Refinding Is Not Finding Again

Robert Capra†, Mary Pinney§, Manuel A. Pérez-Quiñones†
†Department of Computer Science
§Department of Industrial and Systems Engineering
Virginia Tech
rcapra@vt.edu, mpinney@vt.edu, perez@vt.edu

Abstract

A challenging problem for Internet users today is how to *refind* information that they have seen before. We believe that finding and refinding are different user activities and require different types of support. The problem of how to *find* information on the web is studied extensively – new search algorithms, support for natural language queries, and innovative document indexing techniques are common topics in information retrieval research; visualizations of documents, and task support for finding are topics in human-computer interaction. But refinding has only recently begun to receive attention. In this article, we present evidence to support the claim that information refinding is a different activity than information finding. We present results that show how refinding is different from finding and suggest ways to improve web information seeking tools and designs to support refinding information.

Introduction

In today's 'information age', storage and organization of information for re-use is a significant problem [JBD01]. Trying to refind information that was found on the web can be an especially challenging problem for users [GVU98][JBD01]. Specific web pages, entire web sites, and even small pieces of semi-structured information [Abi97] can be difficult for users to re-locate at a later time. Often users do not know ahead of time what information they will need to refind (an aspect known as post-value recall [Wen03]). Even if they do know, studies have indicated that users have difficulties or do not using existing organizational tools such as bookmarks [GVU98] [JBD01] that might help in refinding.

The problem of refinding information is not merely a problem of fine-tuning an information retrieval (IR) algorithm [TAA+04], however. Web search engines receive approximately 550 million search requests per day [Gro03]. But, even among the many good search features that these engines provide, they have only limited resources designed specifically to help a refinding task. For example, search results are ordered by criteria that are independent of user's previous browsing history. The result is that search engines help with finding information but rarely provide extra help for refinding.

Only very recently have search engines attempted to provide tools to aid users specifically in refinding, indicating that the need for refinding tools is beginning to be realized. However, little has been done to understand users' refinding behaviors so that tools can be more effectively designed.

Sidebar: Refinding at Virginia Tech

At Virginia Tech, we have been exploring information refinding. We have developed several prototypes to explore alternative designs for tools to help in information refinding. We have also conducted several controlled studies on refinding, all with the goals of: a) understanding user behavior in information refinding, b) improving information-seeking tools to support refinding, and c) informing designers about factors that can affect refinding. This paper presents some of the results from our laboratory studies.

In study 1, called the '*telephone study*', we asked one person to help another person refind information on the web by talking over a telephone. By observing these conversations, we gained insight into what people remember and how they approach refinding tasks [CPQ03].

For study 2, called the '*general refinding study*', we wanted to take a more extensive view of refinding. We asked people to find and then refind (about a week later) the same or similar information using a web browser on a personal computer for a variety of information-seeking tasks.

For study 3, called the '*post-value recall study*', we conducted a study similar to the general refinding study (#2 above) but focused on the effect it would have on refinding if the users knew ahead of time that they were going to have to refind the information later; half of the participants knew ahead of time that they were going to be asked to refind the same information a week later [Pin04].

Finding versus Refinding

We believe that finding and refinding are two inherently different tasks (see Table 1). Finding information for the first time is an exploratory activity. Users apply knowledge of the web and of information organization, intuition, and foraging strategies together with tools such as search engines to arrive at the information desired. But there is uncertainty in the process of finding information the first time: Is the information desired out there? Where is it? What form is it in? Users often have only partial information about the desired information. They might know some words that can be used in a search engine query, or they might know something about the website where they expect the information to be located. But, do they have the right 'key' or 'clue' to use the existing tools to reach the information? In general, users know some *partial information* that leads them to make an attempt at locating the target information – they know something that leads them to believe that the information can be found on the web.

