

An Overview Study of Personalized Web Search

Chanchala Joshi*, Teena Jaiswal**, Himanshu Gaur***

* Institute of Computer Science, Vikram University, MP, India

** Govt Kalidas Girls College, Vikram University, MP, India

*** Department of Civil Engineering, Dire Dawa University, Ethiopia

Abstract- Personalized web search is one of the growing concepts in the web technologies. Personalization of web search is to carry out retrieval for each user incorporating his/her interests. For a given query, a personalized Web search can provide different search results for different users or organize search results differently for each user, based upon their interests, preferences, and information needs. There are many personalized web search algorithms for analyzing the user interests and producing the outcome quickly; User profiling, Hyperlink Analysis, Content Analysis and collaborative web search are some of the instances for that kind of algorithms. In this paper we are analyzing various issues of personalized web search.

Index Terms- Personalized Web Search, Information Retrieval, User profile.

I. INTRODUCTION

Current web search engines are built to serve all users, independent of the special needs of any individual user. With the exponential growth of the available information on the World Wide Web, a traditional search engine, even if based on sophisticated document indexing algorithms, has difficulty meeting efficiency and effectiveness performance demanded by users searching for relevant information. Personalization of web search is to carry out retrieval for each user incorporating his/her interests. Personalized web search differs from generic web search, which returns identical results to all users for identical queries, regardless of varied user interests and information needs. When queries are issued to search engine, most return the same results to users. In fact, the vast majority of queries to search engines are short and ambiguous. Different users may have completely different information needs and goals when using precisely the same query. For example, a biologist may query “mouse” to get information about rodents, while programmers may use the same query to find information about computer peripherals. When such a query is issued, search engines will return a list of documents that mix different topics. It takes time for a user to choose which information he/she wants. The concept behind personalized search is that by knowing some things about user, a search engine might refine user results to make them more relevant. A teenager searching for music might get different matches than a senior citizen. A man looking for flowers might see different listings than a woman.

II. APPROACHES TO PERSONALIZE SEARCH RESULTS

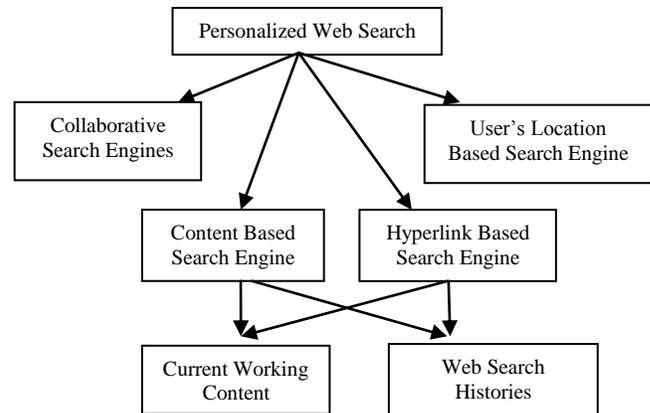


Figure 1: Personalized Web Search Approach

Collaborative Search Engines (CSEs) are an emerging trend for Web search and Enterprise search within company intranets. CSEs let users concert their efforts in information retrieval (IR) activities, share information resources collaboratively using knowledge tags, and allow experts to guide less experienced people through their searches. Collaboration partners do so by providing query terms, collective tagging, adding comments or opinions, rating search results, and links clicked of former (successful) IR activities to users having the same or a related information need. Examples of CSEs are <http://www.scuup.net/>, <http://isedb.com>, and <http://www.eurekster.com>.

Personalized web search can be achieved by checking content similarity between web pages and user profiles. Some work has represented user interests with topical categories. User's topical interests are either explicitly specified by users themselves, or can be automatically learned by classifying implicit user data. Search results are filtered or re-ranked by checking the similarity of topics between search results and user profiles.

Most generic web search approaches rank importance of documents based on the linkage structure of the web. An intuitive approach of personalized web search is to adapt these algorithms to compute personalized importance of documents. A large group of these works focuses on personalized PageRank. The fundamental motivation underlying PageRank is the recursive notion that important pages are those linked-to by many important pages.

There is a huge amount of searches on the Web has local intent, meaning that they are searching for things in a particular

area, like restaurant, job listings, shopping centre, etc. This kind of queries is considered as implicit local intent queries. For such queries, users expect personalized search results that are customized to their locations. Thus, identifying implicit local intent queries and finding out the location information for the users are particularly useful to improve user search experience.

