A Review on Personalized Mobile Search Engine using Inertial Sensors

Jamuna B.S.
Department of Computer science
Bangalore Institute of Technology
Bangalore, India
bsjamuna@gmail.com

Vinodh Kumar S
Department of Computer science
City engineering college
Bangalore, India
selvamvinodh78@gmail.com

Abstract-***Evaluating the importance of relevant search result to the user content and location concepts are considered for personalizing web search results. In this paper the problem of extracting geographic knowledge from the web are studied and try to overcome the amount of personal information being shared in location extraction methods. A proposed methodology uses the concept of Inertial sensors for location extraction and thus presents the relevant results without leaking user privacy parameters.***

Keywords—content; location; search results; Inertial sensors

I. INTRODUCTION

A major problem in mobile search is that the interactions between users and search engines are limited by the small form factors of mobile devices. 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. Observing the need for different types of concepts, a personalized mobile search engine (PMSE), which represents different types of concepts in different ontology's. In particular, recognizing the importance of location information in mobile search, concepts are separated into location concepts and content concepts. Various approaches have been considered for extracting the location and content concepts.

Search engine users are more satisfied than ever with the quality of search results, but many are anxious about the collection of personal information by search engines and other websites. Most search users disapprove of personal information being collected for search results or for targeted advertising.

The Pew Internet & American Life survey included several questions probing how respondents feel about search engines and other websites collecting information about them and using it to either shape their search results or target advertising to them. Clear majorities of internet and search users disapprove of these practices in all the contexts we probed.

This work focuses on extracting the location concepts without leaking the personal information to the third party.

The objectives of this work are as listed below:

1. To trace the location of mobile user by preserving his/her personal details.
2. To incorporate inertial sensors which provides the benefit of location tracking without using GPS (Global positioning system)
3. To integrate content concept and location concept, that provides a uniform solution for mobile environment.
4. To preserve user privacy.

The rest of the paper is organized as follows. Related work reviewed in section 2.In section 3 and 4 existing and proposed work is discussed. System architecture is discussed in section 5 and section 6 concludes the paper.

II. RELATED WORK

User cope up with ambiguous queries based on the premise that user general preferences may help the search engine to learn the true intention of the query. The authors Feng qiu and Junghoo cho [1] proposed how search engine can learn user preference automatically based on her past click history and how it can use the user preference to personalize search results.

In [2] an automatic approach to personalizing Web search given a set of user interests are discussed .The approach is well suited for a workplace setting, where information about professional interests and skills can be obtained automatically from an employee's resume or a database using an IE tool or database queries. The mapping framework automatically maps and resolves a set of user interests with a group of categories in the ODP taxonomy. The framework then uses data from ODP to build text classifiers to automatically categorize search results according to various user interests.

Ben Carterette and Rosie jones [3] proposed a model that leverages the millions of clicks web search engines receive each day to predict document relevance. After an initial training phase using a set of relevance judgments paired with click data, the model can predict the relevance of documents that have not been judged. These predictions can be used to evaluate the performance of a search engine.

The authors Ryen W. White, Peter Bailey and Liwei Chen [4] conjointly worked on Predicting User Interests from Contextual Information. The five contextual information sources used are: social, historic, task, collection, and user interaction. Findings demonstrate that the sources perform differently depending on the duration of the time window used for future prediction, and that context overlap outperforms any isolated source.

[5]Investigates the group behavior patterns of search activities based on Web search history data, i.e., clickthrough data, to boost search performance and propose a Collaborative Web Search (CWS) framework based on the probabilistic modeling of the co-occurrence relationship among the heterogeneous web objects: users, queries, and Web pages.

Search engines maintain larger and larger indexes that results in danger of becoming a victim of their own success. Many common searches can return a vast number of web pages, many of which will be irrelevant to the searcher, and of which only about ten or twenty of the top-ranked results will be browsed. The problem is that while pages returned by a search may be relevant to the keywords entered, the keywords generally give only a partial expression of the searcher's information need. [6] describes personalization concept which takes keywords from the user as an expression of their information need, but also uses additional information about the user (such as their preferences, community, location or history) to assist in determining the relevance of pages [6] also present a novel system which uses a client side user profile to provide a personalized ranking of results from multiple search portals.

The web contains many duplicate and near duplicate documents. To detect these documents many duplication detection algorithms are proposed [7] bridges the gap between query independent redundancy and query independent duplication by click through data also identifies three distinct types of redundancy that commonly occur on the web and show how click data can be used to detect these diff types.

The major problem in mobile based search is the interaction between user and search are control by little numeral factors in the mobile plans. To overcome the problem personalized mobile search engine came into existence [8] introduce an association rule mining algorithm to collect the travel related query pattern and travel patterns from original personalized mobile search engine profile.

[9]Introduced new, robust, probabilistic techniques for interpreting clickthrough evidence by aggregating across users and queries. Results show that predictions of relevance are substantially more accurate than the current state-of-the-art search result ranking that does not consider user interactions. Paper also presented a general model for interpreting post – search user behavior that incorporates clickthrough, browsing, and query features.

