RESEARCH ARTICLE OPEN ACCESS

A Search Engine for Personal Mobile

Priya Panjwani, Nitin Raut, Abdulla Shaik
(Department of Computer Science, Nagpur University, India
(Department of Computer Science, Nagpur University, India
Email:priyapanjwani58@gmail.com),Email: infotech.nitin@gmail.com)
(Department of Computer ScienceNagpur University, India
Email:snanm.shaik@gmail.com)

Abstract
A personalized mobile search engine (PMSE) that captures the users' preferences in the form of concepts by mining their clickthrough data. Due to the importance of location information in mobile search, PMSE classifies these concepts into content concepts and location 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, multifacet 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's to the user's need, four entropies are introduced to balance the weights between the content and location facets. In our design, the client collects and stores locally the clickthrough data to protect privacy, whereas heavy tasks such as concept extraction, training, and reranking 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 Android platform. Experimental results show that PMSE significantly improves the precision comparing to the baseline.
Index Terms—Clickthrough data, concept, location search, mobile search engine, ontology, personalization, user profiling.

I. INTRODUCTION

A major problem in mobile search is that the interactions between the users and search engines are limited by the small form factors of the mobile devices. As a result, mobile users tend to submit shorter, hence, 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. Leung et al. developed a search engine personalization method based on users' concept preferences and showed that it is more effective than methods that are based on page preferences [12].

However, most of the previous work assumed that all concepts are of the same type. Observing the need for different types of concepts, we present in this paper a personalized mobile search engine (PMSE) which represents different types of concepts in different ontologies. In particular, recognizing the importance of location information in mobile search, we separate concepts into location concepts and content concepts.

For example, 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").

To incorporate context information revealed by user mobility, we also take into account the visited physical locations of users in the PMSE. Since this information can be conveniently obtained by GPS devices, it is hence referred to as GPS locations. GPS locations play an important role in mobile web search.

Our proposed framework is capable of combining a user's GPS locations and location preferences into the personalization process. To the best of our knowledge, our paper is the first to propose a personalization framework that utilizes a user's content preferences and location preferences as well as the GPS locations in personalizing search results.

II. RELATED WORK

In this paper, we propose a realistic design for PMSE by adopting the metasearch approach which replies on one of the commercial search engines, such as Google, Yahoo, or Bing, 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, as well as 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. This paper studies the unique characteristics of content and location concepts, and provides a coherent strategy using a client-server architecture to integrate them into a uniform solution for the mobile environment.

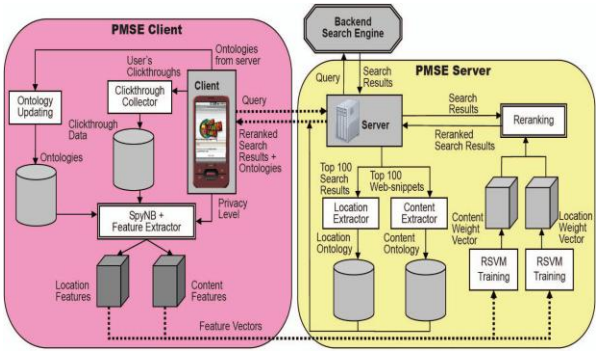

Fig. 1. The general process flow of PMSE.

The proposed personalized mobile search engine is an innovative approach for personalizing web search results. By mining content and location concepts for user profiling, it utilizes both the content and location preferences to personalize search results for a user. PMSE incorporates a user's physical locations in the personalization process. We conduct experiments to study the influence of a user's GPS locations in personalization. The results show that GPS locations helps improve retrieval effectiveness for location queries (i.e., queries that retrieve lots of location information).

III. EXISTING WORK

Most existing location-based search systems, such as [22], 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. 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.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. PROPOSED SYSTEM

In this paper, we propose a realistic design for PMSE by adopting the metasearch approach which replies on one of the commercial search engines, such as Google, Yahoo, or Bing, 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, as well as training and re-ranking 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 PCserver to validate the proposed ideas. We also recognize that the same content or location concept may have different degrees of importance to different users and different queries.

