

# Information Search and Re-access Strategies of Experienced Web Users

Anne Aula, Natalie Jhaveri, and Mika Käki  
Tampere Unit for Computer-Human Interaction  
Department of Computer Sciences  
FIN-33014 University of Tampere, Finland  
Tel. +358-3-2158871  
{anne.aula, natalie.jhaveri, mika.kaki}@cs.uta.fi

## ABSTRACT

Experienced web users have strategies for information search and re-access that are not directly supported by web browsers or search engines. We studied how prevalent these strategies are and whether even experienced users have problems with searching and re-accessing information. With this aim, we conducted a survey with 236 experienced web users. The results showed that this group has frequently used key strategies (*e.g.*, using several browser windows in parallel) that they find important, whereas some of the strategies that have been suggested in previous studies are clearly less important for them (*e.g.*, including URLs on a webpage). In some aspects, such as query formulation, this group resembles less experienced web users. For instance, we found that most of the respondents had misconceptions about how their search engine handles queries, as well as other problems with information search and re-access. In addition to presenting the prevalence of the strategies and rationales for their use, we present concrete design solutions and ideas for making the key strategies also available to less experienced users.

## Categories and Subject Descriptors

H3.3 Information Systems: Information Search and Retrieval – search process. H3.5 Information Systems: Online information services – web-based services. H5.2 Information Interfaces and Presentation: User interfaces.

## General Terms

Design, Human Factors.

## Keywords

Experienced web users, web search, information re-access, questionnaire study.

## 1. INTRODUCTION

The World Wide Web (web) contains enormous amounts of information and search engines are a widely used tool for accessing this information. In the U.S. alone, search engines are used by about 33 million adults on a typical day [14]. In addition to finding information for their current needs, people require methods for re-accessing information they have found earlier.

Our focus is on the information search and re-access strategies utilized by people with considerable web and web search experience. Along with experience, users develop efficient

strategies and make imaginative use of available tools for web information search and management. For example, we have seen them using as many as a dozen browser windows in parallel to manage the search process and to reduce the waiting caused by downloading times [5]. In addition, they e-mail URLs to themselves and add links to their personal web page so that they can access them later from a different computer [21].

Previous studies on experienced users' search and re-access strategies have mostly used observational methods with a small number of users [5],[15],[21],[22],[36]. Though observational studies can provide an understanding of the strategies in context and the rationales behind them, they may over-emphasize incidental findings. Other studies, which are based on log data [17],[19],[31],[34], make it easy to study a large number of people but the approach is perhaps weakened by an ignorance of the context of the use. Log studies are also often limited in scope as they typically gather data in relation to the use of a specific tool or service. In contrast, we applied a questionnaire with both open-ended and closed questions in order to gain data from a large number of users in relation to a pre-defined context of use. Using a questionnaire, we expected to gain a broad understanding of the search and re-access strategies regardless of the tools that people are using.

By reaching tens or even hundreds of people, we can firmly determine the relative importance of the strategies in question along with the rationales for their use. Using this questionnaire, we addressed the following three questions:

1. What are the tools that experienced users use for information search and re-access?
2. How prevalent are the different strategies for searching and re-accessing information?
3. Are there problems in the process of information search and re-access that even experienced users face?

In addition to examining the strategies, we will present concrete interface solutions and design ideas that aim to place the key strategies at the disposal of all web users.

## 2. RELATED WORK

Log studies are common in studying search engine usage. These studies reveal that typical web users formulate short queries, seldom use advanced operators or use them improperly, typically only check the first result page (10 results) per query, and rarely reformulate their queries [17],[18],[19],[31]. Thus, the general public uses search engines in a very simple way, a way which may not be very efficient.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2005, May 10-14, 2005, Chiba, Japan.  
ACM 1-59593-046-9/05/0005.

The information search and re-access strategies of experienced users are expected to be different from those of the general public. The theory of *information foraging* predicts that people modify their strategies in order to maximize the rate of valuable information they gain in a unit of time. As people become more experienced, their strategies will evolve towards the most profitable ones [29]. Good strategies can also be seen as one facet of expertise [33].

In an empirical study [15], Internet professionals searched for information with pre-assigned search tasks. Their queries contained twice as many search terms as those of typical web users. They also used advanced search options commonly (e.g., AND was used in 35.6%, '+' in 29.0%, and phrase search in 24.7% of the queries). Another study [4] supports these findings by showing that the length of the query is correlated with web experience. In the same search tasks, the more experienced web users formulate longer queries than the inexperienced ones.

In addition to searching for new information, web users frequently revisit information found earlier [30]. The average proportion of revisits to web pages was initially found to be 58% [34] and then 81% a few years later [10]. Common tools for revisitation are Bookmarks (also referred to as Favorites or Hotlist), the Back button (only for session-specific revisitation), and the History tool in the browsers. The Back button was found to constitute between 30% [34] and 41% [9] of all navigational acts, while History accounted for less than 1% [34]. Infrequent use of History was also found in [9], where documents were accessed through History in less than 3% of all cases. Although Bookmarks usage is common (94% of respondents in [1] had bookmarks), experienced users are prone to invent their own strategies for saving links for future use. The need to invent new strategies may be due to the difficulties related to bookmark usage (such as invalid bookmarks and cluttering the bookmark collection with possibly irrelevant URLs) [1],[10],[36].

In an observational study [5], researchers in computer science were found to use advanced operators only infrequently, the most common ones being the minus sign and phrase search (in 10% and 4% of the queries, respectively). However, the study showed that the researchers had innovative strategies for information search and re-access: they used several browser windows in parallel, saved links to separate files or folders, copied and pasted search terms from documents, and often iterated their queries. In spite of their expertise, they had misconceptions about the default operator of their primary search engine. They also had a poor understanding of how the results are ranked.

Another study of "high-end information users" [21] found a diverse set of strategies for managing information: sending e-mail to self and others, printing out web pages, saving web pages as files, pasting URLs into a document, adding links to a personal web page, using search engines or directly the URLs for re-accessing information, and adding bookmarks. It has also been noted that participants sometimes keep tested and untested references separated in their bookmark collections [22].

