

Understanding the Brevity of Web Queries

Luanne Freund and Elaine G. Toms

Faculty of Information Studies

University of Toronto. Email: {freund, [toms](mailto:toms@fis.utoronto.ca)}@fis.utoronto.ca.

Introduction

Internet search engines are the most common resource for people seeking answers to questions and access to web-based information, and in most search engines, keyword queries are the primary means of retrieving information. Yet, web queries tend to be remarkably brief and general. Although they have grown slightly in length over time, web queries are still on average about two words in length, which is significantly shorter and less complex than queries used in earlier types of information systems (Jansen & Pooch, 2000; Wang, Berry & Yang, 2003). Popular queries listed on the Google site, such as "weather", "moon," and "nature" serve as very coarse filters for the two billion indexed web pages, and return millions of hits. Given this scenario, it is not surprising that people are often frustrated by web searching, and that search success rates of about 50% are reported (Nielsen, 2001).

Why then, do people submit such non-discriminating queries? This question is particularly relevant in the Internet context, but is not new to information science. The difficulty people have in formulating search statements from information needs has been well documented in the literature on the reference interview (White, 1989) and the online searching interview (Ingwersen, 1992). Various models have been used to explain this phenomenon, including communication models (Eichmann, 1978) and cognitive models of information seeking (Taylor, 1968; Belkin, 1980). Search expertise and the users' mental models of the information system may also play a role in query formulation (Norman, 1986).

This study is part of an ongoing project to better understand the factors at play when people formulate web queries, in order to provide better support for that process. In this study, we analyze interview data of people engaged in web searching to better understand why people use brief queries.

Methods

The study described here was carried out within the parameters of the TREC10 Interactive Track, and a fuller description of the methods is available elsewhere (Toms et al, 2001). In this study, we observed 32 participants search for 4 assigned tasks on the Internet using a slightly modified Google interface. The tasks were from four

domains: research, consumer health, travel and shopping , and half the tasks were left partially open so that participants could personalize them. Additionally, query entry style (Keyword or Sentence) was a condition: we asked participants to use keyword queries for two tasks, and full sentence or question format queries for two tasks.

During the experiment we used a transaction log to record queries and other aspects of search behaviour. Following each task, we conducted a semi-structured interview while re-playing the task using screen capture software. Interviews included questions on how and why queries were formulated and reformulated. The interview data was professionally transcribed and was analyzed using an inductive process of developing codes for the interviews.

Results

The query length varied significantly between conditions. Keyword queries were on average 2.7 terms long, and Sentence queries were 5.7 terms long. The distribution of query length in the two conditions (Figure 1) shows that for Keyword queries, the majority of queries are brief - falling in the 1-3 term range; whereas, Sentence queries are fairly evenly distributed from about 1-10 terms. This is indicative of the fact that even in the Sentence condition many participants chose to enter very brief queries.

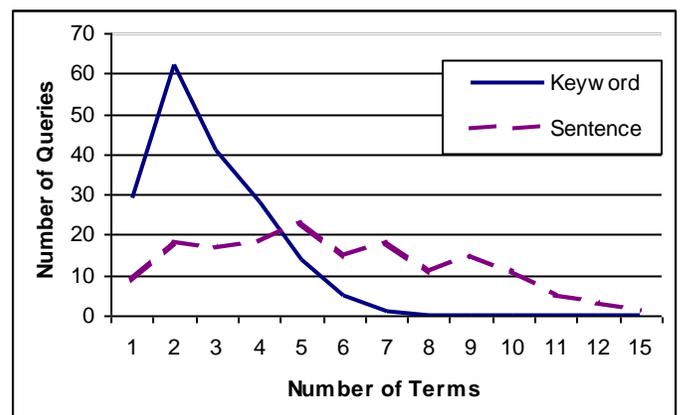


Figure 1: Distribution of Query Length

The interview data suggests that the use of brief and general queries stems from a wide range of reasons:

Queries and Mental Models: Participants seemed to have a strong sense that search engine queries should be brief: “you just get to the point [P09-42]”. Many participants expressed discomfort at being asked to submit queries in sentence format: “I don’t put sentences in these things [P10-21]”. Participants also tended to view the query as a subject classification problem: “what is the base? What is the base topic, and then branch out from there. So the base topic was Titanic, and then the next category would be history [P12-41]”.

