Personalization of Mobile Search Engine using Ontology

Snehal D. Jadhav
PG Student, Department of Information Technology
MIT College of Engineering,
Pune, India.
snehalj90@gmail.com

Vaishali P. Suryawanshi
Assistant Professor,
Department of Information Technology
MIT College of Engineering,
Pune, India.
vaishali.suryawanshi@mitcoe.edu.in

Abstract: **There are trillions of web pages available on the World Wide Web for the user queries to give their search results. However, these results may be relevant to the user queries or they can be irrelevant to the queries. As a solution to this, we are proposing a personalized search engine which will provide the more relevant search results to the user queries. In order to do so, we will be using an ontology concept, according to this we are going to split the user query into two parts: content and location. The location and content concepts will be used to build an ontology profile which in then will be used to capture user interests. Using user queries the feature vectors will be extracted and which in turn will be trained using re-ranking algorithm to rank the future results. We perform an extensive query search experiments to compare the effectiveness of proposed and existing systems.**

Keywords- Personalization, ontology, click though, re-ranking, personalization ranking function

I. INTRODUCTION

In mobile searches, the interaction between the user and the search engine is limited by number of factors such as small screen, limited computing power, less operating memory etc. The search in such case is very limited and it is prime need to lower the user interaction with the engine and provide the most relevant results to user queries. By capturing the user's preferences, it is possible to build a mining server that will be able to give the search results according to user preferences by re-ranking the results according to personalized re-ranking.

Wilfred Ng, et al, proposed a method that provides personalised query suggestions based on personalised concept based clustering technique. This approach used click through data to estimate user's preferences and then provide personalized query suggestions [2]. Dik Lun Lee, et al proposed a personalization technique that captures the user's preferences in the form of concepts by mining their click through data. An ontology-based user profiling is used to capture the user preferences. Location concepts are also given importance in order to give more specific results related to user profile [3]. Three major functions are done in this personalization process: Capturing user preferences, re ranking them and updating user profile. User preferences are captured by using the user click through data and are treated as positive preferences.

II. RELATED WORK

Kenneth Wai-Ting Leung introduced an effective approach that captures the user's abstract preferences in order to provide personalized query indications. Two new strategies were used to achieve this goal. Online techniques that extract concepts from the web-snippets of the search result returned from a query and use the concepts to identify related queries for that query were developed. A new two phase personalized agglomerative clustering algorithm that is able to generate personalized query clusters was introduced. To evaluate the effectiveness of this technique, a Google middleware was developed to collect clickthrough data to conduct experimental evaluation [2].

Dik Lun Lee, et al, proposed a new web search personalization approach that captures the user's preferences in the form of concepts by mining search results and their clickthroughs. As the importance location information plays in mobile search, the concepts are separated into content concepts and location concepts, then arrange them into ontologies to create an *ontology-based, multi-facet (OMF)* profile to precisely capture the user's content and location interests. In turn, it improves the search accuracy. To distinguish different users and queries may have different emphases on content and location information, an idea of content and location entropies to measure the amount of content and location information contained in a query, and *click* content and location entropies to measure how much the user is interested in the content and location information in the results were introduced. According to OMF, to define personalization effectiveness based on the entropies and use it to balance the weights between the content and location facets was proposed. At last, based on the derived ontologies and personalization effectiveness, an SVM is trained to adapt a personalized ranking function for re-ranking of future search [3].

R Divya, et al., proposed a web search personalization approach that captures the user's interests and preferences in the form of concepts by mining search results and their click through. Proposed system is based on a client server model. Major tasks such as training, re ranking are done by the server. To preserve privacy of search only the feature vectors is passed to the server. GPS locations help in re ranking search results. Finally, depending on the derived ontology RSVM is used for re-ranking of future search results [1].

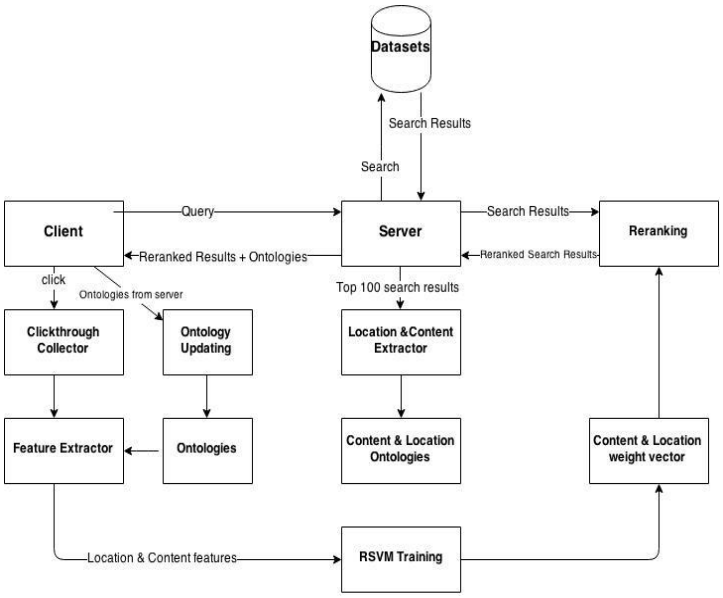

Fig 1: System Architecture

III. SYSTEM ARCHITECTURE

The proposed system consist of an client- server model in which the client fires queries and server will search for the results, arranging them according to user preferences and give back the re-ranked search results. The architecture of the system can be given as below:

A. RSVM Training

Ranking SVM is a modified application of SVM used to solve certain ranking problems. The main use is to improve the performance of the internet search engine. Click through data can be used as the input. It is valid for all document preference pairs possible. It maps the similarities between queries and the clicked pages. Calculates the weights between any two of the vectors that are obtained. Re ranks the search results based on the weights. If a user has visited the GPS location lr, the weight of the location concept is incremented by one unit. Hence, it is assumed that the location that the user has visited a long time ago is less important than the location that the user has recently visited. Thus GPS plays an important role in obtaining location information.

B. Content Ontology and Location Ontology

Concept extraction first starts by extracting the similar words or phrases in the web snippet present in the query. If the keyword or phrase appears repeatedly in the web snippets, then it will be treated as important concept.

In case of location concepts, it is possible that a document contains precise information of the location or none. In such a case a predefined hierarchy of the cities is defined in the form of ontology. Moreover, GPS also helps in providing the user locations.

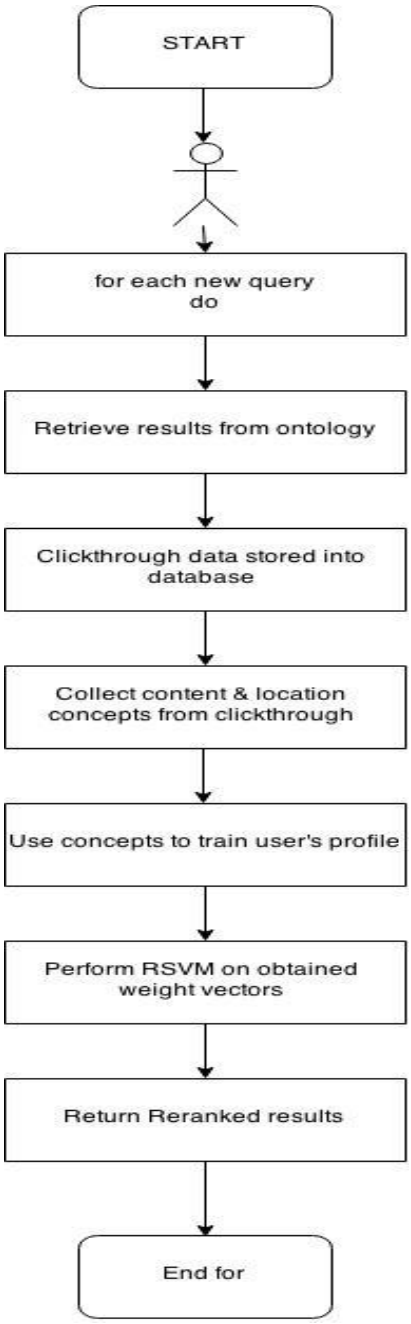

Fig 2: Process Flow

C. Joachim's Method

In Joachim method scanning of results is done from top to bottom. The document which has higher rank is displayed first and other results are displayed after that in descending order of ranking. If a user does not read a document which is at a higher rank and reads on document at lower rank, it means that the user is not interested in document at the higher rank. Thus, Joachim's method concludes that the user prefers di to document dj (denoted as dj <r0 di, where r0 is the user's preference order of the documents in the search result list).

D. SPYNB Method

Similar to Joachim's method, SpyNB learns user behavior models from preferences extracted from click through data.

SpyNB assumes that users would only click on documents that are of interest to them. Then we can treat the clicked documents as positive samples. However, documents that are not clicked are treated as unlabeled samples because they could be either relevant or irrelevant to the user.

E. Extracting Feature Vectors

Two vectors are defined to represent the content and location information present within a document, namely: content feature vector and location feature vector. According to ontology, there exists four relations within these vectors: Similarity, ancestor, decedent and sibling. If the content concept is present in the web snippet then the feature weight vector will be incremented by one, same is the case with location weight vector. If user marks the location as visited then too the location weight vector will be incremented.

F. Evaluation Process

The user only conducts search on the system before personalization as if he/she is using a regular search engine. Then, the user rates the relevance of the search results manually according to his/her search preferences. After these steps, the training of search engine and the measurement of retrieval effectiveness are both conducted offline without the involvement of the user by the server. As a user becomes more skilled with the system, answers of the later queries could become more and more relevant.

