Personalized Mobile Web Search Techniques

Disha Gupta, Nekita Chavhan

Abstract—Volume of information available on web is increasing day by day. People frequently explore web for their various information needs, using web search engines. The recent advancements in the mobile technology have boosted searching and browsing tasks on mobile platforms also. However, the ambiguity in queries and presence of noise hampers the performance of traditional search engines. They provide similar results to all users, irrespective of their context. Thus users require additional effort and time for getting relevant results. In addition to this, the constraints associated with mobile devices pose new challenges for mobile search. Approach es and issues of existing search techniques have been studied in this paper. A framework, incorporating user preferences, has been proposed for mobile search engine personalization.

Index Terms— Search engine, user preferences, personalization, click-through data, ranking, relevancy, user profile.

1 INTRODUCTION

INITIALLY, the World Web was defined as a complete directory of existing servers in the web. It contained a list of webservers, which was edited by Tim Berners-Lee and hosted on the European Council for Nuclear Research (CERN) webserver [1]. As the number of online webservers began to rise, this central list was unable to manage the entire web. Later, other web directories started to appear, having hierarchy of web pages based on their topics. Since these web directories were human-edited, their maintenance became very difficult with fast growth of web. As a result, the information retrieval techniques were put into practice in the web. This was the time when search engines first came into existence.

Today, the search engines serve as a tool to present well-organized search results for user queries from different domains. Web users often explore the digital information resources on web using these tools. As we're getting completely immersed into the Internet era, search engines related research issues are gaining more and more attention. As shown in Fig 1, Search Engine Optimization (SEO) and Search Engine Personalization (SEP) are two such main research areas which affects the results that are displayed by a search engine.

SEO is a methodology consisting of strategies, techniques and tactics used to obtain a high-ranking placement in the search engine's results page (SERP). The main motive behind SEO adoption is the higher a website naturally ranks in results of a search, the greater the chance that that site will be visited by a user. This becomes important for companies for attracting more users on their own website. SEO is done 'for' the search engine.

As a consequence of rapid and continuous expansion of web, extraction of relevant information has become a major challenge [2]. The search engines are often unable to satisfy varying user information needs. Additionally, the noise web documents and ambiguous terms also burden their performance [3]. Incapability of anticipating user's information needs leads to display same results to all, irrespective of their context. SEP is application of personalization process to search engines for relevant information retrieval. Personalization is the ability to provide content and services tailored to individuals, based on knowledge about their preferences and behavior [4].

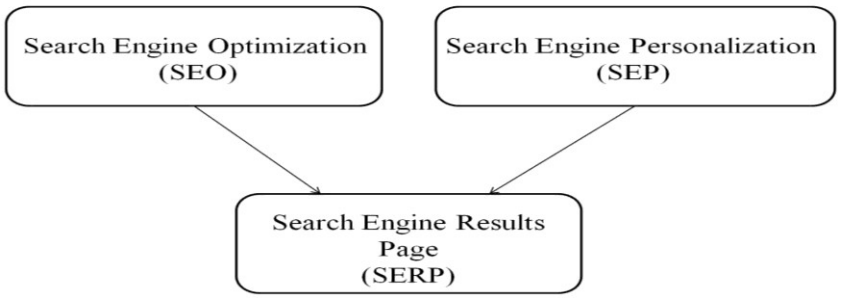

Fig. 1. Research areas in field of Search Engines

- *Disha Gupta, Research Scholar, Department of Computer Science and Engineering, G.H. Raisoni College of Engineering.Nagpur, India. E-mail: disha.g14@gmail.com*
- *Nekita Chavhan, Department of Computer Science and Engineering, G.H. Raisoni College of Engineering.Nagpur, India. E-mail: niki.chavan@gmail.com*

In the past two decades, there has also been a vast advancement in the field of mobile devices [5]. The implementation of web services on a portable platform such as mobile has further enabled the people to instantly search the web, irrespective of physical location, leading to increased use of mobile devices for searching and browsing tasks. Besides the issues with web search, mobile search needs to deal with new challenges due to resource constraints such as limited input modes, comparatively smaller display space, and slow network speeds. These limitations presents the need for mobile search engine personalization, to meet different search goals of users' depending on their information needs.

This paper is organized as follows: Section II describes the process flow of search engine. An overview of existing search strategies is presented in Section III. Adaptive re-ranking approach for mobile web search personalization is discussed in Section IV. Finally, Section V concludes the paper.

2 Process Flow of Search Engine

Working of a typical search engine is illustrated in Fig 2.It shows the flow graph for a searched query by a web user [6]. The process flow of search engines consists of three major functions; Crawling, Indexing and Ranking.

