# A Personalized Mobile Search Engine based on User Preference

Pratibha Rathod
Information Technology,
Terna Engineering College, Mumbai University
Navi Mumbai, Maharashtra, India
pratima.rathod15@gmail.com

Smita Desmukh
Information Technology,
Terna Engineering College, Mumbai University
Navi Mumbai, Maharashtra, India
deshmukhsmita17@yahoo.com

*Abstract* - **Personalized Mobile Search Engine (PMSE) is an effective approach to predict user's preferences from his past rating and clickthrough data for personalized query suggestions. Due to the importance of location in mobile search, PMSE classifies these concepts into content and location concepts. In addition, user's locations (positioned by GPS) are used to supplement the location concepts in PMSE. The user preferences are organized with the help of user's rating, which are used to adapt a personalized ranking function for rank adaptation of future search results. In PMSE, the client collects and stores locally the clickthrough data to protect privacy, whereas heavy tasks such as concept extraction and re-ranking are performed at the PMSE server. Moreover, the privacy issue is addressed here by restricting the information in the user profile exposed to the PMSE server.**

*Keywords* - **Clickthrough data, concept, location search, mobile search engine, personalization, user profiling.**

## I. INTRODUCTION

A realistic design for PMSE is proposed by adopting the metasearch 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 and submitting to the PMSE server, displaying the returned results, and collecting his/her preference with the help of rating 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 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.

It also recognizes that the same content or location concept may have different degrees of importance to different users and different queries. To formally characterize the diversity of the concepts associated with a query and their relevance to the user's need .Similarly, to measure how much the user is interested in the content and/or location information in the results. The results are re-ranked according to the user's preferences before returning to the client.

The proposed personalized mobile search engine is an innovative approach for personalizing web search results and the place search results. By mining search concepts for user profiling, it utilizes the content preferences to personalize search results for a user. Privacy preservation is a challenging issue in PMSE, where users send their user profiles along with queries to the PMSE server to obtain personalized search results. PMSE addresses the privacy issue by allowing users to control their privacy levels.

The Rest of paper is organized as follows: The next section reviews the related work. In section III we present the architecture and the system design of PMSE, Section IV gives an idea about System overview, Section V explains the Experiments which is followed by conclusion of the paper in section VI.

## II. RELATED WORK

In PMSE the Clickthrough data have been used in determining the user's preferences on their search results. The content and the location concepts are extracted from the corresponding results. Many existing personalized web search systems are based clickthrough data to determine users' preferences.

Yokoji [2] in 2001 proposed a location-based search system for web documents. Location information was extracted from the web documents, which was converted into latitude-longitude pairs. When a user submits a query together with 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. T. Joachims [3], states that proposed to mine document preferences from clickthrough data, to mining log files of WWW search engines with the goal of improving their retrieval performance automatically. Later, Ng et al. [6] proposed a new approach to mining a user's preferences on the search results from clickthrough data and using the discovered preferences to adapt the search engine's ranking function for improving search quality. Also combine a spying technique together with a novel voting procedure to determine user preferences. Gan et al. [7], 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

978-1-5386-0814-2/17/$31.00 ©2017 IEEE

search systems designed for location queries have been proposed. Leung et al.[8] introduced an effective approach to predict users' conceptual preferences from clickthrough data for personalized query suggestions. ..Search queries can be classified as content (i.e., non-geo) or location (i.e., geo) queries. Examples of location queries are "Hong Kong hotels," "museums in London," and "Virginia historical sites."

More recently K.W-T. Leung [11] in 2010 introduced the notion of content and location entropies to measure the diversity of content and location information associated with a query and click content and location entropies to capture the breadth of the user's interests in these two types of information.

## III. PROBLEM STATEMENT

In existing system of mobile search, the major problem 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. The existing works on personalization do not address the issues of privacy preservation. Also in some existing location-based search systems users needs to manually define their location preferences (with latitude-longitude pairs or text form), or to manually prepare a set of location sensitive topics.

In PMSE, 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. PMSE addresses the privacy issue by allowing users to control their privacy levels. Empirical results show that the proposal facilitates smooth privacy preserving control, while maintaining good ranking quality. PMSE incorporates a user's physical locations in the personalization process. Experiments conducted to study the influence of a user's GPS locations in personalization, show that GPS locations helps improve retrieval effectiveness for location queries (i.e., queries that retrieve lots of location information).

