



Volume 2, Issue 3, March 2012

ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcse.com

A Survey on Personalized Meta Search Engine

Preeti Naval¹

Dept. of CS&E

G.C.E.T (MTU)

Greater Noida ,India

er.preetinaval09@gmail.com

Priyanka Singh²

Department of C.S.E,

G.C.E.T (MTU)

Greater Noida ,India

Abstract- Information retrieval is a very important technique. Web Mining is the use full field for this technique by using search engine. Search Engines are widely used for information retrieval, but there are lots of WebPages over the internet and a single search engine cannot covers all the web pages .For this purpose we select the Meta Search Engine. Meta Search Engines are the collection of Search Engines. Meta Search Engine send the user query to all the member search engine and retrieve the result from that Search Engine and returned to user. But now a day's user want a personalize their search. So that the result accurately satisfies the users goal. This paper makes the following contribution to the personalize information retrieval: Based on More result returns, Based on the user experience, and Based on the relevance sorted member search engine and so on.

Keywords- Information Retrieval, Web mining, Personalization, Search Engine.

1. INTRODUCTION

Today the use of the internet is very common .Peoples use it for many purpose like e-banking ,e-ticketing, e-billing ,e-business and for many other application. The most important application is that peoples use it most for Information Retrieval. Information Retrieval is the process to find the documents relevant to an information need from a large document set. Web search engines are the most visible applications for the users. A web search engines are designed to search the information on world wide web and FTP server. Search Engine retrieved the information in the form of web pages, images , information and other types of file and returned in a index format. But there is a spectacular growth in web pages, It is estimated that, there is approximately doubling of web pages of each year. The size and the growth of the web is simply too huge and fast for a search engine to keep its index up-to date. Meta-search engines help the situation a bit by having a better coverage of the WWW. A Meta Search engine is a search tool that sends user requests to several other search engines and(or databases) and merge the results into a single list or displays them according to their source. The size of search results returned by a (meta-)search engine per query is overwhelming. To retrieved the result more accurately according to user query we use personalized search. A process of personalization is to storing and gathering information about web searchers. It is based on the analysis, it analyze the

information and deliver the right information to each user at the right time.

WEB Personalization :- We define Web personalization as a process of web page customization which meets the interests of different users. A personalize search tracks a registered users search history, then adjusts the search results based on that user's demonstrated preferences. For example, someone searching for "jaguar" will see results for the animal if they tend to search for information on animals, but will see results for the car if past searches centered on automotive information. It takes the advantage of information that obtained by the analysis of the user's behavior moving around the web sites in correlation with other data collected within the web context.

By the personalization user can benefit by following ways

- 1- By using the personalized search we can save the time through keeping the track of user searches and by eliminating repetitive task.
- 2- Personalization reduce the work that no longer needed.
- 3- Personalization can provide an efficient and faster way to retrieve the results as per the user requirement.
- 4- It eliminate the irrelevant tasks which divert the user.

- 5- By Personalization User get the more accurate and satisfactory result according to their query.

2. RELATED WORK

In this section we present all the work done in the field of personalization of search engine. In the paper [1] captain Nemo define the personalized search engine on The basis of following Parameters

- A. *Participating Search Engines*:-In this parameter User can select their search engine from the member search engine itself. In Some meta search engine this selection is multiple selection and in some other meta search engine it is single selection as per the user requirement.
- B. *Search Engine Weights*:-According to the captain Nemo ,the search engine can participate in the ranking algorithm with different weight and these weights are set by the user .A lower weight search engine indicate lower reliability and importance for that particular engine .The result returned by this search engine will appear lower in the list.
- C. *Number of Results*:-According to the caption Nemo the user can define the number of retrieved web pages per search engine. Example if any user want less then 20 web pages from meta search engine then user can specify the quantity according to their need.
- D. *Search engine timeout*:-User can set the time out option. If user want that the search can complete in any time duration then user set the timeout period ,after this period the search terminated automatically.

According to the Morris [2] found that groupization is one way to enhance the personalization of web search results by using information from a group of related users .Challenge lies in the identification of related groups of people for the data to be used for personalization. Zhengwei, proposed a user interest model, On the basis of which, the corresponding user interest profile can be created. Dow and Wen [3] designed a system to check profile based and click based search strategies and concluded that profile based methods, both long term and short term contexts are important in improving search performance as compared to click based methods.

3. META SEARCH ENGINE PERSONALIZED STRATEGIES

A personalize search tracks a registered users search history, then adjusts the search results based on that user's demonstrated preferences. Personalization retrieved more accurate result ,that satisfies the use query accurately. To achieve effective personalization ,we must focus on the user needs, interests ,preference, expertise , and the nature of task performed by them. Personalization is not a magical phenomenon or crystal ball interpretation. It is purely based on observed patterns, and resulting probabilities.

In this paper we define how the meta search engine can be personalized on the basis of user interest. Each user has a different requirement when they retrieve the information from the internet. Example-some user wants more result as per the query, some wants less results, some wants fast response time and so on. In this Paper we define the following strategies of personalize meta search engine.

