

PERSONALIZED MOBILE SEARCH ENGINE BASED ON MULTIPLE PREFERENCE, USER PROFILE AND ANDROID PLATFORM

Ajit Aher, Rahul Rohokale, Asst. Prof. Nemade S.B.

B.E. (computer) student, Govt. college of engg. & research Awasari(khurd), India

aherajit999@gmail.com

B.E(computer)student, Govt. college of engg. & research Awasari(khurd), India

rahulrohokale18@gmail.com

Asst. Prof of computer engg. Department, Govt. college of engg. Awasari(khurd), India

Sangita_nemade@rediffmail.com

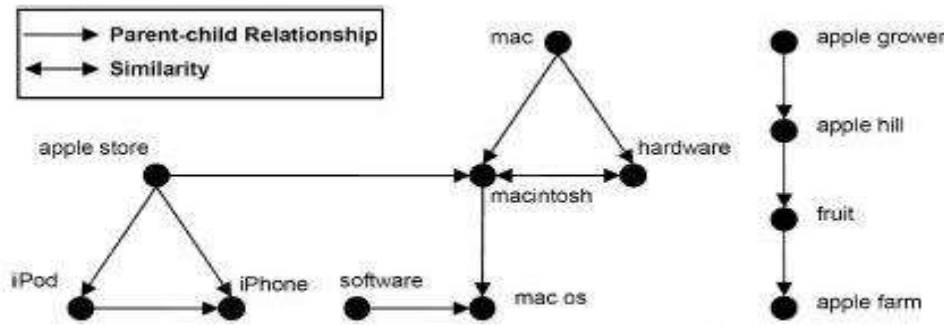
Abstract — Mobile search needs better interaction between user and server usually this interaction is not efficient due to many adopting the meta search approach, click through data, user profiling which is based on client server model. In this model client sends request to the server and server forward this request to the commercial search engine also training and re-ranking is done on server side, we call it PMSE server. We also involve a method to maintain the user's interests over his ongoing search activity and to personalize the search results. The profiles of specific users are stored on the Personalization clients, thus preserving privacy to the users.

Keywords — Clickthrough data, content ontology, location ontology, personalization, user profiling, privacy preservation, spyNB, IR.

1. INTRODUCTION

The proliferation of mobile technologies such as (PDAs and mobile phones) has made access to huge and heterogeneous collection of documents on the web, possible anywhere and anytime. This brings big challenges for researches in the information retrieval (IR) domain. Studies on logs of mobile Internet user queries show that user queries are shorter (thus more ambiguous), that there are fewer requests by session and fewer users who consult farther than the first page of the results list. Furthermore, 72% of the information needs of mobile users are related to contextual factors such as user interests, location and time. So it is very difficult to user to get relevant result or expected result. In our system, we propose User Interest Profile. Each user has its own profile, in the sense which provide user a privacy. When user send query to PMSE server for getting reply, search history is created. So for every user, history is created and it is maintained by ontology DB. This web history will be in use for further query. Personalization aims to alter large amounts of information and returns a view on the information which matches the user's preferences and interests improving therefore the precision of the search results. 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. Previous research shows that

researcher concentrated only on content preference but in our system we are going to use location of user also for better result. We propose our system on android platform so for getting location of user, we can use GPS system.



Example Content Ontology Extracted for the Query “apple”.

Fig.1 Content Ontology Extracted

2. RELATED WORK

Most commercial search engines return roughly the same results to all users. However, different users may have different information needs even for the same query. For example, a user who is looking for a laptop may issue a query .apple. to find products from Apple Computer, while a housewife may use the same query .apple. to find apple recipes. The objective of personalized search is to disambiguate the queries according to the users' interests and to return relevant results to the users. Clickthrough data is important for tracking user actions on a search engine.

Doc	Search Results	c_i	l_i
d_1	Hotels.com	room rate	international
d_2	JapanHotel.net	reservation, room rate	Japan
d_3	Hotel Wiki	accommodation	international
d_4	US Hotel Guides	map, room rate	USA, California
d_5	Booking.com	online reservation	USA
d_6	JAL Hotels	meeting room	Japan
d_7	Shinjuku Prince	facility	Japan, Shinjuku
d_8	Discount Hotels	discount rate	international

Table.1. Clickthrough Data.