## IV. SYSTEM DESIGN

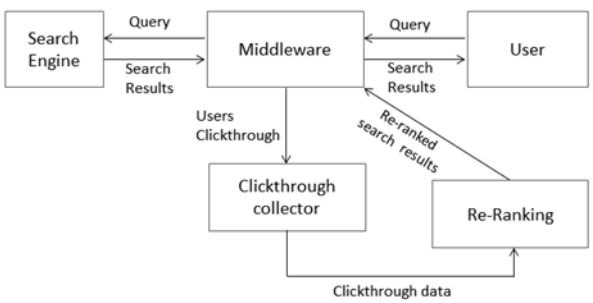


Figure1. The general process flow of PMSE.

Fig. 1 shows the general process flow, which meets the following important requirements. First data transmission between client and server should be minimized to ensure fast and efficient processing of search. Second clickthrough data based on the rating, representing precise user preference on the search results, stored on the client in order to preserve user privacy.

PMSE clients are responsible for storing the user clickthroughs. Simple tasks, such as updating user's rating and creating feature vectors, and displaying re-ranked search results are handled by the PMSE clients with limited computational power. On the other hand, heavy tasks, such as re-ranking 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 re-ranked search results according to the preferences stated in the feature vectors. The data transmission cost is minimized, because only the essential data (i.e., query, feature vectors, and search results) are transmitted between client and server during the personalization process. PMSE's design addressed the issues:
1) Limited computational power on mobile devices.
2) Data transmission minimization.

## V. SYSTEM OVERVIEW

Once the user registration is completed the user logs in into his account. After he logs in successfully, there is an option for Google search, Bing search and place search. Google and Bing search is for normal content searching and place search is for location searching.

### *5.1. Web search:*

This module is done for content-based searches. In this the general results from the Google server are returned to the user's mobile .Or the Bing results returned to the users mobile if user select the Bing search engine . Back-end process: We use Custom Search API for Google and Bing to retrieve the results and apply our re-ranking by using the users rating.

### *5.2. Places search:*

This module is done for location-specific searches. For some searches where the results must be returned based on the particular location (e.g. restaurant, school, hospital etc.) this option is selected by the user. Back-end process: Once this option is selected, the latitude-longitude pair of the user's location is automatically retrieved from the GPS and the results corresponding to that location are only returned. We use Google Places API to retrieve the results and then apply re-ranking based on the Google rating.

*5.3 Algorithmic steps for web search:*

Step 1:- Registering new user by creating his login id and password.

Step 2:- Registering user login with his credentials

Step 3:- Searching User Query

Step 4:- Manual Rating (Scale: 1 to 5; 5 being highest).

Step 5:- Capturing user's preference with the help of their rating

Step 6:- Storing user's search history in his log details.

Step 7:- Searching of the same query by user.

Step 8:- Displaying re-ranked result according to users past given rating.

Step 9:- Manual re-rating by user, if required.

Step 10:- Precision and Recall calculation.

Step 11:- Graph generation.

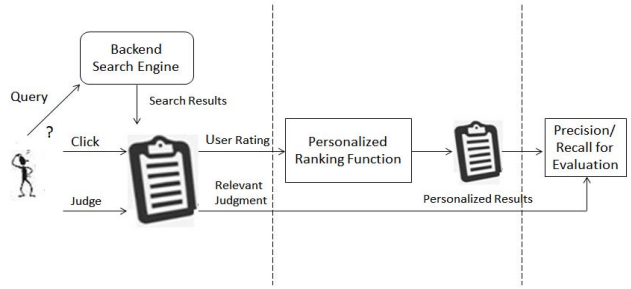


Figure 4.Flow of evaluation process

1. A user submits a query and receives the search results from the back-end search engine (i.e., Google or Bing which the user select) without any personalization.
2. The user then clicks on results, that he/she judges to be relevant to his/her personal interest and give the rating to the link.
3. The clickthrough rated data are used for the personalized ranking function.
4. The results obtained from personalized ranking functions are more relevant for the user.
5. Each user is asked to provide relevant judgment on all of the top results for each query he/she has tested by grading each result with one of the three levels of relevancy - Relevant, Fair and Irrelevant. This rating can be given on the scale of 1 to 5, with 5 being the highest and given to the most relevant result.
6. Perform precision and recall for evaluation based on personalization results.

## VI. EXPERIMENTS

To implement the mobile based application, Google API and Bing API were downloaded and added to the reference library so that the application could fetch URLs from Google and Bing search engines.

As a first step, user registers himself by clicking on 'New User' and fills all the required details. The user thus creates his user id and password.

Once the user registration is completed the user logs in into his account. After he logs in successfully, there is an option for Google search, Bing search and place search.

