ISSN: 2321-7782 (Online)
Volume 2, Issue 4, April 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

SmartPMS: User Profiling Based Re-ranking Approach for Mobile Search Engine

Disha Gupta[1]
Dept. of Computer Science and Engineering
G.H. Raisoni Coleege of Engineering
Nagpur - India

Nekita Chavhan[2]
Dept. of Computer Science and Engineering
G.H. Raisoni Coleege of Engineering
Nagpur - India

Abstract: ***Online information is increasing tremendously day by day. The search engines are often used to retrieve desired information from this web. They serve well in case of focused and to the point queries. However the results retrieved for ambiguous queries are usually a mixture of relevant and irrelevant documents that fails to satisfy the users' information needs. The mobile web search queries tend to have higher ambiguity due to display constraints. This indicates the need of improvement in the current information retrieval systems especially for mobile web search. The paper proposes a personalized user profiling based re-ranking approach. The personalization is provided by mining & integrating users' content & location preferences from previous web interactions. Results represent the benefit of using the implicit feedback in the ranking of search.***

Keywords: Search Engine; Personalization; User Preference; User Profiling; Re-Ranking.

I. INTRODUCTION

Today, internet search engines have become an indispensable part of our lives. People today are able to find all sorts of information instantly from almost anywhere. Search engines have also come to be included within large web sites such as e-commerce sites, corporate sites, and social networking sites. The exceedingly difficult nature of the problem of understanding user intent and matching it with the world's accumulated knowledge stored on the World Wide Web has been the main motivation for researchers [1].

The deterministic approaches of these present day search engines are unable to satisfy different search intents of different people, and returns same generalized results to all who submit the same query at a particular time. This is mainly because the result retrieval process does incorporate user's intent and preferences [2] [3].

In the recent years, an explosive growth has also been witnessed in mobile devices. The modern cell phones are more advanced than those launched few years ago. They are more powerful, provide a much richer user experience and provide users with ubiquitous access to information more than ever before. Mobile Internet has quickly become part of the consumer media experience for millions of people. Unlike early adopters that originally used smart phones primarily for business, most new smartphone owners are using them for mostly personal use [4]. More and more people are searching the web while they are on the move. For instance, one might want to search information about a local restaurant, or need to learn more about a new place they are visiting. The location-awareness of the modern mobile devices is needed to address such queries. These additional features distinguish mobile search engines from the standard web search engines.

Personalization is the ability to provide content and services tailored to individuals based on knowledge about their preferences and behavior [1]. Attempt to provide results which not only satisfy the submitted query but also the user's information needs give the motivation for mobile search engine personalization.

The rest of the paper is organized in the following manner. Section II presents a summary of the related work with respect to search engines ranking. SmartPMS: user profiling based re-ranking approach has been proposed in section III. Section IV demonstrates the experimental setup. A detailed discussion about the results obtained is given in section V. Section VI concludes the paper justifying the use of SmartPMS on mobile devices for enhanced personalization experience.

II. Related Work

The basic process flow of web search engine includes three major tasks, which are, crawling, indexing and ranking [5]. Due to the vast amount of information being continuously updated on web, crawling and indexing of web documents cannot be done on runtime basis for every user. Thus these tasks are performed on periodic manner from time to time. However many research works [6] [7] [8] [9] [10] have been carried out for the variations in ranking techniques to improve search engine effectiveness. The basic foundation of these ranking strategies/techniques lies is Web Mining. It is an application of data mining techniques to web data. It is broadly classified into web structure mining, web content mining and web usage mining.

The initial ranking techniques PageRank [11] [12] and Hypertext Induced Topic Search (HITS) [13][14] belongs to the category of web structure mining. These algorithms have undergone several modifications and are still in use today by the standard search engines. Link structure and popularity are the basis of ranking for algorithms in this category. Based on the fact the more incoming links a document has the more popular and thus the more relevant it is. Such documents are ranked at higher ranks irrespective of their relevancy with the users' search goal.

