Hundreds of Folders or One Ugly Pile – Strategies for Information Search and Re-access

Anne Aula and Harri Siirtola

Tampere Unit for Computer-Human Interaction (TAUCHI),
Department of Computer Sciences, FIN-33014 University of Tampere, Finland
{anne.aula, harri.siirtola}@cs.uta.fi

Abstract. Previous research has identified information search and re-access strategies used by experienced web users, but not the popularity of these strategies. To fill this gap, we collected data from 236 experienced web users via a questionnaire. A cluster analysis of the data revealed three distinct user groups, each having different search and re-access strategies. The group of “professional searchers” was the only one to use Boolean operators in searching and most of the advanced strategies for information re-access. In contrast, the strategies of the other two groups were much simpler. These results show that the previous findings on advanced search and re-access strategies may describe the strategies of information professionals, but not those of other web users, regardless of their extensive experience.

1 Introduction

Previous studies have shown that information professionals or “expert users” commonly utilize advanced strategies for information search and re-access. Most of these studies have used observational methods [6,7,9] and have suggested the following strategies: e-mailing URLs, saving URLs in files or adding them to a website, saving documents as files or printing them out, using a search engine or directly the URL, adding bookmarks, using the History tool, and writing down queries or URLs. The problem with using observational methods is the expense of data acquisition, which typically results in small sample sizes.

The questionnaire approach (previously used by [1,2,3,4]), makes it possible to reach hundreds of users, and to acquire data about the use of all different applications. Although the absolute frequencies of different strategies should not be measured with questionnaires (because this approach depends on the users’ memory), we believe that people can reliably evaluate their relative frequencies.

We compiled a comprehensive list of the search and re-access strategies and conducted a questionnaire study about the usage frequencies of them among experienced web users. The results concerning the strategies of the respondents as a whole are presented in a separate paper [3]; the current paper presents the strategies of three subgroups of experienced users.

2 The Questionnaire and Methods for Data Analysis

To receive responses from experienced web users with different backgrounds (not only information professionals), we sent the URL of the questionnaire 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 distributed it in their organization.

The questionnaire had 7 background questions and 9 questions about computer, web, and search engine use. The respondents were asked to think of a typical work-related information search task when answering the questions. In relation to this task, the frequency of using 14 different search and re-access strategies was asked on a 5 point scale (*almost always*, *often*, *sometimes*, *rarely*, or *never*). In addition, 10 questions addressed bookmark use and the frequency of using operators or modifiers in queries. Three open-ended questions elicited the participants' understanding of their primary search engine, strategies not listed in the questionnaire, and other comments. The questionnaire can be accessed from <http://www.cs.uta.fi/~aula/questionnaire.php>.

The 31 (questions) $\times$ 236 (respondents) table with 6 ratio, 22 ordinal, and 3 nominal scale variables was analyzed. The dissimilarities between observations were computed with algorithm `daisy`. Then, the `hclust` package of statistical system R was used for a hierarchical clustering [8]. For the pairwise comparisons between the clusters, Tukey's multiple comparisons with exact *t*-distribution were used.

3 Results

The cluster analysis resulted in three top-level groups of respondents. In the remainder of this paper these groups will be called *professionals*, *frequent searchers*, and *regular searchers*. In Figures 1 and 2, the boxes denote the region between the first and the third quartile (50% of the responses) and the black dots are the median values. Fig. 1 summarizes the search frequency of these groups. Professionals (37 respondents) were mainly librarians (35.1%) and managers (21.6%) with an average of 11.7 years of experience in their profession. Frequent searchers (111 respondents) were mainly designers (21.1%), researchers (18.0%), and librarians (17.1%), with an average of 6.0 years of professional experience. Regular searchers (88 respondents) were mainly researchers (23.9%), designers (20.5%), and usability specialists (15.9%) with an average of 7.8 years of professional experience.

