RESEARCH ARTICLE OPEN ACCESS

Android Based Effective Search Engine Retrieval System Using Ontology

A. Praveena[1], R. Kalaiselvan[2], R. Anitha Rajalakshmi[3].
[1](M.E computer science and engineering, Parisutham Institute of Technology and Science, Affiliated to Anna University, Chennai, India)
[2](Assistant Professor, Department of computer science and engineering, Parisutham Institute of Technology and Science, Affiliated to Anna University, Chennai, India)
[1](M.E computer science and engineering, Parisutham Institute of Technology and Science, Affiliated to Anna University, Chennai, India)

ABSTRACT
In the proposed model, users search for the query on either Area specified or user's location, server retrieves all the data to the user's computer where ontology is applied. After applying the ontology, it will classify in to two concepts such as location based or content based. User PC displays all the relevant keywords to the user's mobile, so that user selects the exact requirement. The client collects and stores locally then click through data to protect privacy, whereas tasks such as concept extraction, training, and reranking are performed at the search engine server. Ranking occurs and finally exactly mapped information is produced to the users mobile and addresses the privacy problem by restricting the information in the user profile exposed to the search engine server with two privacy parameters. Finally applied UDD algorithm to eliminate the duplication of records which helps to minimize the number of URL listed to the user.
INDEX TERMS- Click through data, concept based, location based search, mobile search engine, ontology, UDD.

I. INTRODUCTION

Android is made up of several essential and dependent parts such as a hardware reference scheme that defines the capabilities required for a mobile device to maintenance the software stack. A Linux operating system kernel that delivers low-level interface with the hardware, memory management, and process control, all enhanced for mobile devices. Open-source libraries for application improvement, including SQLite, Web Kit, OpenGL, and a media executive. A run time used to perform and host Android applications, including the Dalvik virtual machine and the core libraries that run Android-specific functionality. The run time is designed to be trivial and efficient for use on mobile devices. An application framework that agnostically uncovering system services to the application layer, plus the window manager and location manager, content providers, telephony, and sensors. A user interface structure used to host and launch applications. Preinstalled applications transported as part of the stack.

A major problem in mobile search is that the interactions among the users and search engines are imperfect by the small form reasons of the mobile devices. As a result, mobile users be wont to to submit shorter, hence, more uncertain queries compared to their web search counterparts. In order to return highly related results to the users, mobile search engines must be capable to profile the user's comforts and personalize the search results permitting to the users profiles. Detecting the need for different types of concepts, present in this project is personalized mobile search engine (PMSE) which characterizes different types of concepts in different ontology's. In particular, knowing the importance of location information in mobile search, we separated the concepts into location concepts and content concepts. For example, a user who is planning to visit thanjavur may subject the query "hotel," and click on the search results around hotels in Thanjavur. From the click through of the query "hotel" PMSE can study the user's content preference (e.g., "room rate" and "facilities") and location preferences ("Thanjavur"). Consequently, PMSE will favour results that are worried with hotel information in Thanjavur for future queries on "hotel." The overview of location preferences offers PMSE an additional width for catching a user's interest and an opportunity to improve search quality for users. To incorporate context information exposed by user mobility, we also gross into reason the visited physical locations of users in the PMSE. Since this information can be suitably attained by GPS devices, it is hence mentioned to as GPS locations. GPS locations play an vital role in mobile web search
This project expresses the unique features of content and location concepts, and delivers a coherent

scheme using client-server architecture to integrate them into a identical solution for the mobile environment. The proposed personalized mobile search engine is an advanced approach for personalizing web search results. By taking out content and location concepts for user profiling, it utilizes both the content and location preferences to personalize search consequences for a user. PMSE combines a user's physical locations in the personalization development. We conduct tests to study the effect of a user's GPS locations in personalization. The results illustration that GPS locations helps increase retrieval effectiveness for location queries (i.e., queries that retrieve lots of location information).

Privacy preservation is a inspiring issue in PMSE, where users send their user profiles beside with queries to the PMSE server to attain personalized search results. PMSE addresses the privacy issue by tolerating users to control their privacy levels with two privacy parameters, minDistance and expRatio. Practical results show that our proposal facilitates even privacy protective control, while maintaining good ranking feature. And mainly it will used to decrease the duplication of the result for user queries.

II. PROPOSED SYSTEM

In the Proposed Model, users search on the net for query, either Area specified or users location, server recovers all the data to the user's Pc where ontology technique is applied. User Pc shows all the relevant keywords to the users mobile, so that user selects the particular requirement. Ranking arises and finally exactly drawndata is produced to the users mobile and apply UDD algorithm to eradicate the duplication of words as well as we get the feedback from the mobile, which helps to rank the data. So that we can provide the Effective Search Mechanism.

