

Personalized Web Search using User Identity

Nidhi Saxena
Department of IT
SRMGPC, Lucknow, India.

Shalini Agarwal
Department of CS
SRMU, Lucknow, India.

Vinodini Katiyar
Head, Department of IT
D.S.M.N.R.U, Lucknow, India

ABSTRACT

Now a days the websites are available in bulk and a single search can give various different results. There still exist problem of getting results based on user importance in order to save time and complexity while searching. The personalized search built on user unique identification can solve the current problem to large extent. In this paper we have taken a unique personalization approach. We identify user and makes search according to user interest based on previous searches made by him. We present a personalized web search framework UIBP (USER IDENTIFICATION BASED PERSONALIZATION). The comparison of our model with others shows that our search agent will prove more user friendly as it will make the searching fast, easy and provide accurate results. Therefore it is an enhancement in the field of web mining.

Keywords

UIBP, User Search History, Web Search Log, Ranking, Personalization, User Profile.

1. INTRODUCTION

The web is medium through which variety of information can be accessed in the whole world. The data on the web is growing rapidly as the user has variety of information need. The increased amount of data on web, lead to the problem of finding relevant information in easy way. Different user prefer different result based on interest. The query like “crane” have two meanings and therefore search results must vary according to one’s interest. Therefore personalized search is a solution to all these problems. On the other hand personalized search is a solution to information overhead by building, handling and representing information for individual users.

There are various queries made to search engine which are short and not properly specified. Different people have different intensions for same query. For example two persons making same query “Ram”. One is a religious person and having interest in Lord Ram , may want to read some article on Lord Ram while the other need to know about RAM “ Random Access Memory . Both uses short query but will get results related to all possible meanings of RAM. Therefore there is a necessity to provide full description while making query each time. The searching time as well as accuracy of results both affect.

To consider different interest and to customize results based on user interest there is a necessity of personalization in search engine.

With the growing demand of personalization, various search engines developed which provides user interest based results. Some of them were explicitly collecting user interest and some implicitly store user searches in search log in order to find user interest.

Unluckily, it was found that explicit collection of user feedback is not supported by all users. On the other hand the implicit approaches for finding interest proved better. In the present scenario there are methods in personalization which results by ranking the link according to their visits and time spend. These methods make distinct queries and gives relevant results but there exists a major problem of unique identification. Till now users are identified on the basis of system used for searching and not the person.

Our system is an enhancement to personalization by uniquely identifying the user and then recording in search log. There by giving results based on user interest.

2. BACK GROUND

Personalization is the method of providing information to the user on the basis of user’s interest. User’s interest can be collected explicitly by feedback or it can be implicit that collect information based on user behavior. Such data are retained in user profile, analyzed and used as a sample for future search results.

Previous work on personalization includes both explicit and implicit profiling techniques. Several approaches were developed in Explicit personalization . PERSONA [1] where user profile relies on relevance feedback. Each positive and negative feedback serves two function. First is to refine the set of searches and re-rank the results. Second is to build the user’s profile.

Query expansion and using social media for personalization [2] method uses user profile created by taking the tags and the web documents. Work is carried on developing user profile based on semantic extraction from news article. It can be applied to Social web system and has impact on personalization [3]. There are personalization approaches based on explicitly creating user profile by collecting some information from user [4]. For profile developing some cases document present at user machine is used for personalization [5]. The assumption is that if a operator keeps a document on his/her machine there is a strong prospect that the user is involved in those documents.

Explicit construction of user profile has a drawback of incorrect and insufficient information.

Implicit personalization was started earlier with the creation of user profile.

There is a research based on search history which investigate three conditions [6]–

Session- All previous work in current search session.

Historic- All previous work apart from current session.

Aggregate- All previous work before the current query.

One of the technique was to find conceptual similarity between each document and user interest[7].When Web server

logs and web contents were used for describing user navigation patterns and predicting user's future requests[8]. It calculate importance degree of a webpage to a user. Two measures were introduced 'Frequency' and 'Duration'. Further primary goal of Automatic Identification[9] is to learn the user's topic preference vector from past click history and use this vector to personalize search ranking,. Search engine called (UCAIR)[10] developed that can perform eager implicit feedback, e.g., query extension based on previous queries and instant outcome reranking based on clickthrough information. Weighted Association Rules[11] assign a significant weight to each page built on the time spent by user on each page and visiting occurrence of each page, taking in to account the amount of interest instead of binary weighting. STRank [12] takes use of semantic relevance and time frequency for website ranking.