Web content mining category includes ranking techniques proposed in [15] [16] [17]. In addition to the link structure of the documents these strategies also rely upon the similarity between the submitted query and the content of available on the web document. It proves to be more efficient in satisfying the users' information needs when unambiguous query is explicitly submitted by the user. As it solely relies on similarity with respect to the query, it fails to address users' search goals for ambiguous queries.

The recent research work [18] [19] [20] has been carried out in the category web usage mining. The ranking methodologies in this category focus on utilizing the web interactions for inferring users' context with respect to the query. Various studies [21] have exploited different explicit and implicit usage parameters to achieve personalized ranking of the results.

III. SMARTPMS: User Profiling based Re-Ranking Approach

Smart **P**ersonalized **M**obile **S**earch (SmartPMS) aims at building a personalized approach for mobile web search. Figure 1 shows the implementation of the system design via client server architecture. Collection of the search query and display of the personalized re-ranked search results to the user are the main tasks carried out at any SmartPMS client. Besides this, all the adaptation required to provide personalization to the user is done at the SmartPMS Adaptation Manager (SAM), which acts as a server.

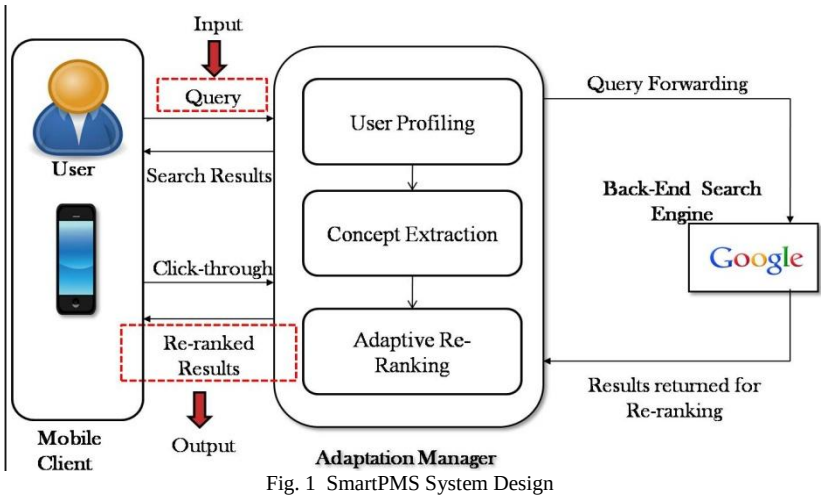


Fig. 1 SmartPMS System Design

The development of personalization process has been carried out with the help of following steps:

A. Components of Smart PMS

SmartPMS provides personalized re-ranked results for ambiguous queries by integrating users' content and location preference as per his/her choice. The SmartPMS system architecture can be divided as three major modules as shown in the Figure IV-2. First is the Mobile Client, second is SmartPMS Adaption Manager (SAM) and third is a commercial web search engine, which is also referred as third party search engine (TPSE) or as back-end search engine. Functions of each module are discussed as follows

1). Mobile Client: It provides an interface to user for entering his/her search query and for browsing the displayed search results. Considering the memory and power constraints of mobile devices, intensive tasks are avoided at the client-side.

2). Smart PMS Adaptation Manager: It is responsible for performing the computational tasks and personalized re-ranking of results. The server side pre-processing of results removes the device dependency for the user i.e using his/her credentials the user can get the same personalized search experience on any mobile device.

3). Third Party search Engine: This returns the search results of the submitted search query to SAM for personalized re-ranking process. The acquired results are in the form of a mixture of relevant and irrelevant results which are then further arranged based on their relevancy to users' information needs.

B. Phases of SMARTPMS

The SAM implements the personalization approach in three phases as mentioned below:

- User Profiling
- Concept Extraction
- Adaptive Re-ranking.the colors used in each figure contrast well

Figure-2 represents the personalization approach diagrammatically.

1). Phase-I: User Profiling: This is the first phase of the Personalization Process which involves namely three main tasks as shown in Figure 2. These tasks are described as follows:

- Maintaining a log of registered users who have created their account in the system.
- Keeping a track of whether submitted search query is seen or unseen for each specific user.
- Recording of clicked URLs of a registered user.

2). Phase-II: Concept Extraction: This is the second phase of the Personalization Process. Its process flow is represented through Figure 2. The main task of this phase includes.

