Analysis of An Expert Search Query Log

Yi Fang†, Naveen Somasundaram†, Luo Si†, Jeongwoo Ko‡, Aditya P. Mathur†
† Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA
‡ Google Inc, Mountain View, CA 94043, USA
†{fangy, nsomasun, lsi, apm}@cs.purdue.edu; ‡jko@google.com

ABSTRACT

Expert search has made rapid progress in modeling, algorithms and evaluations in the recent years. However, there is very few work on analyzing how users interact with expert search systems. In this paper, we conduct analysis of an expert search query log. The aim is to understand the special characteristics of expert search usage. To the best of our knowledge, this is one of the earliest work on expert search query log analysis. We find that expert search users generally issue shorter queries, more common queries, and use more advanced search features, with fewer queries in a session, than general Web search users do. This study explores a new research direction in expert search by analyzing and exploiting query logs.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation

General Terms

Experimentation, Measurement

Keywords

Expert Search, Query Log Analysis

1. INTRODUCTION

As an important information retrieval application, expert search (also known as expert finding) has received substantial attention in the IR research community. Rapid progress has been made in modeling, algorithms and evaluations in the recent years [1] since the launch of TREC Enterprise track in 2005 [2], but we have very little understanding about how people actually use the expert search systems. On the other hand, query logs provide an excellent window for gaining valuable knowledge about how a search engine is used and what the users' interests and preferences are since they form a complete record of what users searched for in a given time frame. The past few years have seen an increased interest in the research of query log analysis. However, most of the related work is conducted on general Web search [5, 6] or with a special group of users on Web search such as children search [7]. There is no prior work on query log analysis for expert search. As expert search users have a specialized type of information need, it is desirable to understand their characteristics and to provide them with suitable IR systems. In this study, we collect a user query log of an expert search system and present an analysis of individual queries and sessions. Our findings support the conjecture that expert search users generally issue shorter queries, more common queries, and use more advanced search features, with fewer queries in a session, than general Web search (as well as children search) users do.

2. AN EXPERT SEARCH QUERY LOG

The query log that we analyze is extracted from INDURE[1], which is an expert search system containing over 20,000 researchers from four universities in Indiana [3]. Besides the basic search function, INDURE also provides an advanced search interface[2] where users can refine their query by using filters or restrictions. The query log data is captured over a period of 90 days from October 19, 2010, to January 16, 2011 and it includes a series of requests. A request may consist of a new query or a new result screen for a previously submitted query or a click on one returned result. Each request includes the following fields: visitor IP, requested page, session ID, timestamp, and query terms. In the query log, the 6,434 unique visitors posted 14,503 queries with totally 104,030 requests. Among them, there are 851 empty queries and 2,923 unique queries of which 1,297 are advanced queries. The total number of sessions is 85,215.

3. FINDINGS

The analysis is carried out at the query level and at the session level, respectively. Query length and frequency are considered for the former, and session length and duration for the latter. The results are compared with those of Web search [5] and children search [7]. Table 1 presents the comparison with the statistics in [5] and [7] (other literatures such as [6] shows similar values of the statistics for Web search and children search).

3.1 Analysis of Queries

The number of terms in a query is an indicator of the complexity of the query and the difficulty of the user to express information needs using keywords. Figure 1 (a) shows the frequencies of queries with different length. As shown in Table 1, the average query length here is 1.96 which is lower

---

[1]https://www.indure.org/
[2]https://www.indure.org/search/advanced.cfm
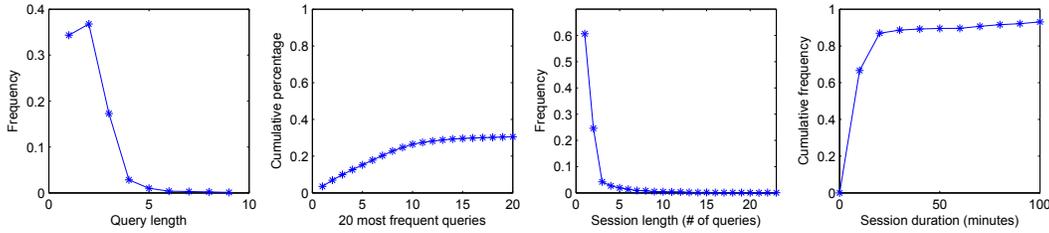
Figure 1: Statistics of the INDURE query log. (a) Query length frequency; (b) Cumulative percentage of frequent queries; (c) Session length frequency; (d) Cumulative percentage of session duration

