Exclusively Your’s: Dynamic Individuate Search by Extending User Profile

Harshit KUMAR,\textsuperscript{1,2} Sanggil KANG\textsuperscript{3}
\textsuperscript{1}University of Suwon, South Korea
\textsuperscript{2}Digital Enterprise Research Institute, National University of Ireland, Galway, IRELAND
\textsuperscript{3}INHA University, 402-751, KOREA
hkumar.arora@gmail.com, sgkang@inha.ac.kr

Received 15 December 2008
Revised manuscript received 5 August 2009

Abstract A universal search engine is unable to provide a personal touch to a user query. To overcome the deficiency of a universal search engine, vertical search engines are used, which return search results from a specific domain. An alternate option is to use a personalized search system. In our endeavor to provide personalized search results, the proposed system, Exclusively Your’s, observes a user browsing behavior and his actions. Based on the observed user behavior, it dynamically constructs user profile which consists of some terms that are related to user’s interest. The constructed profile is later used for query expansion. The goal of research work in this paper is not to provide all the relevant results, but a few high quality personalized search results at the top of ranked list, which in other words means high precision. We performed experiments by personalizing Google, Yahoo, and Naver (widely used search engine in Korea). The results show that using Exclusively Your’s, a search engine yields significant improvement. We also compared the user profile constructed by the proposed approach with other similar personalization approaches; the results show a marginal increase in precision.

Keywords: Search Engine, Personalization, Summarization, Weighted Index, Anchor Text, Hyperlink.

\textsuperscript{*} A part of this research work was carried out while the first author was with DERI, National University of Ireland.
§1 Introduction

The first step in the direction of search engine development was taken by Gerard Salton; he, along with his team members, developed the SMART information retrieval system. The first pre web search engine was Archie, which allowed searching for file names of a database. The search engines developed in the initial stage retrieved results from their indexed database and displayed the cached pages based on keyword match and similarity measures. Traditional indexing methods worked quite well for database or structured information, but later it was discovered that they are not sufficient enough for indexing unstructured information such as World Wide Web. The search engines for searching WWW were based on simple indexing technologies, for instance Lycos, Alta Vista etc. Google made an innovative page ranking system which revolutionized the use of search engines. Page rank used the citation graph of the web along with introduction of link analysis in the search engine systems.

The latest statistics at searchenginewatch.com shows that the Google is the frontrunner with 52% number of searches followed by Yahoo (21%) and msn (13%). In another article, it is reported that Google search engine satisfaction ratings has dropped by 3.7% whereas yahoo search engine ratings has increased by 3.9%. These results show that the users seem to be discontented with the search result which is due to the polysemous nature of queries. By polysemy, we mean a word has different meanings in different contexts, hence it is difficult for a search engine to judge what the user intent is by the query he/she types. A search engine returns the most relevant results matching a user query. For instance, if a user requests a query ‘window,’ he will surely receive results related to ‘windows operating system’ and not about ‘windows’ and ‘doors.’ This is due to the nature of queries. There are two aspects associated with a user query: one is user interest and the other is query intent. In this paper, we have tried to capture user interest. And our experiments suggest that to get much better results, a search engine should be able to judge query intent. Judging the query intent will bring in semantic aspects i.e. annotation and classification. There are some evidences of use of ontology for personalization.

To improve the quality of search results returned by a search engine, many solutions have been proposed: first is to use a Vertical Search Engine for specific information needs, second is the use of a personalized search engine, and third is to introduce innovative algorithms. There is definitely a limit up to which a search engine results can be improved. After that, the only solution left is to personalize search results. In, it has been reported that although search engines provide the best possible result set, they are not satisfied at individual user level. Using Topic Distillation and ARC, Chakrabarti et al. made a good attempt to enhance the quality of search results by extending Kleinberg work on authorities and hubs. Another such attempt was made by Haveliwala, who used hub vectors limited to 16 for calculation of topic sensitive page rank. These approaches improve search results, but they will not provide different results for different users. Using a Vertical Search Engine is not appropriate in all the cases as they have an inherent limitation; they are