



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Personalizing the Web Search for Effective Results Filtered on the Basis of User Needs

Suchita Suresh Dange¹, Grishma Sharma²

M Tech Student, Dept. of Computer Engineering, K.J Somaiya College of Engineering, Vidyavihar, Mumbai, India

Asst. Professor, Dept. of Computer Engineering, K.J Somaiya College of Engineering, Vidyavihar, Mumbai, India

ABSTRACT: World Wide Web contains the largest collection of data. People are completely dependent on the web for retrieving the information they need. The desired result should be at the top of the list is the requirement of all the users. But with the huge amount of data available related to the query keywords, the user has to find the needed information from this huge list. To make this work easier, personalization of the web search according to individual user can be incorporated. This helps to save a lot of time and effort of the user. The study shows that there are various factors which can be considered for getting the personalized result of the web search. These factors mainly include the user profile creation, search history, feedback, dwelling time on a particular page, collaborative search, capturing user's context through semantics and so on. This paper describes the fundamental concepts of personalization, various techniques and methods that can be used to filter the normal search result according to the respective user so that he gets the personalized result of his query. Here major factors such as user profile, search history, feedback and collaborative search results will be provided so that the user can get the desired result belonging to any of this category at any time.

KEYWORDS: Personalization, User profile based search, Search History, User Feedback, Collaborative Search.

I. INTRODUCTION

The growth in the technologies has led to the growth in the need of information. This information can be obtained through World Wide Web (WWW). It is a digital world which is rapidly and continuously growing on all aspects. It is a massive, huge, diverse and dynamic unstructured data repository. Search engines are most widely used for this purpose. Current web search engines are built to serve all users, independent of the special needs of any individual user. With the exponential growth of the available information on the World Wide Web, a traditional search engine, even if based on sophisticated document indexing algorithms, has difficulty meeting efficiency and effectiveness performance demanded by users searching for relevant information. A user fires a query relative to his need to the search engine which after processing it returns the result to the user. Personalized search is nothing but retrieving the results confined to a particular user.

A major deficiency of generic search engines is that they follow the "one size fits all" model and are not adaptable to individual users. That means different users have different context at different point of time, but the results always remain the same and for all. Therefore to save time by providing relevant information according to individual's need, user personalized web search seems to be helpful. For a given query, a personalized Web search can provide different search results for different users or organize search results differently for each user, based upon their interests, preferences, and information needs.

Personalized web search differs from generic web search, which returns identical results to all users for identical queries, regardless of varied user interests and information needs. When queries are issued to search engine, most return the same results to users.

For example, a biologist may query "mouse" to get information about rodents, while programmers may use the same query to find information about computer peripherals. When such a query is issued, search engines will return a list of documents that mix different topics. It takes time for a user to choose which information he/she wants. The concept



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

behind personalized search is that by knowing some things about user, a search engine might refine user results to make them more relevant.

II. RELATED WORK

Now a days information means the web. That means for almost all the information we refer the internet. This information is available on the World Wide Web. The only task is to retrieve the necessary data from it. This process is done by search engines. But the results provided is common to all. The user has to select from the given list which is relevant to his query. This is overcome by the personalized web search which provides the results by taking into consideration the user's context. Various methods are implemented to achieve personalized search results. Some of the previous work done is presented here.

In [1] ranking is based on multiple factors that includes reachability, value and user feedback. The paper focused on three things, Web search, User feedback, Web Markov skeleton process. In the proposed methodology the user is made to register himself with details regarding his area of interest, job, etc. This data is stored in the server using which the search is personalized. In this the user is provided with two options of search, general search and user profile based search.

General Search: The user when opts for a general search he/she is considered as any other user on the web. The results are retrieved from the server and displayed according the ranking proposed. The user's profile is not taken into consideration.

User Profile search: When a particular user opts for user profile based search and gives a query, the query is mapped with all the keywords related to data gathered from the profile of that particular user.

User Feedback: The user is provided with the like button. If the user finds the information useful he will click on that button. This feedback is a very essential component or factor as it would help to exactly know how useful that particular page was for the user and how useful it can be for the other users of similar interest.

