PERSONALISED WEB SEARCH USING USER BEHAVIOUR

Swati

Assistant Professor, Dept of CS, C.R.M. Jat PG college, Hissar, Haryana (India)

ABSTRACT

There are many existing studies of user behavior in simple tasks in a short duration of one to two queries. In this paper we present and empirically analyze a user study using a web search logs and eye tracking to measure user behavior during a query session, that is, a arrangement of user queries, results page views and content page views, in order to find a specific piece of information. From this we are able to identify a number of different behavior patterns for successful and unsuccessful users, and different trends in user activity during the query session. We find that a user behaves differently after the first query formulation, when we compare with the second formulation (both queries being for the same information item). The results can be used to improve the user experience in the query session, by recognizing when the user is displaying one of the patterns we have found to have a low success rate, and offering contextual help at that point. The results may also contribute to improving the design of the results page.

Keywords: Web Search, Query Session, User Behaviour, Searching Experience, Web Mining, Personalised Search.

I. INTRODUCTION

Public domain search engines are one of the most utilized tools on the Internet for accessing online information. However, when users are unable to find what they are looking for, they feel frustrated and this affects their user experience. These reasons, among others, are the motivation for incorporating changes in the search engine's functionality and interface which will enhance user support. With the development of World Wide Web, web search engines have contributed a lot in searching infonl1ation from the web. They help in finding information on the web quick and easy. But there is still room for improvement. Current web search engines do not consider specific needs of user and serve each user equally. It is difficult to let the search engine know what we the user actually wants. Generic search engines are following the "one size fits all" model which is not adaptable to individual users. When different users searching a web contains give same keywords, same result will be returned by an existing search engine, no matter which user submitted the query without knowing its domain knowledge and searching history. This might not be appropriate for users which require different information. While searching for the information from the web, users need information based on their interest. For the same keyword two users might require different piece of information. This fact can be explained as follows: a biologist and a programmer may need information on "mouse" but their fields are is entirely different. Biologist is searching for the "mouse" that is a small mammal and programmer is searching for the computer mouse which is a pointing device. For this type query, a number of documents on distinct topics are returned by generic search engines. Hence it becomes difficult for the user to get the relevant content. Moreover it is also time consuming. Personalized web search is considered as a promising solution to handle these problems, since different search results can be provided depending upon the choice and information needs of users. It exploits user information and search context to learning in which sense a query refer. The main source that search engines have to know what the users are doing in the results and contents pages are weblogs, from which a wide variety of information can be obtained. For example, typical descriptive variables include: the terms that were used to formulate a query, how much time was spent in the Search Engine Results Page, which results were selected, how long the user took to return to the Search Engine Results Page from the content page (if they return), whether the original query was reformulated or another result selected, and how long they took to do this. However, although much work has been done on search engine usage, it has mainly focused on the analysis of large anonymous web search logs, or on specific user studies which analyses the user's reaction to specific items. In this work we offer a novel approach, in which we study the user behavior in the context of a query session whose substructure consists of different Search Engine Results Page views and query formulations in order to find a given piece of information. The data captured and made available for analysis consists of the web search log data together with the data which is obtained by an Eye Tracking device, which enables us to obtain statistics about ocular activity, such as fixation rate and fixation duration.

II. RELATED WORK

Thus, in this paper we derive two aspects of utility for Human-Computer Interaction in Search Engine Results Page design and online help: **(i)** a behavior model which classifies the user with respect to their navigation pattern; **(ii)** key trends of specific search log and eye tracker variables which have been shown to be statistically significant. This paper proposes architecture for constructing user profile and enhances the user profile using background knowledge. This Enhanced User Profile will help the user to retrieve focused information. It can be used for suggesting good Web pages to the user based on his search query and background knowledge.

Our study is related to three areas of existing work: web search user behavior with eye-tracking data; search task and its effects to user behavior; and search session. We review each area below. Early studies of web search user behavior are mostly based on the analysis of large-scale query logs. These studies described what real life web searches are like and how users interact with search engines at the query level, but they do not provide details of user behavior on a Search Engine Result Page, such as how users examine result abstracts. The first work using for web search user behavior discovered how users browse a Search Engine Result Page, examine abstracts, and click results. They found decayed visual attention on results as the rank of the result increases and biased clicks on the top ranked results. Lorigo et al. showed that task type and gender may result in differences in search behavior and browsing style [1]. Later studies with eye-tracking [2, 3, 4, 5] further confirmed that users behave diversely in different tasks. They also showed that users may react distinctly to different outlooks of search result abstracts [6, 7] and a Search Engine Result Page elements other than result abstracts, such as ads and related searches [8, 9]. More recently [10] used eye tracking studies to verify models. However, due to the limited accessibility of devices, user behavior studies with eye-tracking data are limited.

