

# A Survey on Web Search Results Personalization

<sup>1</sup>Blessy Thomas, <sup>2</sup>Jilson P Jose

PG Scholar, Asst. Professor

Department of CSE

Vimal Jyothi Engineering College

**Abstract:** Web is a huge information repository covering almost every topic, in which a human user could be interested. As the size and richness of information on the web increases, diversity and complexity of the tasks users tries to perform also increases. With the overwhelming volume of information on the web, the task of finding relevant information related to a specific query or topic is becoming increasingly difficult. Now a day's commonly used task on internet is web search. User gets variety of related information for their queries. To provide more relevant and effective results to user, Personalization technique is used. Personalized web search refer to search information that is tailored specifically to a person's interests by incorporating information about query provided. Two general types of approaches to personalizing search results are modifying user's query and re-ranking search results. Several personalized web search techniques based on web contents, web link structure, browsing history, user profiles and user queries. This paper is to represent survey on various techniques of personalization

**Keywords:** Data mining, Information retrieval, Personalized web search, Search queries, User profile, User model

## I. INTRODUCTION

Data Mining is the process of extracting information from large data set and transform into data set which is understandable and useful. Information retrieval is the techniques of obtaining information relevant to an information needed from a collection of information resources .World Wide Web (WWW) is largest , commonly used and most accessible source of information. Day-by-day the web pages on Internet are growing rapidly. Web structures are large as well as sophisticated and users often miss the goal of specified queries or receive ambiguous and sometimes unwanted results when they try to navigate through them. The search engines finds out the relevant web pages according to the query specified by user. The web search engine is become very popular and important way to search useful information on the internet .While searching users might experiences failure when search engines returns unwanted as well as irrelevant results that do not meet their searched query expectation. Such type of irrelevance of results is largely because of the enormous variety of user's contexts and backgrounds, the ambiguity of texts and some type of confusion of queries Data Mining is the process of extracting information from large data set and transform into data set which is understandable and useful .Information retrieval is the techniques of obtaining information relevant to an information needed from a collection of information resources .World Wide Web (WWW) is largest , commonly used and most accessible source of information. Day-by-day the web pages on Internet are growing rapidly. Web structures are large as

well as sophisticated and users often miss the goal of specified queries or receive ambiguous and sometimes unwanted results when they try to navigate through them. The search engines finds out the relevant web pages according to the query specified by user. The web search engine is become very popular and important way to search useful information on the internet .While searching users might experiences failure when search engines returns unwanted as well as irrelevant results that do not meet their searched query expectation. Such type of irrelevance of results is largely because of the enormous variety of user's contexts and backgrounds, the ambiguity of texts and some type of confusion of queries.

Personalized web search (PWS) is a one of the category of search techniques which provides better search results and the results which are tailored for individual user needs. User specified information has to be collected and analyzed to find out the user intention and goals behind the issued query. Personalized Web Search solutions can be categorized into two types; one is click-log-based methods and other profile-based ones. The click log based methods are based on simply impose bias to clicked pages in the user's query history. Profile based method improve search experiences with complicated user-interest models which are generated from user profiling techniques. The profile-based PWS is more effective in improving the quality of web search. Now-a-days search with increasing usage of personal information to profile its users, which is gathered implicitly from query history, browsing history, click-through data, bookmarks, user documents, and so on.

## II. 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
- Search history
- 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. 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.

## III. REVIEW

### A. Personalized Search Based on User Search Histories

M Speretta and S Gauch et al.,[2] explores the use of a less-invasive means to gather user information for personalized web search. User profiles, descriptions of user interests, can be used by search engines to provide personalized search results. Many approaches to creating user profiles capture user information through proxy servers (to capture browsing histories) or desktop bots (to capture all activities on a personal computer). These both require participation of the user to install the proxy server or the bot. Particularly build user profiles based on activity at the search site and study the use of these profiles to provide personalized search results. User profiles are created by classifying the collected information into concepts in a reference concept hierarchy and then these profiles are used to re-rank the search results and the rank-order of the user-examined results before and after re-ranking are compared. Finds that personalized re-ranking results in a 34% improvement in the rank order of the user-selected results.

### B. Personalized Mobile Search Engine

Leung et al. [3] proposed a personalization method for mobile search engines (PMSE) that capture the user preferences in the form of concepts by mining their clickthrough data. Due to the importance of location information in mobile search, PMSE classifies these concepts into content concepts and location concepts. In this work users' locations (positioned by Global Positioning System, GPS) are used to supplement the location concepts in PMSE. The user preferences are organized in an ontology-based, multifaceted user profile, which are used to adapt a personalized ranking function for rank adaptation of future search results. To characterize the diversity of the concepts associated with a query and their relevance to the users need, four entropies are introduced to balance the weights between the content and location facets. Based on the client-server model, this paper presented a detailed architecture and design for implementation of PMSE. In this design, the client collects and stores the click-through data locally to protect privacy, whereas heavy tasks such as concept extraction, training, and re-ranking are performed at the PMSE server.