The data for the current study had already been collected when Bruce et al. [7] published closely related results of a survey study about information keeping and re-finding methods. Their findings are similar to the findings of the current study, for example, their most commonly mentioned re-finding methods were creating bookmarks, searching the material again and directly accessing pages via the URL. However, there are some differences in the results. For example, our data shows that saving documents as

files was as common as using search engines to find material again, whereas the results by Bruce et al. [7] ranked saving to be considerably less important. One explanation for this may be as simple as the wording used in the question: we referred to the material as *documents*, whereas Bruce et al. used the term *web page*. Thus, their respondents possibly only reported the frequency of saving HTML files.

### 3. METHODS

To gain a broad and yet detailed understanding of the experienced users' activities, preferences, and understanding of the tools, we used a questionnaire as our research method. It needs to be acknowledged that questionnaires rely on people's own evaluation and memory on the issues being asked. However, due to the context of our research questions and through careful design of the questionnaire, we have lessened the possible effects of these concerns. In addition, questionnaires have been previously applied successfully in similar circumstances [1],[7].

#### 3.1 Developing the Questionnaire

The questionnaire was developed based on previous findings (e.g., [5],[15],[21]), existing guidelines [24],[25],[32], and our own questionnaire about casual web users' methods of web page revisitation. Initially, the questionnaire was pre-tested with 5 people and, after modifying it accordingly, we ran a pilot study by administering the questionnaire to the personnel of the Department of Computer Sciences at the University of Tampere. 30 people responded to the pilot questionnaire. Based on the pilot test, the questionnaire was slightly modified, for example, by adding a couple of questions based on the answers to the "Other strategies" question, and re-wording questions that had been misinterpreted by the respondents.

#### 3.2 Final Questionnaire

The final questionnaire had 7 background-related questions and 9 questions related to computer, web, and search engine use. In the main part of the questionnaire, we asked the respondents to think of a typical work-related information search task (e.g., finding information related to their area of expertise) and to imagine doing it for a couple of hours with their primary search engine and web browser. In relation to this task, we listed 14 different strategies (see Figure 3) for information search and re-access and asked the respondents how often (almost always, often, sometimes, rarely, or never) they would use each in the above-mentioned search task. In addition, there were 10 questions related to Bookmark usage and frequency of advanced operator usage in queries. The questionnaire also contained 3 open-ended questions to elicit the participants' understanding of the functionality of their primary search engine, to allow them to list unmentioned strategies, and for free form comments. The questionnaire can be found from [www.cs.uta.fi/~aula/questionnaire.php](http://www.cs.uta.fi/~aula/questionnaire.php).

The URL of the final questionnaire, along with a cover letter (also available at the above mentioned URL) was sent to CHI-WEB and SIGCHI-Finland mailing lists in August 2004. In addition, the URL was sent to seven personal contacts from a large IT company who were asked to send the URL also to their colleagues, if possible. The questionnaire was available for two weeks.

#### 3.3 Respondents

The above-mentioned mailing lists were chosen because we wanted to have responses from experienced computer and web users. Thus, unlike most previous studies, our sample went beyond the "knowledge worker" confines, those being individuals

who manipulate information as their main profession. Instead, we also requested the responses from individuals who use web-based information to support their primary work tasks, such as programmers and designers.

Originally, we received 239 responses, 3 of which had to be rejected due to unanswered questions. Thus, complete responses from 236 people (50.6% males, 49.4% females) were analyzed. 67.4% of the respondents were from CHI-WEB and 25.0% from SIGCHI-Finland mailing list. 7.6% received the questionnaire by e-mail through personal contacts. The respondents were divided into groups based on their profession, the largest groups being designers (21.6%), researchers & lecturers (19.1%), librarians (16.1%), usability specialists (12.7%), and managers (11.0%). On average, the respondents had worked in this profession or with similar tasks for 8.2 years (SD = 6.6).

All of the respondents used computers and the web daily or almost daily. They had used computers for 16.7 years (SD = 6.0) and the web for 9.2 years (SD = 2.6), on average. The web and web search engines were frequently used for work-related information search: 94.9% used the web and 90.3% used web search engines several times a week or more for this purpose. The average rating the respondents gave for their own web search skills was 8.3 (SD = 1.3) on a scale from 1 (novice) to 10 (expert).

## 4. RESULTS

### 4.1 Browsers and Search Engines Used

Altogether, 12 different web browsers were mentioned (Figure 1 shows the most common ones). 62.3% of the respondents use Internet Explorer (IE) as their primary browser, while each of the others is used as primary browser by less than 15% of the participants. 92.4% of the respondents use IE to some degree (as the primary or other browser). Although the use of browsers other than IE appears marginal, it should be noted that browsers supporting tabbed browsing, in which several documents are presented on tabbed panes within one window, are popular among the respondents. At the time, the most popular versions of Opera, Mozilla, Mozilla Firefox, and Apple Safari supported tabs, whereas IE did not. Netscape version 7.2, which was released in the time of data collection, also supports tabs. Although browser versions differ in tab support, it is safe to say that over 60% of the respondents are using tab-supporting browsers as their primary or other browser.

95.3% of the respondents use Google as their primary search engine. When the primary and other search engines are considered together (Figure 2), Google is used by 99.2% of the respondents,

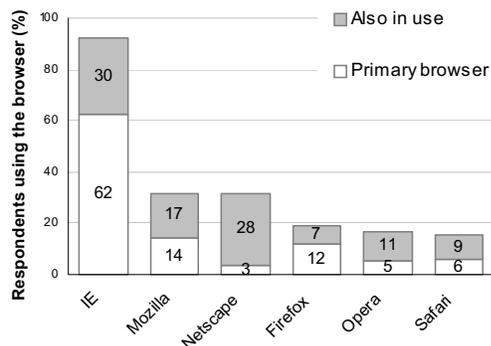


Figure 1. Most common browsers in active use. IE stands for Microsoft Internet Explorer.

meaning that only 2 respondents did not mention using Google at all. Yahoo! and AltaVista are each used by nearly 20% of the respondents, while the others are clearly less common. Altogether, 57 different search facilities were mentioned.

### 4.2 Strategies for Search and Re-access

Conducting work-related search tasks was common among the participants: 37.8% engage in this type of a task daily, 42.2% weekly, 14.3% monthly and only 1.3% less frequently than that.

