Personalised Web Search using Browsing History and Domain Knowledge

Mayur Bhosale [1] , Navnath Jagdale [2], Shailesh Gawali[3] ,Virendrakumar Dhotre[4]

[1]Savitribai Phule Pune University,
H.S.B.P.V.T.COE, Kashti, Shrigonda 414701,Ahemednagar.
mayur.bhosale63@gmail.com

[2]Savitribai Phule Pune University,
H.S.B.P.V.T.COE, Kashti, Shrigonda 414701,Ahemednagar.
navnathjagdale143@gmail.com

[3]Savitribai Phule Pune University,
H.S.B.P.V.T.COE, Kashti, Shrigonda 414701,Ahemednagar.
shaileshgawali111@gmail.com

[4]Savitribai Phule Pune University,
H.S.B.P.V.T.COE, Kashti, Shrigonda 414701,Ahemednagar.
virendrakumar@gmail.com

Abstract: Different users have different needs when they submit a query to a web search engine. Personalized web search is able to satisfy individual's information needs by modeling long-term and short-term user interests based on user past queries, actions and incorporate these in the search process. A Personalized Web Search has various levels of effectiveness for different contexts, queries, users etc. Personalized search has been a most important research area and many techniques have been developed and tested, still many challenges and issues are yet to be explored. This paper proposes a framework for building an Enhanced User Profile by using user's browsing history and improving it using domain knowledge. Enhanced User Profile is used for suggesting relevant web pages to the user. The results of experiments show that the suggestions provided to the user using Enhanced User Profile are better than those obtained by using a User Profile.

Keywords: *Personalized Web Search, User Modeling, Enhanced User Profile, Domain Knowledge.*

*****

1. INTRODUCTION

Personalization is an effort to uncover most relevant documents using information about user's target, browsing history, domain of interest, query context etc. that gives higher value to the user from the large set of results. Generic search engines are following the "one size fits all" model which is not adaptable to individual users.

When different user fires same query, then same result will be displayed by a typical search engine, no need to know which user submitted the query. This may not be useful for users which require different information. When we searching for the information related to any query from the web, users need information based on his interest. For the same keyword two users might require different pieces of information. Personalized web search is considered as a favorable solution to solve these problems, since different search results can be provided depending upon the information needs and the choice of users. It utilizes user information and search context to learning in which sense a query refer.

It is very difficult for users to define their own interests accurately. Another complex method is based on implicitly observing user's browsing activities and adapting the system according to them. Information stored in user profile can be used to disambiguate or to infer user's query context. Studies in personalized search include [6] which provided the searcher with different search topics and monitored clicked search results so as to learn user's current interests and re-order web search result accordingly. Another approach in [5] models user interests as a vector of weighted terms from visited URLs, and apply a snippet scoring method to re rank search results. In [7] the user profile is composed of each submitted query with its clicked URLs and their corresponding topics, then re-rank is achieved by boosting results with similar topics to topics of queries in the profile that are relevant to current query. Our project proposes an architecture for constructing user profile and enhances the user profile using background knowledge. This Enhanced user profile will help the retrieve focused information. It can be used for suggesting good web pages to the user based on his search query and background knowledge.

2. LITERATURE SURVEY

Literature survey presents different approaches and the related work done in the field of personalized web search. For providing personalized web results, Micro Speretta et al.,[1] implemented a wrapper around the search site that collects information about user's search activity and builds user profile by classifying collected information (query or snippets). They have used the sepro files which will be used to re-rank the search results and the rank-order of the user-examined results before and after re-ranking were compared. They found that user profiles which will be based on queries and user profiles which will be based on snippets both were equally effective and re-rank provides 34% improvement as compare to rank-order.

Fang Liu et al.,[2] identified that current web search engines do not consider the special needs of user or interest of user and proposes an ovel technique which uses search history

of user to learn user profiles. This work user's search history for learning of user profile and category hierarchy for learning of a general profile and then combines both profiles to categorize user's query to represent user's search in mention and to disambiguate the words used in query.

Chunyan Liang [3] also identifies that different users may have need of different special information, when they use techniques of personalized web search can be used to solve the problem effectively.They uses three approaches which are Rocchio method, K- nearest method and support Vector Machines have been used in [3] to build user profile to present an individual user's preferences and found that K- nearest method is better than others in terms of its efficiency and robustness.

Rakesh Kumar et al., [4] proposed framework for personalized web search by considering user's interest to suggest the relevant pages. For this, Rakesh Kumar et al., [4] considers user's profile and domain knowledge. By using data of web usage of the user User profile is created. To store information about different categories and different domain the Domain knowledge is used. Information which is obtained from user profile is classified according to categories and domain. DMOZ directory is used as source for preparing domain knowledge.