- Analysis of clicked URLs, by examining their web snippets.
- Extraction of user preferences as concepts, based on its frequency of appearance in relation to the query.
- Filtration of duplicates and stop.

3). Phase-III: Adaptive Re-ranking: This is the third and the final phase of the Personalization process. As shown in the Figure-IV # the implementation of final adaptive re-ranking is done here. It involves

- Computation of relevancy score utilizing the extracted content concepts
- Re-ranking of results based on preferred location concepts computed relevancy score.

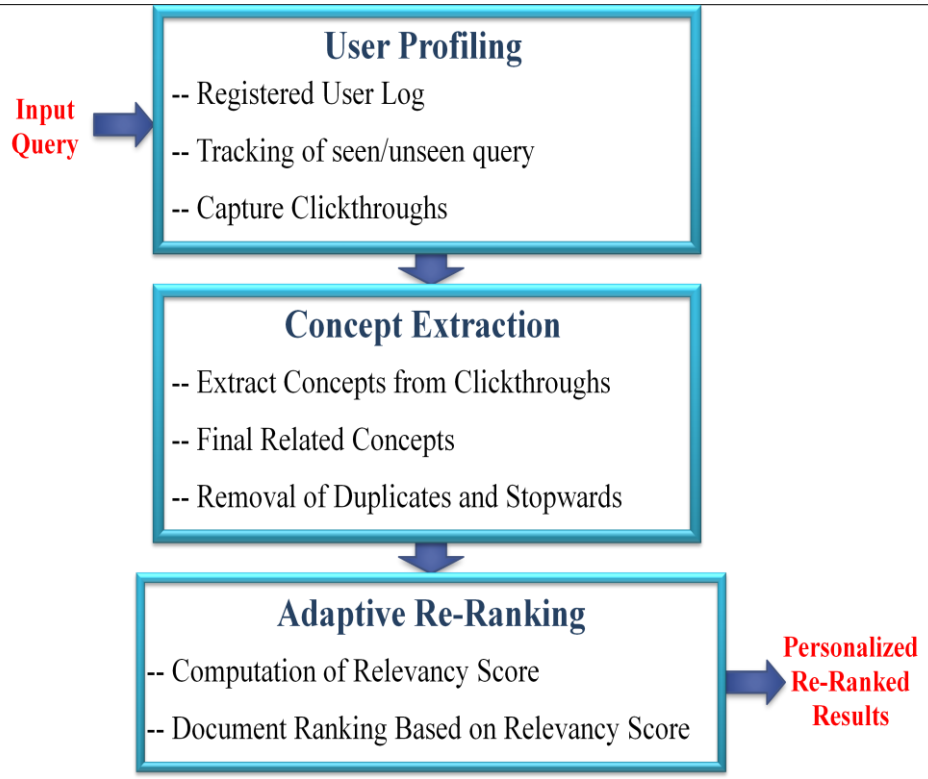


Fig. 2 Phases of Personalization Process at SAM

C. SMARTPMS Personalization Algorithm

Concept Extraction and Re-ranking Algorithms used in SmartPMS Personalization are discussed briefly in following sections

1) *SmartPMS Concept Extraction Algorithm*: The concept extraction algorithm has been integrated in the SmartPMS personalization process with the intention to infer users' search goal. The description available in the web snippets of the documents acts as the source for concept extraction. Since this description generally gives limited information about the document, the no. of terms available for the extracting concepts gets less. Based on the fact that frequency of term repetition in short description is low and also to avoid missing of any user interest the threshold used in concept extraction algorithm is set at minimum count of 2. This algorithm attempts to find out user preferences by their occurrence frequency, thus they need to be filtered further for the removal of duplicates and stop words. The filtered terms are treated as concepts with respect to the query.

Consider,

q ---- Submitted query

C ---- Set of concepts extracted for query q

Ci --- ith concept extracted for query q, 1<j<k

m ----total concepts extracted for query q

D ---- Set of documents returned for query q

dj ---- jth document returned for query q, 1<i<m

n ----total number of documents returned for query q

t-----total terms present in description of document d

1. do
2. do
3. If occurrence frequency of term > 2

4. Concept= Term
5. Endif
6. while t exists
7. Enddo-while
8. while n exists
9. enddo-while
10. do
11. Removal of duplicates and Stop words from Concepts
12. while C != φ and n exists
13. enddo-while".

