluation of Refinding Behaviors

Robert G. Capra III
Virginia Tech
Center for Human-Computer Interaction
Department of Computer Science
rcapra@vt.edu

Manuel Pérez-Quiñones
Virginia Tech
Center for Human-Computer Interaction
Department of Computer Science
perez@cs.vt.edu

ABSTRACT

In this paper we describe difficulties involved in studying PIM behaviors and factors that can affect how people approach the refinding process. Factors that have been studied in the context of information seeking behaviors are summarized and their application to PIM and refinding are discussed. Finally, we note the importance of task in evaluating information refinding and reuse behaviors and discuss several task taxonomies that have been proposed that have applicability to the evaluation of PIM.

INTRODUCTION

Increasingly, organizing and managing electronic information is a problem for many computer users. Several recent research efforts are focusing on understanding not just how people search for and save information when seeking it the first time, but also on how they subsequently try to refind and reuse the information.

Many aspects of PIM involve users refinding information that has been seen before. Research from the information science and information retrieval communities has examined information seeking and searching behaviors, but refinding behaviors have been examined less.

In this position paper, three aspects of the study of PIM are discussed. First, we describe reasons that PIM and refinding are difficult to study. Second, we outline important factors that may affect how people approach PIM and information refinding. We discuss previous research – much from the information science community – on the effects of these factors on information seeking behaviors and relate these to the study of refinding and PIM. Finally, we discuss the importance of task in evaluation and summarize several taxonomies of tasks that may be useful in the study of PIM and information refinding.

REFINDING IS DIFFICULT TO STUDY

Refinding can be difficult to study. For example, the nature of refinding requires that a person: 1) has found some information, and 2) has a need or motivation to relocate this previously found information. Refinding is often motivated by unpredictable, external events. In addition, the process of finding information is often opportunistic and can evolve over time. These aspects make it difficult to observe refinding episodes as they occur and difficult to recreate in a laboratory setting. Interview and diary studies have been adapted and applied to study the unpredictable nature PIM and refinding by asking users to recall recent events (see [17] for an example of a modified diary technique). Studies of data logs (such as the Web logs studied in [16]) have been used to explore aggregate web behaviors, but logs often do not provide the rich insights and details from direct observation or interviews. Studying refinding in a laboratory setting has the benefit of being able to better control external factors and to isolate factors of interest. However, laboratory studies can lose important aspects of real-world activity and care must be taken to insure ecological validity.

Across the research community, a combination of study approaches are needed to gain a full understanding of refinding tasks, user needs, and behaviors.

FACTORS THAT AFFECT REFINING

Many factors play a role in how people find (i.e. search for) information. Studies from information retrieval, information science, cognitive science, human-computer interaction, and digital libraries communities have explored effects based on task type and task domain, individual differences, searcher experience, searcher familiarity with the topic of the search, and interface/system differences.

Information refinding is a fairly new area of study, with little prior research that has focused directly on the factors that affect it, the strategies that people use, and the differences between finding and refinding. In this section, several factors that have been studied in the context of information finding are reviewed and their relation to refinding is discussed.

Task Type, Domain, and Complexity

Task type and task domain are important factors in human-computer interaction and have been found to be a significant factor in studies of Web information *finding* [23] [18]. Task domain is considered here as a related, but different concept than an individual's knowledge of the domain being searched, which is often called *domain knowledge* or *topic familiarity*.

In information science, task complexity is often considered in terms of: repetitiveness, *a priori* determinability, the number of alternative paths to the information, number of goals, uncertainty, number of inputs, and other factors [5, p.193] [4, p. 582]. The predeterminability of task outcomes is often considered as a defining characteristic of the complexity of a task [5] [19] [22].

Many of the tasks addressed in PIM and especially in refinding are directed information seeking tasks – in other words, they are concerned with finding specific information, the structure and topic of which is probably known to the searcher. Often these tasks are fairly simple in concept, although the information itself may be difficult to locate. For example, a user may know that she needs to refind John Doe's phone number, but remember where it was stored or how it was originally found. A complicating aspect is that information is often *found* in the course of addressing more complex or open-ended tasks, making the *refinding* process more difficult.

In HCI and WWW research, information-seeking behaviors are often considered as *directed* or *undirected* – searching or browsing. Searching is concerned with looking for a known item, while browsing is a process of looking to see what is available [13]. Searching and browsing have complimentary benefits – searching is good for finding pages with specific terms quickly and browsing is useful in situations where a good search query cannot be constructed (such as relativistic comparisons) [15]. For complex tasks both searching and browsing may be beneficial, although it can be difficult to combine the two approaches [15]. Research on exploratory search techniques (see the April 2006 issue of the Communications of the ACM for a set of feature articles on exploratory search) is examining ways to support searching and browsing. These approaches have important application to PIM and are being used in interfaces such as Phlat [9].