This paper has studied the unique characteristics of content and location concepts, and provided a coherent strategy using client-server architecture to integrate them into a uniform solution for the mobile environment. User profiles are created using content concepts and location concepts and the search result re-ranking is done based on user profiles.

### C. Mining the Search Engine Query Logs

Sharma A.K et al. [4], proposed a novel result optimization technique based on learning from historical query logs, which predicts users' information needs and reduces their navigation time within the result list. The method first performs query clustering in query logs based on a similarity function and then captures the sequential patterns of clicked web pages in each cluster using a sequential pattern mining algorithm. Finally, search result

list is re-ranked by updating the existing Page Rank values of pages using the discovered sequential patterns.

#### D. Query Expansion based on User Interactions

Hang et al. [5], proposed a new method for query expansion based on user interactions recorded in user logs. The central idea is to extract correlations between query terms and document terms by analyzing user logs. These correlations are then used to select high-quality expansion terms for new queries. Queries to search engines on the web are usually short. They do not provide sufficient information for an effective selection of relevant documents. As a consequence, the documents returned by search engines are not relevant to the user information need. To solve this problem, query expansion techniques are used, because query expansion helps the user to formulate better queries. Query expansion involves supplementing the original query with additional words and phrases. The two key aspects in any query expansion techniques are the source from which expansion terms are selected and the method to weight and integrate expansion terms. A query expansion method based on global analysis usually builds a thesaurus to assist user to reformulate the queries. A thesaurus can be automatically established by analyzing relationships among documents and statistics of term co-occurrences in the documents. From the thesaurus constructed in this way, one will be able to obtain synonyms or related terms for the given user query. Thus, these related terms can be used for supplementing users' original queries. This work contributed a new query expansion method based on user logs which record user interactions with the search systems. User logs are exploited so as to extract implicit relevance judgments they encode. In this approach, the researchers assumed that the documents that the user chooses to read are "relevant documents".

The log-based query expansion overcomes several difficulties of local analysis because can extract a large number of user judgments from user logs, while eliminating the step of manually collecting feedbacks from users. Probabilistic correlations between terms in the user queries and the documents can then be established through user logs. With these term-term correlations, relevant expansion terms can be selected from the documents for a query.

#### IV. CONCLUSION

The remarkable development of information on the web has forced new challenges for the construction of effective search engines. The only input to the search engine is the keyword and it searches the whole WWW to provide the relevant information to the users. However, most of users are likely to use only a few keywords to convey their information requirements, and thus the search queries usually do not correspond to what the users want specifically. In addition, with the huge development of the information presented on the web, it is very complicated for web search engines to satisfy the user information requirement only with a short ambiguous query. To overcome such a basic difficulty of information retrieval, personalized search, to provide the customized search results to each user, is a very promising solution.

#### V. REFERENCE

- [1] Ji-Rong Wen, Zhicheng Dou, Ruihus Song, "Personalized Web Search", Microsoft Research Asia, Beijing, China, 2009.
- [2] M Speretta and S Gauch , "Personalized Search Based on User Search Histories", Proceeding Of International Conference on Web Intelligence, pp. 622-628,2005.
- [3] K. W. Leung, D.L.Lee, and W. Lee, Pmse: A personalized mobile search engine, IEEE Transactions on Knowledge and Data Engineering, vol. 25, 2013.
- [4] Sharma A.K, Aggarwal, and Neelam Gupta, Web search result optimization by mining the search engine query logs, Methods and Models in Computer Science (ICM2CS), 2010 International Conference on, IEEE, 2010, pp. 39–45.
- [5] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma, Query expansion by mining user logs, IEEE Transactions on Knowledge and Data Engineering, vol. 15, 2003.
- [6] J. Lai and B. Soh, "Personalized Web search results with Profile comparisons," Third International Conference on Information Technology and Applications (ICITA 2005), Vol. 1, Pp. 573 – 576,2005.
- [7] B. Smyth, "A Community-Based Approach to Personalizing Web Search," IEEE Journals, Computer, Vol. 40, No. 8, Pp. 42 –50, 2007.
- [8] M. Kutub, R. Prachetaa and M. Bedekar, "User Web Search Behaviour," 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET), Pp. 549 –554, 2010.
- [9] Xuwei Pan, Zhengcheng Wang and Xinjian Gu, "Context-Based Adaptive Personalized Web Search for Improving Information Retrieval Effectiveness," International Conference on Wireless Communications, Networking and Mobile Computing, Pp. 5427 – 5430, 2007.