Table 1. Finding vs. Refinding Tasks

Finding	Refinding
Uncertainty: Is the information available at all? Do I know the right keyword to use in a search engine?	Certainty: I have seen the information already, but where? What was the keyword that I used in the search engine?
Recognition: Is this the information I am searching for?	Recognition & Recall: Where did I see that? Is this where I saw it? Context is very important.
Strategies: Intuition, search tools, foraging, browsing	Strategies: Directed (focused) navigation

In contrast, refinding is a more directed process. Users know that the information is out there because they have already seen it at least once. This makes the search for the information a focused task, one of *getting back* to the information. While finding relies on recognition ('Is this the information that I was looking for?'), refinding relies both on recognition and recall ('Where did I see that?'). There is some certainty about the task, the users *know* the information was there, and they know something about the previous time they saw the information. But even with this certainty, the user also needs to deal with partial information: if the user remembered all details, s/he would just go directly to the desired information. Thus, context is very important to the refinding process – users may recognize a key link or website based on seeing it in the same context they did the first time it was found. But even memory recall fails at times, as it is easily influenced by other factors.

Recent/Related Work

The problem of how to support users' information refinding needs has started receiving considerable attention from both industry and the academic research community.

Industry Forays into Refinding

Microsoft and Apple are launching tools to support refinding into their operating systems. Microsoft is investigating refinding in its 'Stuff I've Seen' research project [DCC+03] and their new WinFS filesystem is being designed with a goal of making it easier to find and use information [Riz04]. Apple is incorporating a personalized search tool called 'Spotlight' into its next OS release (Tiger) (http://www.apple.com/macosx/tiger/spotlight.html). Spotlight will include a feature that allows users to make temporal references like "today" or "last week". This is Apple's approach at including user history into the search process. Systems such as these are moving in a direction to help users refind information they have seen on their computer by indexing multiple sources of information (e.g. emails, web pages, documents) and providing access from a single, unified interface.

Several search engines are beginning to explore refinding tools as well. As of this writing, Ask.com is beta testing a personalized version of their search that allows for saving and organizing searches. This feature will automatically save the query terms searched for, and URLs that users would like to keep from the search. Amazon.com has

recently provided a search engine website, A9.com, that also provides tools for refinding. It can save a user's search history and page visits from that search. According to A9, it also includes a feature that allows a user to search through previously viewed pages.

Academic Research on Refinding

Academic researchers are also tackling the refinding problem (examples include [TAA+04] [CGJ+03] [JBD01] [TG97]). Findings from the 'Keeping Found Things Found' project (http://kftf.ischool.washington.edu/) at the University of Washington show that people go to great lengths to make existing tools work for their information refinding and re-use needs. For example, people email URLs to themselves, save web pages to a local disk, and print out Web pages [JBD01]. Users rarely use bookmarks and history lists [JBD01], possibly an indication that the tools do not provide the right level of support for the task of refinding information. More research is needed to gain a better understanding of what users' needs and strategies are for refinding information.

The Haystack group at MIT (http://haystack.lcs.mit.edu/) is developing a system that indexes and makes connections between large amounts of a users' data including email, documents, and web pages viewed. The system constructs RDF databases using meta-data from these sources and infers connections between information with a goal of making it easier to re-locate and re-use. In addition, researchers in the Haystack group have examined how people use existing tools to refind information [TAA+04]. One of their major findings is that while search engines are important, they were only used in 39% of the instances they studied where people were looking for information in their personal information spaces.

Our research group at Virginia Tech has been exploring information refinding for the last few years (see sidebar 'Refinding at Virginia Tech'). The remainder of this paper presents some results of our research as evidence that refinding and finding are different user activities. Our focus has been primarily on how people refind information seen on the web and to study refinding activities in a laboratory setting so that we can control and measure factors that may effect refinding.