A) Algorithms and Techniques

1) ONTOLOGY:

Ontology represents information as a set of concepts within a domain and the relations between those concepts. It can be used to typical a domain and support thoughtabout concepts. Study of relativities. If the user provides the query as hotel, automatically it will bring together the relative answer for the hotel such as hotel location, reservation, facilities, and restaurant etc. Example for ontology is given below

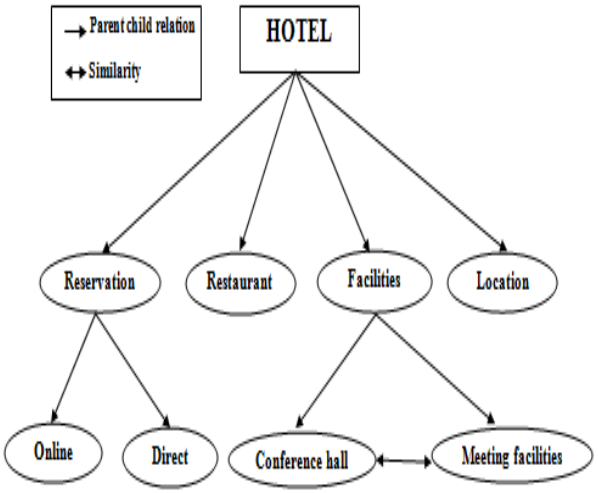

Fig1: ontology technique

2) RSVM (REDUCED SUPPORT VECTOR MACHINE)

In algorithm is projected which generates a nonlinear kernel-based separating surface that needs as little as 1% of a large dataset for its clear evaluation. To generate this nonlinear surface, the whole dataset is used as a control in an optimization problem with very few variables equivalent to the 1% of the data kept. The remainder of the data can be terrified away after solving the optimization problem. This is attained by making use of a rectangular kernel that greatly decreases the size of the quadratic program to be resolved and simples the categorization of the nonlinear separating surface. Here, the m rows of A represent the new m data points while the m rows of A represent aim mordantly reduced m data points. Computational results show that test set correctness for the reduced support vector machine (RSVM), with a nonlinear separating surface that hang on on a small randomly selected portion of the dataset, is enhanced than that of a conventional support vector machine (SVM) with a nonlinear surface that openly depends on the entire dataset, and much improved than a conventional SVM using a small casual trial of the data. Computational era, as well as memory usage, are much lesser for RSVM than that of a conventional SVM using the whole Dataset. Support vector machines have originate to play a very dominant role in data classification via a kernel-based linear or nonlinear classier. Two major problems that challenge large data classification by a nonlinear kernel are the pure size of the mathematical programming problem that needs to be solved and the time it proceeds to solve, even for temperately sized datasets.

The need of the nonlinear separating surface on the whole dataset which creates heavy storage problems that stops the use of nonlinear kernels for anything but a small dataset. For example, even for a thousand opinion dataset, one is confronted by a fully compact quadratic program with 1001 variables and 1000 constraints bring about in constraint matrix with over a million entries. In contrast, our projected approach would typically decrease the problem to one with 101 variables and a 1000 constraints which is cheerfully resolved by a smoothing technique as an unconstrained 101-dimensional minimization problem. This generate a nonlinear separating surface which depends on a hundred data points only, as a substitute of the conventional nonlinear kernel surface which would depend on the entire 1000 points. In [24], an approximate kernel has been projected which is based on an eigen value decomposition of a randomly selected subset of the training set. However, unlike our approach, the whole kernel matrix is generated within an iterative linear equation solution procedure. We note that our data-reduction method should effort just as well for 1-norm based support vector Machines, chunking methods as well as Platt's consecutive minimization Optimization (SMO) .We brief outline the contents of the paper now. In thiSection we describe Kernel-based classification for linear and nonlinear kernels. In Section 3 we outline our compact SVM approach. Section 4 give computational and graphical result that display the electiveness and control of RSVM. Ranking SVM is active in our personalization approach to learn the user's preferences. For a given query, a set of content

Concepts and a set of location concepts are take out from the search result as the document features. Since each document can be signified by a feature vector, it can be preserved as a point in the feature space. Using click through data as the input, RSVM purposes at finding a linear ranking function, this holds for as several document preference pairs as possible.

3. *UDD (UN* SUPERVISED DETECTION ALGORITHM)

Data of high quality is the requirement for the feat of data mining. Data cleaning is vital for improving data quality in data addition. To detect and reject duplicate records is a key stage for data cleaning, and is also an vital problem for improving data quality. Replica records are the records that represent the same object in the real world while are not recognized by DBMS due to different data format or misspell. The resolution of duplicate record detection is to match, combine and remove the fired database records that denote the same entity while with dissimilar data expression and produce the Exact data to user without replication of the data as the user wish. It is mainly used in the search engine system.