2) SmartPMS Re-ranking Algorithm: Adaptive Re-ranking is the main and the final phase of the SmartPMS personalization approach. Its preceded by extraction of concepts from the documents, returned for the search query. Once the concepts get extracted, relevance score is computed for all the documents based on the amount of similarity between concepts and the documents. Then the documents are ranked by sorting in descending order of their relevance score.

Consider,

q ---- Submitted query

C ---- Set of concepts extracted for query q

Ci --- ith concept extracted for query q, 1<j<k

m ----total concepts extracted for query q

D ---- Set of documents returned for query q

dj ---- jth document returned for query q, 1<i<m

n ----total number of documents returned for query q

1. do
2. do
3. con_sim (q, ci, dj)= ∑ 〖No.of occurrence of ci in dj 〗
4. while m exists
5. Enddo-while
6. Relevance Score (q,dj)=∑m (i=1) Con_sim(q, ci, dj〗)
7. while n exists
8. Enddo-while
9. Sort n documents in descending order of Relevance Score.

IV. EXPERIMENTAL SETUP

This section presents the performance evaluation of SmartPMS. Mobile Client has been prototyped on Google Android Platform, whereas Web Service is hosted at Windows Server Platform. The role of commercial search engine in the system design has been substituted through standard Google Search API. The experiment has been performed in a controlled environment. The number of users involved and the set of queries to be considered is pre-defined. The users are required to synthesize their information needs from the given queries and conduct their searches accordingly. The positive experimental results discussed in the following sections appear as the strong evidence of SmartPMS effectiveness.

A. Experimental Procedure

25 users, who were graduation students in computer science department, were asked to use SmartPMS for evaluation. Each user is asked to randomly select 10 test queries from the given set. Whenever an unseen query is submitted the user clicks are captured and analyzed for preference mining. On re-submission of the same query SmartPMS adapts the personalized approach and presents re-ranked results to the user. The user click-through is again captured for the new re-ranked results. Considering the fact, that user clicks those particular URL that appear in relation to his/her information needs, these URLs are considered relevant for the user. The original ranking obtained from the Google API and the re-ranked list of search result is then evaluated. User-wise ARR and Average Precision metrics of relevant documents are the measured and compared for each ranking approach.

Statistics of Experiment

The statistics of experiment conducted for evaluation of SmartPMS ranking quality is shown in the Table as follows:

TABLE I
Experimental statistics of smartpms evaluation

Parameters	**Values**
#Users	25
#Queries	35
#Queries Per User	10
#Total URLs	1580
# Clicked URLs	375
#Concepts Extracted	452
#No of URLs re-ranked per Query	10

List of test queries available for random selection to the users is given in the Table II.

TABLE II
Test Query Set

Apple	Android	Application	ATM	Orange
Ticket	Sun	Oracle	Java	Microsoft
Cloud	Drive	Canon	Framework	Photoshop
ODP	Virtualization	Function	Network	Call
Script	Language	Platform	Canon	Software
Developer	Hypervisor	Remote	Batching	Cache
Security	Attack	Redhat	Package	Ruby
Hotels	Restaurants	Hospitals	Colleges	Multiplex

V. RESULTS AND DISCUSSION

The efficiency of SmartPMS has been analyzed using two metrics, namely Average Relevant Rank (ARR), Average Precision. These metrics are discussed in detail as follows:

B. Average Relevant Rank (ARR)

Average Relevant Rank (ARR), is the average rank of the relevant documents, for which a lower value indicates better ranking quality.

$$AAR = \sum \text{Rank of Relevant Document} / \text{Total Number of Relevant Document}$$

The average relevant rank for Google's Ranking and SmartPMS Ranking is shown in the Figure 3. Personalized SmartPMS has comparatively lower ARR than Google i.e SmartPMS is able to achieve improved ARR over Google.