- Crawling: The web crawlers are programs that search the web for new web pages.
- Indexing: These web pages are indexed and stored in search engine databases.
- Ranking: When a user makes any query to the search engine it displays the all the indexed pages for that particular keyword as per some pre-defined ranking scheme. Hence, though different users submit same query with different intentions they get to view the results in same order.

3 Overview of Exsisting Serach Strategies

With the growing information on web, even the simple search queries are accompanied with large volume of results and scanning through all the results is time-consuming for the users. Often, as it has been shown by some studies [7], most of users refer only to first few pages of the result. Hence, the ranking methodologies play a vital role in search engine performance, as it determines the order of importance for the retrieved search results.

The search strategy incorporated in a search engine is usually categorized by the ranking technique used by it.

Web mining is application of data mining techniques to extract useful information from Web data. It can be broadly divided into three distinct categories, according to the kinds of data to be mined i.e content mining, structure mining and usage mining. Exploiting these categories, the web search engines employ different ranking algorithms which are broadly classified as content-based, structure-based and usage-based techniques respectively as shown in Fig. 3.

PageRank and Hypertext Induced Topic Search (HITS) are two well-known link analyses ranking algorithms. Brin and Page proposed the PageRank [8] algorithm based on the observation that not all the links, pointed to a document, have equal importance. This used by the Google search engine. Its ranks the hyperlinked set of documents, assigning a numerical weight or rank to its each page, to measure its relative importance within a set for particular query. If the sum of the back links to a document is large then it is allotted a higher rank. The algorithm is incapable of reflecting user's view of importance of results. HITS [9] proposed by Kleinberg, presented a different way to exploit the hyperlink structure of the documents. While PageRank computed the page ranks on the entire web graph, the HITS algorithm tried to distinguish between hubs and authorities within a subgraph of relevant pages. But since, it is query dependent, current search engines find it infeasible to use for handling their millions of queries. The link-structure based ranking algorithms were found to provide inadequate satisfaction for user's varying information needs. This lead towards the need for developing new personalized search approach.

Acquiring user's real-time information requirements is the key issue for personalized search. The search query submitted by the user to the search engine, servers as an important source for this purpose. Clearly specified queries help to satisfy user search goals to great extent. However, since many of the mobile devices tend to have constrained display abilities, queries are generally seen to have characteristics of shortness, ambiguousness and incompleteness.