V. USER INTEREST PROFILING

PMSE uses "concepts" to model the interests and preferences of a user. Since location information is important in mobile search, the concepts are further classified into two different types, namely, content concepts and location concepts. The concepts are modeled as ontologies, in order to capture the relationships between the concepts. We observe that the characteristics of the content concepts and location concepts are different. Thus, we propose two different techniques for building the content ontology.

5.1 Content Ontology :

Our content concept extraction method first extracts all the keywords and phrases (excluding the stop words) from the web-snippets2 arising from q. If a keyword/phrase exists frequently in the web-snippets arising from the query q, we would treat it as an important concept related to the query, as it coexists in close proximity with the query in the top documents.

We adopt the following two propositions to determine the relationships between concepts for ontology formulation:

Similarity : Two concepts which coexist a lot on the search results might represent the same topical interest.

Parent-child relationship : More specific concepts often appear with general terms, while the reverse is not true.

5.2 Location Ontology

Our approach for extracting location concepts is different from that for extracting content concepts. We observe two important issues in location ontology formulation. First, a document usually embodies only a few location concepts, and thus only very few of them co-occur with the query terms in web-snippets. To alleviate this problem, we extract location concepts from the full documents. Second, the similarity and parent-child relationship cannot be accurately derived statistically because the limited number of location concepts embodied in documents. Furthermore, many geographical relationships among locations have already been captured as facts.

VI. SYSTEM DESIGN

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 the issues:
1) limited computational power on mobile devices, and 2)data transmission minimization. PMSE consists of two major activities:
1. 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.

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 concepts are stored in the clickthrough database on the client. The clickthroughs are stored on the PMSE clients, so the PMSE server does not know the exact 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.

VII. EXPERIMENTAL EVALUATION

Experiment Setup :

Methodology-

The experiment aims to answer the following question:

Given that a user is only interested in some specific aspects of a query, can PMSE generate a ranking function personalized to the user's interest from the user's clickthroughs? To answer this question, we need to evaluate the search results before and after personalization. The difficulty of the evaluation is that only the user who conducted the search can tell which of the results are relevant to his/ her search intent. Another difficulty of evaluating personalized search systems is that since relevance judgment is highly dependent on the users, care must be taken to ensure that the users' behaviors are not affected by experimental artifacts.

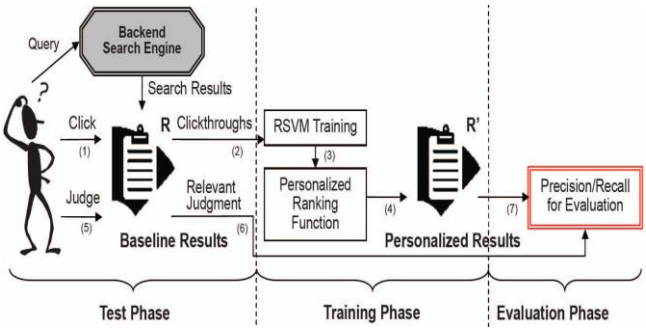

Fig. 3. Flow of the evaluation process.

Limitations

While the methodology tries to minimize the user's involvement in the experiment, it is nevertheless a controlled experiment and thus has some limitations. First, the number of users and queries in the experiments are small. This means that the results from the experiments cannot be construed as representative in diverse situations. Second, since users are given with predefined queries and topical interests, they have to synthesize their information needs from the given queries and topical interests and conduct their searches correspondingly.

VIII. CONCLUSION

We proposed PMSE to extract and learn a user's content and location preferences based on the user's clickthrough. 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.

ACKMOWLEDGMENTS

I would like to thank Prof. Nitin Raut for his guidance and all the supportive environment of NCET, Nagpur to give me opportunity and facilities to carry out this work.

REFERENCES

[1]Appendix,http://www.cse.ust.hk/faculty/dlee/tkde-pmse/appendix.pdf, 2012.

[2] Nat'l geospatial, http://earth-info.nga.mil/, 2012.

[3] svmlight, http://svmlight.joachims.org/, 2012.

[4] World gazetteer,http://www.world-gazetteer.com/, 2012.

[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] K.W. Church, W. Gale, P. Hanks, and D. Hindle, "Using Statistics in Lexical Analysis," Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon, Psychology Press, 1991.

[9] 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.

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

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

[12] 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.