#### 4.2.1 Prevalence of Strategies

In relation to the use of different strategies, the respondents were asked to consider using their primary browser and search engine. As Figure 3 shows, having multiple browser windows or tabs open while searching is very common (median frequency of use often and almost always, respectively). For re-accessing information, the respondents most commonly use a search engine to find the information again, directly type in the URL, or save documents as local files. All of these strategies are used at least sometimes. Bookmarking and printing out documents is also rather common (median frequency sometimes). However, their frequency of use varies a lot – many respondents only use these strategies rarely, while there are equally many who use them often.

The use of the browser's History tool is not very common (sometimes), nor is the strategy of sending URLs in an e-mail to somebody else (sometimes). However, it is more common to send URLs to others than to oneself (rarely) as many respondents never send URLs to themselves. Saving URLs in a document, adding URLs to a website, and writing down URLs are all used rarely. The least popular strategy is writing down queries (never).

#### 4.2.2 Advanced Operators and Modifiers

Figure 4 shows that among the advanced operators or query modifiers, quotation marks (denoting phrase search) are used most frequently (often). The use of the other modifiers and Boolean operators is rare, although there are respondents who sometimes use plus and minus signs (to include and exclude terms from the documents) and the OR operator to broaden their query. The NOT operator (the same as the minus sign) is used very infrequently.

The most common operator mentioned in addition to the ones already listed was the site-operator (available in Google). This operator restricts the search to a specific domain (site). In addition to the site operator, a wide variety of others were mentioned, such as '~' for synonym search in Google, '\*' in search engines that allow wild cards or truncation, NEAR to look for terms appearing close together, define for a definition of the

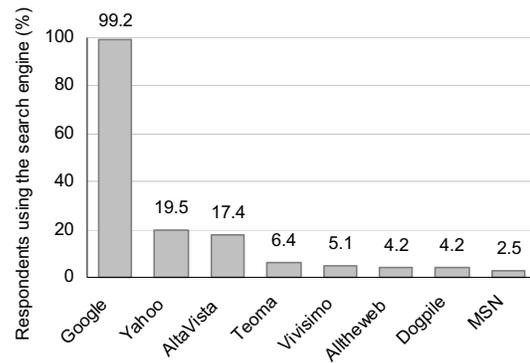
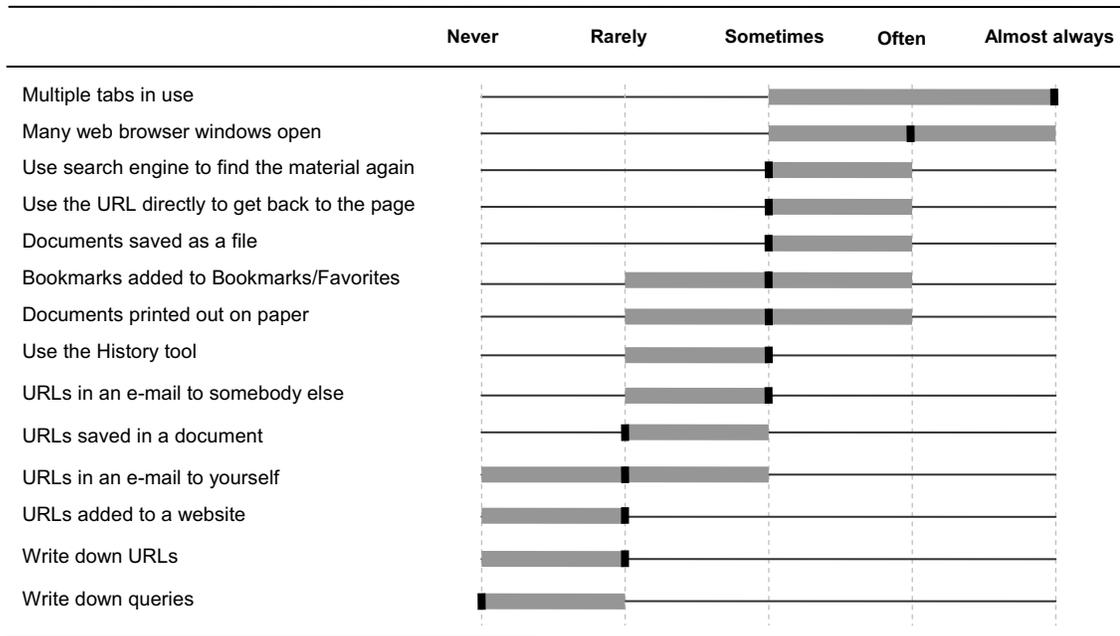


Figure 2. Search engines in active use (as primary or other search engine).



**Figure 3. The information search and re-access strategies. The grey bars denote the region between the first and the third quartile (50% of the responses) and the black dots are the median values. In all of the strategies, the values ranged from *never* to *almost always* (thin black lines).**

words typed into the query field in Google, and link to see which pages have links to the specific page.

#### 4.2.3 Bookmark Usage

A separate question asked the respondents whether they use the Bookmarks tool of their primary browser. They were also asked to give the number of links and folders in their collection (or estimate the numbers). 92.4% of the respondents indicated using Bookmarks in their primary browser. The size of the bookmark collections varied greatly, an average collection included 220 links (SD = 327.4) and 29.7 folders (SD = 47.3). 6.4% of the respondents did not have any bookmarks, 14.4% had less than 50, 63.1% between 51 and 300, and 16.1% more than 300 bookmarks. The largest collection included 2589 links and 425 folders.

### 4.3 Understanding the Search Engine

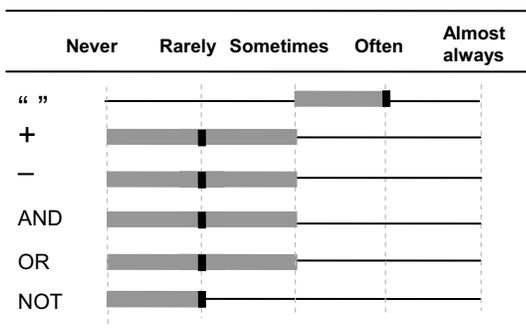
When examining how the respondents understand the functionality of their primary search engine, we only analyzed the

data from the respondents using Google as their primary search engine. This was done because Google is the most popular search engine and because some search engines do not reveal their default operator nor explain their ordering of the results. The respondents who gave multiple search engines as their primary search engine were also left out of this analysis as we did not know which search engine they were referring to. Thus, this analysis is based on the data from 220 respondents.