When user search a query 'apple' and select Google search, then results are retrieved from backend Google search engine. The user gives rating to few of the relevant results According to the users rating that is if the user gives 5 rating for one of the search result and 4 rating to other search results. These ranking and preference will be useful while user searches for the same query in future. The highest rank result will appear on the top of other search results while a user searches for the same query.

In this example 10 results are shown out of which 2 search results are found to be relevant and hence they were rated. Using these relevant and retrieved results we calculate precision, recall and F measure for the searched query 'Q'.

$$\text{Precision}(q) = \frac{Q\_relevant \cap Q\_retrieved}{Q\ retrieved}$$

$$\text{Precision}(q) = \frac{2}{10} = 0.2$$

$$\text{Recall}(q) = \frac{Q\_relevant \cap Q\_retrieved}{Q\_relevant}$$

Here, every time the recall value is 1so we add the constant 2 in denominator so that the results changes every time.

$$Recall\ (q) = \frac{2}{4}$$

$$= 0.5$$

$$F\_measure = 2.\frac{(Precision * Recall)}{(Precision + Recall)}$$

$$F\_measure = 2.\frac{(0.2 * 0.5)}{(0.2 + 0.5)}$$

$$= 0.28571$$

Graphs for the precision, recall and F-measure are plotted using these values.

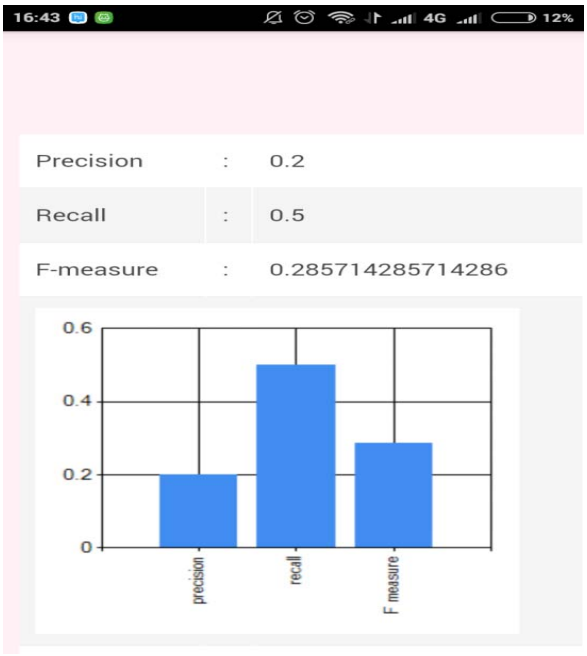


Figure 4: Precision and Recall for single query of Google search

When user search a query 'apple' and select Bing search, then results are retrieved from backend Bing search engine. In the above10 results are shown out of which 3 search results are found to be relevant and hence they were rated. Using these relevant and retrieved results we calculate precision, recall and F measure.

$$Precision\ (q) = \frac{Q\_relevant \cap Q\_retrieved}{Q\_retrieved}$$

$$Precision\ (q) = \frac{3}{10}$$

$$= 0.33$$

$$Recall\ (q) = \frac{Q\_relevant \cap Q\_retrieved}{Q\_relevant}$$

$$Recall\ (q) = \frac{3}{5}$$

$$= 0.6$$

$$F\_measure = 2.\frac{(0.33 * 0.6)}{(0.33 + 0.6)}$$

$$F\_measure = 2.\frac{0.198}{0.93}$$

$$= 0.4258$$

Graphs for the precision, recall and F-measure are plotted using these values

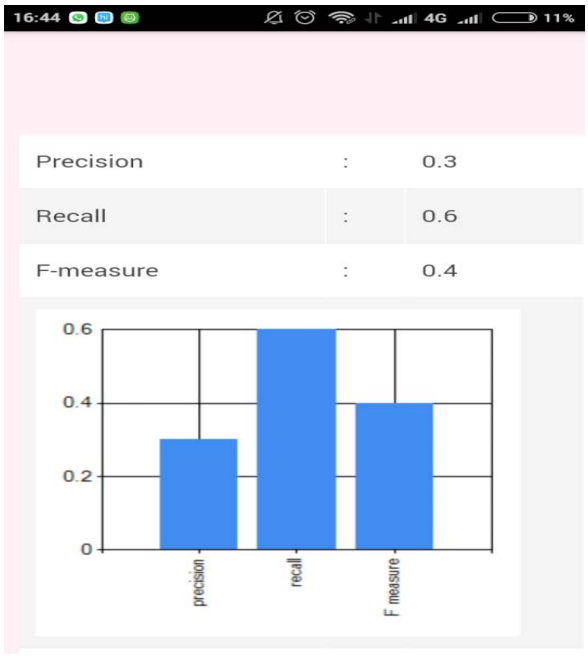


Figure 5: Precision and Recall of single query of Bing search

If a user search different queries, and gives the rating, the average rating of relevant results of that particular user is plotted and also the average Precision and recall plotted below,