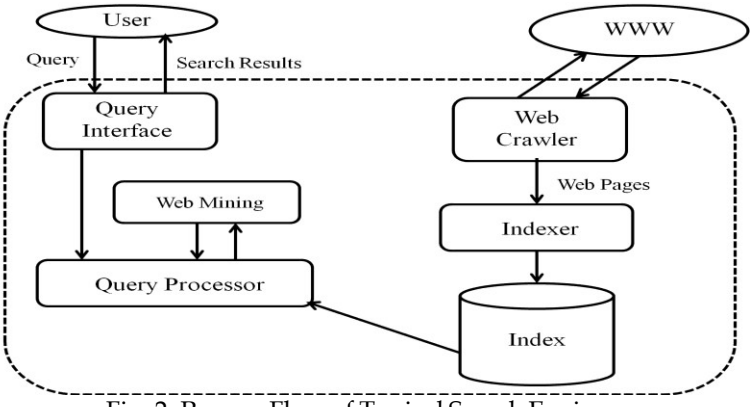

Fig. 2. Process Flow of Typical Search Engine

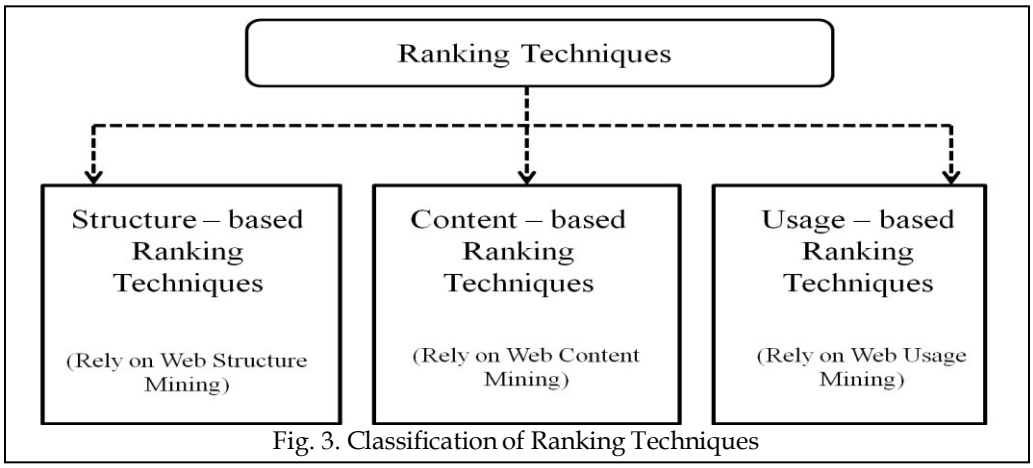

Fig. 3. Classification of Ranking Techniques

As the search query is the first source to evaluate user's information requirement, the above mentioned characteristics highly influence the quality of results provided by the search system. Thus, designing a search system which achieves user's requirement only from the query proves to be insufficient for personalization.

Effective personalization requires understanding different user search goals, especially in case of ambiguity. Ambiguous query terms have many different meaning in different contexts and the anticipated context can be determined by the user alone. Personalized search systems were built upon user feedback approach. Some systems [10] [11] relied on the explicitly specified preferences and interests of the users, to provide them, intended search results. However, due to the requirement of the additional effort and time, the users are generally unwilling to specify their interest information explicitly.

Recently developed approaches, focused on deriving user's interest information implicitly, with no extra effort from the user. These approaches combined the content and usage mining techniques to provide relevant results depending on the users search goals. There are huge amounts of search and browse log data being gathered everyday at the search engines. Mining these logs provides various ways to collect user requirements for enhancing the search process [12]. Search logs are responsible for recording the all interaction details between the user and the search engine. These details usually incorporate the information about the user, query, clicked URLs, results returned by the search engine for a query. Besides this, browse logs are collected using client-side browser plug-ins. These include user's interaction details with web pages other than the search result page. Browse logs gives a broader view of user's search and browse behavior as compared to search logs alone. Although they contain the data about the browsed URLs but the results returned by search engines are not included here. The volume of logs and the presence of noise in the logs pose major limitations for utilizing them in personalization process. However, to overcome these, only certain parts of the logs are processed and considered for analysis, but this pre-processing is also a challenging task.

Some personalized systems, [13], [14], [15], [16], [17] [18] focused on learning user preferences by mining only the implicit clickthrough data from logs. Preferences extracted from click-through data are classified either to be document-based or concept-based, thus different personalization methods have been developed to use them respectively. Document based methods focused on mining users' document preferences. It gives information about users' interest in a particular document rather than interest in a particular concept. Based on the assumption that user would scan the search result list from top to bottom, Joachims [19] first proposed to extract user clicking preferences from the click-through data. These were then applied for learning ranking function that best fits the user's preferences. Tan et al. [18] proposed RSCF algorithm, extending the ranking SVM [19] using a co-training framework [20]. Further, Agichtein et al. [13] proposed to exploit click-through data for learning users' clicking and browsing behavior by utilizing RankNet [21]. A combination of spying technique with a novel voting procedure was proposed by Ng et al. for mining users' document preferences.

Concept-based methods concentrate on extracting users' topical preferences from users' browsed documents. The intention is to discover the topic in which the user is actually interested rather than in what document. Exploring the content of the browsed documents of the user [15] proposed a method to automatically derive user's topical preferences. User profiles based on user –browsed content and predefined taxonomy, open directory project (ODP) was proposed by Speretta and Gauch in [22]. Shen et al. [23] used the ODP taxonomy to propose web query classification for personalization. Sieg et al. proposed building user profiles based on the ODP taxonomy for personalizing search process [24]. Li et al. [25] proposed a similar technique, but the user profiles were constructed using Google Directory as predefined taxonomy. Most of concept based methods are found to be dependent on predefined taxonomies to extract user's topical preferences. Maintenance of these taxonomies requires additional human efforts.

Gan et al. [26] suggested that search queries can be classified into non-geo (i.e, content) or geo (i.e, location) queries. Based on the fact that users' location information plays a significant role in mobile search, several location-based search systems have been developed. Yokoji [27] employed a parser to mine location information from the web documents and convert it into latitude-longitude pairs and later utilized it to retrieve documents for the location corresponding to user's query. Chen et al. [28] contributed to study of efficient query processing, based on query footprint which specified the geographical area of interest of the user. Thus, it has been observed that existing location-based search systems are dependent on users for explicit location preferences. Methods to automate this process are not yet popular

4 Adaptive Re-Ranking Approach For Search Engine Personalization

The proposed framework consists of three modules: the Mobile Device (client), Adaptation Manager (Server) and Third-party Search Engine (TPSE). Interaction between the modules is explianed briefly in Fig.4. The user is required to submit the search query at the mobile device, which acts as the client and forwards it to the Adaptation Manager. This in turn does the heavy computational tasks and gets the results from the TPSE. Based on these tasks, it learns user preferences and adaptively re-ranks the TPSE results as per the concept relevancy of the user. These re-ranked results are then forwarded to the client, where they are displayed to the user.