Google explains that “By default, Google only returns pages that include all of your search terms. There is no need to include “and” between terms.” (<http://www.google.com/help/basics.html>). When the respondents were asked whether they know how a query containing multiple terms is handled by Google, 14.1% responded simply “Yes.” and the correctness of their understanding could not be determined. 33.6% correctly explained that the pages must include all the terms. This leaves at least 52.3% of the respondents with an incorrect understanding or no idea of the default operator. Nearly all of the respondents with an incorrect understanding thought that the default operator is OR. However, they thought that Google orders the result listing so that the first results contain all of the query terms and the next ones all except one term etc.

The understanding about the ranking of the results was more complex to analyze. Although it is told in Google’s web page (<http://www.google.com/technology/index.html>) that PageRank™ [6] is in the heart of it, the whole ranking algorithm is too complex to explain in a questionnaire. Or as one respondent commented: “The only people who know how ranking truly works, work for the search engine companies.” Thus, we only analyzed the answers to the ranking question by categorizing them into three groups. These groups are presented below with the proportion of respondents falling into each.

1. No explanation: blank, “I don’t know”, or “By relevance” (without any explanation) 30.6%



**Figure 4. The frequency of modifier and operator use. The grey bars denote the region between the first and the third quartile and black dots are the median values. All values ranged from *never* to *almost always*.**

2. PageRank™: answers mentioning PageRank™ (and possibly others) 45.4%
3. Other: answers listing other ranking mechanisms (but not PageRank™) 24.1%

## 5. DISCUSSION

In this section, the results of the questionnaire study are discussed along with comments from the respondents. The comments are used as examples of the rationales behind the strategies used. In addition, they highlight the problems involved in using some of the strategies. This discussion focuses on the most helpful and prevalent strategies revealed by the questionnaire responses.

### 5.1 Key Strategies during the Search Process

**Experienced users manage the search process with multiple windows and tabs.** By using several browser windows or tabs in parallel during the search session, users can leave tracks of their browsing history and easily return to earlier pages. Additionally, this strategy enables the user to do something else, for example, go through the result list while slow pages download (also noted in [8]). The benefit of tabs is that, unlike multiple browser windows, they do not clutter the workspace. In Opera, for example, tabs are even saved between sessions.

*I often use tabs to stagger the tasks of opening and loading results and looking at them. So I click 3 results on the Google page, then read them.*

*I like to do searches with two browser windows open. I use the first window to initiate the search, and I use the second window to drag links into from the search results list.*

*Because sometimes a search can be (very) lengthy and I'm the lazy sort of guy it's useful that in Opera you can save the session (which tabs are open etc) and return to it later (no need to write down anything!).*

**Experienced users see the benefits of categorized information.** For the “Other comments or strategies” question, some respondents explained the benefits of using categorizing (or clustering) search engines along with their primary search engine. Categorizing helps the user by providing an overview of the result set and generally about the topic, and also by supplying additional search terms. Categories also provide access to results that are further down the result list, which is useful especially when the topic of the search is unfamiliar. In those cases, users tend to formulate queries with general terms and the ranking algorithm does not necessarily get the best document in the top of the list – that is, if the user even knows what the best documents for her vague information need are.

*I'll sometimes start in Vivisimo to get a relative idea of the use of terminology for a topic, then use the terms I find to search more narrowly in Google.*

*The main issue and flaw in Google's results that I and numerous other information professionals have pointed out is that they're basically unstructured. Clustering of search results some search engines such as Vivisimo use would improve the ability of the end user to locate relevant information and further limit or expand their search.*

*I use Google primarily for known item finding - when I know that the answer that I want exists and I am happy finding any site with the answer. I use Vivisimo when exploring, when not sure how to phrase my search or what I will find.*

**5.2 Key Strategies for Information Re-access**  
**Experienced users use search engines for information re-access, but have problems with this approach.** Although the respondents seem to frequently rely on using a search engine to re-access material, this strategy is also problematic:

*I think my main problem in web searches is nowadays that I can't remember which were the terms that I used when I found a relevant site.*

Finding relevant information is often an iterative process, especially for experienced users [5]. Because several queries are sometimes needed for finding information, it can be almost impossible to remember the exact query that was used when a specific piece of information was found. It can also be difficult to use search engines to re-access information that was originally found when browsing [36].

**Experienced users use Bookmarks frequently despite the associated burdens.** A clear majority of the respondents use Bookmarks and, since the average collection has 220 links (similar numbers are presented in [10]), many do so plentifully. Yet, it is well-known that large bookmark collections are difficult to organize and require continuous maintenance [1],[10]. The use of sophisticated bookmark organization methods is more typical with more experienced users [1]. However, even experienced users, who unquestionably have the skills needed for organization, struggle with the tools provided for this task. Several comments related to the “painful organization” of bookmarks can be seen as signs of serious usability problems with existing Bookmark tools.

*IE makes it so hard to organize favorites that I leave them all in an ugly pile and don't rely on them as much as I'd like.*

*Re-org is a pain. The simple tree of the bookmark manager hides nooks and crannies.*

People also add bookmarks even though they are not sure whether the information will be used again. The disadvantage of this is that the bookmark collection becomes cluttered. This, in turn, makes organizing and using the collection even more difficult.

*Many of the URLs I bookmark or pages I download are not subsequently reviewed. I save things that look like they may be relevant (now or later) but I know that I don't refer to them again - other than if I run a similar search and remember that I've saved information, in which case I may search my hard drive (using Index Server on WinXP).*

The fact that Bookmarks can only be from one computer makes their use difficult for a large number of computer users:

*(...) I would like to have them always accessible, independently from location and machine. During meetings or seminars, I would like to go back to one of the web resources I've stored on my computer or show something.*

One solution would be to have Bookmarks integrated to the search engine (also suggested in [1]) or some other web-based tool:

*Ideally, I think a web user should get a web-based tool that could centralize bookmarks on a device-independent area, which should be available from everywhere.*

We were interested in knowing more about the benefits of using large bookmark collections. Thus, we e-mailed 10 respondents with a bookmark collection of more than 500 files and asked them about their experiences. Most of these respondents appear very

active in their use of Bookmarks and they use a large proportion of them regularly, although usually only one, e.g., a project-related folder, is in active use at a time. They clean up their bookmarks from time to time by deleting unnecessary files and folders (invalid links are a well-known problem with the bookmarks [10]), archive links, etc. Nevertheless, cleaning up was seen as an activity that should be done more often.