III. BASICS OF PERSONALIZED SEARCH

A. *Creation of User Profile*

To provide personalized search results to users, personalized web search maintains a user profile for each individual. A user profile stores information about user interests and preferences. It is generated and updated by exploiting user-related information. Such information may include:

- Information about the user like age, gender, education, language, country, address, interest areas, and other information.
- Search history, including previous queries and clicked documents.
- Other user documents, such as bookmarks, favorite web sites, visited pages, and emails.

B. *Server-Side and Client-Side Implement*

Personalized web search can be implemented on either server side (in the search engine) or client side (in the user's computer or a personalization agent) [1]. For server-side personalization, user profiles are built, updated, and stored on the search engine side. User information is directly incorporated into the ranking process, or is used to help process initial search results. The advantage of this architecture is that the search engine can use all of its resources, for example link structure of the whole web, in its personalization algorithm. Also, the personalization algorithm can be easily adapted without any client efforts. This architecture is adopted by some general search engines such as Google Personalized Search. The disadvantage of this architecture is that it brings high storage and computation costs when millions of users are using the search engine, and it also raises privacy concerns when information about users is stored on the server. For client-side personalization, user information is collected and stored on the client side (in the user's computer or a personalization agent), usually by installing a client software or plug-in on a user's computer. In client side, not only the user's search behavior but also his contextual activities (e.g., web pages viewed before) and personal information (e.g., emails, documents, and bookmarks) could be incorporated into the user profile [2]. This allows the construction of a much richer user model for personalization. Privacy concerns are also reduced since the user profile is strictly stored and used on the client side. Another benefit is that the overhead in computation and storage for personalization can be distributed among the clients. A main drawback of personalization on the client side is that the personalization algorithm cannot use some knowledge that is only available on the server side (e.g., PageRank score of a result document). Furthermore, due to the limits of network bandwidth, the client can usually only process limited top results.

IV. CONTENT V/S HYPERLINK BASED ANALYSIS

A. *Content Based Personalized Search*

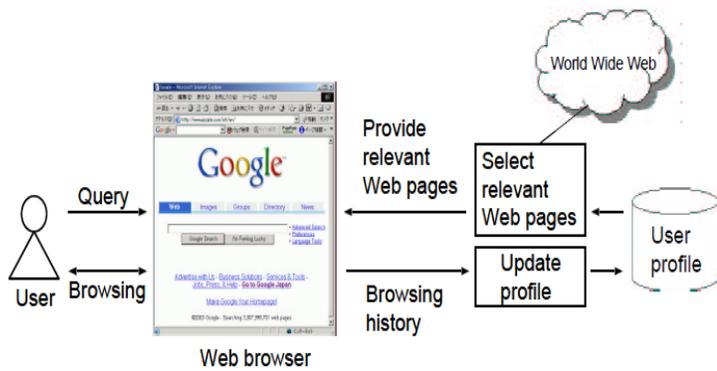
By checking content similarities between web pages and user profile personalized search can be improved [3]. User's interests can be automatically learned by classifying implicit user data. Search results are filtered or re-ranked by checking the similarity of topics between search results and user profiles. User-issued queries and user-selected documents are categorized into concept hierarchies that are accumulated to generate a user profile. When the user issues a query, each returned result is also classified. The documents are re-ranked based upon how well the document categories match user interest profiles. Chirita et al. [4] use the ODP (Open Directory Project, <http://www.dmoz.org/>) hierarchy to implement personalized search. User favorite topics nodes are manually specified in the ODP hierarchy. Each document is categorized into one or several topic nodes in the same ODP hierarchy. The distances between the user topic nodes and the document topic nodes are then used to re-rank search results.

B. *Hyperlink Based Personalized Search*

Hyperlink Analysis significantly improves the relevance of the web search results so that all major search engines claim to use some type of hyperlink analysis. Web information retrieval mainly focuses on hyperlink structures of the Web, like with Web search engine Google. In personalized Web searches, the hyperlink structures of the Web are also becoming important. The use of personalized PageRank to enable personalized Web searches was first proposed in [6], where it was suggested as a modification of the global PageRank algorithm, which computes a universal notion of importance of a Web page. The computation of (personalized) PageRank scores was not addressed beyond the original algorithm. Experiments[7] concluded that the use of personalized PageRank scores can improve a Web search.

Crawling (process of gathering the web pages by the search engine) and ranking are the main uses of hyperlink analysis. In this approach, web crawler which is a software program to browse WWW in automated methodical manner find more and more web pages linked to the source page with the assumption of nearly all the linked web pages are on same topic. This process repeats for each set of web pages until no more linked pages. Then crawler of the search engine orders the web pages by the quality. To judge high quality pages, hyperlink analysis is used. In this case, search engine assumes that the source page pointed to many pages is of higher quality than a source page pointed to few numbers of pages. For this ranking process, hyperlink analysis involves with connectivity-based ranking, PageRank and HITS (Hyperlink-Induced Topic Search) algorithms etc.