Quick Query to Start: The most common response when we asked people why they use brief queries was that they tend to try a quick, short query first, and then expand it if necessary. This first query is often, “...just sort of a general start to see what it would pull up...[P14-13]”. Such queries are considered to be an easy first step that does not require much thought: “deciding what to query is not the challenging part for me. Figuring out how to refine sometimes is....Starting isn’t the problem [P06-33].” General queries also tend to be described as a “safe” approach, in that they do not eliminate too many results: “I think that “second hand smoke,” maybe its too narrow to find in the directory, so I go to “smoke” firstly and try to find something...[P22-11]”.

Query to Portal: Another approach that results in brief queries, is use of the search engine to find known websites or large information portal sites: “when I approach certain topics, I choose to go to a specific website and do a search there, if I think its going to be more efficient..” [P09-24]”. Of the 1 and 2 term queries, about 14% were clearly seeking websites of well-known organizations, such as “Futureshop”, “Lonely Planet” and “British Museum”. This approach was most common for shopping queries, but the interviews revealed that many other brief queries, such as “titanic” and “global warming” were also aimed at finding good “overview” websites, rather than specific information.

Query to browse: Some participants suggested that they used general queries because they did not want or need a highly focused search. These participants seem to prefer to look through a large number of results, rather than relinquish those decisions to the search engine: “actually... I want to decide, I want to sift through everything. Even though it’s overwhelming... [P10-21]”.

Discussion

This study provides further evidence that querying is a complex and interactive process. The use of quick and broad opening queries that are refined based on feedback accords with communication models of information seeking, which emphasize the social aspects of interaction and the establishment of a common understanding (Eichman, 1978). Further, it seems that web queries play different roles than traditional IR queries, which is one

reason for their brevity. In some cases, queries are intended to serve only as coarse filters to define a browsing space or starting point. In other cases, web queries are being used to retrieve specific, known websites. Finally, the interview data suggest that many users have a general sense that brief, subject oriented queries are the “correct” way to interact with search engines. This mental model seems to underlay the search behaviour of the “Google generation”, and will have to be addressed as we seek ways to support more effective query formulation for web based information retrieval.

Acknowledgements

This work was partially funded by an IBM Centre for Advanced Studies Student Fellowship to the first author and a Natural Sciences and Engineering Research Council of Canada grant to the second author.

REFERENCES

- Belkin, N.J. (1980) Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, vol. 5: 133-143.
- Eichman, T. L. (1978). The complex nature of opening reference questions. *RQ*, 212-222.
- Ingwersen, P. (1992). *Information Retrieval Interaction*. London: Taylor Graham.
- Jansen, B. J. & Pooch, U. (2000) A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology* 52, 235-246.
- Nielsen, Jakob. (2001, May 13) Search: visible and simple. Retrieved October 10, 2002 from <http://www.useit.com/alertbox/20010513.html>
- Norman, K. L. (1986). Cognitive engineering. In D. A. Norman, & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 31-61). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Taylor, R. S. (1968). Question negotiation and information seeking in libraries. *College and Research Libraries*, 29, 178-194.
- Toms, E., R. Kopak, J. Bartlett, L.S. Freund, J. Heaton & A.Olsen (2001). Selecting Versus Describing: the Efficacy of Categories in Exploring the Web. *Proceedings of the 10th Text Retrieval Conference*, Gaithersburg, MA, 64-70.
- Wang, P., Berry, M.W. & Yang, Y. (2003). Mining longitudinal web queries: trends and patterns. *Journal of the American Society for Information Science and Technology* 54(8), 743-758.
- White, M. D. (1989). Different approaches to the reference interview. *Reference Librarian*, 25-26, 631-646.