Existing System:-

When different user fire some query same result will be displayed by a typical search engine, no issue which user submitted the query. This might not be appropriate for the user which required different requirement / information .While searching for the information from the web, user needs information based on their needs /interest. For the same keywords user may require different information. This fact can be explained as fallows –

A programmer and a biologist may need information based on **"VIRUS".** But they are from entirely different fields. A programmer is searching for the **'malicious software'** and Biologist is searching for the virus that is **'microorganism' .** Generic search engine provides a number of documents for different queries.

3. PROPOSED SYSTEM

For personalized web search we propose a framework which considers individual's interest into mind and enhances the traditional web search by suggesting the relevant pages of user's interest. We have proposed a simple and efficient model which ensures good suggestions as well as promises for effective and relevant information retrieval. We have also implemented the proposed framework for suggesting relevant web pages to the user.

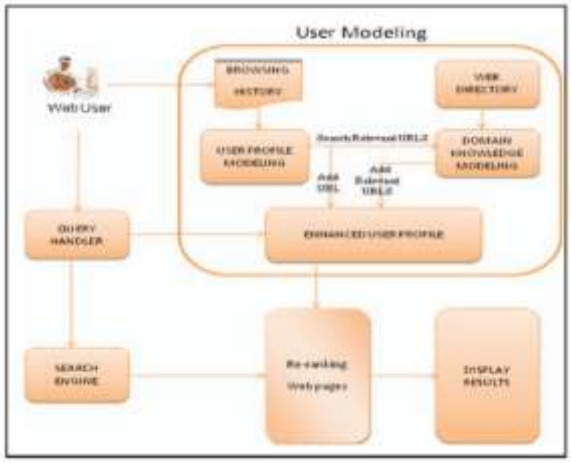

Fig.(1) Proposed Method Architecture.

3.4 Algorithm

Following steps explain the process of preparing the Enhanced User Profile. Perform the following steps for each document (URL) in user profile :

STEP 1- Select the URL from the User Profile.

STEP 2- Add the URL to the Enhanced User Profile.

STEP 3- Find the cosine similarity of this URL with the URLs present in user specific categories from the Domain Knowledge base.

STEP 4- Rank the URLs on descending order of cosine similarity.

STEP 5- Retrieve top 20 URLs.

STEP 6- Calculate the average of the cosine similarity of these top 20URLs.

STEP 7- Add only those URLs to the enhanced user profile whose similarity value is above the average value and which are top 20 URLs.

To summarize the process, for each URL (form user profile) most relevant URLs from the user specific Domain Knowledge category are added to prepare enhanced user profile. The cosine formula used for the similarity of the URL u in User Profile to each web pages dj in Domain Knowledge is as follows:

$$\text{Cosine}(dj,u) = \frac{<dj*u>}{||dj||*||u||}$$

A cosine similarity measure is the angle between the web page in User Profile u and the document vector dj.

4. RECOMMENDING PERSONALIZED PAGE RANKING STRATEGIES

There are several ways to retrieve the documents relevant to the query. The research efforts on re-ranking web search results are categorized into the following classes of strategies.
(i) Explicit relevance judgments

(ii) Implicit relevance judgments
(a) Content-based implicit measurestim
(b) Behavior-based implicit measures

4.1 Explicit Relevance Judgments

The trouble-free way to verify whether a result retrieved for a query is relevant to the user is to explicitly ask that user. Explicit judgment allows us to examine the uniformity in relevance assessments across judges in a controlled setting. Advantage of this method allows us to examine the consistency in relevance analysis across judges in a controlled setting. Following are the limitations:
(i) It is inconvenient for people to give explicit judgment because it consumes additional time and effort from the users.
(ii) It is difficult to gather sufficient data to generalize across a broad variety of people, tasks and queries.
(iii) It is captured outside an end-to-end search session.

4.2 Implicit Relevance Judgments

Implicit data can be generated by users' interaction with their service. Implicit measures are easier to collect and allow us to explore many queries from vast variety of searchers.

5. CURRENT APPROACHES ON THE SEARCH PERSONALIZATION:

Current information retrieval and data mining research have explored the enhancement of user's web experience from several directions. One direction is to create a better structural model of the web, such that it can interface more efficiently with search engines [16]. Another approach is to model user behavior as to predict user's interests better [3]. In this paper, we review relevant studies with the most significant research and focuses on analyzing the user's behavior to construct the user profile for personalization search results.