III. EXPERIMENTS

We conducted an experiment to collect user behavior data in search sessions for completing complex tasks. We collected users' queries, their browsing history page navigation rate of Search Engine Result Pages and their clicks of search results. In addition, we deliberately asked users to perform different types of search tasks. We built an experimental search system providing modified Google search results. First, all ads and sponsors' links were removed. Second, we showed 9 results each page (rather than the usual 10) to make sure that users do not need to scroll down to see all of the result items. This change made it much simpler to analyze eye-tracking data. However, previous studies [11] also reported that scrolling down affected browsing patterns on results shown below the screen cutoff of a search result page. This change will miss such effects. We adopted this change because Joachims et al. [11] showed that most of the users' attention is still focused on the top ranked results which are visible before scrolling down. Third, if Google provided query suggestions (i.e., "related searches") for a query (usually shown below the search results), we moved them to the right side of the search results, again, to eliminate scrolling pages. The system looks very similar to existing search engines except a few places specifically desig ed for our search tasks. It shows the task descriptions at the top of the search result page. This is because we found in our pilot study that, without showing the task description, users might constantly switch between search result pages and another page showing the task description, because they forgot details of the task. We believe this would cause greater issues to the collected data (e.g., more constantly switching of pages) than showing task description on the search result page. In addition, the system has a highlighted "finish task" link if the session exceeds the time limit (but not before the limit is reached). Although many systems for user studies (e.g. Liu et al.'s systems [12]) allow users to bookmark relevant results at search time, we did not adopt such settings because it may affect users' browsing behaviors. Instead, relevance judgments were completed after search.

IV. FRAMEWORK FOR PROPOSED SYSTEM

We propose a framework for personalized web search which considers individual's interest into mind and enhances the traditional web search by suggesting the relevant pages of his/her interest. We have proposed a simple and efficient model which ensures good suggestions as well as promises for effective and relevant information retrieval. In addition to this, we have implemented the proposed framework for suggesting relevant web pages to the user.

