A SEMANTIC WEB APPROACH FOR EFFICIENT DATABASE STORAGE USING ONTOLOGY

Sunandhini.S#1 Sangetha.M#2 Manjula.S#3 Arulmozhi.N#4 and Mrs.M.A.Gunavathie #5

#1 B.Tech,Information Technology, Panimalar engineering college, Chennai, India
#2 B.Tech,Information Technology, Panimalar engineering college, Chennai, India
#3 B.Tech,Information Technology, Panimalar engineering college, Chennai, India
#4 B.Tech,Information Technology, Panimalar engineering college, Chennai, India
#5 Assistant Professor, Information Technology, Panimalar engineering college, Chennai, India.

Abstract—Everybody use mobile phones in our daily life for various purpose. Making our mobile to accustom to our usual practices will make it more easier to use and time saving also. We extensively use the search engine for our queries and we get a lot of results that are both relevant and non-relevant our needs. To make web search result more relevant to the user mobile search engines must be able to profile the user’s interest and personalise the searchresults according to user profile. The user has their own personalised web search option that records the click through data with location preferences which improves personalisation more effectively. It is an act of filtering the results you get and giving out the best ones to the user. Hence user need not spend more time in choosing which result to open and search for more relevant result. The user search results are ranked according to their interest or the searchable links are re-ranked according to user-defined concepts for obtaining the desired information faster. Personalized search engine provides more user-preferred information. During user search process the content (user query), location (using GPS) parameters are collected to improve the search accuracy. User interests are classified into content based and location based concepts. The user preferences are organized in an Ontology-based user profile, which are used to adopt a personalised Ranking function for rank adaptation which is used in future. The system architecture is designed to extensively support web services and client applications. The efficiency of multiple profile collection and application of interested user contents for different profiles are incorporated with a view to achieve maximum efficiency.

Keywords—Click through data, location search, search engine, ontology, personalization, user profiling.

I. INTRODUCTION

Search engines are used to search the required data or information. Most of the prominent Web search engines return search results based on a user’s query but don’t take into account the user’s specific interests and needs. Therefore, the similar query from different users who used it will generate the same set of results displayed in the same way for all users. Users have cellphones which allows them to be permanently online anywhere, anytime. The user needs a fast but also almost accurate answer to his query as the large amount of information available causes information overloading, the demand for personalized search for information access increases. Personalized systems address the overload problem by managing customized result display for individual users. This customization helps in getting faster answers that seem to be more relevant to the user query and his interests based on his previous search history. Semantic web assigns the purpose of uniqueness to the information we retrieve and this is a high end technology that allows both humans and machines to understand information. Semantic Web express semantics in a standardized method. Everyone using mobile phones which will be connected to the mobile network and GPS which is present in it, helps to get user preferred output is ranked using PMSE which uses location ontology will rank the web pages based on the user’s location with content requested by user. Ontology is utilized to represent user profiles. Ontology makes intelligent agents to get Web information for users in knowledge-based Web gathering. Search engines are used to search the required data or information. Most of the prominent Web search engines return search results based on a user’s query but don’t take into account the user’s specific interests and needs. Therefore, the similar query from different users who used it will generate the same set of results displayed in the same way for all users. Users have cellphones which allows them to be permanently online anywhere, anytime. The user needs a fast but also almost accurate answer to his query as the large amount of information available causes information overloading, the demand for personalized search for information access increases. Personalized systems address the overload problem by managing customized result display for individual users. This customization helps in getting faster answers that seem to be more relevant to the user query and his interests based on his previous search history. Semantic web assigns the purpose of uniqueness to the information we retrieve and this is a high end technology that allows both humans and machines to understand information. Semantic
Web express semantics in a standardized method. Everyone using mobile phones which will be connected to the mobile network and GPS which is present in it, helps to get user preferred output is ranked using PMSE which uses location ontology will rank the web pages based on the user's location with content requested by user. Ontology is utilized to represent user profiles. Ontology makes intelligent agents to get Web information for users in knowledge-based Web gathering.

