STUDY AND IMPLEMENTATION OF A PERSONALIZED MOBILE SEARCH ENGINE FOR SECURE SEARCH

Mr. R.A.Sanadi[1], Mr. Abhijit Patil[2],
Mr. Akshay Patil[3], Mr. Aniket Sutar[4],
Mr. Akshay Kondekar[5]

[1]Department of Computer Science and Engineering, Professor at Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.
sanadiraj@gmail.com

[2]Department of Computer Science and Engineering, Student at Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.
abhijit.patil55555@gmail.com

[3]Department of Computer Science and Engineering, Student at Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.
akshaypatil4831@gmail.com

[4]Department of Computer Science and Engineering, Student at Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.
sutaraniket924@gmail.com

[5]Department of Computer Science & Engineering
Student at Dr. J. J. Magdum College of Engineering, Jaysingpur, Maharashtra, India.
akshaykondekar81194@gmail.com

---***--

Abstract - *As we know, the amount of Web information is growing rapidly, so the Search engines must be able to provide information according to the user's preference. In this paper, we propose Ontology Based Personalized Mobile Search Engine (PMSE) that captures user's interest and preferences in the form of concepts by mining search results and their clickthroughs. PMSE profile the user's interest and personalized the search results according to user's profile. PMSE classifies these concepts into content concepts and location concepts. In addition, user's locations are used to supplement the location concepts in PMSE. The user's location is achieved by using GPS. The user preferences are organized in an ontology-based user profile, used to achieve a personalized ranking function which in turn used for rank adaptation of future search results. We propose to define personalization based on the entropies and use it to balance the weights between the content and location facets. PMSE is based on client-server model. In our design, the client collects and stores locally the clickthrough data to protect privacy, whereas heavy tasks such as concept extraction, Searching, training, and reranking are performed at the PMSE server. PMSE distribute the task to each individual component to decrease the complexity.*

Key Words: Clickthrough data, concept, mobile search engine, concepts, ontology ,user profiling, personalization.

1.INTRODUCTION

In today's world, most of people are using smart phones and internet. Also they want fast acces to the internet. But one of the most important difficulties in mobile search engine is the interfacing among the users and search engine. The mobiles having less computing power. The interaction is restricted due to the small form factors of mobile devices. On the World Wide Web large amount of information is available which is copious and personal. In the personalized mobile search engine various kind of concepts are used with diverse ontologies. Mobile search engine in which location of the user is one of the important key factor as well as information related with particular location is having same importance. Personalized search engines catches user location and gives information associated to that location. For consistently optimizing the retrieval priority of search engines by using clickthrough data the information related to Clickthrough data is categorize as content and location concepts. For example, a user who wishes to visit Cold places in India may submit query as Cold places. From that query keyword , PMSE understand user's content preference ("cold places") and the location preference ("india"). Also if a user who is planning to visit Japan may issue the query hotel, and click on the search results about hotels in Japan. From the clickthroughs of the query hotel", PMSE can learn the user's content preference (e.g. room rate and facilities) and location preferences (Japan).

The user preferences are prepared by ontology-based, comprehensive user profile, which are used to adapt a

personalized ranking function for rank adaptation for search results. The personalized mobile search engine server is trustworthy and very useful for completion of the heavy tasks .Server preparing and re-ranking the search results according to the users content and location preferences sooner than they come back to client side. The personalized mobile search engine clients keep records of the specific user's profile and maintain user's privacy. Personalized mobile search engine profiles content and location preferences in the ontology based user profiles, which are automatically achieved from the clickthroughs as well as GPS data.

TABLE 1
Clickthrough for the Query "hotel"

Doc	Search Results	Ci	Li
d1	Hotels.com	room rate	International
d2	JapanHotel.net	reservation, room rate	Japan
d3	Hotel Wiki	accommodation	International
d4	US Hotel Guides	map, room rate	USA, California
d5	Booking.com	Online reservation	USA
d6	JAL Hotels	Meeting room	Japan