In addition to produce a quality and relevant web results, hyperlink analysis have several advantages like finding mirrored hosts, web page categorization and identify the geographical scope of the search etc. But in this approach, search engine has to deal with more details consist even with unnecessary stuffs also. It becomes wastage of the resources.



V. TECHNOLOGIES AND ENVIRONMENT FOR PERSONALIZED WEB SEARCH

A. 5.1 Web Search engine technology

The main purpose of search engine is that searching web resources from Internet and present a list of them to the user. Web crawling is one of the most important operations of the search engine. Web crawler follows the resources of WWW in an automated way or orderly fashion. It copies the all the visited pages for searching rapidly in future. Another functionality of search engine is indexing which collects and stores data to optimize the speed of information retrieval for a given a search query. Most of search engines support full-text, natural language data, audio, video and graphics also.

B. 5.2 PageRank

In 1998, Larry Page and Sergey Brin who were the founders of Google introduced a new linking analysis method named as PageRank. PageRank is a probabilistic distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page [9]. Main advantage of this PageRank analysis is usability for collections of documents of any size. One of the main goal of PageRank is to improve the quality and scalability of search. Google makes efficient use of storage space to store the index. This allows the quality of the search to scale effectively to the size of the Web as it grows.

VI. CONCLUSION

There are number of researches which are conducting to minimize some of the drawbacks of Personalized Web Search.

Hyperlink based personalization algorithms work only for repeated queries, they are simple and stable. The topical interest-based personalized search algorithms implemented were not as stable as the click-based. They could improve search accuracy for some queries, but they harmed performance for more queries. Personalized Web search yields significant improvements over generic Web search for queries with high click entropy. For the queries with low click entropy, personalization methods performed similarly or even worse than generic search. As

personalized search had different effectiveness for different kinds of queries, queries should not be handled in the same manner with regard to personalization. No personalization algorithms can outperform others for all queries. Different methods have different strengths and weaknesses.

The violation of privacy is also a well know issue in Personalized Web Search approach. It generates ethical and security problems. Another limitation in this method is that users' needs are not static, it changes continuously. As well as there are some occasions, users do searching for others needs also. So there is problem to search engines to distinguish these scenarios.

Some of experiments consider doing Personalized Web Search by considering geographical location of the users and use social tagging to query expansion by using social networks and collaborative tagging system.

REFERENCES

- [1] Ji-Rong Wen, Zhicheng Dou, Ruihus Song, "Personalized Web Search", Microsoft Research Asia, Beijing, China, 2009
- [2] P.A. Chirita, C. Firan, and W. Nejdl, "Summarizing Local Context to Personalize Global Web Search", Proc. ACM Int'l Conf. Infor. and Knowledge Management (CIKM), 2006.
- [3] Zhongming Ma, Gautam Pant and Olivia R. Liu Sheng, "Interest-Based Personalized Search," ACM Transactions on Information Systems, Vol 25, Issue 2, Article 5, February 2007.
- [4] Chirita P.A., Nejdl W., Paiu R., and Kohlschütter C. "Using ODP metadata to personalize search". In Proc. 31st Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 2005, pp. 178-185.
- [5] Aktas M. S., M. A. Nacar and F. Menczer, 2004. "Using hyperlink features to personalize web search". Adv. Web Mining Web Usage Anal., 3932: 104-115.
- [6] Page L., Brin S., Motwani R., and Winograd T. "The Pagerank Citation Ranking: bringing order to the Web", Technical report, Stanford Dig. Lib. Tech. Project, 1998.
- [7] K. Sugiyama, K. Hatano, and M. Yoshikawa. "Adaptive web search based on user profile constructed without any effort from users". In Proceedings of the 13th international conference on World Wide Web, pages 675-684, 2004.
- [8] Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Computer Networks and ISDN Systems Journal, 30, pages 107-117, 1998.

AUTHORS

First Author – Chanchala Joshi, Master of Philosophy in Computer Science, Institute of Computer Science, Vikram University, Ujjain, M.P. India, chanchala.joshi@gmail.com

Second Author – Teena Jaiswal, Master of Science in Computer Science, Govt. Kalidas Girls College, Ujjain, M.P. India, teenajaiswal99@gmail.com

Third Author – Himanshu Gaur, Master of Technology in Construction Management, DireDawa University, Ethiopia, himanshugaur82@gmail.com.