II. RELATED WORKS

Applying Co-training to Clickthrough Data for Search Engine Adaptation by Qingzhao Tan, Xiaoyong Chai, Wilfred Ng, Dik-Lun Lee propose a novel algorithm technique, Ranking Support Vector Machine in a Co-training Framework (RSCF). Basically, the algorithm considers the clickthrough data which includes the things in the search result that used or clicked by a user as an input, and produces adjudicate rankers as an output. Analyzing the clickthrough data, RSCF distinguish the data as the labeled data set, containing the objects that scanned already, and the unlabelled data set containing the objects that not scanned. The labelled data is then improved as compare with the unlabelled data to achieve a superior data set for training the rankers. Here the proposed method is capable to get better retrieval superiority of search result by learning from clickthrough data and algorithm does not put in any burden to the users for the duration of the process of web searching. To enhancing the ways to recognize sessions of clickthrough data into log files also to provide individual needs in accumulation to adapting the search engine to users. This is a concept where clickthrough data is analysed and then the search takes place in the search engine here.

Optimizing the Search Engine by Clickthrough Data:

Joachims introduce a method, which mechanically optimize the retrieval accuracy and efficiency of search engines by using clickthrough data. Naturally, a strong information retrieval method giving most relevant or most relatable data first in the ranking followed by less amount of relevant data. The goal is to build a technique that utilize the clickthrough data to the log record of the query search engine along with the log of links on which user clicked on in the existing ranking. The clickthrough data is accessible in great quantity and can be recorded at small cost by using the Support Vector Machine mechanism, a process for learning retrieval functions. In the proposed method we make use of a Support Vector Machine algorithm which helps in widening to non linear ranking functions. In this framework it might also be possible to explore mechanisms that make the algorithm robust against spamming and duplication. The idea is to customise the search results by referring to a user profile that reveals an individual information goal and needs by utilizing theory search patterns. Many personalized web search systems are based on analyzing users' click through. By T.Joachims presented an approach to learn retrieval functions by analyzing which links the users click on in the through data, developed a search personalization method based on users' concept preferences and showed that it is more effective than the other methods that are based on page preferences and paid ranking. By finding the need for different types of concepts, we present a personalized user model which represents different concepts in different ontology bases. By recognizing the importance of location information in user search, we divide the concepts into location based concepts and content based concepts. For example, a user who is planning to travel may issue the query “trekking” and click on the search results about tourism in India. From the click through of the query “trekking,” user can learn the user’s content preference (e.g., “waterfalls” and “museums”) and location preferences (“south india”). The proposal is capable of combining a user’s location by GPS and location preferences into the personalization process.

III. SYSTEM ARCHITECTURE

1. Click through collection at PMSE client

The ontology returned from the PMSE server contain the concept space that models the relationships between the concepts obtained from the search results. They are stored in the ontology database on the client. When the user clicks on one of the search results, the click through data along with the associated content and location concepts are stored in the click through database on the PMSE client. The PMSE server does not know the set of documents that the user has clicked on during the search. This design allows user privacy to be preserved in certain degree. SQL Database is used for database maintenance of clickthroughs. The clickthroughs are stored on the clients, so the server does not know the exact amount of documents that the user has clicked on. Separate databases are maintained for content-based searches and location-based searches, in which each click through data are stored accordingly.

2. Re-ranking the search results at PMSE server

When the user invokes a query to the PMSE client, the query is forwarded to the PMSE server. It obtains the search results from the back-end search engine. The content and location concepts are extracted from the search results and organized into ontologies by connecting the relationships between the concepts with the keywords. The search results are then re-ranked considering all the linking factors. Finally, the re-ranked results and the extracted ontologies for the personalization of future queries are again stored for future references. Re-ranking is performed in two ways: i) Based on higher number of clicks in case of different click values. ii) Displaying top results in case of same number of click values

The number of queries displayed to the user will also differ according to the algorithm and user preference. The user can limit the amount of results he wants to see with this application. Hence the effective re-ranking takes place based on the user profile, clickthroughs and location if only the user requests for it.

3. User Interest Profiling

PMSE uses concept models to categorize the interests and preferences of a user. The concepts are further classified into two different types, namely, content concepts which is used to filter and rank the relevancy of similar search content the Back-end process in content concept: We use Custom Search API to retrieve the results and apply our re-ranking algorithm using the clickthrough data and location concept is done for location-specific searches. For some searches where the results must be retrieved based on the particular location which is tracked by the GPS in the device (e.g. schools, theatres, malls, clinic etc.). This option is selected by the user
from given options. Once this option is selected, the latitude-longitude of the users location are automatically retrieved from the GPS present in the mobile phone and the results corresponding to that location only are returned. We use Places API to retrieve the results and then apply re-ranking algorithm.