Web Markov Skeleton Process: Web Markov Skeleton Process (WMSP) is a random process that contains a Markov chain as its skeleton. It is used for finding the importance of the page which is done by finding the mean time spent on particular page.

In [2] the main focus on the interaction time of the user. Interaction time of the user is nothing but the time spent by him in a particular web page.

The steps followed in this paper are as follows:

1. User submits the query.
2. On the basis of submitted query a number of URL retrieves.
3. for each of the URL web page access
4. Compute access time of each of the visited URL web page.
5. Create web log data for each of the visited web page with their respective access time.
6. Compare the access time with the minimum and maximum threshold access time.
7. If the access time is greater than max threshold time set the access time as maximum threshold time.
8. If the access time is less than min threshold time set the access time to zero.
9. Compute page ranking based on the access time of each web page.

In [3] a dynamic approach of capturing user's interest is presented. Here dynamic refers to the ability to capture user's interest and recommend end results while the user navigates through retrieved documents. For this purpose, SIM- Search Interest Model is used. It involves three main jobs, SIM creation, SIM update, Results re-ranking and recommendation.

SIM creation: The user's interest is represented and maintained during the search activity. Such representation is made by the SIM, which uses the classic Vector Space Model (VSM) and the Term Frequency Inverse Document Frequency (TF-IDF) heuristic to associate the textual information present in the issued query and in the documents related to the captured feedbacks.

SIM Update: The SIM update step occurs whenever the user informs explicitly or implicitly a relevant document from the retrieved search space. This step is responsible for composing the information obtained from the feedback to the current model.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Results Re-ranking and Recommendation: The purpose of re-ranking is to provide a rank that meets the user's interests and that allows to derive an effective recommendation from the retrieved results. Thus, the re-ranking process should sort the documents according to their similarity to the current SIM, representing the degree of proximity to the user's interests.

In [4] special attention is paid to the user's context by actually taking the sense of the query fired by him. Here an alternate query is generated to capture all the senses of main query and assist the user with alternate query. Further personalization is based on user profile, click history, last action performed by user. Alternate query is generated by taking the synonyms of the keywords entered by the user with the help of the word net. This helps to capture the different sense along with the fact that the user might not always know the correct word for his requirement. Along with this personalization through user profile, history is also taken into consideration to obtain good results.

III. PROPOSED APPROACH

Personalization of web search is to carry out retrieval of the required information from the web for each user differently by incorporating his/her interests as a constraint while returning the results to the fetched query. This tells us that every user is treated separately and the results are unique according to the interests of the users. Hence personalized web search take into consideration the user's context rather than just a query. Unlike in traditional web search the keywords from the query are given importance rather than the context behind those words.

A. Main Approaches

There are three basic approaches to be followed in order to get the results personalized that is according to the user's need. They are- re-ranking, filtering and query expansion/modification. These three approaches play a very important role in personalization of the web search.

1. Re-ranking: Here the results obtained from the search engine are re-ranked according to the user interests. In this approach, highest rank is given to the page having the most matching content as per the user's interest. User's interests can be collected directly through user profile or indirectly by learning them through the previous history (links visited by the user). The problem which could arise here is that the quality of the results to be re-ranked should be good.
2. Filtering: In this process the results obtained by the traditional search process are processed. This means that the obtained results are examined and the pages which are irrelevant to the user are pruned. The usability of the page before pruning is determined in the same way as in the re-ranking process i.e. with the use of user profile and/or user's history. The con found here is the overhead of removing irrelevant pages and also at times the potential to remove the relevant pages.
3. Query expansion/modification: Here the query submitted by the user is expanded or modified according to the context in which the user is asking about. The context of the user can be known through the keywords found in the user profile/ history. Modification of the query can be done by taking semantics into consideration. This can be done by finding the alternate words with different meanings through a dictionary. This process needs to be done carefully as it can result into fetching irrelevant results.