Table I is an example clickthrough data for the query. It consists of the search results of a user's query and the results that the user has clicked on by bold. c_i 's are the content concepts and l_i 's are the location concepts extracted from the corresponding results. Many personalized web search systems are based on analyzing users clickthroughs. Joachims proposed to use document preference mining and machine learning to rank search results according to user's preferences. Later, Agichtein et al. proposed a method to learn users' clicking and browsing behaviour from the clickthrough data using a scalable implementation of neural networks called Rank Net. Gan et. al suggested that search queries can be classified into two types, content (i.e., non-geo) and location (i.e., geo). Typical examples of

geographic queries are hotels, football ground. A classifier was built to classify geo and non-geo queries, and the properties of geo queries were studied in detail. It was found that a significant number of queries were location queries focusing on location information. Hence, a number of location-based search systems designed for geo queries have been proposed. These include Yokoji et al., who proposed a location-based search system for web documents. A parser was employed to extract location information from web documents, which was converted into latitude longitude pairs or polygons. When a user submits a query together with the location information specified in a latitude longitude pair, the system creates a search circle centered at the specified latitude-longitude pair and retrieves documents containing location information within the search circle.

The differences between our work and existing works are:

- Existing works such as require the users' to manually define their location preferences explicitly (with latitude-longitude pairs or text form). With the automatically generated content and location user profiles, our method does not require users to explicitly define their location interest manually.
- Our method automatically profiles both of the user's content and location preferences, which are automatically learnt from the user's clickthrough data without requiring extra efforts from the user.
- Our method uses different formulations of entropies derived from a query's search results and a user's clickthroughs to estimate the query's content and location ambiguities and the user's interest in content or location information. The entropies allow us to classify queries and users into different classes and effectively combine a user's content and location preferences to rerank the search results.
- In Existing works there was nothing about users privacy and profile, but in our system we are going to maintain user profile.
- 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. PROPOSED SYSTEM

Most of the previous work assumed that all concepts are of the same type. We separate concepts into location concepts and content concepts to recognize information importance. So far there have been many papers written & researched on search engines. There is tremendous evolution in this field. In this paper, we propose a realistic design for PMSE by adopting the metaearch approach which relies 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 clickthrough 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. Studies the unique characteristics of content and location concepts, and provides a coherent strategy using client-server architecture 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.

4. SYSTEM DESIGN

1. Weight vector- content weight vector and user weight vector describes the user interests based on the user's content and location preferences extracted from the user clickthroughs respectively.

2. Feature vector- feature vector is a n dimensional vector of numerical feature that represent some object. feature vector are often combined with weights using a dot product in order to construct a linear predictor function that is used to determine a score for making prediction.

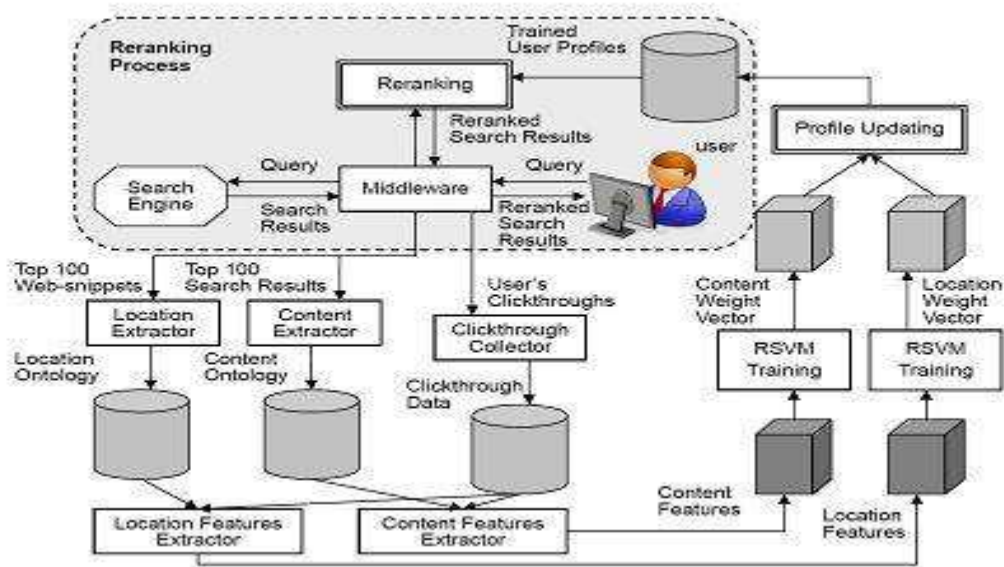


Fig. 2. System Design

Content ontology - if a keyword/phrase exists in web-snippets arising from the query, we would treat it as an important concept related to the query.

Location ontology - extract location concepts from full documents. The predefined location ontology is used to associate location information with the search results. All of the keywords from the documents returned for query are extracted. If a keyword or key-phrase in a retrieved document matches a location name in our predefined location ontology, it will be treated as a location concepts.