The work [10] studies efficient query processing in geographic web search engines. A framework for ranking search results based on combination of textual and spatial criteria is proposed for efficiently executing ranked queries on very large collections.

III. EXISTING SYSTEM

In existing works, content concept and location concept are considered to give the relevant results to the users. Various content extraction methods and location extraction methods are incorporated to filter the search results with respect to user interests.

The major lacunae comes in the location extraction concept is where the users were requested to manually specify the location information like latitude, longitude etc.

To overcome the overhead at the client side, recent researchers have proposed GPS technology.

Lacunae are in GPS:

1. GPS works well once your phone finds 3 or four satellites, however it will take an extended time, or not happen at all if you are inside or in an "urban canyon" of buildings that replicate satellite signals
2. Each time he/she seeks for an address or use secondary services on a provided route (all the Starbucks, for example), he/she is unknowingly providing information to the app developer. The data does not personally determine a candidate, however it will be sold to advertisers to make a composite profile of one's demographic kind (where you search, eat, hang around etc.). Firms do not perpetually reveal what is included in the location information they collect, nor do they reveal how they use it.
3. Each time he/she moves with a location service, they are pinging very little bits and bytes of personal information to corporations in order that they will use it for selling functions.

IV. PROPOSED SYSTEM

In order to overcome the lacunae's faced in the previous works inertial sensors technology is incorporated which is used to trace the location of mobile user without using mobile GPS technology. Inertial sensors can keep track location based on other inputs. Nowadays most smart phones come with three inertial sensors:

1. Compass
2. Accelerometer
3. Gyroscope

1. Compass- To help determine the direction.
2. Accelerometer- To report how fast your phone is moving in that direction.
3. Gyroscope-To sense turning motion.

Delivering accurate location while indoors is holy grail for developers of positioning systems. In order to deliver accurate information for indoor location most existing systems rely on expensive fixed infrastructure or multiple radio signals which are often unrealiable.This problem can be overcome by using low power, low cost inertial sensors and algorithms that interpret the motion of the user to trace their path through a building.

A. Approach

The crucial factor was making the best use of the available data through the design and implementation of smart algorithms. The sophisticated mathematics in these algorithms compensates for the increased noise in smaller and cheaper sensors. These algorithms: Harness our understanding of inertial sensors and the mechanics of human motion to classify the user's motion - and use this information to track their path while indoors. Perform data fusion using Bayesian statistics to incorporate any additional information available to improve performance. This potentially allows integration with any other electronic or map-based location system which allows to achieve the same performance that previously required expensive bulky equipment in a prototype that is small enough to be clipped onto a belt.

B.BENEFIT

The resulting positioning system is able to provide indoor location estimates even in the absence of external references, such as GPS or radio signals, which are needed for existing systems. Accuracies to within a few percent of the distance travelled with no external references have been achieved in a range of trials. Potential applications of the technology include locating fire-fighters inside smoke-filled buildings or pinpointing the nearest doctor in a large hospital during an emergency. Consumer applications open up the possibility of using map apps to find the person's location even when GPS is unavailable; also opportunities for tracking of pets, children or the elderly emerge.

V. ARCHITECTURE

Module 1: Client

The operations carried out at the client side should be minimal as subjected to limited computational power of the mobile devices. Clients are responsible for storing the user clickthrough and the ontology's derived from the PMSE server. Simple tasks, such as updating clickthrough and ontology's, creating feature vectors, and displaying reranked search results.

Module 2: Server

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. The data transmission cost is minimized, because only the essential data (i.e., query, feature vectors, ontology's and search results) are transmitted between client and server during the personalization process.

Module 3: Reranking

When a client submits the query to the PMSE server, the query is sent to Backend Search engine .This module matches the search results and feature extractor from client side and sends the results back to the Client.

Module 4: Search results

Results from search engines and extracted features from the client side are manipulated and sent back to the client. Ontology stores this information at PMSE server.