The personalization approach is incorporated at the Adaptation Manager. It based on combining the derived document and concept preferences. The concept preferences are classified into content and location concepts. Efforts are made to utilize the location services of the mobile devices for efficient handling of the location sensitive queries. Besides, considering only the click-through logs the approach also tries to exploit the users browsing dwell time to enhance learning of the user behavior.

5 Conclusions

The effectiveness of search technologies is reduced by the ambiguity of the user' s query and the diversity of their information needs. Mobile, being a constrained device, mobile search also introduces new challenges not present in traditional search systems. The proposed framework addresses the limitations of mobile search, employing an adaptive re-ranking function in accordance with user information needs. It provides an integrated solution which captures users' preferences with associated context, to save user's efforts and time.

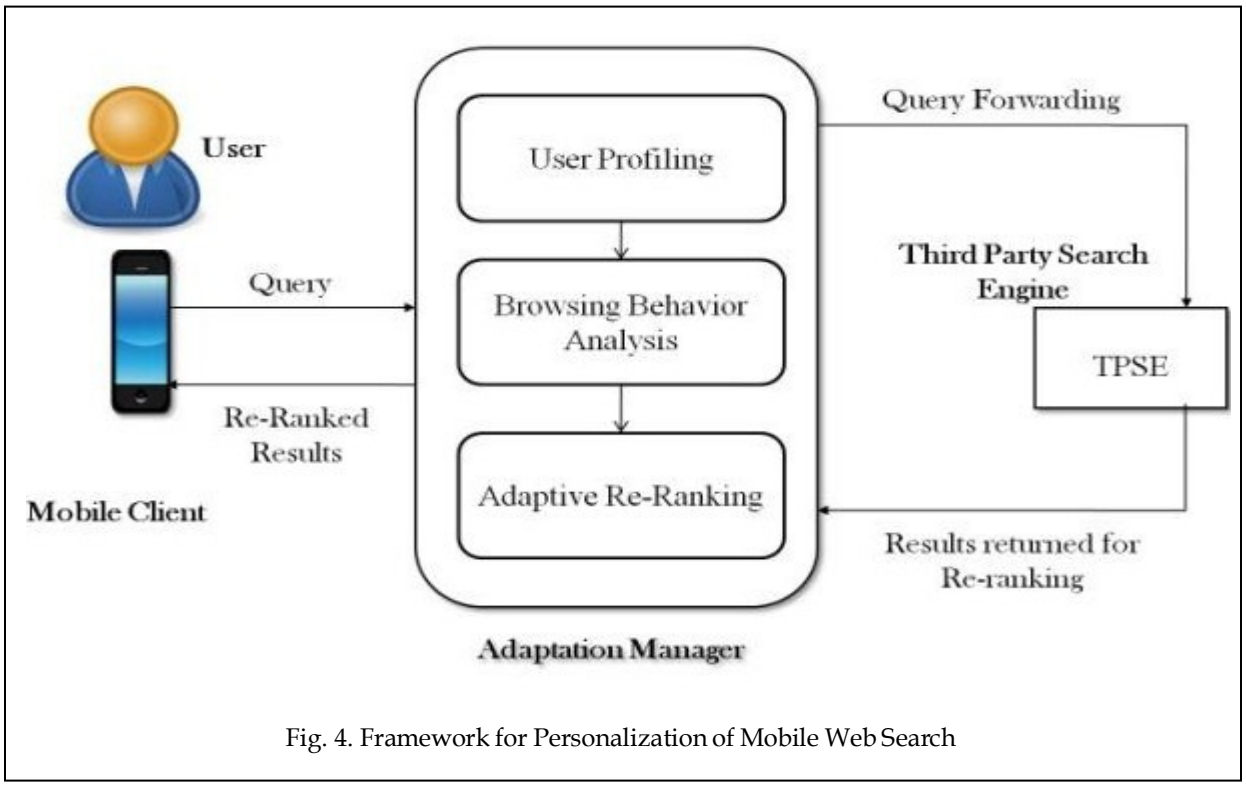

Fig. 4. Framework for Personalization of Mobile Web Search

REFERENCES

[1] David Green,"The evolution of Web searching", Online Information Review", Vol. 24 Iss: 2, pp.124 – 137,2000.