6. RERANKING SEARCH RESULT

Page [8] proposed the first personalized web search by modifying the global PageRank algorithm with the input of bookmarks or homepages of a user. Their work mainly focuses on global "importance" by taking advantage of the link structure of the web. Brin et al. [9] suggested the idea of biasing the PageRank computation for the purpose of personalization, but it was never fully explored. Bharat and Mihaila [10] suggested an approach called Hilltop, which generates a query-specific authority score by detecting and indexing pages that appear to be good experts for definite keywords, based on their links. Hilltop is designed to enhance results for popular queries; however, query terms for which experts were not found will not be handled by the Hilltop algorithm. Haveliwala[11] used personalized PageRank scores to enable "topic sensitive" web search. They decided that the use of personalized PageRank scores can improve web search, but the number of hub vectors (e.g., number of interesting web pages used in a bookmark) used was limited to 16 due to the computational requirements. Kamvar[12] determined that PageRank could be computed for very large subgraphs of the web on machines with limited main memory. Jeh and Widom [13] scaled the number of hub pages beyond 16 for finer-grained personalization.

7. RESULTS

In the absence of standard benchmark datasets which is suitable for our problem, we have designed our own dataset. In our Experiment, we have used the browsing history of 10 different users from our collage, 6 from Computer department and 4 from Mechanical department. Our Experiment is carried out for 50 queries of which 35 queries from Computer domain and 15 queries from Mechanical domain.

8. CONCLUSION

We have proposed a framework for personalized web search using Domain Knowledge and User Profile. Based on the Domain Knowledge and the User Profile, the system keeps on updating the user profile and thus builds an enhanced user profile. For suggesting relevant web pages to the user this Enhanced user profile is then used.

The proposed framework has been implemented by conducting few experiments. These experiments displays that the performance of the system using enhanced user profile is better than those which are obtained through the simple user's profile. Our work is important as it enhances the overall search effectiveness, catering to the personal interest of the user's.

In future this framework may be applied for re-ranking the web pages retrieved by search engines on the basis of user importance. For personalized web search we also apply collaborative filtering in our framework.

9. REFERENCES:

[1] M Speretta and S Gauch, "Personalized Search Based on User Search Histories",Proceeding Of International Conference on Web Intelligence, pp. 622-628, 2005.
[2] F Liu, C Yu and W Meng, "Personalized Web Search for Improving Retrieval Effectiveness", IEEE Transactions On Knowledge And Data Engineering, pp. 28-40, Volume 16, 2004.
[3] C Liang, "User Profile for Personalized Web Search", International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1847- 1850, 2011.
[4] Rakesh Kumar and Aditi Sharan," Personalized web search using Browsing history and domain Knowledge", 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT).
[5] Matthijs, N.and Radlinski, F. (2011) Personalized Web Search using Long term Browsing History. *Proc.4th ACM Int. Conf. Web search and data mining*, Hong Kong, China, 9-12 February, pp. 25-34, ACM Press, New York, NY, USA.
[6] Hoeber, O. and Massie, C. (2009) Automatic Topic Learning for Personalized Re-Ordering of Web Search Results. *Proc. 6th Atlantic Web Intelligence Conf.*, Prague, Czech Republic, 9-11 September, pp.105-116, Advances in Intelligent and Soft Computing, Vol. 67, Springer-Verlag, Berlin.
[7] Li, L. Yang, Z. and Kitsuregawa, M. (2009) Rank Optimization of Personalized Search. *Journal of the Database Society of Japan*, 8, pp.35-40.
[8] Page, L., et al., *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report, (1999).
[9] Brin, S., R. Motwani, and T. Winograd, What can you do with a web in your pocket, *Data Engineering Buletin*, **21**(2), (1998), pp.37-47.
[10] Bharat, K. and G.A. Mihaila, When experts agree: using non-affiliated experts to rank popular topics. *ACM Trans. Inf. Syst*., **20**(1), (2002), pp.47-58.

[11] Haveliwala, T.H., et al., *Evaluating strategies for similarity search on the web*, in *Proceedings of the 11th nternational conference on World Wide Web*, ACM: Honolulu, Hawaii, USA, (2002), pp. 432 - 442.
[12] Kamvar, S., T. Haveliwala, and G. Golub, "Adaptive methods for thecomputation of PageRank". *Linear Algebra and its Applications*, 386: pp.51-65, (2004).
[13] Jeh, G. and J. Widom, "Scaling personalized web search", *Proceedings of the 12th international conference on World Wide Web*, ACM: Budapest, Hungary, (2003), pp.271-279.

Author Profile

Mayur Bhosale pursuing the Bachlor Degree in Computer Science Engineering from H.S.B.P.V.T.COE,Kashti under SPPU.

Navnath Jagdale pursuing the Bachlor Degree in Computer Science Engineering from H.S.B.P.V.T.COE,Kashti under SPPU.

Shailesh Gawali pursuing the Bachlor Degree in Computer Science Engineering from H.S.B.P.V.T.COE,Kashti under SPPU.

Virendra Dhotre, he has done his MTech from Walchand college of engineering, Sangli. He is currently pursuing the PhD in Pune university.