4. Diversi ty and Concept Entropy
PMSE consists of a content based and a location based facets. In order to seamlessly integrate the preferences in these two facets into one coherent personalization framework weights of content preference and location preference based on their effectiveness is analysed in the personalization process. The notion of personalization effectiveness is the diversity of the content and location information in the search results. After a extraction process The PMSE consists of a content and a location features it select the situations based on the entropy. that can designate the uncertainty associated among information available from the user’s view. We use the following algorithm and a step by step approach to fine tune the result display process to the user.

IV. ALGORITHM
Input : {S} set of queries {S1, S2, ..., Sn}
{f} set of components {f1, f2, ..., fn};
{H} set of histogram;
Ti time session;
ζ threshold;
D data points
Output: set of detected events by mean shift clustering.

Steps
BEGIN
//Transform query sessions to polar space
For each session Sa Є{S}
Compute radius
T(Sa)-min(T(Sb))
ra=maxb (T(Sa) - min(T(Sb)))
where T(Sa) is the occurring time of query sessions Sa
ra takes value in the range [0,1]
//semantic similarity between two query sessions
let S1 = (Q1, P1) and S2 = (Q2, P2) as
for each principle component fa Є{f}
calculate
θi is restricted to [0, π/2].
//Estimate subspaces of query sessions
For given data point xa, weight wa assigns 1
Wa= 1 + (s(NNxa) + n(NNxa))
Where as
(NNxa) is the variance of xi’s K nearest neighbors along the subspace direction
n(NNxa) is the variance of its neighbors along the orthogonal direction of the subspace.
For all xi Є D && hi Є {H};

//SubspacePruning
Calculate
I(sa) = 1 – [−p alog ha−(1−p) alog va]

//Cluster generation
If
(I(Sa) < threshold
Drop data items
else
Event list ( mean_shift(event_detect=meanshift(D))
END

The algorithm used here is a fair outline for how the queries are aligned and keywords are used to get the most relevant results. The clustering of the results plays a part in getting the top ranked results by data stored and location detected.

V. RESULTS AND DISCUSSIONS
One of the most important evaluation criteria is the Response time. The response time of an algorithm is the time taken to perturb the requests. Response time is a performance measure. Number of users requesting from different queries from different locations. following table shows the time need for getting precise response from server for serving different number of bunch of users. Apart from response time, the accuracy of the response also increases as shown in the graph. Along with the accuracy graph, another component of our search engine is also tested, the GPS search result. The search is carried on with turning the GPS on in our mobile device and the searching for things that we need based on our location. Hence the results we get will be based on the location the GPS gets (ie) the latitude and longitude points. Based on the results we obtained, we get the graph below. Average relevant rank is taken to calculate the graph proceedings. Comparisons done between a regular search engine and our personalized mobile search engine here.
The graphs show the significant variation in performance compared to other peers.

VI. CONCLUSION

The ontology concept is the one that is used to group the data as per the related domain to the user query. So that, the user can search the most relevant information that is required in specific domain they are requesting, using short queries. We proposed PMSE to extract and learn users query content and location preferences based on the user's click through data. To adapt to the user mobility, we incorporated the users GPS locations in the personalization process. We observed that GPS locations help to improve effectiveness in data retrieval, especially for location queries. The system extracts the location concept from users click through data. Hence the semantic web approach for efficient database storage using the ontologies that improve the efficiencies and improve clustering efficiency.

REFERENCES

[10] PMSE: A Personalized Mobile Search Engine Using Content and Location Concept Deepa S1, Arun P S2 1,2 Department of Computer Science & Engineering, Sree Buddha College of Engineering, Alappuzha, Kerala, India
“Personalized Web Search For Improving Retrieval Effectiveness” University of Illinois at Chicago, Chicago, IL 60607

FIRST AUTHOR

S.SUNANDHINI.
B.Tech, Information Technology, PanimalarEngineering College,
Chennai, India

SECOND AUTHOR

M.SANGETHA.
B.Tech, Information Technology, PanimalarEngineering College,
Chennai, India

THIRD AUTHOR

N.ARULMOZHI.
B.Tech, Information Technology, PanimalarEngineering College,
Chennai, India

FOURTH AUTHOR

S.MANJULA.
B.Tech, Information Technology, PanimalarEngineering College,
Chennai, India