2.RELATED WORK

As shown in Table 1, Clickthrough data have been used in determining the user's preferences on their search results. As shown, Ci is the content concepts and Li is location concept. A major problem in mobile web search is that the interaction between the users and interactions between the users and search engines. Mobiles have low computing power. As a result, mobile users tend to submit shorter, more ambiguous queries compared to their web search counterparts. In order to return highly relevant results to the users, mobile search engines must be able to profile the users' interests and personalize the search results according to the users' profiles. A practical approach to capturing a user's interests for personalization is to analyze the user's clickthrough data. Clickthrough data have been used in determining the users' preferences on their search results. Many existing personalized web search systems [1],[2], [3], are based on clickthrough data to determine users' preferences. In paper [4], proposed a method that provides personalized query suggestions based on a personalized clustering technique. In [5] presents a new approach for situation aware personalized search. Case-Based Reasoning [CBR] approach is used. In [6] proposed a system based in a semantic context-aware framework, which helps the user to build personalized search queries by means of an auto completion mechanism. However, most of the previous work assumed that all concepts are of the same type. In this paper separate concepts into location concepts and content concepts to recognize information importance. So far there have been many papers written and researched on search engines. Most commercial search engines return roughly the same results to all users. However, different information needs even for the same query. PMSE profiles both of the user's content and location preferences in the relation based user profiles, which are automatically learned from clickthrough and GPS data without requiring extra efforts from the user.

In PMSE propose a realistic design for PMSE by adopting the metaearch approach which depends on one of the commercial search engines, such as Google, Yahoo etc. to perform an actual search. The client is responsible for receiving the user's requests, submitting the requests to the PMSE server, displaying the returned results, and collecting his/her clickthroughs in order to derive his/her personal preferences. The PMSE server, on the other hand, is responsible for handling heavy tasks such as forwarding the requests to a commercial search engine, training and reranking of search results before they are returned to the client. The user profiles for specific users are stored on the PMSE clients, thus preserving privacy to the users. PMSE has been prototyped with PMSE clients on the Google Android platform and the PMSE server on a PC server to validate the proposed ideas. The client-server architecture provides a coherent strategy to integrate them into a uniform solution for the mobile environment. By mining content and location concepts for user profiling, it utilizes both the content and location preferences to personalize search results for a user.

The differences between existing work and PMSE using content and location concept are:

1. Propose and implement a new and realistic design for PMSE. This helps to train the user profiles quickly and efficiently.

2. Most existing location-based search systems require users to manually define their location preferences or to manually prepare a set of location sensitive topics. PMSE profiles both of the user's content and location preferences in the ontology based user profiles, which are automatically learned from the clickthrough and GPS data without requiring extra efforts from the user.

3. SYSTEM DESIGN

The PMSE's architecture meets two important requirements. First, computation-intensive tasks, such as RSVM training, should be handled by the PMSE server due to the limited computational power on mobile devices. Second, data transmission between client and server should be minimized to ensure fast and efficient processing of the search. In the PMSE's client-server architecture, PMSE clients are responsible for storing the user clickthrough and the ontologies derived from the PMSE server. Simple tasks, such as updating clickthoughs and ontologies, creating feature vectors, and displaying reranked search results are handled by the PMSE clients with limited computational power. On the other hand, heavy tasks, such as RSVM training and reranking of search results, are handled by the PMSE server.

Moreover, in order to minimize the data transmission between client and server, the PMSE client would only need to submit a query together with the feature vectors to the PMSE server, and the server would automatically return a set of reranked search results according to the preferences stated in the feature vectors. Fig.1 shows PMSE's client-server architecture, which meets three important requirements. First, computation-intensive tasks, such as RSVM training, should be handled by the PMSE server due to the limited computational power on mobile devices. Second, data transmission between client and server should be minimized to ensure fast and efficient processing of the search. Third, clickthrough data, representing precise user preferences on the search results, should be stored on the PMSE clients in order to preserve user privacy. PMSE's design addressed two issues:

1. Limited computational power on mobile devices.
2. Data transmission minimization.

PMSE consists of two major activities:

Reranking the search results at PMSE server:

When a user submits a query on the PMSE client, the query together with the feature vectors containing the user's content and location preferences (i.e., filtered ontologies according to the user's privacy setting) are forwarded to the PMSE server, which in turn obtains the search results from the back-end search engine (i.e., Google). The content and location concepts are extracted from the search results and organized into ontologies to capture the relationships between the concepts. The server is used to perform ontology extraction for its speed. The feature vectors from the client are then used in RSVM training to obtain a content weight vector and a location weight vector, representing the user interests based on the user's content and location preferences for the reranking. Again, the training process is performed on the server for its speed. The search results are then reranked according to the weight vectors obtained from the RSVM training. Finally, the reranked results and the extracted ontologies for the personalization of future queries are returned to the client.