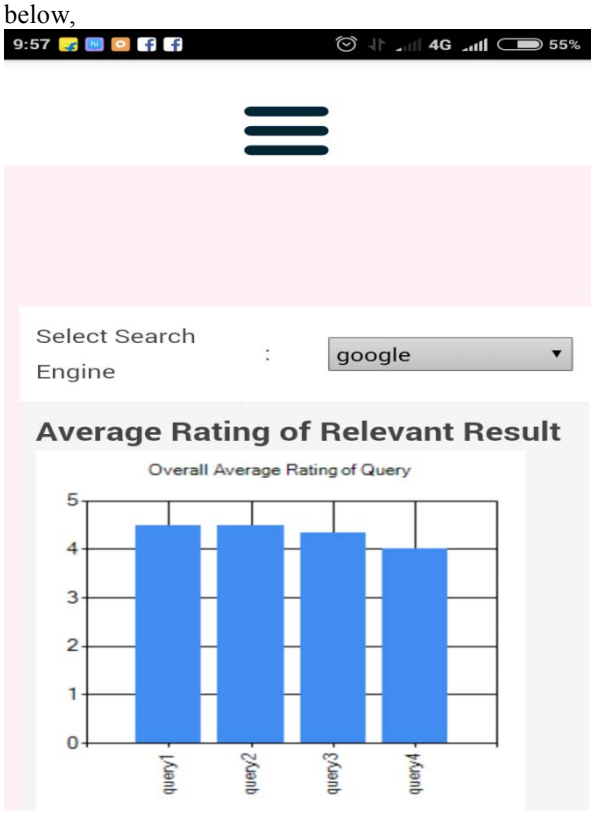


Figure 6: Average Rating of Relevant Result of Google Search

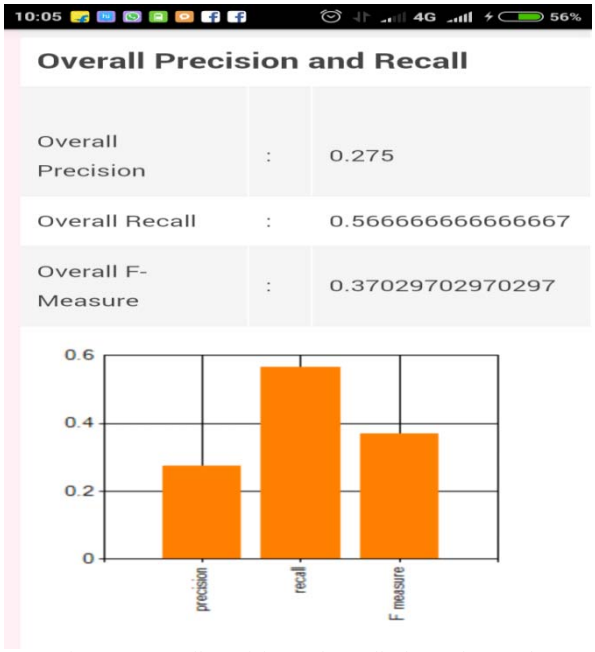


Figure 7: Overall Precision and Recall of Google Search

If a user search different queries from Bing backend search engine and gives the rating ,the average rating of relevant results of that particular user is plotted and also the average Precision and Recall is plotted below,