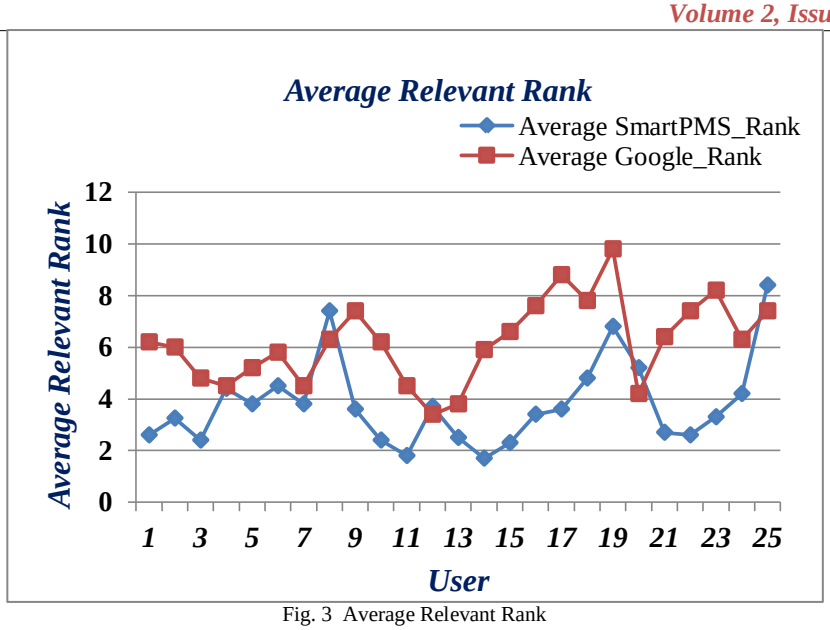


Fig. 3 Average Relevant Rank

C. Average Precision

Average Precision, is the average ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. Higher values for average precision indicate good ranking quality.

Average Precision = ∑ Number of Relevant Document / Total Number of Relevant Document

The graph for average precision comparison of SmartPMS and Google is shown in Figure 4. It represents that higher average precision can be achieved personalized re-ranking of results as compared to the Google generalized ranking.

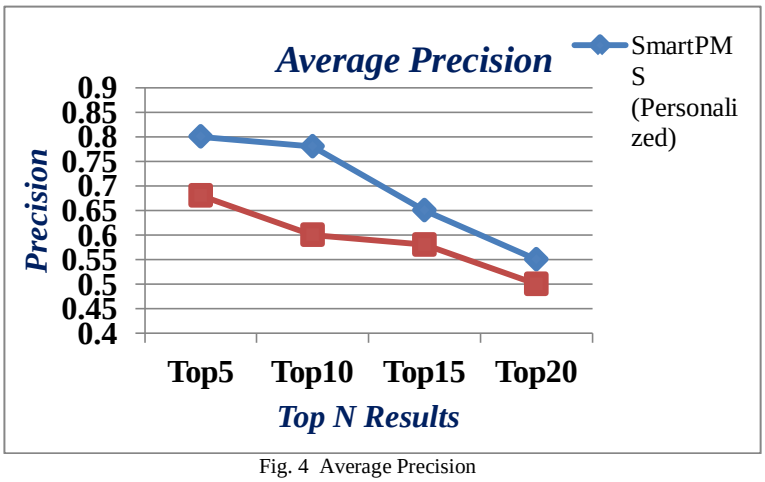


Fig. 4 Average Precision

VI. CONCLUSION

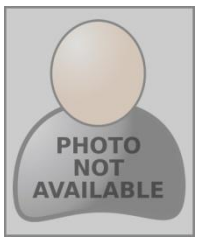
SmartPMS provides an approach for mobile search engine personalization, which adapts the means of personalization as per the user information needs. It utilizes the implicit feedback present in the user's web interaction in the form of click-through. Considering the significant role of location, the context based adaptive re-ranking technique, integrates the users' content as well as location preference. This aids in resolving the query ambiguity and enhancing the location based services, a distinct feature of mobile web search. The conducted experiments reveals the benefit of using implicit feedback to improve results quality by combining both personalized and original web search results. The results obtained highlight the need of personalized approach based on user preferences and search goals. The further work can be done to extend personalization to image search and for other information retrieval systems.