B. Techniques used

There are various techniques used for personalization of the web search. These techniques mainly comprise of using the user related information through the use of User Profile, User History, Feedback, Collaborative search, Semantic search. Fig. 1 shows the various techniques used for personalization.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

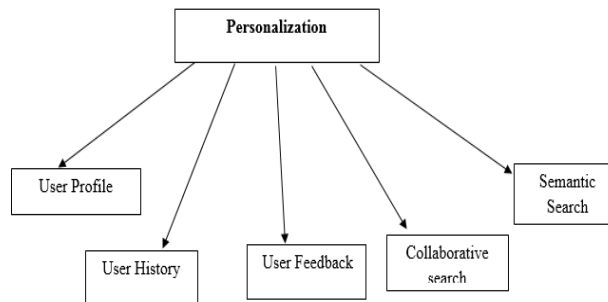


Fig. 1 Techniques used for Personalization

1. User profile

Unlike traditional search process which returns same results to all users corresponding to the query, personalized search provides the results which are according to the interests of the users. In order to understand the user interest, the information about the user is needed. This information about the user will be helpful to find out the relevant results for the user query. For this purpose a user profile is created in which all the information about the user will be stored. Personalized web search can be achieved by checking content similarity between web pages and user profiles. The information stored in the user profile will differ from person to person. This information can include user's age, gender, education, occupation, language, country, address, interest areas, and other information. By considering this information, it is believed that the user might search the related information about his education, occupation, or places nearby his location and so on. User information can be taken implicitly or explicitly.

2. User History

When the user fires query to obtain the results, sessions can be maintained. This session will record all the searches performed by the user. Perfect name for such information is User's History. The user needs to login with his unique credentials and then his search data will be recorded accordingly. By analyzing the history of the user useful information such as which links were visited by the user previously can be found out. Also which links are frequently visited are found. Dwelling time that is time for which the user stayed on the page can be recorded. These information can help to personalize the web the results according to the taste of the result. And hence the results will be altered from the normal search results which will indeed save the user's efforts as well his time.

3. Feedback

Another way of taking the user likes, dislikes or simply interests is by taking the feedback from the user himself [1]. There are websites which give an option to the user after clicking the links at the end of liking the page. User's view is taken into consideration whether he /she liked the information and if he found it interesting. So such feedbacks can be recorded and we can make out which websites or pages display the information that may be liked by the user. Feedback helps us to know whether the user got the information that he wanted. Hence if at all the user wants to read the same information again, the personalized search will show the same pages ranked first which the user has given positive feedback. Feedback from other users can also be considered and it can act as a review for the user.

4. Collaborative search

The growth of the Internet has made it much more difficult to effectively extract useful information from all the available online information. The overwhelming amount of data necessitates mechanisms for efficient information filtering. Collaborative filtering is one of the techniques used for dealing with this problem. The motivation for collaborative filtering comes from the idea that people often get the best recommendations from someone with tastes similar to themselves [13].

Collaborative filtering encompasses techniques for matching people with similar interests and making recommendations on this basis. That means people who have same likings tend to match each other's recommendation. This process involves looking for users who share the same rating patterns with the active user (the user whom the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

prediction is for). Use the ratings from those like-minded users found to calculate a prediction for the active user. This can work great in an organization where one colleague has to learn something new which is already learned by his seniors. By just referring to only links studied by the seniors saves the user's time and efforts to choose from thousands of links available on the internet.

5. Semantic Search

The user may not always frame the query correctly or he can use different words at times for the same results. Words have different meanings when framed in different ways. The context of the user can be something different from what he writes in the query. Hence in personalized search, taking user's context is important. This can be done by taking semantics into consideration. Alternate query can be generated for theonyms for the keywords. User profile can be taken along for better quality of semantic results.

Flow of the process: The user is first checked for registered ones if not he is supposed to register first. Then he has to login before the beginning of his search session, both normal i.e. impersonalized as well as personalized search results are displayed and the respective details of clicked links are stored in the database.

