PMSE: Personalized Mobile Search Engine Using Location Concept

Miss. Belge Priyanka, Miss. Chauhan Manisha, Miss. Dhubale Yogita, Prof: Patil S.S. (Guide)
Department of computer engineering, AAEMF’S & MS, College of Engineering, Koregaon Bhima, Pune,
Savitribai Phule Pune University, Maharashtra, India

Abstract
We propose a personalized mobile search engine, PMSE, that captures the users’ preferences in the form of concepts by mining their click through data. Observing the need for different types of concepts, we separate concepts into location concepts and content concepts. In addition, users’ locations (positioned by GPS) are used to supplement the location concepts in PMSE. The user preferences are organized in an ontology-based, multi-facet user profile, which are used to adapt a personalized ranking function for rank adaptation of future search results.

Based on the client-server model, we also present a detailed architecture and design for implementation of PMSE. In our design, the client collects and stores locally the click through data to protect privacy, whereas heavy tasks such as concept extraction, training and are performed at the PMSE server.

Moreover, we address the privacy issue by restricting the information in the user profile exposed to the PMSE server with two privacy parameters. We prototype PMSE on the Google platform.

Keywords— Clickthrough data, content concept, location search, mobile search engine, ontology, personalization, user profiling

I. INTRODUCTION
The small form factors of the mobile device limits the interaction between the mobile user and the search engines. In order to get a highly relevant result, user profiling is recommended. User profiling is nothing but the mobile search engines capturing the user interest to personalize. User interest personalization is based on the concept preferences. Previous works are based mostly on same types, this proposed enhanced personalized search engine works on different types of concepts in different ontologies.

In personalized mobile search engine (PMSE) that captures the users’ preferences in the form of concepts by mining their click through data. Knowing the importance of the location information in mobile search, this search engine captures users preferences in the form of concepts viz., content concept and location concept. Location information are supplement to the location concept. User can also submit the location by simply typing it on a particular column or GPS helps. The user preferences are organized in an ontology-based, multi facet 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.

Search results have become increasingly complex and that trend is likely to continue. The traditional model of 10 blue links and rank checking is no longer accurate as users are receiving results that are increasingly customized to them. As results are becoming more personalized, its valuable to better understand how personalized search results are being content concept and location concept.

II. OBJECTIVE FOR PMSE
1. We have to develop a classifier to classify geo and non-geo queries.
2. It was found that a significant number of queries were location queries focusing on location information.
3. In order to handle the queries that focus on location information, a number of location-based search systems designed for location queries have been proposed.

III. PROPOSED SYSTEM
Many existing personalized web search systems are based click through data to determine users preferences proposed to mine document preferences from click through data. Later, Ng, et. al. proposed to combine a spying technique together with a novel voting procedure to determine user preferences. More recently, Leung, et. al. introduced an effective approach to predict users’ conceptual preferences from click through data for personalized query suggestions. Search queries can be classified as content or location queries. Examples of location query are hongs hotels, “museums in london and virginia historical sites. In developed a classifier to classify geo and non-geo
queries. It was found that a significant number of queries were location queries focusing on location information. In order to handle the queries that focus on location information, a number of location-based search systems designed for location queries have been proposed. proposed a location-based search system for web documents. Location information were extracted from the web documents, which was converted into latitude-longitude pairs.

The differences between existing works and ours are:
1. Most existing location-based search systems, require users to manually define their location preferences (with latitude-longitude pairs or text form), 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.
2. We propose and implement a new and realistic design for PMSE. To train the user profiles quickly and efficiently, our design forwards user requests to the PMSE server to handle the training and reranking processes.
3. Existing works on personalization do not address the issues of privacy preservation. PMSE addresses this issue by controlling the amount of information in the client’s user profile being exposed to the PMSE server using two privacy parameters, which can control privacy smoothly, while maintaining good ranking quality.

IV. ARCHITECTURE

Figure 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. In the PMSE’s client-server architecture, PMSE clients are responsible for storing the user clickthroughs and the ontologies derived from the PMSE server. Simple tasks, such as updating clickthroughs 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.

PMSE consists of two major activities:
1. Reranking: 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.
2. 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.1 When the user clicks on a search result, the clickthrough data together with the associated content and location.

Query and User Classes:

To characterize queries and users with the proposed content and location entropies, we employ K-Means to cluster the queries and users into different classes, and evaluate the performance of PMSE on the different classes.
To classify the 250 queries into different classes, we compute their content and location entropies and display them on a scatter plot with location entropy as x-axis and content entropy as y-axis. K-Means is then employed to cluster the queries into four classes, marked with different colors. We characterize the four query classes as follows:
1. Explicit queries
2. Content queries
3. Location queries
4. Ambiguous queries.

As with the four query classes, we display the queries on a scatter plot with click location entropy as x-axis and click content entropy as y-axis. Again, the test queries are clustered into five classes with K-Means according to their click entropies. The five query classes are
1. Low click entropies.
2. Content-seeking.
3. Location-seeking.
4. Medium click entropies.
5. High click entropies.

V. CONCLUSION
To adapt to the user mobility, we incorporated the user’s GPS locations in the personalization process. We observed that GPS locations help to improve retrieval effectiveness, especially for location queries. We also proposed two privacy parameters, minDistance and expRatio, to address privacy issues in PMSE by allowing users to control the amount of personal information exposed to the PMSE server. The privacy parameters facilitate smooth control of privacy exposure while maintaining good ranking quality. For future work, we will investigate methods to exploit regular travel patterns and query patterns from the GPS and clickthrough data to further enhance the personalization effectiveness of PMSE.

REFERENCES