The personalization process consist of three major phases: test phase, training phase and evaluation phase.

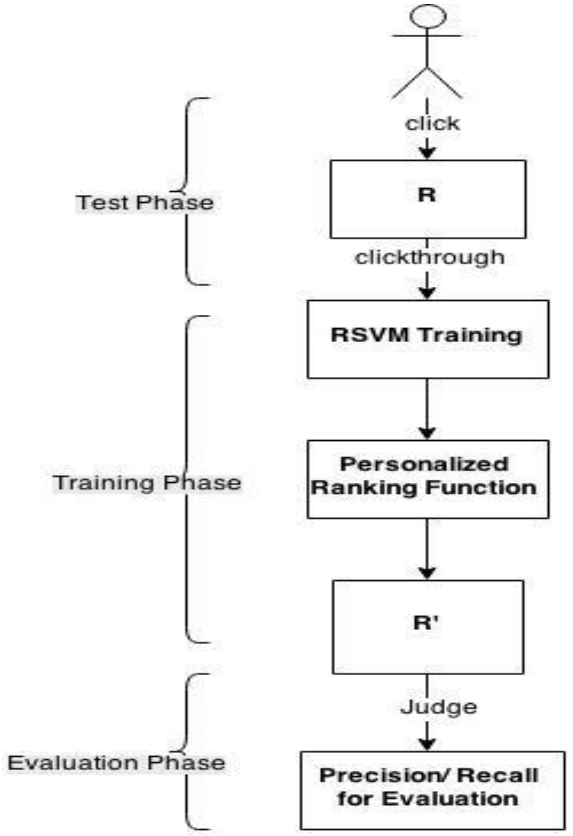

Fig 3: Personalization process

In the test phase, a user submits a test query and receives the top 100 search results R from the search engine without any personalization. The user then clicks on any of results (one or more) that he/she judges to be relevant to his/her interest in the same way that a standard search engine would have been given. The clicked results from the test phase are considered as positive training samples in RSVM training. The click through data, the extracted content and location concepts are employed in RSVM training to obtain the personalized ranking function. The evaluation phase is then performed to decide whether the personalized ranking function can return more relevant results. The users were asked to provide relevance judgment by grading each result with one of the three levels of relevancy ("Relevant," "Fair," and "Irrelevant").

IV. EXPERIMENTAL EVALUATION

…Considering our proposed system, we use two straightforward evaluation measures, namely, precision and recall. Different queries may induce from the search results different concept spaces which are different in both sizes and diversities. Additionally, different users may have different interests and preferences on the search results. The difference between variety of content and location information of a query can be found from the extracted content and location concepts, while the behaviour of a user can be found by his click through behaviour. In information retrieval, precision is the fraction of retrieved documents that are relevant to the search.

$$Precision = \frac{\{relevant\ documents\} \cap \{retrieved\ documents\}}{\{retrieved\ documents\}}$$

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$Recall = \frac{\{relevant\ documents\} \cap \{retrieved\ documents\}}{\{relevant\ documents\}}$$

Precision and recall are defined in terms of a set of retrieved documents (e.g. the list of documents produced by a web search engine for a query) and a set of relevant documents (e.g. the list of all documents on the internet that are relevant for a certain topic).

V. CONCLUSION

In this paper, we are proposing a personalization approach to extract the user preferences through their click through by means of content and location concepts for building the ontologies. The motive of this system is to extract the features of content and location and re rank the search results according to user interests. For future scope we are planning to study the frequent travel and query patterns of the users in order to provide more relevant search results.

REFRENCES

[1] Dr. C. R .Rene Robin and R Divya, "Onto-Search: An Ontology Based Personalized Mobile Search Engine", 2014

[2] K.W.-T. Leung, W. Ng, and D.L. Lee, "Personalized Concept-Based Clustering of Search Engine Queries," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 11, pp. 1505-1518, Nov. 2008.

[3] K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "Personalized Web Search with Location Preferences," Proc. IEEE Int'l Conf. Data Mining (ICDE), 2010.

[4] K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "PMSE: A Personalized Mobile Search Engine", IEEE Trans. Knowledge and Data Eng., vol. 25, no. 4, April 2013
[5] E. Agichtein, E. Brill, and S. Dumais, "Improving Web Search Ranking by Incorporating User Behavior Information," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006
[6] E. Agichtein, E. Brill, S. Dumais, and R. Ragno, "Learning User Interaction Models for Predicting Web Search Result Preferences," Proc. Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2006.
[7] 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.
[8] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002
[9] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. Int'l Conf. World Wide Web (WWW), 2007
[10] S. Yokoji, "Kokono Search: A Location Based Search Engine," Proc. Int'l Conf. World Wide Web (WWW), 2001