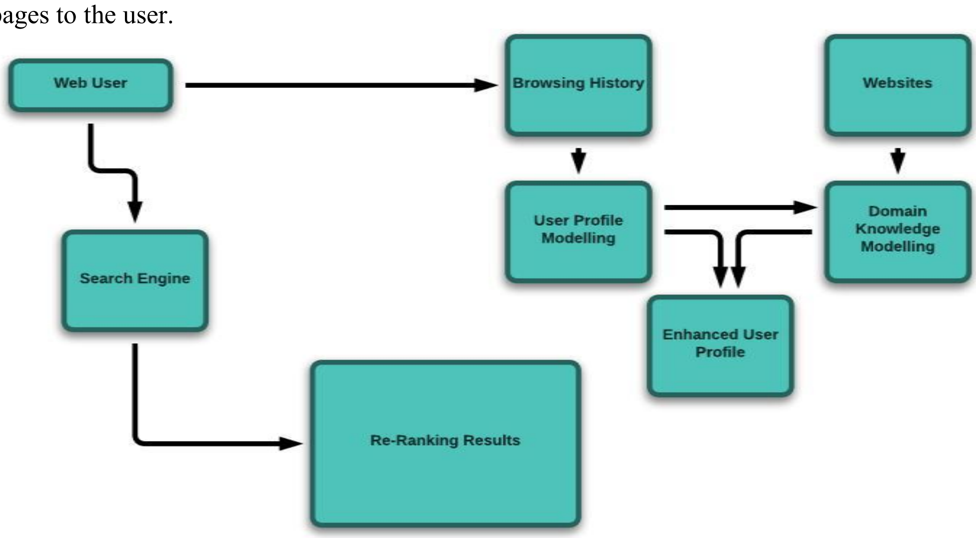

Fig: Architecture of Proposed Framework

Our system considers user's profile (based on user's web log navigation browsing history) and Domain Knowledge in order to perform personalized web search. Information obtained from User Profile is classified into these specified categories. The learning agent learns user's choice automatically through the analysis of user navigation/browsing history, and creates/updates enhanced User Profile conditioning to the user's most recent choice. Once the user inputs query, the system provides good suggestions for personalized web search based on enhanced user profile. Further our model makes good use of the advantages of popular search engines, as it can re-rank the results obtained by the search engine based on the enhanced user profile.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Users interact with a search engine mainly in two ways. First, they proceed through the search process by issuing queries. Second, for each query, they examine result abstracts on the Search engine result page and may click on results. Therefore, we first compare the "search activeness" of users in terms of how frequently they search and how often they examine and click results. This helps us understand the diverse weight of the two types of interactions in different tasks. Then we look into details of search engine result page browsing. Using the information of user browsing history and domain knowledge, we create an Enhanced User Profile. Once the Enhanced User Profile is created, we take the user query and suggest the relevant web pages with respect the query. In our Experiment, we have used User Profile as a base case for suggesting the relevant pages and compared the results with the pages suggested from Enhanced User Profile. For each query, we suggest top 20 relevant documents from User Profile and for the same query we also suggest top 20 relevant documents from Enhanced User Profile. In order to compare the efficiency of the result, we compared the similarity of suggested documents with the user query. We collected user behaviors from 80 search sessions on 20 unique tasks. During a search session, the participants on average issued 4.9 queries, examined 16.1 unique result abstracts, and clicked 9.3 unique results. The average length of query was 3.96 words (without removing stop words). As with most search engines, if the participant clicks a result, the experiment system left the current search engine result page (SERP) and switched to a new tab of the browser showing the result webpage. The participants needed to switch between the SERP and result webpages. This resulted in multiple views for SERP. We refer to the duration from showing a SERP to switching to other webpages as a "SERP view". In our experiment, participants had 3.6 views for a SERP on average.

For each session, the participant on average judged 20.1 results and left 13.3 un-judged. In total this resulted in 992 unique un-judged task-URL pairs. In order to evaluate search performance of sessions, we asked an annotator (not an author of this paper) to assess the relevance of the un-judged results. To evaluate the agreement between the annotator and the participants on relevance judgments, we also sampled 100 unique judged results for the annotator to assess. The annotator was not aware of which result has been judged by the participants. If we merge "highly relevant" and "somewhat relevant" into one class, the annotator agreed with the participants on 77% of the cases.

We further compare the four types of tasks based on the users' clicking behaviors. Whether or not a result is clicked depends on two factors. First, whether the user examined the result abstract or not (though possible, it is very unlikely that uses blindly open a result without examining it). The previous section examined that factor. This section focuses on the second one: after examining a result, what is the chance a user clicks on it? The results show that users do not click every result abstract they examined. The chance of clicking varies by task, by relevance of results, and by whether the result has been visited previously.

IV. CONCLUSION

We have proposed a model of user search behavior which consists of 5 possible navigation patterns, in terms of the route that the users follow from the first results page through to the end of the session. The descriptive variables consist of non-ocular data (web log) together with ocular data (eye tracker). Based on the User Profile and the Domain Knowledge, the system keeps on updating the user profile and thus builds an enhanced user profile. This Enhanced user profile is then used for suggesting relevant web pages to the user. The proposed framework has been implemented by performing some experiments. Our work is significant as it improves the overall search efficiency, catering to the personal interest of the user's. Thus, it may be a small step in the field of personalized web search. In future this framework may be applied for re-ranking the web pages retrieved by search engines on the basis of user priorities. We may also apply collaborative filtering for personalized web search in our framework.

ACKNOWLEDGMENT

We would like to express our appreciation to our parents, all the teachers and lecturers who help us to understand the importance of knowledge and show us the best way to gain it.

REFERENCES

[1]. Lorigo, L. et al. 2006. The influence of task and gender on search and evaluation behavior using Google. Information Processing & Management. 42(4): 1123-1131.

[2]. Buscher, G. et al. 2010. The good, the bad, and the random. In Proc. SIGIR’10: 42-49.

[3]. Cutrell, E. and Guan, Z. 2007. What are you looking for? In Proc. CHI’07: 407-416.

[4]. Dumais, S.T. et al. 2012. Individual differences in gaze patterns for web search. In Proc. IIiX’10: 185-194.

[5]. Thomas, P. et al. 2013. What Users Do: The Eyes Have It. In Proc. AIRS’13: 416–427.

[6]. Clarke, C.L.A. et al. 2007. The influence of caption features on clickthrough patterns in web search. In Proc. SIGIR’07, 135-142.

[7]. Cutrell, E. and Guan, Z. 2007. What are you looking for? In Proc. CHI’07: 407-416.

[8]. Buscher, G. et al. 2010. The good, the bad, and the random. In Proc. SIGIR’10: 42-49.

[9]. Dumais, S.T. et al. 2010. Individual differences in gaze patterns for web search. In Proc. IIiX’10: 185-194.

[10]. Moffat, A. et al. 2013. Users Versus Models: What Observation Tells Us About Effectiveness Metrics. In Proc. CIKM’13.

[11]. Joachims, T. et al. 2005. Accurately interpreting clickthrough data as implicit feedback. In Proc. IGIR’05: 154-161.

[12] Liu, J. et al. 2010. Search behaviors in different task types. In Proc. JCDL’10: 69-78.