*I probably clean up bookmarks once or twice a year. I bet at this point, it's been at least a year since I've done a cleanup.*

*When I find a link that seems to be obsolete, I try to remember to delete the bookmark. But, I am often in too much of a hurry or too lazy to do it then.*

These heavy-users of bookmarks also carefully organize bookmarks, typically with two or three levels of folders. When asked about the successfulness of their bookmark organization, all of them were happy with it. For these people, Bookmarks had become an indispensable tool:

*I have spent lots of time thinking about the organization in order to find the ones I need as quickly as possible. I have several folders (and subfolders) named based on the bookmark content, for example Music, Work, Usability, eLearning, Studies, News etc. I'm quite satisfied with the organization - there could be somewhat less folders, though.*

*Yes it is highly successful for my needs over the last ten years! I literally have hundreds of folders.*

*(...) it is a really helpful thing and I would be totally lost without the favorites folder!*

Thus, it seems that Bookmarks is a valuable tool for people who are willing to use the extra time necessary to keep the collection organized. For others, the problems outlined above significantly reduce the utility of the Bookmarks tool.

### 5.3 Struggling with Strategies

There is always the chicken and egg problem with the use of different strategies [11]: limitations of existing tools might prevent or discourage users from using beneficial strategies. Infrequently used strategies and the misunderstandings related to the functioning of Google are discussed next.

**Experienced users rarely rely on the History tool.** The results showed that the respondents use the History tool infrequently. As with Bookmarks, one problem with History is that it is only available on one computer.

*Also I use so many different computers during the day that certain browser's history information won't help me.*

There are also other possible reasons for the infrequent use of the History tool. Tauscher and Greenberg [34] suggested that the stack model used in the History tools might not optimally support the user's task. In addition, History tools rely on page titles which often poorly represent the contents of the page [10]. Another problem is that the history list is inevitably cluttered: it saves the URLs of both the pages that, though visited, were actually irrelevant to the user along with the pages that were very important to the user. These reasons compromise the usability of History and result in at least experienced users seeking for alternative ways to support information re-access. One way encountered includes the temporary use of bookmarks:

*Sometimes I'll just throw a bookmark in and use it for a little while and then delete it. I don't usually put those temporary Bookmarks in a folder. (Hmm, maybe I should start a "temp" folder.)*

#### **Experienced users rarely e-mail URLs to themselves or save links to a web page.**

There are a couple of possible rationales for e-mailing URLs to self [21]. First, e-mailing the URLs provides the possibility for accessing it in another computer, which is also one rationale behind adding URLs to a website. Secondly, people sometimes use the incoming mail as a reminder for using the information. Although e-mailing URLs to oneself and adding them to a website serve useful functions, this study suggested that their use is not common. There are several possible explanations for their infrequent use: it is possible that respondents do not need to access work-related URLs at home and home-related URLs at work, thus, the need for computer-independent access may be small. The use of laptops also decreases the need for these strategies: if people are always using the same computer, they can access the URLs, for example, by using Bookmarks. On the other hand, although providing clear benefits through computer independent access, these strategies may still be too troublesome and thus, infrequently used. Mailing URLs requires the user to have two applications (the web browser and the e-mail application). In addition, it also requires the user to save the URL at the target computer, if the user does not want to use the e-mail system as additional bookmark storage. Adding URLs to a web page can be equally cumbersome.

*On my home computer I have a link (on the links bar) to my work bookmarks file. Of course, it is not as convenient to use the bookmarks from the html file. I really should transfer the work bookmarks folder to my home computer and set up something to synchronize the two.*

#### **Queries of the experienced users resemble those of typical web users, though imaginative use of search engines increases with experience.**

Previous results concerning the experienced users' use of advanced operators are mixed; some claim that their use is frequent [15], while others have shown their infrequent use [5]. The current results support the latter view. Of the advanced operators, only the phrase search is used frequently, while the use of the others is rare. The reason why experienced users formulate such simple queries might be optimization: they use strategies that take little time and effort and still deliver satisfying results.

*I used to work for a search service and was fairly sophisticated in my use of Boolean operators. Because I have been pretty happy with Google results (...) I have virtually abandoned using operators or engines that let me control the search query.*

In other aspects as well, experienced users resemble the "typical searchers" studied in the log studies:

*With Google, I don't care if I get thousands of results because I usually only look at the first few pages of results.*

Although the queries by more experienced users resemble those of less-experienced users, they may still be more successful. The more experienced users may, for example, choose more suitable terms by using the words that are likely to appear in a relevant page. Currently, we are analyzing data from 22 searchers to determine whether such differences in term selection exist.

*I choose search terms based not specifically on the information I want, but rather how I could imagine someone wording a site that contains that information.*

*Google can be quite fast and accurate, if you just know the right way to present your question (that is to 'reverse the keywords' from the imaginary result page - if the keywords don't give you the right answer, you'll just try to figure out another way how the thing might be presented on a web page).*

Web users rarely check results beyond the 10<sup>th</sup> or 20<sup>th</sup> position (the 1<sup>st</sup> and 2<sup>nd</sup> result page) in the search engine's result list [17]. Experienced users may not check any more result pages, but the pages they check may have 100 results. They may even sporadically check deeper into the listing. So, they seem to be aware that even the best ranking algorithms have limitations.

*I also use different strategies with the results I get from Google - sometimes I jump to pages like 14 or 37 to check how it affects the results.*

*I rarely look at the second or third result pages (but I have set [the search engine] up so that 100 results are shown)*

**Experienced users have misconceptions about the functioning of Google.** Over half of the respondents had misconceptions about the default operator of Google. Despite this erroneous understanding, it seems that people can still make successful queries:

*We have our own search engine (that searches a relational database) and it's interesting to note that I know how our search works (ANDs, ORs, etc.) but have not researched how Google works. I think because it works so well with the basic "just type in the words" search, I haven't needed much else.*

However, a comment by one user from our earlier study [5] clearly shows one problem caused by this misunderstanding:

*In a way, the selection of the query terms is almost random when I copy search terms from documents. Here, for example, the term yellow does not have anything to do with the topic.*

As this very experienced user thought that Google does not require all the terms in the results, she carelessly entered terms into her query, including terms that had nothing to do with the topic. As a consequence, the results she received, which all contained this unrelated term, were skewed. It is likely that she missed several important documents because of this harmless-looking misunderstanding.