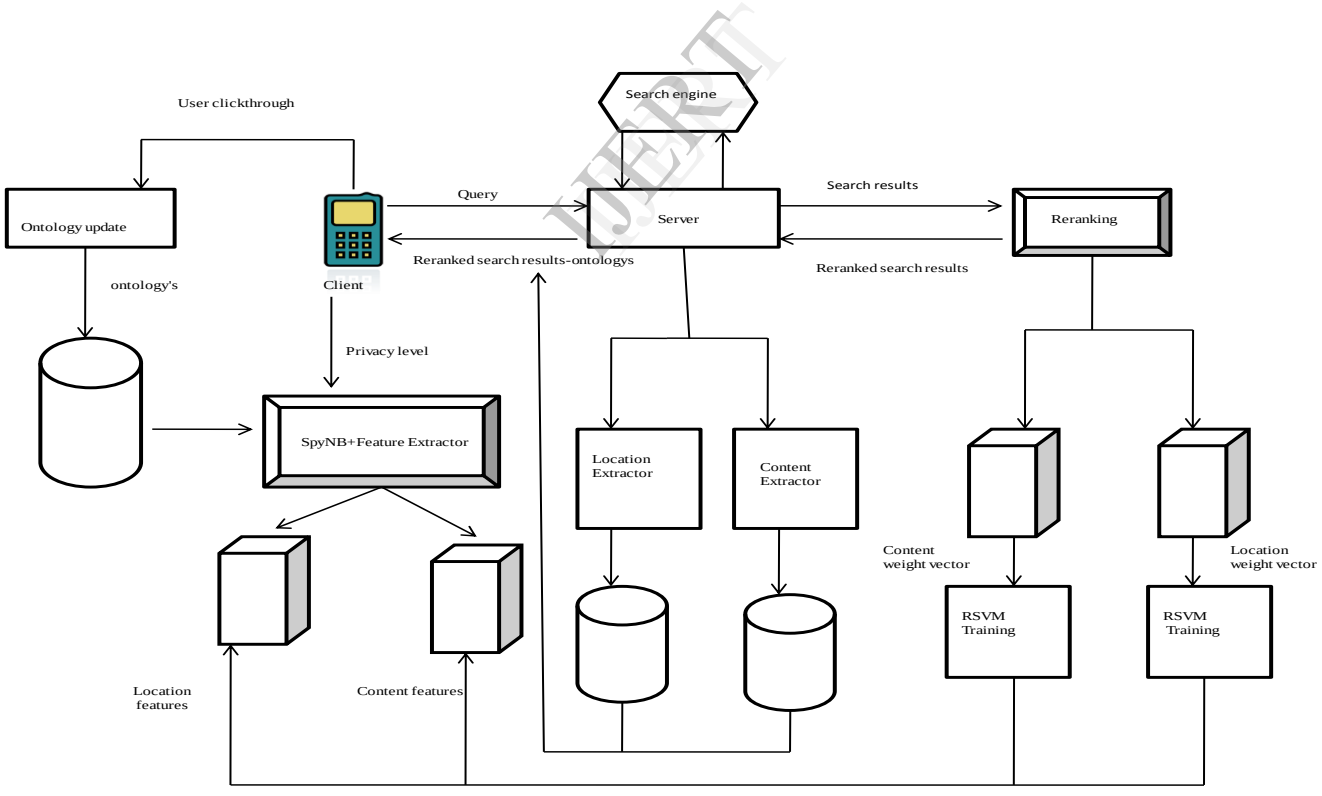

Fig 1: System architecture

Module 5: Ontology

User clicks called click through data along with the extracted location and content concept are stored in the ontology database at the client side. The collected click through data from ontology's is sent to RSVM training on server side. Client can control the amount of personal information transmitted to the server. By defining Privacy parameters.

VI. CONCLUSION

This paper discuss the concept of personalized mobile search engine which represents different types of concepts in different ontology's that helps in recognizing the location and content concepts and also prevent the leakage of user's personal information to third parties at various stages of the process. The paper also discuss the lacunae's in GPS such as privacy can't be maintained and cannot identify or trace the location of mobile number within building.To overcome the above problems the concept of inertial sensors is used which provides the benefit of location tracking and able to provide indoor location estimates even in the absence of external references such as GPS or radio signals.

REFERENCES

[1] Feng Qiu and Junghoo cho. Automatic Identification of User Interest for Personalized Search. *WWW2006, May 22–26, 2006, Edinburgh, UK*
[2] Zhongming ma, Gautam pant, and Olivia r. liu sheng The University of Utah: Interest-Based Personalized Search
[3] Ben Carterette and Rosie Jones: Evaluating web search engines using Click through data.
[4] Ryen w.white, Peter Bailey and Liwei chan: Predicting user interests from contextual information.*SIGIR'09, July 19–23, 2009, Boston, Massachusetts, USA.*
[5] Jian-Tao Sun, Xuanhui Wang, Dou Shen, Hun-Jun Zeng, Zeng Chen: Mining click through data for collaborative web search. *WWW 2006, May 22–26, 2006, Edinburgh, Scotland.*
[6] Kevin Kenoy and Mark Levene: Personaliztion of web search.
[7] Filip Radlinski, Paul N. Bennett, Emine Yilmaz: Detecting Duplicate Web documents using click through data. *WSDM'11, February 9–12, 2011, Hong Kong, China.*
[8] P. Karuppuswamy, Mrs.T.Laksmipriya: Personalized Query Based Search Techniques using Association Rules. *http://www.ijettjournal.org.*
[9] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Rgno: Learning User Interaction Models for Predicting Web Search Result Preferences.*SIGIR'06, August 6-11, 2006, Seattle, Washington, USA.*
[10] Yen-Yu Chen, Torsten Suel, Alexander Markowetz: Efficient Query Processing in Geographic Web Search Engines. *SIGMOD 2006, June 27-29, 2006, Chicago, Illinois, USA*

Indoor location estimates even in the absence of external references such as GPS or radio signals.

As privacy can't be maintained and we cannot identify or trace the location of mobile number in indoors. To overcome the above problems in this paper we use the concept signals.