References

1. Gao, B. J., Buttler, D., Anastasiu, D. C., Wang, S., Zhang, P., & Jan, J. "User-Centric Organization of Search Results". Internet Computing, IEEE, 2013.
2. Nidhi Grover, Ritika Wason,"Comparative Analysis Of Pagerank And HITS Algorithms", International Journal of Engineering Research & Technology (IJERT),Vol. 1 Issue 8, October, 2012.
3. Chirita, P., Nejdl, W., Paiu, R., And Kohlshuetter, C., "Using ODP Metadata To Personalize Search". In Proc. Of ACM SIGIR Conference, 2005.
4. Woerndl, Wolfgang, and Hubert Kreuzpointner. "Issues in Mobile User Modeling and Search Personalization." SIGIR 2008 Workshop on Mobile Information Retrieval. 2008.
5. 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.
6. Hawalah, Ahmad, and Maria Fasli. "A Hybrid Re-Ranking Algorithm Based On Ontological User Profiles." 3rd. IEEE Computer Science and Electronic Engineering Conference (CEEC), 2011.
7. Thanh-Trung Van and Michel Beigbeder, "A Comparison of Re-ranking Methods in Digital Libraries using User Profiles", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.
8. Mukhopadhyay, Debajyoti, Pradipta Biswas, and Young-Chon Kim. "A Syntactic Classification based Web Page Ranking Algorithm." arXiv preprint arXiv:1102.0694, 2011.
9. Desikan, P., Srivastava, J., Kumar, V., & Tan, P. N., "Hyperlink Analysis–Techniques & Applications", Army High Performance Computing Center Technical Report, 2002.
10. Dou, Z., Song, R., And Wen, J.-R., "A Largescale Evaluation And Analysis Of Personalized Search Strategies." In Proc.of WWW Conference, 2007.
11. Richardson, Matthew, and Pedro Domingos. "The Intelligent Surfer: Probabilistic Combination Of Link And Content Information In Pagerank.", Advances in neural information processing systems 2, 2002.
12. Borodin, A., Roberts, G. O., Rosenthal, J. S., & Tsaparas, P. "Finding Authorities And Hubs From Link Structures On The World Wide Web", InProceedings of the 10th international conference on World Wide Web ACM, April 2001.
13. Haveliwala, Taher H. "Topic-sensitive pagerank." Proceedings of the 11th international conference on World Wide Web. ACM, 2002.
14. Ding, Chris, Xiaofeng He, Parry Husbands, Hongyuan Zha, and Horst D. Simon. "PageRank, HITS and a unified framework for link analysis." In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval,. ACM, 2002.
15. Shikha Goel ,Sunita Yadav, "Search Engine Evaluation Based on Page Level Keywords", 3rd IEEE International Advance Computing Conference (IACC),2013.
16. Liu, F., Yu, C., And Meng, W. "Personalized web search by mapping user queries to categories." In Proc. of CIKM Conference, 2002.
17. 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.
18. Joachims, T., "Optimizing Search Engines Using Clickthrough Data" In Proc. of ACM SIGKDD Conference, 2002.
19. D. Poo, B, Chng and J.Goh, "A Hybrid Approach For User Profiling", Proceedings of 36th International Conference on System Sciences, 2003.
20. Ng, L. Deng, and D.L. Lee, "Mining User Preference Using Spy Voting for Search Engine Personalization," ACM Trans. Internet Technology, Vol. 7, No.4, 2007.
21. K.Leung, D.Lee, "Deriving Concept-Based User Profiles From Search Engine Logs", IEEE Trans. Knowledge And Data Engg., Vol. 22, No.7, July 2010.

AUTHOR(S) PROFILE

Disha Gupta is pursuing Master of technology in Computer Science & Engineering discipline from G. H. Raisoni college of Engineering (Autonomous), Nagpur, Maharashtra.



N.A.Chavhan is currently working as Assistant Professor in Computer Science & Engineering Department, G. H. Raisoni college of Engineering, Nagpur, Maharashtra.