## 6. DESIGN IMPLICATIONS

For typical web users, the advanced information search and re-access strategies may be too difficult because they are not directly supported by current search engines and web browsers. In this section, we will first briefly present two of our research platforms, *Session Highlights* and *Findex* (Figure 5). These tools have been designed to support users with various levels of expertise in information search and re-access tasks. The importance of the ideas behind the tools was confirmed by the results of this study, but a number of ideas for developing the tools further were also raised. These ideas will be discussed after presenting the tools.

### 6.1 Research Platforms

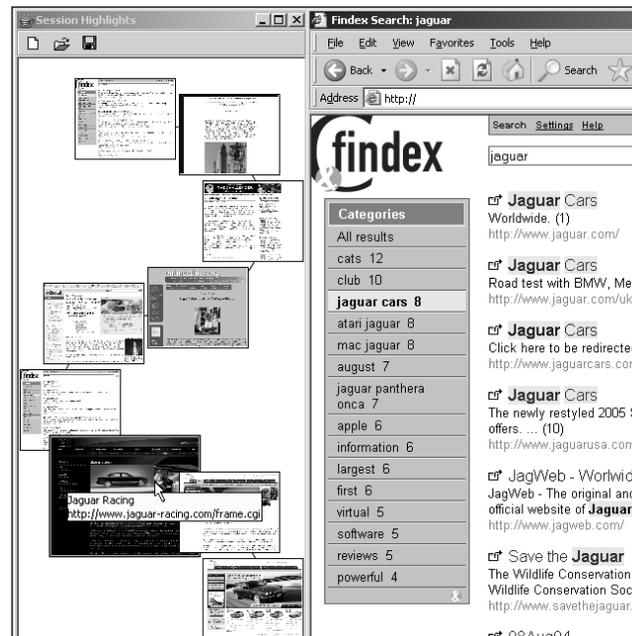
#### 6.1.1 Session Highlights

*Session Highlights* is a browser-independent web companion in which URLs of interest can be collected. A URL is added by

dragging the Address icon (in the address field of the browser) and dropping it anywhere on the workspace. The URL is added to the chronological list and represented as a web page thumbnail. When the mouse cursor is placed over the thumbnail, it is enlarged and a tool tip displaying the title and URL of the page appears. The resulting collection can be saved and later re-opened for accessing the collected URLs or further supplementing the collection. Although the basic functionality of *Session Highlights* resembles that of bookmarks, its approach of visual presentation, automatic chronological ordering, and continuous presence, changes the nature of the tool. As users can easily store intermediate search results, the tool ideally supports various stages of the search process. Interfaces with related motivations exist (e.g., [2],[28]); however, the designs and approaches differ greatly from that of this tool. For a detailed description of *Session Highlights*, the reader is referred to [20].

#### 6.1.2 Findex – Categorizing Search Interface

*Findex* is a search interface that enhances Google searches with automatically generated categories. For example, for the query 'jaguar', *Findex* will present categories such as: 'jaguar cars', 'jaguar panthera onca', and 'mac jaguar', and 'atari jaguar', meaning that the categories present different contexts where the word 'jaguar' is used in the results. The categories act as filters so that upon category selection, *Findex* will only show the results that contain the terms presented in the category name. Similar approaches have been suggested earlier (e.g., [13],[38]), but our solution is simpler: the categorization is based on word and phrase frequencies and thus, the logic behind the categorization is easy to understand. In contrast to the hierarchical categorization of some commercial categorizing search engines, such as Vivísimo [35], iBoogie [16], or WiseNut [37], *Findex* produces a simple flat categorization. For a detailed description of *Findex* and the results from two user studies, refer to [26] and [27].



**Figure 5. Session Highlights (left) and Findex (right). The outcome of the search (a search result list, 4 relevant links, another search result list, and then 3 relevant links) is collected in Session Highlights.**

## 6.2 Easy Access for Advanced Strategies

### 6.2.1 Multiple Tabs or Windows

Our results showed that multiple tabs or web browser windows are often used in parallel during the search session. These strategies provide the users with tracks of their search history as well as an easy access to the previously visited pages. Although both of these strategies provide the same benefits, multiple browser windows have the disadvantage of cluttering the workspace. In addition, multiple windows may be less intuitive to use for less experienced users as other applications than web browsers do not usually allow multiple instances of the application to be open at the same time. Thus, we think that browsers should rather support tabbed browsing than only opening multiple windows.

When using Session Highlights, the use of multiple windows or tabs is no longer necessary. By collecting pages of interest, rather than opening them in new windows or tabs, the user has a constant visual summary of the key pages. As recognizing is typically much easier than recalling [3],[12], using the thumbnails is less cognitively demanding than remembering which web pages are open in the background. Additionally, it is known that people remember visual information well [3] and can use visual cues from thumbnails to recognize web pages [23]. Thus, it is conceivable that if the user needs to access open pages by using only the page titles (which can be misleading or missing altogether) from the browsers' Window menu or names in the tabs, the performance in the recognition task is poorer than when also the thumbnails are presented.

### 6.2.2 Using a Search Engine to Re-access Material

Using search engines to get back to the previously found information is a widely used strategy. However, this strategy has problems as it is difficult to remember the exact search terms used to find the material in the first place. To alleviate this problem, we have planned to include a category to Findex that would show the user those documents among the result list that s/he has visited recently. In practice, Findex will maintain a history list and every time the user submits a query, Findex compares the history list with the URLs of the result set. If recently visited documents are found, Findex shows a category "Recently visited documents" to the user in addition to the normal categories. Inside this category, the results are further organized temporally so that the most recent visits will be on the top of the list. This approach facilitates revisitation when the relevant (previously visited) result is somewhere in the result list, but not among the first results. This situation may happen when the search engine updates its database and the ranking order of the results changes. In addition, when the user does not remember the exact query terms she used when the document was found for the first time, the rank of the relevant document may be considerably different than what it was earlier.