Duplicate record detection has expected many disquiets and research. Rule-based technique manually made full rules according to domain-specific application and users. To achieve high accuracy, large quantity of human energy is required. To solve the problem of the boundaries of rules, some methods based on supervised learning are offered, which detect replica records by learning concealed rules and knowledge from data models. The drawback of this method is that it is tough to get sample data. Unsupervised learning is the development for this problem, which has been talk over in related works. M. Elfeky proposed TAILOR method, which use unsupervised k-means method to develop three groups (matched, possible matched and unmatched) and become a decision tree classifier typical with matched and unmatched clusters. Then the typical is used to handle the other records to discover out all matched record sets.

This method cannot assurance the precision of the training models for the decision tree, which moves the results for duplication record detection. LiangGU proposed a better clustering decision model to detect replica records, which added unclear regions for users to control the clustering result of matched and unmatched record sets, concluding better result for three clusters. However, this process only considers the clustering and does not more combine classifier for duplicate record recognition.

The statement of using web search engine example is to highpoint the essential for an algorithm that can switch huge amounts of data and be capable to derive a single set that is most significant to the user query.

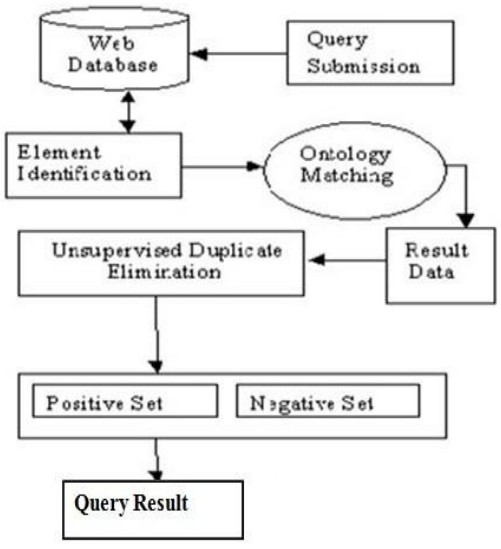

Fig2: UDD algorithm

B. System Architecture

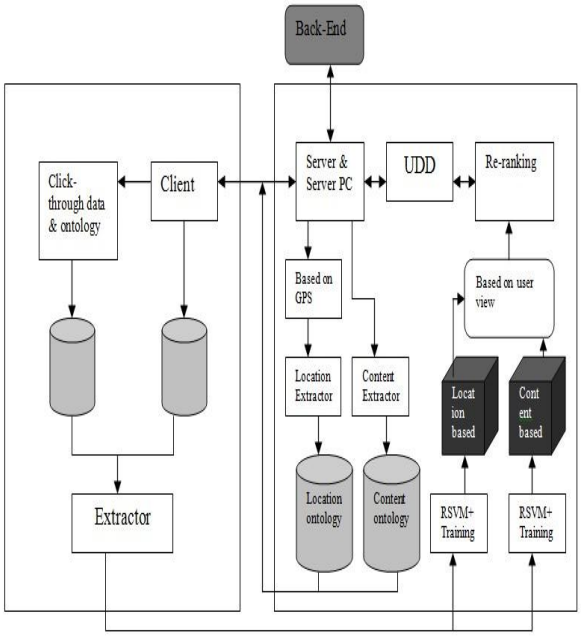

Fig3: system architecture for proposed system

This is the architecture of search engine which is based on the android system. Once user enter in to search engine it will move to the server and server PC and collect those from the database. Then again asked question to the mobile user. After received the answer from the user it will segregate the content as concept based or location based. Then ontology is applied on those concepts and retrieved the data from database. Then RSVM technique is also applied on the result and gives the training to the answer. it will separate the concept such as concept based and location based. Location also classified in to two such as based on user specification or gps based. If the user mentioned the location it will produce the result based on that otherwise it will drag the user location and give the result for the user queries.

III. CONCLUSION

In this PMSE to abstract and absorb a user's content and location predilections based on the user's click through. To adapt to the user mobility, we combined the user's GPS locations in the personalization process. Detected that GPS locations help to improve recovery effectiveness, especially for location requests. And also proposed two privacy parameters, min-distance and expRatio, to report privacy issues in PMSE by permitting users to control the amount of personal information visible to the PMSE server. The privacy parameters enable smooth control of privacy coverage while conserving good ranking quality. For future work, will investigate methods to deed regular travel patterns and query designs from the GPS and click through data to extra enhance the personalization effectiveness of PMSE.

IV. FUTURE WORK

To acclimate to the user mobility, merged the user's GPS locations in the personalization process..Explore methods to abuse regular travel patterns and query patterns from the GPS and click through data to further increase the personalization efficiency of PMSE. Voice based credit for users input instead of manually typing the queries to the mobile by this way we can further improve the efficiency of users input and make it more user friendly to the user. Then also find the result while user in traveling mode

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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 (Loc Web), 2008.

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

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

[8] H. Li, Z. Li, W.-C. Lee, and D.L. Lee, "A Probabilistic Topic-Based Ranking Framework for Location-Sensitive Domain Information Retrieval," Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2009.