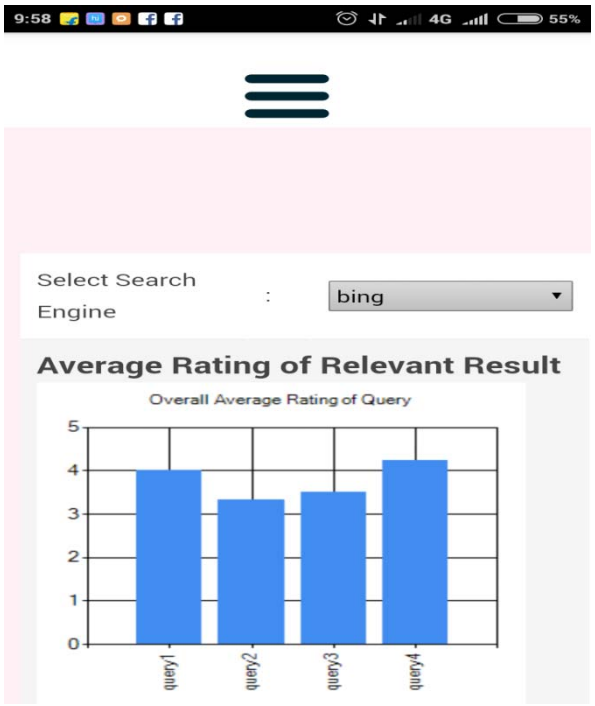


Figure 8: Average Rating of Relevant Result of Bing Search

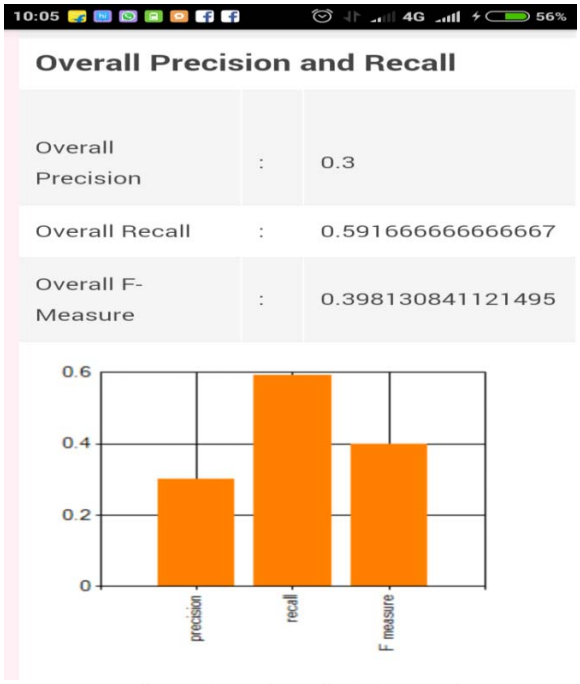


Figure 9: Overall Precision and Recall of Bing Search

From the above Google and the Bing average rating of the relevant results here show the comparison between them,

Comparison between Average Rating of Relevant Results of Google and Bing

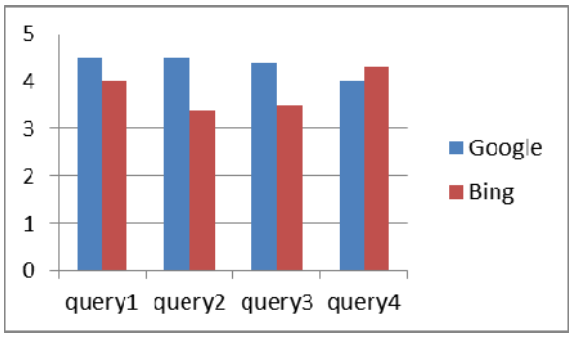


Figure 10: Comparison between Average Rating of Relevant Results of Google and Bing

Comparison Between overall Precision and Recall of Google and Bing

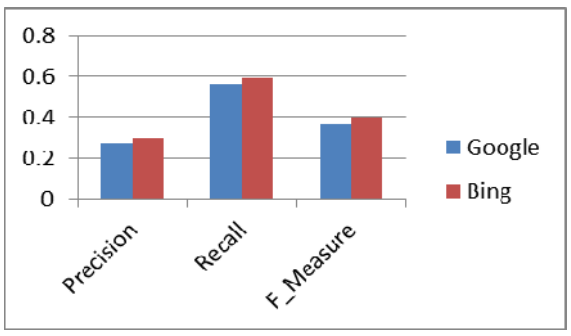


Figure 11: Comparison of overall Precision and Recall of Google and Bing

## VII. CONCLUSION

PMSE extracts and learns user's searand location preferences based on the user's clickthrough. The GPS trajectories are used to adapt the user mobility. GPS locations help to improve retrieval effectiveness, especially for location queries. Two privacy parameters, minDistance and expRatio are proposed. The privacy parameters facilitate smooth control of privacy exposure while maintaining good ranking quality.

## *References*


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

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

[3] 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.

[4] 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.

[5] 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.

[6] K.W.-T. Leung, D.L. Lee, and W.-C. Lee, "Personalized Web Search with Location Preferences", Proc. IEEE Int'l Conf. Data Mining (IC0DDE), 2010.

[7] Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang-Chien Lee, "PMSE: A Personalized Mobile Search Engine", Ieee Transactions On Knowledge And Data Engineering, Vol. 25, No. 4, April 2013