Individual

It is well known that individual differences affect many aspects of human-computer interaction and how individuals deal with information. Models of information seeking include components for individual factors [5]. Factors such as cognitive style, self-efficacy, gender [10], and experience have been studied for their effect on information search and retrieval.

Topic familiarity refers to the searcher's general knowledge of the topic of the information search and is also referred to

as domain knowledge. Familiarity with a topic or task can be an important factor in information search and retrieval. Wildemuth [21] presents good review of prior research on the effects of domain knowledge. The concepts of domain knowledge and topic familiarity seem to apply equally as well to refinding as they do to finding. However, application to refinding does raise some additional research questions. How does the evolution of domain and task knowledge affect refinding strategies? How does a short amount of exposure or experience with information affect a user's knowledge of the area to help them refind information? In a study of refinding, we found that a single experience with a task to find a piece of information on the Web increased refinding performance versus finding performance when a user's prior familiarity with a task was not high and the time gap between finding and refinding was short (one week) [7]. More research is needed to understand how these factors affect refinding so that tools can better model and leverage user's existing knowledge and recall of cues.

Related to the concept of topic familiarity is that of *task familiarity*. Whereas topic familiarity addresses how familiar a searcher is with a particular topic area, task familiarity measures knowledge about how to do a specific task. In our study of refinding, we observed that search engine use differed significantly based on prior task familiarity [6].

In a strict sense, information refinding by an individual can be viewed dichotomously – the information being sought has either been seen before and the current task is refinding, or this is the first time the information is being looked for and the task is finding. In a more relativistic sense, refinding can be considered in terms of frequency – how often does the searcher look for this information? For very frequent tasks, a user may have highly developed methods of access and may be more likely to “teleport” directly to the information. For medium frequency tasks, the refinding process may involve a mixture of recall and recognition. For very infrequent tasks, refinding may truly be “finding again”.

In addition, expertise may be considered in terms of two additional dimensions: domain expertise and search expertise [11]. Domain experts are knowledgeable about the particular domain in which information is being sought. Search experts have specialized knowledge and experience with the tools or databases being used in the search. For example, librarians may have knowledge about how to search particular databases and Web experts may know special techniques for searching the Web.

Time

Elapsed time is known to be an important factor in forgetting [1]. In addition to forgetting, additional stimuli may interfere with existing memories [1], especially of similar items. The impact of how forgetting and memory interference affect PIM and refinding is an area that needs

additional exploration, especially in terms of how time affects the strategies used to refind information. A related factor is how time affects the value/importance of information. Information that is important now may be less so in the future. Similarly, information that does not stand out as being valuable now may be realized to be very important at a later time. Wen [20] investigated “post-valued recall” (PVR) Web pages – that is, pages “whose value is not recognized until some time after its initial retrieval.” [20, p. 185] and observed that users had difficulties relocating PVR pages.

EVALUATION AND TAXONOMIES

As mentioned earlier, task is an important factor in the study of PIM and refinding. Taxonomic classification of tasks and information seeking behaviors can provide valuable insights for the development of information design guidelines and tools for PIM/refinding. Several taxonomies of tasks performed on the Web have been proposed [3] [14] [2] and are summarized below.

The taxonomy that Bryne et al. [3] proposed is useful for examining the methods used to accomplish higher-level tasks and may be considered at the levels of methods and operators in GOMS [8][12]. For example, in a task to find and print the current weather for Blacksburg, VA, a user might go to a page (www.noaa.gov), locate on a page (find a text box to enter a city or zip code), provide information (zip code = 24061), and configure the browser (for printing the page). A taxonomy such as the one described by Bryne et al. could be useful in describing PIM behaviors and looking for common patterns in refinding (i.e. high-level refinding strategies).

Broder [2] identified three types of motivations for a Web search: navigational, informational, and transactional. Navigational searches have a goal of finding “a particular site that the user has in mind” [2, p. 5] and “is sometimes referred as ‘known item’ search in classical IR” [2, p.5]. Informational searches are traditional searches for information on a particular topic. These queries may be broad in nature or quite narrow. Transactional queries are performed to find a site where additional interaction will occur, such as shopping. Broder conducted two studies of searches on AltaVista and found that 20-25% of queries were Navigational, 40-50% were Informational, and 22-36% were Transactional. This classification of search engine use could be used to help streamline interactions for refinding and management of found information.

