PERSONALIZED MOBILE WEB SEARCH ENGINE

Adarsh Swaminathan. B*, Vinod Kumar. G1, R. M. Gomathi1

1Department of Information Technology, Faculty of Computing, Sathyabama University, Chennai, Tamil Nadu, India.

Email: adarshswaminathan@gmail.com

Received on 27-04-2016 Accepted on 20-05-2016

Abstract

Mobile and Internet plays an important role in our day-to-day life. The knowledge of our surroundings is immense so that we will be having a lot to know about. In case, we need to know anything, we will obviously surf them through the Internet and search for the best results. Since a decade, the access to the Internet among the citizens has increased enormously hence providing a lot of results during the search for a particular information. The main objective of this project is to provide an efficient personalised web search engine which provides ontology based results. This includes data extraction from the user’s browsing history and re-ranking process is done in order to get a better search results in the future as per the user’s preferences.

Key words: Ontology, Location Concepts, Semantics, SpyNB, Stemming Algorithm, Re-Ranking Algorithm.

Introduction:

Development of web services is so rapid in recent years and applications like e-commerce, integration of enterprise application and others, plays a significant role with the help of web services. The contribution of web search engines is enormous for retrieval of information from the web. The information obtained from the web is more accurate and rapid with the help of web search engines. But there is more possible ways to improve the web-searches. The existing web search engines do not consider the exact needs of user rather it serves every user the same. It is hard for the search engines to know about the user's exact preferences. Generic search Engine follows "one size fits all" model but it is not flexible for every individual user.

When many users try to give the same query, the results retrieved by the search engines will be same irrespective to which user submits the query but this is does not fit for the users who need different information. The information that is retrieved from the web should be based on user's interest. For Example, if two users are searching for the same
keyword but they may need different information on that keyword. The above situation can be explained by assuming that two users "User-A" and "User-B" need information on the keyword "CELL". But User-A and User-B might be searching the keyword from different perspective. User-A might need the information on the keyword "CELL" as a Biological Term, Whereas User-B might need the information about the battery cells using the same Keyword. When these users give their search query, Lots of documents of various topics based on the user's keyword will be retrieved. Hence it will be very difficult for the users to retrieve the relevant data and it consumes more time in order to search for the user preferred data.

Therefore the introduction of Personalised Web search engine is considered to be a method in order to solve the above situation which retrieves the information based on the user preferences. It analyses the user's information and the search contents to know the about what the user's query refers to.

A model based on user's preferences has to be built in order to get a customized user web search. The process of constructing user profile is an important in order to provide customized web searches. The User profile model is built in order to retrieve relevant data for the user based on the user's web usages.

This system carries a framework for building an efficient user profile that helps in enhancing the user profile by using the background knowledge. Thus, the enhanced user profile will focus on retrieval of more relevant data as per the user's need. This helps in retrieval of relevant web pages with the help of background knowledge and user's previous search query.

**Related works**

The framework of Customized personalised web search engine contains of user profile model that is built on the basis of previous searches or browsing history of the user. This context helps in personalising the web search. This segment displays different methodologies and related works done in this field.

Query processing is the most important part of Generic Web Search Engines. A Lot of the machines uses Query processing in order to do efficient work with their major search engines. The Location Based search engine is based on query processing differs as it requires several combinations of text and it utilizes large volumes of geo-data. There are many algorithms that have been introduced in order to perform effective query processing in Location based search engines. These are integrated to the present web search query processing, and are evaluated on huge collections of real data and query traces. Hence the issue of providing versatile Location based search engines is solved using productive query processing [1].
A new Approach is introduced in order to retrieve search results based on user's preferences by click-through data.

An Efficient search is done by acquired preferences of the user that adapt to the search engine’s Ranking Function. A new technique on SpyNB is developed on the basis of preference mining and it is also based on the assumptions of the results that are clicked by the user which reflects the user's preference. Conclusions cannot be made on the results which the user did not refer to [2].

A new algorithm, Ranking SVM in a Co-training Framework (RSCF) is introduced to retrieve the results based on the click-through data which contains the data about the search results and the consideration of inputs is based on the clicks by the user. Thus, this provides an adaptive ranked search result as an output. RSCF classifies the data as labelled dataset by examining the methodology of click-through data. The labelled dataset includes the data that has been previously scanned whereas the unlabelled dataset includes the data that are not scanned. In order to train the rankers to obtain large data sets, labelled datasets and unlabelled datasets are integrated [3].

Privacy plays an important role for every user accessing the web. The quality user profiles can be built automatically in a more scalable way. These profiles organize the user's interest in a hierarchical manner with accordance to a particular user interest. Two frameworks are used to specify the requirements for the privacy that are proposed to help the user in order to choose the content and also the details that have been exposed to the search engine. Thus it concludes the user profile has improved search quality in comparison with the Standard MSN Rankings [4].