Refinding is a different user activity from finding

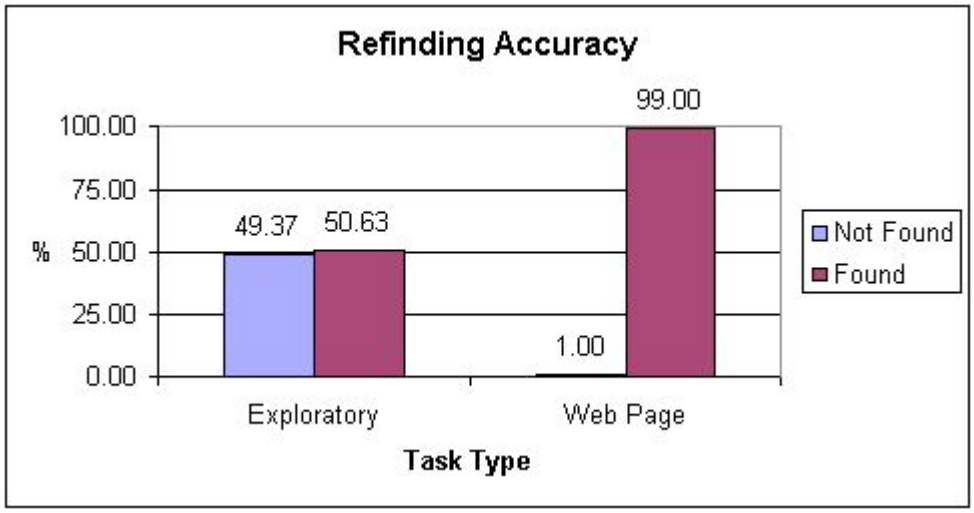
This section presents results from our research, organized according to the following dimensions: type of task, recall and recognition, post-value recall, and strategies for refinding.

Task Types

In our post-value recall study (see sidebar), we asked users to find and then, a week later, to refind information on the web. The study included tasks that were: 1) exploratory tasks that involved looking for information on a general topic (e.g. tourist attractions in San Francisco), and 2) tasks that involved looking for a specific web site/page such as a company's homepage. The results of the study showed that while users were almost always able to refind a specific web page, they had mixed results when trying to refind information from the exploratory tasks (Figure 1). For the web page tasks, users were able to refind the correct page 99% of the time, but for the exploratory tasks, users were

only able to refind the information about 51% of the time (these results are statistically significant at $p<0.05$).

Table 1. Refinding Accuracy for Post-Value Recall Study



These results indicate that the type of task influences how effective refinding might be. For refinding episodes where the user is trying to get back to a particular web page, often users remember URLs and other cues that make the refinding task as easy as the original finding task. However, when the finding and refinding task are more exploratory in nature, the refinding task can be significantly harder.

As an example, a task in our study asked users to find some sweatshirts they would like to buy for themselves. In trying to refind the sweatshirts the next week, if the users could not remember the name of the store or the name of the sweatshirt, they generally had trouble refinding the item. Search engines were not helpful in this case, likely because users were unsure of what search terms to use.

Not all refinding tasks are more difficult than the corresponding finding task. As noted above, the tasks that involved refinding specific websites had very high success rates. Our colleague, Dr. Naren Ramakrishnan, uses the term 'information addressability' to refer to many the different ways that we describe how to get to information. We have noticed an interesting trend to talk about information accessibility in terms of a search query that would refind the information. People talk about how to refind information by saying things like "search on Google for the term *X*." Search engines have become another way to explain the 'address' for a piece of information, but our results indicate that they are not enough for all cases.

Recall and Recognition

Despite the difficulties that users may have refinding information from complex finding tasks, a defining difference between finding and refinding is that when refinding, the user has found the information before. This difference plays a role in recall and recognition of information. Both finding and refinding depend highly on *recognition* of information.

With finding, users continually try to recognize if they are making the right choices to lead to their goal. Recognition in this case is based on related experiences of looking for similar types of information. With refinding, users attempt to recognize specific cues (pages, text, URLs, pictures, etc.) that they saw when they first found the information.