4.1 MODULES

- 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 and location ontology. The ontologies indicate a possible concept space arising from a user's

queries, which are maintained along with the clickthrough data for future preference adaptation. In PMSE, we adopt ontologies to model the concept space because they not only can represent concepts but also capture the relationships between concepts. Due to the different characteristics of the content concepts and location concepts.

- Diversity and Concept Entropy

PMSE consists of a content facet and a location facet. In order to seamlessly integrate the preferences in these two facets into one coherent personalization framework, an important issue we have to address is how to weigh the content preference and location preference in the integration step. To address this issue, we propose to adjust the weights of content preference and location preference based on their effectiveness in the personalization process. For a given query issued by a particular user, if the personalization based on preferences from the content facet is more effective than based on the preferences from the location facets, more weight should be put on the content-based preferences; and vice versa.

- User Preferences Extraction and Privacy Preservation

Given that the concepts and clickthrough data are collected from past search activities, user's preference can be learned. These search preferences, inform of a set of feature vectors, are to be submitted along with future queries to the PMSE server for search result re-ranking. Instead of transmitting all the detailed personal preference information to the server, PMSE allows the users to control the amount of personal information exposed. In this section, we first review a preference mining algorithms, namely SpyNB Method, that we adopt in PMSE, and then discuss how PMSE preserves user privacy. SpyNB learns user behavior models from preferences extracted from clickthrough data. Assuming that users only click on documents that are of interest to them, SpyNB treats the clicked documents as positive samples, and predict reliable negative documents from the unlabeled (i.e. unclicked) documents. To do the prediction, the "spy" technique incorporates a novel voting procedure into Naïve Bayes classifier to predict a negative set of documents from the unlabeled document set. The details of the SpyNB method can be found in. Let P be the positive set, U the unlabeled set and PN the predicted negative set ($PN \subset U$) obtained from the SpyNB method.

- Personalized Ranking Functions

Upon reception of the user's preferences, Ranking SVM (RSVM) is employed to learn a personalized ranking function for rank adaptation of the search results according to the user content and location preferences. For a given query, a set of content concepts and a set of location concepts are extracted from the search results as the document features. Since each document can be represented by a feature vector, it can be treated as a point in the feature space. Using the preference pairs as the input, RSVM aims at finding a linear ranking function, which holds for as many document preference pairs as possible. An adaptive implementation, SVM light available at, is used in our experiments. In the following, we discuss two issues in the RSVM training process:

- 1) how to extract the feature vectors for a document.
- 2) how to combine the content and location weight vectors into one integrated weight vector.

5. ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to my guide Assistant Professor Mrs. Nemade S.B. madam who gave me the golden opportunity to write this important paper on the topic personalized mobile search engine, which also helped me in doing a lot of Research and I came to know about so many new things, I am really thankful to them.

6. CONCLUSION

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. The possible outcome will improve retrieval effectiveness for location queries (i.e. queries that retrieve lots of location information). 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:

- [1]. Varun Misra, Premarayan Arya, Manish Dixit, "Improving Mobile Search through Location Bases context and personalization", IEEE Transactions on Communication Systems and Network Technologies, 2012.
- [2]. Ourdia Boui dhagen, Lynda Taime, Mohand Boughanem, "Coext-Aware user's Interest for personalizing Mobile Search", IEEE Transactions on Knowledge and Data Engineering, 2011.
- [3]. K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "Personalized Web Search with Location Preferences," IEEE Transactions on Data Mining , 2010.
- [4]. K. Arabshian, E. Brill, "A framework for Personalized Context –Aware search of Ontology based tagged data, IEEE Transactions on Services Computing, 2010.
- [5] W. 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, article 19, 2007.
- [6] 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.
- [7] E. Agichtein, E. Brill, and S. Dumais, Improving web search ranking by incorporating user behavior information," in Proc. Of ACM SIGIR Conference, 2006.
- [8] T. Joachims, "Optimizing search engines using clickthrough data," Proc. Of ACM SIGKDD Conference,2002.
- [9] S. Yokoji, "Kokono search: A location based search engine," Proc. Of WWW Conference,2001.
- [10] Schoeld, E., Kubin, G. On Interfaces for Mobile Information Retrieval, Mobile HCI (2002), LNCS 2411, pp. 383-387, (2002).
- [11]T. Joachims, Optimizing search engines using clickthrough data, in Proc. of ACM SIGKDD Conference, 2002.
- [12] K. W.-T. Leung, W. Ng, and D. L. Lee, Personalized concept-based clustering of search engine queries, IEEE TKDE, vol. 20, no. 11, 2008