An analysis of querying behaviors of domain knowledgeable users and novice users

Xiangmin Zhang, Wayne State University

Abstract

Users with different amount of domain or topic knowledge tend to find different results, when searching on the same topic. We were interested in finding out how domain or topic knowledge is associated with users' querying behaviors, which would result in different results. In this study, thirty-five students participated in a user experiment, conducted searches from genomics data set that was used by TREC Genomics track 2004. The participants' background domain knowledge and the knowledge of each individual topics was assessed, before they performed the search tasks. Their querying behaviors were recorded. The results demonstrate that domain knowledge level is significantly associated with the number of search terms in queries: the more knowledgeable, the more search terms. Further analyses are planned to be conducted.

1   Introduction

Querying is the critical step in the information searching process. It determines the success or failure of a search. It is generally acknowledged that the experts in a subject domain are more capable of finding relevant documents than a domain novice user is, given the domain expertise the more knowledgeable user has. There has been much research on the difference between novice and experts in terms of information seeking. Previous studies have found that users with higher level of domain knowledge have different search tactics (querying behaviors), performance (result accuracy, etc.), and time spent on task accomplishment and document reading, and so on (e.g., Wildemuth, 2004; White, Dumais, & Teevan, 2009; Zhang, Anghelescu, & Yuan, 2005). Stamou & Ntoulas (2009) discussed possibly different queries among users because of diverse backgrounds. In this research, we were interested in finding out if the two types of users would exhibit different querying behaviors, and therefore hoping to provide evidence for search system designs to design help features for domain novice users.

To investigate querying behaviors, we used the following measures to describe users' querying behavior: the number of words in a query (query length, or average query length over more than one query for a given search topic), and number of queries for a search topic (i.e., per topic session). Further analyses are planned to be conducted on the number of search words in the controlled vocabulary for the domain (MeSH), the specificity of controlled vocabulary in the vocabulary hierarchy, and the similarity of queries between domain knowledgeable users and novice users.

2   Method

Experiment Settings (system and dataset)

A controlled laboratory experiment was conducted, using an experimental search system based on Indri, in the genomics domain, using TREC dataset and topics. Thirty-five student participants (including post-docs) were invited individually to an interaction lab, working on 4 out of 5 topics in assigned task orders. Users were asked to rate their familiarity levels of the selected MeSH terms that were in the topic areas. Questionnaires were used before each task to elicit users' self-judged knowledge on a topic. Users' interaction with the search system was logged.

Domain Knowledge (DK) Assessment

The participant's DK level was calculated as follows. We first calculated the average of the MeSH term ratings for each participant. Then correspondingly the average of the pre-task topic familiarity and expertise ratings were calculated. Because the MeSH ratings and pre-task questionnaire ratings used different scales, the average ratings for MeSH and pre-task questions were transformed into Z scores. These two Z scores for each participant were consequently averaged as DK for each participant:

$$DK\_user=((Z(MeSH-K) + Z((familiarity+expertise)/2))/2$$

1.

3   Results
Descriptive Statistics

The 35 participants' searches on 5 search topics resulted a total of 140 cases or observations. Each participant did 4 of 5 search topics. Because the topic orders were randomized for each individual participant, and a participant's domain knowledge level varied from topic to topic, we first analyzed the 140 cases data as a whole data set, without distinguishing between different topics. The statistics on the data is listed in Table 1 below.

Table 1. Descriptive Statistics of Measures (n=140)

Variables	Mean	Median	SD	Min	Max
DK level	0.000357	-0.1550000	0.781187	-1.37	2.03
Average query length	4.666	4.330	1.74	2.5	14
Number of queries	3.171	3	1.73	1	8

General Results

As presented in Table 1, the participants' DK level scores on all search topics ranged from -1.37 to 2.03. The higher the score, the more knowledgeable the participant was on a particular search topic.

To investigate the relationships between a user's DK level and the query length, and separately the number of queries per search topic session, we paired each of the latter two variables with DK, and conducted Kendall's tau test on the two pairs of data. While no significant association was found between DK and number of queries per search topic, the results showed that DK is significantly associated with the average query length, at p=0.001153.

The results demonstrate that, more knowledgeable about the search topic a user is, the more search terms the user tend to use in a query. This result is consistent with findings from previous studies (e.g., Zhang et al. 2005). However, the number of queries for a search topic did not change with the level of domain knowledge. Intuitively, more knowledgeable users would issue more queries to search on a topic simply because the user knows more aspects of the topic and more terms or words to describe the topic. This study did not find out a significant association between DK and the number of queries issued. One explanation might be that knowledgeable users would not need more queries to find relevant documents. Another reason might be that the experimental time was limited (each user had about 2 and half hours in total), which put a constraint on the participants' behavior.

4   Further Analyses
The result reported in this paper was the first step. Further analyses are planned on the following tasks:
- Because topics were identified as difficult ones and easier ones, analysis for each topic will be conducted to examine the relationships between DK and average query length, as well as the number of queries in each topic
- Based on the DK scores, the participants will be divided into two groups. Those whose scores are above the median (value is -.155) will be put in a high DK group and those below will be in a low DK group. Differences between the groups' querying behaviors will be compared
- Examining the semantics of queries of the two groups: the number of controlled vocabulary terms used in queries, and the breadth and depth of the terms used. Query similarities between different groups' of users will also be calculated, to find out differences.

5   Limitations of the Study

There are various ways to estimate a person's domain knowledge level. The method we used in this study has limitations. We used a selected subject field, the specially implemented experimental search system with just one source of documents (the PUBMED database) that the participants could access. These may have limited the accuracy of the estimation. For example, it has been acknowledged that people with high level domain knowledge tend to visit or access more professional websites, therefore, the user-visited websites can be used as a predictor for the user's domain knowledge level, as demonstrated by White, Dumais, & Teevan (2009). However, this variable was not available in the current study because the participants were given only one system to accomplish their tasks. The application of the method, which used the MeSH terms, is also limited to other non-medical fields where no MeSH terms are available. We hope in the future studies to improve our method.

6   Conclusion

This paper reports the initial results of a study on the relationship between a user's domain knowledge level and the user's querying behaviors. The results demonstrated that the more knowledgeable a user is, the more terms/words the user would use in a query. More data analyses are planned, and the complete results will be reported at a later time.

Acknowledgement

References

Sofia Stamou & Alexandros Ntoulas (2009). Search personalization through query and page topical analysis. User Model User-Adap Inter (2009) 19:5–33

White, R., Dumais, S.T., & Teevan, J. (2009). Characterizing the influence of domain expertise on Web search behavior. Proceedings of WSDM 2009.

Wildemuth, B. (2004). The effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science and Technology, 55(3), 246-258.

Zhang, X., Anghelescu, H. G. B. &Yuan, X. (2005). Domain knowledge, search behavior, and search effectiveness of engineering and science students. Inf. Res., 10(2), 217.