Refinding may also depend heavily on *recall*. Users may recall the sequence of steps they followed when they originally found the information, or they may try to recall key information (such as a URL or search phrase) that will help them "jump" closer to the information. Users are more likely to recall pictures than text from a web site [Wen03], but current refinding tools do little to help a user translate a recalled picture into a search. For example, Google has an image search facility, but it is intended as a tool to search for images, not to help the search of non-image information that is contained on a page that had an image.

Waypoints

Although users may not remember the exact sequence of steps that lead to the information they are trying to refind, the recall or recognition of "waypoints" along the path [MB97] often plays an important role in refinding. Waypoints are web pages or sites along the path the user took when they first found the information.

In our telephone refinding study [CPQ03], we observed extensive use of waypoints to aid refinding. In the study, one person called another person on the telephone to get help refinding information on the web. Waypoints were used in 20 of the 26 conversations we observed (76.9%). Three main categories of waypoints were observed: Page/Site Titles (e.g. "the Outback Steakhouse website"), URLs ("Fandago dot com"), and Page Descriptions ("it's kind of like a Yellow Pages kind of thing"), illustrating differing levels of specificity. It is our belief that tools for refinding, especially for complex, exploratory tasks, will need to help users recognize waypoints from previous searches.

Recall Can Affect Refinding Approach

Recalling key details can greatly assist refinding. In the post-value recall study, we observed that when people knew the web page that they needed to get back to, they were more certain about the path that they needed to take to refind the web page. For example, if they remembered that they typed the company's name into Google and clicked on the link for the first result, they were able to go into the refinding task with high certainty that this strategy would work a second time. In some instances, when users remembered the URL for a web site, they were able to shorten the path taken by typing the URL directly into the address bar for the web browser, jumping past the step of using a search engine.

However, refinding is not always easier using existing tools. In the exploratory tasks in our post-value recall study, people did not generally refind the information more quickly. Our results showed that there was no significant difference in the number of page visits between finding and refinding. Several things could lead to this – 1) the person could remember some details but not be able to use these recollections to help the refinding with existing tools, or 2) the person might be forced (either from 1 or because they did not remember many details) to begin browsing sites again in an attempt to recognize

something they had seen before that would lead them back to the information. This could lead to browsing many different sites and could possibly take more steps than finding the information the first time. This again leads us to believe that tools to support refinding should help users recognize waypoints from prior finding activities.

Importance / Post-Value Recall

Often we find ourselves needing to refind information that we were not expecting to need later. People use a variety of different techniques to organize information for reuse, especially if they expect that the information might have future value [JBD01]. We often think to ourselves, "If only I had known I would need that, I could find it now." At times, the information in question is information that was available only on the periphery of our attention focus. For example, we are searching for hotels for an upcoming trip, and along the way see an advertisement for some sporting event in the destination city of our trip. We later find out that the sporting event might be a significant one, such as the baseball game that will determine a post-season team. Now, the information that originally was on the periphery becomes the focus of a refinding activity. This notion of the future value or importance of information is the essence of post-value recall.

In our post-value recall study, we informed to half of our participants that they would have to refind the same information a week later. We discovered in our results that knowing ahead of time of the need to refind did not have a significant effect on refinding the information. Regardless of whether users knew that they would need the information at a later time during the original finding tasks, they still had difficulty locating the information a second time for exploratory tasks. Motivating the future value of information is difficult in controlled testing conditions, so it is possible that post-value recall has an effect that we were not able to measure in our laboratory testing. We are currently exploring how best to "motivate" future value of information in a controlled setting to further explore this issue.

Tools and Strategies

Users employ interesting and varied strategies to help organize and refind information. For example, users rarely use bookmarks and history lists [JBD01], possibly an indication that current tools do not provide the right level of support for the task of refinding information. What *strategies* are used for finding and refinding and how they differ is a large, open question that is being investigated by the research community. In this paper, we present some preliminary observations about search engine use and user's models of the web from our studies.