Morrison, Pirolli, and Card proposed three separate taxonomies based on the purpose, method, and content of a Web search [14, p. 164]. The first taxonomy is based on the purpose of the search: “What was the primary reason for the respondent’s search?” The second taxonomy is based on the method used: “How did the respondent find the information?” Third is a content-based taxonomy: “For what type of information is the responded [sic] searching?” Future studies of PIM and refinding behaviors could benefit

from using these taxonomies to examine if there are patterns of access based on the taxonomic classification of the task being performed.

CONCLUSION

PIM and refinding are difficult to study and require a variety of experimental approaches to explore. Factors such as task (type, domain, and complexity), individual factors (familiarity with the task and domain, expertise) and time (elapsed time between finding and refinding, and the changing value of information over time) all may affect PIM behaviors in ways that are not currently well understood. Research from information seeking and searching behaviors may help understand these factors, but specific studies of factors affecting PIM and refinding behaviors are needed. Because task can have a large effect on users approaches to refinding and reuse of information, the development and use of task taxonomies in studies of PIM is important.

REFERENCES

1. Anderson, J. R. (1980). *Cognitive Psychology and Its Implications*. San Francisco: W.H. Freeman.
2. Broder, A. (2002). A Taxonomy of Web Search. *ACM SIGIR Forum*, 36(2):3-10.
3. Byrne, M., John, B., Wehrle, N., Crow, D. (1999). The Tangled Web We Wove: A Taxonomy of WWW Use. In *Proceedings of the 1999 Conference on Human Factors in Computing Systems*. ACM Press, New York, 544-551.
4. Byström, K. (2002). Information and Information Sources in Tasks of Varying Complexity. *Journal of the American Society for Information Science and Technology*, 53(7):581-591.
5. Byström, K. and Järvelin, K. (1995). Task Complexity Affects Information Seeking and Use. *Information Processing and Management*, 31(2):191-213.
6. Capra, R., and Pérez-Quiñones, M.A. (2005). Using Web Search Engines to Find and Refind Information. *IEEE Computer* 38(10): 36-42.
7. 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, Virginia, USA.
8. Card, S.K., Moran, T.P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, NJ.
9. Cutrell, E., Robbins, D., Dumais, S., and Sarin, R. 2006. Fast, flexible filtering with phlat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006)*. CHI '06. ACM Press, New York, NY, 261-270.

10. Ford, N., Miller, D., and Moss, N. (2001). The Role of Individual Differences in Internet Searching: An Empirical Study. *Journal of the American Society for Information Science and Technology*, 52(12):1049-1066.
11. Jenkins, C. Corritore, C., Wiedenbeck, S. (2003). Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise. *IT&Society*, 1(3):64-89.
12. John, B., and Kieras, D. (1996). Using GOMS for User Interface Design and Evaluations: Which Technique? *ACM Transactions on Computer-Human Interaction*, 3(4):287-319.
13. Jul, S. and Furnas, G. W. (1997). Navigation in Electronic Worlds: A CHI 97 Workshop. *SIGCHI Bulletin* 29(4).
14. Morrison, J., Pirolli, P., and Card, S. (2001). A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions. In *Proceedings of the 2001 Conference on Human Factors in Computing Systems (CHI '01) Extended Abstracts*. ACM Press, New York, 163-164.
15. Olston, C. and Chi, E. (2003). ScentTrails: Integrating Browsing and Searching on the Web. *ACM Transactions on Computer-Human Interaction* 10(3):177-197.
16. 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(1):97-137.
17. Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. (2004). The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems*. ACM Press, New York, 415-422.
18. Toms, E., Freund, L., Kopak, R., and Bartlett, J. (2003). The Effect of Task Domain on Search. In *Proceedings of the 2003 Conference of the Center for Advanced Studies on Collaborative Research*. IBM Press, 303-312.
19. Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information Processing and Management* 35(6):819-837.
20. Wen, J. (2003). Post-Valued Recall Web Pages: User Disorientation Hits the Big Time. *IT & Society*, 1(3):184-194.
21. Wildemuth, B. 2004. The Effects of Domain Knowledge on Search Tactic Formulation. *Journal of the American Society for Information Science and Technology*, 55(3):246-258.
22. Wildemuth, B., Barry, C., Luo, L. Crystal, A., Oh, S. 2006. Establishing a Research Agenda for Studies of Online Search Behaviors: A Delphi Study. Accessed on January 27, 2006 from: http://ils.unc.edu/sig_use_delphi/
23. Woodruff, A., Rosenholtz, R., Morrison, J., Faulring, A., and Pirolli, P. (2002). A Comparison of the Use of Text Summaries, Plain Thumbnails, and Enhanced Thumbnails for Web Search Tasks. *Journal of the American Society for Information Science and Technology*, 53(2):172-185.