[2] Zheng Lu, Hongyuan Zha, Xiaokang Yang, Weiyao Lin, Zhaohui Zheng, "A New Algorithm for Inferring User Search Goals with Feedback Sessions", IEEE Trans. ON Knowledge and Data Engineering, Vol. 25, No. 3, March 2013.

[3] Mingyang Sun et. Al., "FoSSicker: A Personalized Search Engine by Location-Awareness", International Conference on Computing, Networking and Communications, Communication Software and Service Symposium, 2012.

[4] Nauman, Mohammad Khan and Shahbaz, "Using Personalized Web Search for Enhancing Common Sense and Folksonomy Based Intelligent Search Systems," International Conference on Web Intelligence (IEEE/WIC/ACM), Pp. 423 –426, 2007.

[5] Wolfgang Woerndl, Hubert Kreuzpointner, "Issues in Mobile User Modeling and Search Personalization", ACM SIGIR Workshop on MobIR., 2008.

[6] Dilip Kumar Sharma et al.,"A Comparative Analysis of Web Page Ranking Algorithms", (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 08, 2010.

[7] A.Broder, "Web Searching Technology Overview", Advanced school and Workshop on Models and Algorithms for the World Wide Web, 2002.

[8] S.Brin, L.Page, "The anatomy of a large-scale hypertextual web search engine", Proceedings of the 7th International World Wide web Conference, 1998.

[9] K.Berberich, M.Vazirgiannis, G.Weikum, "T-rank: Time-aware authority ranking", Proceedings of 3rd International Workshop on Algorithms and Models for the web-Graph, 2004.

[10] Chirita, P., Nejdl, W., Paiu, R., And Kohlshuetter, C., "Using odp metadata to personalize search". In Proc. Of ACM SIGIR Conference., 2005.

[11] Liu, F., Yu, C., And Meng, W. "Personalized web search by mapping user queries to categories." In Proc. of CIKM Conference, 2002.

[12] Daxin Jiang, Jian Pei, Hang Li, "Mining Search and Browse Logs for Web Search: AvSurvey", ACM Transactions on Computational Logic, Vol. V, No. N, April 2013.

[13] Agichtein, E., Brill, E., Dumais, S., And Ragno, R. "Learning user interaction models for predicting web search result preferences." In Proc. of ACM SIGIR Conference, 2006.

[14] Dou, Z., Song, R., And Wen, J.-R., "A largescale evaluation and analysis of personalized search strategies." In Proc.of WWW Conference, 2007.

[15] Gauch, S., Chaffee, J., And Pretschner, A., "Ontology-based personalized search and browsing." ACM WIAS 1, 3-4, 2003.

[16] Sun, J.-T., Zeng, H.-J., Liu, H., And Lu, Y. "Cubesvd: A novel approach to personalized web search" In Proc. Of WWW Conference, 2005.

[17] Ng, W., Deng, L., And Lee, D. L., "Mining user preference using spy voting for search engine personalization." ACM TOIT 7, 2007.

[18] Tan, Q., Chai, X., Ng, W., And Lee, D., "Applying co-training to clickthrough data for search engine adaptation." In Proc. of DASFAA Conference, 2004.

[19] Joachims, T., "Optimizing search engines using clickthrough data." In Proc. of ACM SIGKDD Conference, 2002.

[20] Blum, A. And Mitchell, T., "Combining labeled and unlabeled data with co-training." In Proc. of COLT Workshop, 1998.

[21] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., And Hullender, G. "Learning to rank using gradient descent." In Proc. of ICML Conference, 2005.

[22] Speretta, M. And Gauch, S., ". Personalized search based on user search histories." In Proc. of IEEE/WIC/ACM Conference, 2005.

[23] Shen, D., Sun, J.-T., Yang, Q., And Chen, Z. "Building bridges for web query classification." In Proc. of SIGIR Conference, 2006.

[24] Sieg, A., Mobasher, B., And Burke, R., "Web search personalization with ontological user profiles." In Proc. Of CIKM Conference, 2007.

[25] Li, L. And Kitsuregawa, M. "Personalizing web search via modelling adaptive user profile." In Proc. of DEWS Conference, 2007.

[26] Q. Gan, J. Attenberg, A. Markowetz, and T. Suel, "Analysis of Geographic Queries in a Search Engine Log," Proc. First Int'l Workshop Location and the Web (LocWeb), 2008.

[27] Yokoji, "Kokono Search: A Location Based Search Engine," Proc.Int'l Conf. World Wide Web (WWW), 2001.

[28] Y.-Y. Chen, T. Suel, and A. Markowetz, "Efficient Query Processing in Geographic Web Search Engines," Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.