Search Engine Use

It is tempting to think that general web search engines are the answer to all our refinding problems. Why do we need tools for refinding when there are powerful general-purpose search engines? The short answer is that while search engines are powerful for some tasks, there are many other tasks (both finding and refinding), for which people choose strategies other than using a search engine.

In our general refinding study, web search engines were used in 43.35% of all the tasks completed by all the participants on the first day (n = 346). On the second day (about one week later), search engines were used in 44.60% of all tasks (n = 278). While there does not appear to be a significant difference in search engine use across day, these results do indicate that *a majority of tasks on both days were done without web search engines*.

In our post-value recall study, the percentage of total user actions that were searches using a web search engine was only 16.6% for the first day and 16.2% for the second day. Total user actions in this study counted actions such as clicking on a hyperlink, typing in a URL, entering a search string, and pressing the back button on the browser.

These results are consistent with those in a study done by the MIT Haystack group in which they examined people searching for information in their own personal information spaces. They found that keyword search was used in only 39% of the instances they studied [TAA+04]. We believe that this suggests that tools for refinding need to go beyond traditional keyword search and leverage users' recall and recognition (of things such as waypoints) to help support refinding information.

Although we have found plenty of evidence to show the finding and refinding are different tasks, the types of web tools used currently are similar for both. As we saw from our studies, the percentage of searching done on both days was similar. However, the result from our post-value recall study showed that for exploratory tasks users failed to refind the information about 50% of the time. This indicates that with existing tools, the strategies being employed by users do not always work.

Mental Models of the Web

We observed many instances of users having a model of how web sites are designed and features they expected to be available. Other research [Nie04] has also shown that users have models and expectations of the web. Examples of such features include having a search box on a web home page and a company logo in the upper right hand corner of a page [Nie04]. A striking example occurred when we asked students at Virginia Tech to search for the web site of a professor at a different university. Instead of using a web search engine, many students went to the web site for the other university looking for a "people finder" feature to be present as it is on the Virginia Tech web site. Based on their experience using the Virginia Tech people finder, these students appeared to have a mental model that all university web sites have "people finders", and followed that pattern in their approach to locate the information.

Conclusions

In this paper we have shown evidence we have gathered in several studies to support the argument that refinding is not just finding again. For some tasks (e.g. finding a website), refinding information may get easier, since once the user finds the desired website, it may be easier to remember the URL, or because of a URL-completion feature available in the web browser. For other types of tasks, however, refinding can be more difficult. Other researchers have shown that users have a hard time remembering the full path to

information; they often remember only waypoints and need to browse around to jump to the next waypoint. Furthermore, and in what might be the least supported task in IR, users need to get to information that when they saw it at first, they did not know that it was an important piece of information. Thus the recollection of that information is often about the information itself (e.g. I know I saw it when I was looking for something else, and it was a page that had a large image), but not about how they ran into the information during the finding episode.

Research into the refinding problem touches on topics from research areas including: Human-Computer Interaction, Information Retrieval, Cognitive Psychology, Human Factors, Digital Libraries, Semantic Web, and Data Management. It requires an understanding of what users remember from web pages, how remembered information is used to refind other information, what strategies are employed to organize and retrieve information, and how IR tools can be designed to support these activities.

IR has studied how the use of "histories" can be used to improve information retreival effectiveness. Some of these approaches need to be revisited from a refinding context point of view. We have shown that information refinding can be difficult for some types of tasks. This has implications for the IR algorithms and priority of results returned to the user. For example, when refinding, users might prefer to have a search results prioritized by pages that have been seen before.