Fig. 2 shows the frequency of using operators and modifiers, as well as the use of information re-access strategies. Quotes were the most common modifier, and they were used *often* in all groups. The use of the other modifiers and operators was *rare*, only professionals used all of them *sometimes*. In their use of operators, regular searchers were the least active of the three groups as most of them *never* used the operators OR and NOT, and also the use of AND, +, and - was *rare*. Frequent searchers seemed to use the operators and modifiers slightly

Fig. 1. Search frequency in the three groups

Fig. 2. Frequency of using query operators and strategies for information re-access

more often, the median usage frequency being *rarely* for all of them. Statistically, the only significant difference between the frequent and regular searchers was in the use of the NOT operator.

In information re-access strategies, statistically significant differences were found between the two less experienced groups (frequent and regular searcher) and professionals. The professionals email URLs to themselves, add URLs to web-sites, and print out documents more often.

Bookmark collection sizes seem to increase with experience. Professionals had collections of 400 files, on average; the collections of frequent and regular searchers had an average of 213 and 167 files, respectively. The number of folders pointed to the same direction with professionals having an average of 33 folders, and frequent and regular searchers having 25 and 31 folders (difference *n.s.*).

4 Discussion

The three groups differed in their use of the strategies. Professionals used the advanced strategies most frequently, whereas the use of them was relatively rare in the other groups. The differences in the strategies of the two less experienced groups were subtle, but their search frequencies differed significantly.

The use of some strategies was rare in all of the groups (e.g., writing down web addresses or queries). As most of the strategies were used frequently by

professionals only, it seems that these strategies are truly advanced strategies – only used by information professionals or others with similar needs. However, the other user groups might also benefit from the advanced strategies, but are currently unable to use them efficiently. *“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.”* Professionals, on the other hand, may be in a position where they must go through some trouble for making information useful for them. *“I have organized bookmarks in subject alphabetical based folders [sic] with subfolders and links below... Yes it has been highly successful for my needs ... I literally have hundreds of folders.”*

All of the groups use quotes *often*. Otherwise, the queries are simple. The regular searchers, in particular, employ a “just-type-in-words” strategy. In contrast, professionals use all of the listed operators *sometimes*. Frequent searchers were in the middle, using them *rarely*. Thus, it seems that professionals are the only notable exceptions to the just-type-in-words strategy.

The strategies of professionals are not unquestionably better than those of less experienced users. However, as there is evidence that the effectiveness in information search tasks increases with experience [5], we feel safe to make this assumption. Thus, the focus in developing interfaces for information search and re-access should be in making these strategies *accessible* for all users and for the time being, set aside the question of whether they will be used by them. If we do not try and make the strategies accessible, the “lazy sort of guy”, as one respondent described himself, will never get to see their possible benefits.

References

1. Abrams, D., Baecker, R., and Chignell, M. Information archiving with bookmarks: personal web page construction and organization. In P. Isaías and N. Karmaker (eds.), Proc. CHI 1998, ACM Press (1998), 41–48.
2. Aula, A. Query Formulation in Web Information Search. In Proc. WWW/Internet 2003, IADIS Press (2003), 403–410.
3. Aula, A., Jhaveri, N., and Käki, M. Information search and re-access strategies of experienced web users. Proceedings of WWW 2005, May 10-14, 2005, pp. 583–592.
4. 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, (2004).
5. Hölscher, C., and Strube, G. Web search behavior of internet experts and newbies. In H. Maurer & R.G. Olson (eds.), Proc. 9th Int. WWW conference, (2000), 337–346.
6. Jones, W., Bruce, H., and Dumais, S. Keeping found things found on the Web. In H. Paques et al. (eds.), Proc. Information and Knowledge Management, (2001), 119–126.
7. Jones, W., Bruce, H., and Dumais, S. How do people get back to information on the web? How can they do it better? In M. Rauterberg et al. (eds.), Proc. INTER-ACT’03, IOS Press (2003), 793–796.
8. R Development Core Team and R Foundation for Statistical Computing. R: A language and environment for statistical computing, (2004).
9. Wen, J. Post-valued recall web pages: User disorientation hits the big time. IT & Society, 1, 2 (2003), 184–194.