- A. *Basis of the Result returned* :-In this strategy we firstly read the recent n result of each useful search engine returning from database ,and calculate the average of these returning result. Then short the useful search engine on the basis of these average of returning result.

According to this average ,send the query to the first N member s of returning most result . The average is calculated as follow

$$a=(a_1+a_2+a_3...+a_n)/n$$

- B. *Basis on the User experience*:- In this Strategies user can select the member search engine on basis of personal experience.
- C. *Group Based Behavior Based Personalization* – Serve you personalized results based on the actions of all who have made similar queries in the past. Example-If 90% of people who click on a certain search result in response to a specific search query abandoned it after 5 seconds (while others hold user attention much longer), then there's a high probability that the listing in question is not highly relevant to the search query and the search results should be reorganized as a result.
- D. *Virtual Personas/Assistants Based Personalization* – serve you search results through a virtual persona you created.

Example- Think of Star Trek or Terminator if you will. When you talk to a computer, it feels uncomfortable. We often feel the need to assign a personality and persona to it. Now imagine we can customize the appearance and sound of that persona. User can choose the accent, the sex, the age, and much more And sound of that persona. user can choose the accent, the sex, the age, and much more.

- E. *State evaluation strategy of member search engines-* The abnormal information of each search engine and average relevance of return documents decided that the state of member search engines is good or bad. If certain search engine appears sequential abnormal, in this situation a need to amend relevance and calculate average relevance of returning document is required. Evaluate the search engine working condition by average relevance of each search engine results returned and abnormal information[4].
- F. *Based on the User Behavior-* This Search is based on the behavior of user, such as their prior interactions with search result lists, to select what is relevant. to select the relevant result click through is a common behavior based proxy. We collected click-through data to use for this purpose by analyzing the fully anonymized logs of queries issued to Live Search. For each query instance, the logs contained a unique user ID, time stamp, and list of clicked results. While the manually collected relevance data only represents a few hundred queries, because we were able to collect the behavior-based dataset implicitly, it includes information about millions of queries issued by millions of users. Using this dataset we are able to study many different users' interactions with the same self-generated queries in a way that would be infeasible with explicit data. For each Web search result we computed the behavior-based similarity score based on a user's previous history of interaction using the simple domain matching algorithm[6].
- G. *Time Based Personalization* – serve you personalized results based on times it perceives you are typically working. Example- searching for "Trains" has a different meaning at the office (think ... book train tickets to Montreal), then when

you're at home with the kids (think Thomas the Train). Perhaps Google has identified the pattern that most of your queries are business oriented between 9:00am and 6:00pm and more child oriented in all other hours.

- H. *Content Based Personalization-* context based search use a textual representation of people's interests to infer which results are relevant to their current need .There are many ways of representing people's interests, including explicit user profiles, implicit profiles based on previous query history, and richer implicit profiles based on the full content of documents. We consider the use of a very rich interest profile based on the frequencies of terms in previously viewed documents. Such a representation can be obtained from a desktop index such as Google Desktop Search, Mac OS X Spotlight, Windows Desktop Search, X1 or Yahoo! Desktop Search. The system we used to collect relevance judgments based on content-based profiles indexes all of the information created, copied, or viewed by an individual. Indexed content includes Web pages that the individual viewed, email messages that were viewed or sent, calendar items, and documents stored on the client machine. We implemented content-based Web search personalization by modifying BM25 a well-known text-based probabilistic weighting scheme for information retrieval. BM25 assigns weights to individual terms in a document based on their frequency of occurrence in the document and the corpus, and uses these weights to estimate the probability that the document is relevant to a query. When relevance information is available, term weights can be modified by giving additional weight to terms that discriminate relevant documents from irrelevant documents, a technique known as relevance feedback. Relevance information is typically obtained by explicitly asking users which documents are relevant; this kind of feedback can be considered a very simple and short-term content-based user profile, based on documents the user has selected as relevant to the particular query.

CONCLUSION

According to the above description we can say that the user make their search more personalize for the better result. Personalization can improve

the efficiency of search engines ,increase the number of result as per the user requirement, Provide member search engine selection on the basis of user experience. All these strategies improve the searching .

REFERENCES

- [1] Captain Nemo: A *Metasearch Engine with Personalized Hierarchical Search Space*, Stefanos Souldatos.
- [2] Barry Smyth and Evelyn balfe, “*Anonymous personalization in collaborative web search*”, Information Retrieval,2006,Volume9, Number 2, pages 165-190.
- [3] Dou, Z., Song, R., and Wen, J. R. (2007) “*A large-scale evaluation and analysis of personalized search strategies*”. In Proc. of WWW '07, 581-590.
- [4] *Excalibur : A Personalized Meta Search Engine*, Leo Yuen, Matthew Chang, Ying Kit Lai and Chung Keung Poon
- [5] *An overview of personalization in web search*, Indu Chawla.
- [6] *Potential for Personalization* ,Jaime teevan, susan t. Dumais, and eric Horvitz.