Table 1: Comparison of statistics among Web search [5], children search [7] and the INDURE expert search. "N/A" means the statistics is not available in the reference

	Web	Children	Expert
Average query length	2.35	3.23	1.96
Average query frequency	3.97	N/A	5.65
Average session length	2.02	8.76	1.81
Average session duration	N/A	20.38	18.67

than that of Web search (2.35 words per query) and much lower than that of search used by children (3.23 words per query). On the other hand, out of 2,923 distinct queries, there are 1,297 advanced queries. This proportion (44.4%) is significantly higher than that in Web search (i.e., 21.4% of the distinct queries with operators [5]). This indicates that expert search users are more willing to use advanced search features to articulate their information needs.

Table 2 lists the 8 most frequent queries. Figure 1 (b) shows the cumulative percentage of the top 20 queries. The top 20 queries represent only 0.68% of all unique queries, yet they account for around 30% of all search queries as shown in Figure 1 (b). The average query frequency is 5.65 in INDURE and is 3.97 in Web search (Table 1). These statistics indicate that expert search users tend to use more common queries than Web search users do.

Table 2: 8 most frequent expert search queries

1) cancer	2) energy
3) electrical engineering	4) nanotechnology
5) engineering education	6) design
7) aerospace	8) biomedical engineering

3.2 Analysis of Sessions

A session is meant to capture a single user's attempt to fill a single information need. Figure 1 (c) shows the session length frequency (represented by the number of queries in a session). We can see that most sessions are short with over 80% of sessions consisting of only one or two queries. As shown in Table 1, the average session length is 1.81, which is shorter than Web search session (2.02) and much shorter than children session (8.76). The long average session length in children search can be explained by the nonlinear navigation style [7]. Considering the average session length in Web search is mostly between 2 and 3 [6], the gap between

expert search and Web search is not insignificant. This may be explained by the fact that users in expert search know more exactly about what they search for and thus it is often easier for them to fulfill their information needs with shorter average session length.

Figure 1 (d) shows the distribution of session duration in minutes. The average session duration is 18.67 minutes, which is shorter than that of children search (20.38 minutes) while the difference is small. This may indicate that users require more time to explore and complete information needs associated to children content [7].

4. CONCLUSIONS AND FUTURE WORK

We present a query log analysis for expert search. To the best of our knowledge, this is one of the first query log analyses targeting on expert search. The findings can help improve user interface design for expert search. As a lot of work has been done on query log analysis for Web search, there is an abundance of research opportunities in the same analysis for expert search. In the future work, we will conduct deeper analysis of the expert search query log. We will also exploit the query log to develop ranking models for expert search by learning from the implicit user feedback [4].

5. ACKNOWLEDGMENT

6. REFERENCES

[1] K. Balog. People search in the enterprise. In *PhD Thesis*. University of Amsterdam, 2008.
[2] N. Craswell, A. de Vries, and I. Soboroff. Overview of the trec-2005 enterprise track. In *TREC*, 2005.
[3] Y. Fang, L. Si, and A. Mathur. Discriminative probabilistic models for expert search in heterogeneous information sources. *Information Retrieval*, 14(2):158–177, 2010.
[4] C. Macdonald and R. White. Usefulness of click-through data in expert search. In *SIGIR*, pages 816–817, 2009.
[5] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large web search engine query log. In *SIGIR Forum*, pages 6–12, 1999.
[6] A. Spink, D. Wolfram, M. Jansen, and T. Saracevic. Searching the web: The public and their queries. *JASIST*, 52(3):226–234, 2001.
[7] S. Torres, D. Hiemstra, and P. Serdyukov. An analysis of queries intended to search information for children. In *Third Symposium on Information Interaction in Context*, pages 235–244. ACM, 2010.