Session Highlights also provides ways to overcome the difficulties related to using search engines for information re-access. When conducting a search, users may add search result pages as well as results to their workspace, thereby preserving their successful queries and their key findings. As a collection can be saved, Session Highlights enables search session continuation at a later time. Thus, the user does not need to recall the specific query terms or even the search engine she used, as the whole result page with both the query terms and the results can be saved. In addition, if the browser history is left intact for subsequent sessions, the link colors will indicate which URLs in the result list

were already visited. Altogether, the need to recall queries and repeat lengthy search processes is abolished.

### 6.2.3 Storing Documents and Bookmarks

Saving documents as files and printing documents on paper are also commonly used strategies among the experienced users. Both of these strategies remove the risks related to the instability of the web contents. When using the current tools, such as Bookmarks, there is always a risk that the page is unavailable when needed for the next time and the information is lost. To remove the risk of losing important information, browsers could automatically save a copy of the page to the hard disc when the user indicates that the page is important by bookmarking it. Although the copies of the document require space, the cost of disk space is low when compared to the price of losing important information.

In addition to securing the information, print-outs also make it possible to access the information independent of the location. However, printing out the documents is costly and thus, mechanisms for accessing documents in a location-independent manner without having to print them out would be beneficial. To achieve this, search engines could provide a possibility for saving Bookmarks as well as copies of the bookmarked documents. In practice, users could be provided a service for storing their Bookmarks and web pages. The search engine would also search this personal collection thereby alleviating some of the problems related to the organization of Bookmarks and revisitation. This functionality allows the users to access their Bookmarks from different computers without cumbersome procedures of adding URLs to a webpage or e-mailing them to oneself.

Session Highlights also addresses some of the reported problems of bookmark collection cluttering and management. It promotes behavior whereby a working set of URLs can be first collected, leaving their evaluation as a second phase. After having been evaluated, the URLs of key importance can be added to a bookmark collection, a document, or an e-mail. Thus, the user can focus on the search task without being distracted with concerns of bookmark management.

### 6.2.4 Helping Users Understand their Queries

Surprisingly, most of the experienced users did not know how the search engine handles queries with multiple terms and it is presumable that novices have even less understanding of the issue. To alleviate the problems related to this misunderstanding, we have implemented a query explanation feature which will be integrated into Findex. Table 1 presents a couple of examples of the query explanations. This tool translates queries into natural language phrases by using a query parser and explanation templates. In practice, default operators and more elaborate queries (operator precedence, mistakes in using operators etc.) are

Table 1. Examples of the query explanations

Query	Explanation
atari jaguar	Matching documents contain both of the words atari and jaguar.
atari jaguar game	Matching documents contain all of the words atari, jaguar, and game.
atari jaguar OR game	Matching documents contain the word atari. In addition, the documents contain either the word jaguar or the word game.

translated into natural language and thus, the correctness of the query is easy to check. Although the advanced operators or term modifiers are not commonly used in web queries, the natural language explanations will help the users understand the default functioning of the search engine (how does it handle queries without any operators). In addition, one reason for the users not using operators may be that they do not know how to use them correctly, and thus, do not benefit from their usage. The natural language explanations are also expected to help in this problem.

### 6.2.5 Evaluating and Filtering the Result Set

The experienced users reported using categorizing search engines when they needed to get an overview of the result set and the topic of their search in general. In addition, the categories were used as additional search terms. The categories of Findex provide the same benefits: the users can both evaluate the success of the query easier and add the category names as search terms to the query. In addition, categories provide easy access to relevant results further down the list. This is an important functionality because typical web users normally check only the first 10 or 20 results [17], while the more experienced users sometimes check also results ranked over 30 or have set the search engine to show 100 results in one page. In addition, the queries of less experienced users are typically short and broad, so the ranking of the search engine cannot necessarily position the most relevant results on the top of the list (even the user may not know which results are the most relevant ones). The categories help with this problem as the user can easily access also the results that are not high in ranking.

## 7. CONCLUSIONS

Past research has mostly identified information search and re-access strategies of experienced users either by means of observational studies and interviews or through log data. To build on previous findings, we compiled a comprehensive list of information search and re-access strategies and identified the relative importance of each among a varied group of 236 experienced users. The questionnaire approach made it possible to gain a broad understanding of the strategies regardless of the tools that people are using. In addition, the responses provided valuable information about the rationales behind the different strategies as well as revealed some new strategies. The better understanding of the strategies arms the designers with greater support for designing tools for information search and re-access. Additionally, the open-ended questions clearly showed that even experienced users have difficulties in finding and re-using information on the web. This fact has not received much attention previously.

It is not certain that all the strategies of experienced users are actually the most effective and efficient ones, but at least they seem to be more successful than the strategies of less experienced users [15]. Thus, we believe that these advanced strategies would also benefit the users having less experience. Furthermore, our solutions do not force people to use strategies that they do not find beneficial. The support for the advanced strategies is simply an added possibility for the less experienced users.

In response to our three research questions, the conclusions were the following:

- The respondents' most common browser is Internet Explorer, but over 60% of them use others that support tabs as either the primary or other browser. As a search engine, Google is clearly the most frequently used, but they are also using a large number of other search facilities.

- The most frequent advanced strategy is to use multiple browser windows or tabs in parallel while searching. Common strategies for re-accessing information are search engine usage, using URLs directly, or saving the document to the computer. Bookmarks are also commonly used, but their frequency of use varies a lot. Saving URLs in a document, e-mailing URLs to self, adding URLs to a website, and writing down URLs or queries are all infrequently used.
- In spite of the advanced strategies, several points were found where information search and re-access are problematic: users (even highly experienced ones) do not know how their search engine really functions. Bookmarks are laborious to organize, but without organization, they are very difficult to use and at times, even useless. Using search engines is, in theory, a simple method for re-accessing information. In practice, however, it is very difficult to remember the query terms used when finding the information in the first place.

To build on these findings, we presented design ideas that would make advanced strategies accessible to less advanced users, while also alleviating the complications associated with information search and re-access for all users.

## 8. ACKNOWLEDGMENTS

We thank all of the respondents for their participation. We would also like to thank Professor Kari-Jouko Rähkä and anonymous reviewers for their valuable comments on this work. This study was financially supported by the Graduate School in User-Centered Information Technology, the Academy of Finland (project 178099), and the National Technology Agency (project 20685).