An effective approach is introduced in order to get the best personalised query results which are incorporated from user's preferences. This goal can be achieved by two concepts. First, the online Techniques are developed in order to extract the information from the web-snippets of the search result that is returned from a query and this concept is used find out the relevant queries for the given query. Second, Personalised agglomerative clustering algorithm having two phase is introduced that generates clusters of personalised queries [5].

A new methodology is introduced that personalizes the web search based on the user's preferences and user interests by mining the search results of the user and their click-through’s. Location information plays an important role in retrieval of data in mobile search. Content concepts and Location concepts are organized into ontology which creates an ontology-based multi-facet (OMF) profile and thus it enhances the search accuracy acquiring user's content and location interest.

But, considering the fact that various users and queries might have different importance on content and location information, the opinion of content and location information is introduced to find the amount of content and location
information that is associated with the query. The interest of the user in the content and location information in the result is found out based on click content and location entropies [6].

Framework for proposed system:

The proposed system will refine the searches based on the ontology or semantics related to user’s previous access to the web. This system is a realistic design to train the user profile more quickly and efficiently with the help of re-ranking processes. This system is more secure by controlling the amount of user profile’s details at the client side being passed to the server side, thus providing privacy and maintaining a good ranking quality.

The Figure.1 indicates the proposed system’s process. This system is now implemented in mobile. We use the client and server connection through the static IP address. Whenever the user gives the desired search keyword for the first time. The keyword and the search results will get stored at the server that is connected to the mobile through the static IP. Thus, assuming that the user refers the same relevant data more number of times. We use Stemming Algorithm in order to group the keywords that are frequently used by the user.

Figure 1: Architecture of proposed system

During the retrieval process of data for the second or later time for the same relevant keyword will be more accurate and user needy than the previous normal results. While retrieving the data, Re-Ranking process will be done in order to classify the data with respect to the more number of visits to a particular URL of the search result.

The URL’s in this system are stored based on unique user profiles. Every user will be given their own unique ID in order to access the internet so that their searches will be stored uniquely.

These are the steps involved in retrieval of results for each user profile:

- First, the URL is selected from the User Profile.
Then add the URL in order to categorise them to the Next level User Profile.

Analyse the user search browsing history (Stemming Algorithm).

Update the enhanced User Profile with the user’s most recent data reference.

Retrieve URLs by using re-ranking process (Attribute based clustering algorithm).

Repeat the same process and retrieve URLs for each query of different domains searched by the user.

**About Stemming Algorithm:**

A stemming algorithm is a way of linguistic normalisation, in which the alternative forms of a word are reduced to common form, example. "Engineering", "engineers", "engineered", "engineer" Stemming generally means to cut off characters from the end of the word.

These four words would not be similar if they were tested for equality, however by stemming these words we can decrease them to a more fundamental form,

- engineering --> engineer
- engineers --> engineer
- engineered --> engineer
- engineer --> engineer

Now the words are stemmed they will match for equality, so now if user try searching using the word for engineer, documents on engineering, engineers and engineered would be given back from a stemmed index/database.

**About Re-Ranking Algorithm:**

A re-ranking algorithm is a process of ranking the data according to the user’s relevance. The most or frequently visited URL will be ranked depending upon the number of clicks on that particular URL. The number of times that the user visits the particular URL is recorded and will be ranked with respect to the most frequent or most visited URL by the user.

Thus the result obtained after the process will be having more user desirable and relevant data.

The Fig. 2(a) represents the normal search done for the keyword “Apache”. Thus the user will get several results based on the keyword.

Here, the user refers to “en.wikipedia.org” link in the Fig. 2(a) and later by using search based on user history will tend to retrieve more relevant results based on the previous search.
Figure 2: Screenshots of output of existing and proposed system.

![Figure 2(a)](image1.png)  ![Figure 2(b)](image2.png)

Figure 2(b)

Figure 3: Bar Graph of retrieval of relevant data

![Figure 3](image3.png)

The Fig.3 represents the graphical representation of retrieval of relevant data from the search. The search based on user history is having more relevant results when compared to the normal search. Thus this helps user to get the results faster in a more efficient way. It also reduces the energy consumed for searching and also the time taken to retrieve the most relevant data with the help of Re-Ranking process.

Conclusion:

This is a project which will provide us a better Web search providing enhanced user profiles that uses ontology and suggests the most related web results to the user. This will improve the overall efficiency of the searches with respect to the personal interest of the user. It helps in decreasing the energy consumed by the web browser and also provides privacy to the user by not exposing the user profile details to the server.

Acknowledgement:

We thank Sathyabama University for providing us various resources and unconditional support for carrying out this work.
References


5. Leung, Kenneth Wai-Ting Ng, Wilfred Lee “Personalized Concept-Based Clustering of Search Engine Queries” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 11, NOVEMBER 2008.


Corresponding Author:

Adarsh Swaminathan. B*,

Email: adarshswaminathan@gmail.com