Ontology update and clickthrough collection at PMSE client:

The ontologies returned from the PMSE server contain the concept space that models the relationships between the concepts extracted from the search results. They are stored in the ontology database on the client. When the user clicks on a search results, the Clickthrough data together with the associated content and location concepts are stored in the Clickthrough database on the client. The clickthroughs are stored on the PMSE client, so the PMSE server does not know the set of documents that the user has clicked on. This design allows user privacy to be preserved in certain degree. Two privacy parameters, *minDistance* and *expRatio* are proposed to control the amount of personal preferences exposed to the PMSE server. Depending on the user, the privacy level can be set to high or to low. If privacy level set to high, only limited personal information will be send to PMSE server through the feature vectors and if privacy level set to low then the PMSE server can use the full feature vectors.
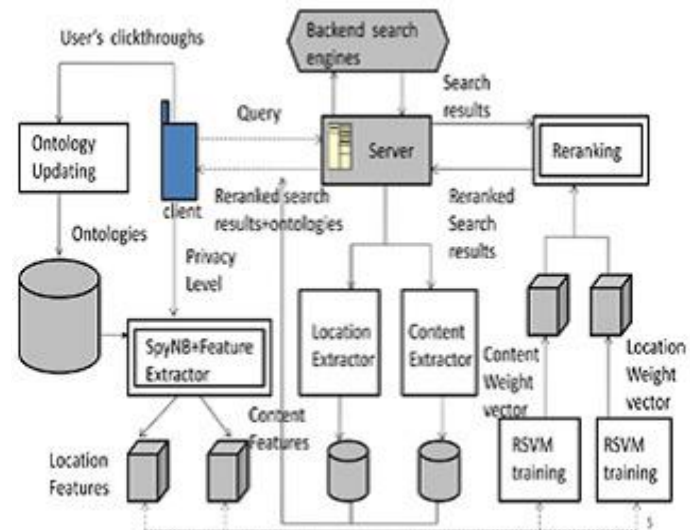

Fig -1: System Design of PMSE

4. PROPOSED METHODOLOGY

1. Registration of User/Login:
         This will allow the user to register on the device. This information shall be stored on the server side database. He/she shall input the basic information.

2. Extraction of query concept:
         The user shall input the query word and the keyword shall be sent to the server. The server sends the query to the internet and retrieves the snippets of the keyword. The snippets are cleaned of html and the keywords and their frequencies are calculated. They are further stored as the users preferred keywords on the server side database.

3. Clickthrough data of the User :
         On the client side the url on which the user has clicked is stored along with the users location data. This data is then sent to server again to be stored in database.

4. Ranking Algorithm:
         The clickthrough data and the keyword frequencies are combined into a vector called a feature vector. The feature vector is provided as input to the RSVM to find the probability of the url's and are sorted in descending order. These ranked pages are then provided to the user

5.CONCLUSION

We represent A Personalized Mobile Search Engine to separate and learn user's content and location preferences based on user's clickthroughs. Also we used GPS locations of user's to adapt user mobility. results shows that the GPS l ocations help to improve retrieval effectiveness. We also proposed two privacy parameters to address privacy issues in PMSE. In future work, we will investigate some methods to fast and more secured access to further enhance the personalization of PMSE.

REFERENCES

[1]Q. Tan, X. Chai, W. Ng, and D. Lee, "Applying Co-Training to Clickthrough Data for Search Engine Adaptation," Proc. Int'l Conf. Database Systems for Advanced Applications (DASFAA), 2004.

[2] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, 2002.

[3] 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.

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

[5] "Ontology Supported Personalized Search for Mobile Devices" Daniel Aréchiga, Jesús Vegas and Pablo de la Fuente Redondo, proc.int'l conf. Web search and Data Mining,2011.

[6] "semantic context aware framework for personalization" Y. Xu, B. Zhang, and Z.Chen,2oo7.