## 9. REFERENCES

- [1] Abrams, D., Baecker, R. and Chignell, M. Information archiving with bookmarks: Personal web space construction and organization. In Proc. CHI 1998, ACM Press (1998), 41-48.
- [2] Amento, B., Terveen, L., Hill, W. and Hix, D. TopicShop: enhanced support for evaluating and organizing collections of Web sites. In Proc. UIST 2000, ACM Press (2000), 201-209.
- [3] Anderson, J.R. Learning and Memory: An Integrated Approach. John Wiley & Sons, Inc. (2000).
- [4] Aula, A. Query Formulation in Web Information Search. In Isaias, P. & Karmakar, N. (Eds.) Proc. IADIS International Conference WWW/Internet 2003, Volume I, 403-410. IADIS Press (2003).
- [5] Aula, A. and Käki, M. Understanding Expert Search Strategies for Designing User-Friendly Search Interfaces. In Isaias, P. & Karmakar, N. (Eds.) Proc. IADIS International Conference WWW/Internet 2003, Volume II, 759-762. IADIS Press (2003).
- [6] Brin, S. and Page, L. The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems, 30, 1-7(1998), 107-117.
- [7] Bruce, H., Jones, W., and Dumais, S. Keeping and re-finding information on the web: What do people do and what do they need? In Proc. ASIST 2004, Chicago, IL, Information Today, Inc., October, 2004.

- [8] Byrne, M.D., John, B.E., Wehrle, N.S., and Crow, D.C. The tangled web we wove: A taskonomy of WWW use. In Proc. CHI 1999, ACM Press (1999), 544-551.
- [9] Catledge, L. and Pitkow, J. Characterizing browsing strategies in the World-Wide-Web. In Proc 3<sup>rd</sup> International World Wide Web Conference. <http://www.igd.fhg.de/www/www95/papers/>
- [10] Cockburn, A., Greenberg, S., Jones, S. McKenzie, B. and Moyle, M. Improving web page revisitation: Analysis, design and evaluation. *IT & Society*, 1, 3 (2003), 159-183.
- [11] Cockburn, A. and McKenzie, B. What do web users do? An empirical analysis of web use. *Int. J. Human-Computer Studies*, 54, 6 (2000), 903-922.
- [12] Dhamija, R. and Perrig, A. Déjà Vu: A user study using images for authentication. In Proc. 9<sup>th</sup> USENIX Security Symposium, 2000.
- [13] Dumais, S., Cutrell, E., and Chen, H.: Optimizing Search by Showing Results in Context. In Proc. CHI 2001, ACM Press (2001), 277-284.
- [14] Fox, S. Search engines: A Pew Internet project data memo (2002). Available at: <http://www.pewinternet.org/reports/toc.asp?Report=64>
- [15] Hölscher, C., and Strube, G. Web search behavior of internet experts and newbies. In Proc. 9<sup>th</sup> International WWW conference. Amsterdam, The Netherlands, 337-346 (2000).
- [16] iBoogie Metasearch Clustering Engine. <http://www.iboogie.com>
- [17] Jansen, B.J. and Pooch, U. A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*, 52, 3 (2001), 235-246.
- [18] Jansen, B.J. and Spink, A. (in press) How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, in press.
- [19] Jansen, B.J., Spink, A., and Saracevic, T. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36, 2 (2000), 207-227.
- [20] Jhaveri, N. and Rähä, K.-J. The advantages of a cross-session web workspace. To appear in Proc. CHI 2005, ACM Press (2005).
- [21] Jones, W., Bruce, H., and Dumais, S. Keeping found things found on the Web. In Proc. Tenth International Conference on Information and Knowledge Management, 119-126 (2001).
- [22] Jones, W., Bruce, H., and Dumais, S. How do people get back to information on the web? How can they do it better? In Proc. INTERACT 2003, 793-796.
- [23] Kaasten, S., Greenberg, S., and Edwards, C. How people recognize previously seen WWW pages from titles, URLs and thumbnails. In X. Faulkner, J. Finlay and F. Detienne (Eds.) Proc. of Human Computer Interaction 2002, BCS Conference Series, 247-265.
- [24] Kitchenham, B.A. and Pfleeger, S.I. Principles of survey research part 3: Constructing a survey instrument. *Software Engineering Notes*, 27, 2 (2002), 20-24.
- [25] Kitchenham, B.A. and Pfleeger, S.I. Principles of survey research part 4: Questionnaire evaluation. *Software Engineering Notes*, 27, 3 (2002), 20-23.
- [26] Käkä, M. Findex: Search result categories help users when document ranking fails. To appear in Proc. CHI 2005, ACM Press (2005).
- [27] Käkä, M. and Aula, A. Findex: Improving search result use through automatic filtering categories. To appear in *Interacting with Computers*.
- [28] OmniGroup. OmniWeb. <http://www.omnigroup.com/applications/omniweb/>
- [29] Pirolli, P. and Card, S. Information foraging. *Psychological Review*, 106, 4 (1999), 643-685.
- [30] Sellen, A.J., Murphy, R. and Shaw, K.L. How knowledge workers use the web. In Proc. CHI 2002, ACM Press (2002), 227-234.
- [31] Spink, A., Wolfram, D., Jansen, B.J., and Saracevic, T. Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52, 3 (2001), 226-234.
- [32] Straub, D.W. Validation in information systems research: A state-of-the-art assessment. *MIS Quarterly*, 24, 1 (2001), 1-16.
- [33] Sutcliffe, A. and Ennis, M. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10 (1998), 321-351.
- [34] Tauscher, L. and Greenberg, S. Revisitation patterns in World Wide Web navigation. In Proc. CHI 1997, ACM Press (1997), 99-106.
- [35] Visimo Clustering Engine. <http://www.vivisimo.com>
- [36] Wen, J. Post-valued recall web pages: User disorientation hits the big time. *IT & Society*, 1, 2 (2003), 184-194.
- [37] WiseNut. <http://www.wisenut.com>
- [38] Zamir, O. and Etzioni, O. Web Document Clustering: A Feasibility Demonstration. In Proc. SIGIR'98, ACM Press (1998), 46-54.