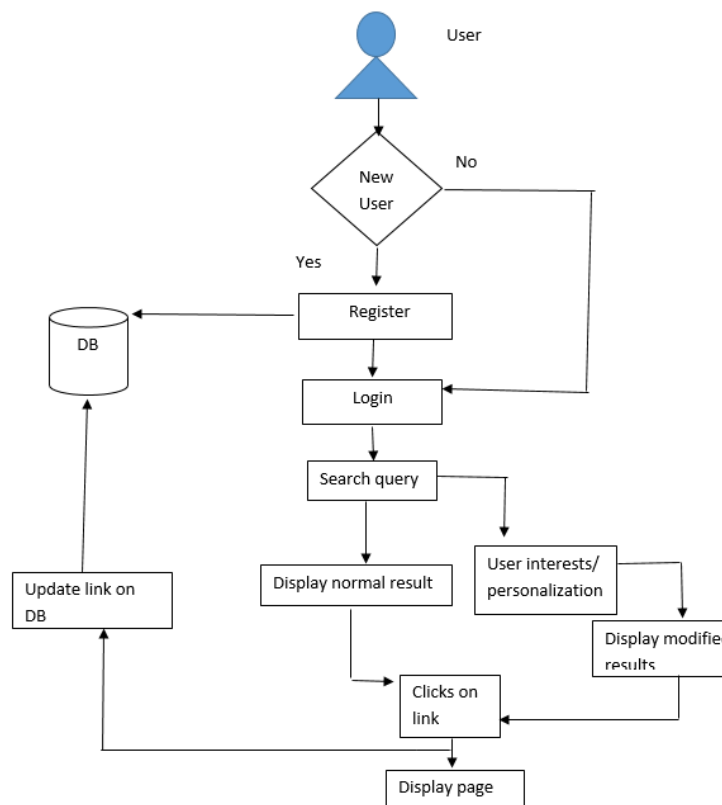


Fig.2 Flow of the process

IV. CONCLUSION AND FUTURE WORK

Personalization plays an important role in finding out the results which are according to the user's context. Hence in this project various factors like user profile, feedback, search history and collaborative search are used to obtain



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

personalized result to the users. By using the profile of the user, his interests are taken into consideration as he might be searching the information related to his interests or hobbies.

User search history is observed to find out which the links that the user visits frequently and to understand what type of content he surfs. Feedback from the user gives direct information whether the result is useful or not while collaborative search works on the basic principle that like people have like interests.

The user is provided with the traditional well as personalized result side by side hence giving him an option of both the search results. Personalization can be useful as it saves time and efforts both of a user in finding the expected results. Providing security along with personalization can be obtained in future work.

REFERENCES

1. Mercy Paul Selvan, A.Chandra Shekar, Deepak R Babu, A. Krishna Teja, "Efficient Ranking based on Web Page Importance and Personalized Search" IEEE ICCSP 2015 conference.
2. Nandnee jain, Upendra Dwivedi, "Ranking Web Pages Based on User Interaction Time" ICACEA 2015.
3. Lucas Pupulin Nanni, Val'eria Delisandra Feltrim, "DESiRe: A Dynamic Approach for Exploratory Search Results Recommendation" IEEE 2015.
4. Shilpa Sethi, Ashutos h Dixit, "Design of Personalised Search System Based On User Interest and Query Structuring" IEEE 2015.
5. Anjali, Ankita Sadhwani, Nidhi Saxena, "A New Approach to Ranking Algorithm Custom Personalized Searching" 2nd International Conference on Computing for Sustainable Global Development (INDIACom) 2015.
6. Priyanka C. Ghegade, Prof. Vinod Wadane, "A Survey of Personalized Web Search in Current Techniques" JCSIT 2014. Vol. 5 (6) , 2014, 7945-7947shima
7. Xiao Han, Leye Wang, Soochang Park, A´ngel Cuevas, Noe´l Crespi, "Alike People, Alike Interests? A Large-scale Study on Interest Similarity in Social Networks" IEEE 2014.
8. Ivan Marcialis, Emanuela De Vita, "SEARCHY: An Agent to Personalize Search Results" IEEE 2008.
9. Durgesh Sawkhedkar, Sonal Patil, "A Survey Paper on Personalized Web Search" IJARCCCE 2016. Vol. 5, Issue 2, February 2016.s
10. Venkata Udaya Sameer, Rakesh Chandra Balabantaray, 'Improving ranking of webpages using user behavior, a Genetic algorithm approach' IEEE 2014.
11. Niclaas Matthijs, Filip Radlinski, "Personalizing Web Search using Long Term Browsing History", Proc. of WSDM'11, ACM, 2011.
12. Chanchala Joshi, Teena Jaiswal, Himanshu Gaur, "An Overview Study of Personalized Web Search" International Journal of Scientific and Research Publication 2013.
13. <https://en.wikipedia.org>