Studying refinding can be a tricky problem that involves balancing attempts to isolate factors that may affect refinding and gaining insight into 'real-world' refinding episodes. Our focus has been to study refinding 'in the lab' so that we can control and examine factors that may affect refinding. However, it can be difficult to re-create 'real-life' settings and motivations in the lab. Other research groups have studied refinding using observational and self-reporting methods to try to gain insight into day-to-day refinding activities. All these different techniques for studying refinding are important and are essential to gaining an overall picture of how people go about refinding information, what is easy and what is difficult about refinding, and how tools can be designed to better support users refinding needs.

With the proliferation of mobile computing devices, cell phones, and PDAs, refinding takes on an even more complex dimension that is a topic for future research: cross-platform refinding. Currently, browsing histories are stored predominantly at the web browser and are not shared across devices. This can make it even more difficult to refind information on a different device than the one on which the information was initially found. Services such as the A9.com and Ask.com browsing histories mentioned earlier address this problem by storing history information on the servers of the service provider. Future tools to support refinding will need to consider such cross-platform issues, including information privacy and protection concerns as well as issues discussed in this article.

References

[Abi97] Abiteboul, S. (1997). Querying Semi-Structured Data. In F. N. Afrati and P. Kolaitis (Eds.), Proceedings of the 6th International Conference on Database Theory (ICDT-97), Delphi, Greece, January 8-10, 1997. (pp. 1-18). Springer.

[BS04] Boardman, R., and Sasse, M. A. (2004). "Stuff Goes into the Computer and Doesn't Come Out" – A Cross-tool Study of Personal Information Management. Proceedings of the 2004 Conference on Human Factors in Computing Systems. ACM Press, pp. 583-590.

[CPQ03] Capra, R. and Pérez-Quiñones, M.A. (2003). Re-Finding Found Things: An Exploratory Study of How Users Re-Find Information. Technical Report cs.HC/0402001, Computing Research Repository (CoRR).

[CGJ+03] Cockburn, A., Greenberg, S., Jones, S., Mckenzie, B. and Moyle, M. (2003) Improving Web Page Revisitation: Analysis, Design And Evaluation, IT & Society, V1, N3, WINTER 2003, PP. 159-183.

[DCC+03] 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. Proceedings of SIGIR 2003.

[Gro03] Grossman, L. (2003). "Search and Destroy: A gang of Web-search companies is gunning for Google", Time Magazine, December 22, 2003, pp. 46-50.

[GVU98] GVU Center at Georgia Tech. (1998). GVU's 10th WWW User Survey (October 1998). Retrieved March 19, 2002, from http://www.gvu.gatech.edu/user_surveys/

[JBD01] Jones, W., Bruce, H. and Dumais, S. (2001). Keeping Found Things Found on the Web. In Proceedings of the Tenth International Conference on Conference on Information and Knowledge Management. ACM Press, New York, NY, 119-126.

[MB97] Maglio, P. and Barrett, R. (1997). How to Build Modeling Agents to Support Web Searchers. In Proceedings of the Sixth International Conference on User Modeling (UM97, June 2-5, 1997). Springer Wien New York, 5-16.

[Riz04] Rizzo, T. (2004). WinFS 101: Introducing the New Windows File System. Microsoft Developer Network Article, March 17, 2004. Microsoft Corporation, http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwinfs/html/winfs03112004.asp, web page accessed October 1, 2004

[Nie04] Nielsen, J. Alertbox, September 13, 2004. Web page, http://www.useit.com/alertbox/20040913.html, accessed October 1, 2004.

[Pin04] Pinney, M. (2004). Finding and Refinding Information on the Web: Factors that Influence Strategies and Performance. Masters Thesis, Department of Industrial and Systems Engineering, Virginia Tech, Fall 2004.

[TG97] 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.

[TAA+04] 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. Proceedings of the 2004 Conference on Human Factors in Computing Systems. ACM Press, pp. 415-422.

[Wen03] Wen, J. (2003), Post-Valued Recall Web Pages: User Disorientation Hits The Big Time, IT & Society V1, N3, WINTER 2003, PP. 184-194.