Later on after 2010 the work was carried on further on profile creation . Some algorithms keep record of each page visited by the user and time devote on that page[13] [14][15] . The algorithms are effective for web page prediction. The algorithm provides efficient and optimized web path traversal for various users based on their path navigation behavior. More optimized version of earlier algorithm was when click event on the each page by the user is also stored and used for efficient ranking [16]. The algorithm shows that as the number of parameter increases, the effectiveness also increases. A new ranking algorithm Ratio Rank [17] was introduced in 2013 in which in links weights and out link weights are used with the consideration of number of visit count which is a better approach for personalization. New Enhanced page rank algorithm [18] considers link of the webpages. Here the relevancy of the webpages displayed is high, as the user behavior is also considered to rank the webpages

3. USER IDENTIFICATION BASED PERSONALIZATION (UIBP)

Our model is an extension to the existing user profile creation method used in implicit personalization. Apart from creating record of web page visited and time spend , the unique identification record is also maintained in search log history . This record distinguishes two user's search results on the same machine. The outcomes so received will be more specific and ranking efficiency is also improved. In section A we first describe the user profile data. Section B contains the framework of the search engine.

The idea is taken in order to make two persons as different users on same machine. Suppose a system in a home or office is accessed by two different users. The results will be same in earlier approaches but it differs as we have extended the ranking with user identification.

3.1 User Profile data

The system is working on user identification and

maintaining log based on identification. For each user on the system a separate profile is maintained . It covers user identification record along with query given by user, searched web page and time devoted on them .

This idea is taken in order to make two person as different users on same machine. Earlier IP address were recorded but they were considering a user as a machine.

Suppose a system in home or office accessed by two different people. The results were same in earlier approaches no matter the search being made by different person, but it

differs as we have extended the ranking with user identification.

3.2 Frame work of UIBP

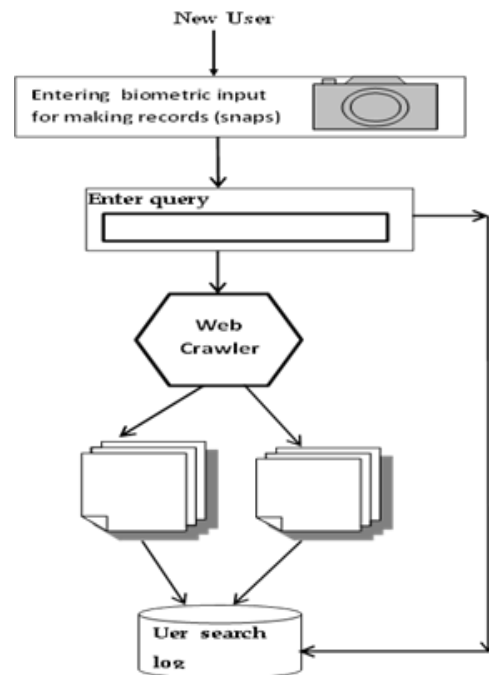


Fig 1: New User Search

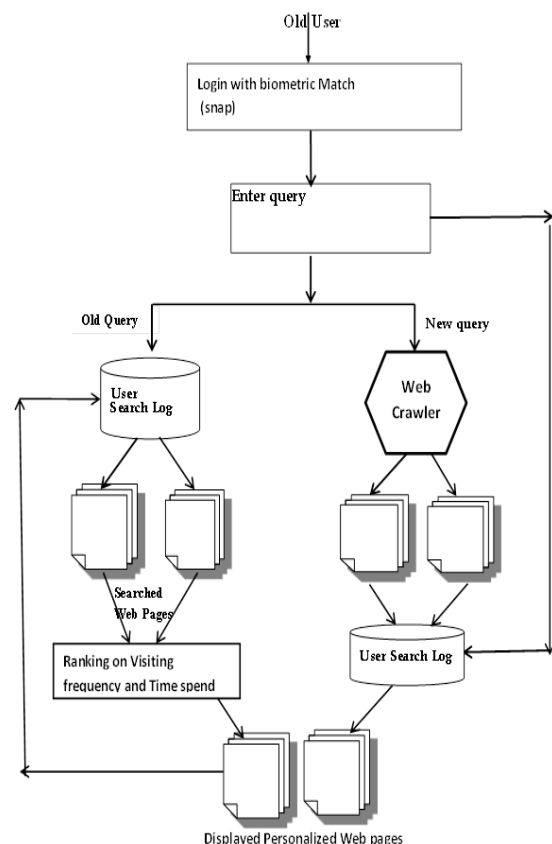


Fig 2 :.Old User Search

The given models illustrates the working of New user and Old user to our search engine separately. The new user who enters the system first makes it records with his snap which is kept for further recognition and thereby personalization. The user enters the query and the web crawlers crawls the web to find the matching web pages. The selected pages by user are stored in web log along with user identification record.

In the old user search the first step is to recognize the user with the help of face recognition. After that there can be situations-

1. New Search
2. Old Search

In the case of new search the results will be similar to new user and selected results will be stored in web log. But the user is old and the search is the one which was made earlier by him once or more. Then personalization algorithm is applied for visiting frequency and time spends and the Ranking would vary each time the similar search made by him.

4. RESULTS AND EVALUATION

In order to properly test the information filter, the raw stream data collected from user search logs must be cleaned . Test are performed taking a user identification match and the records of search made by him taking data from personal web log on machine.

4.1 SEARCH SPACE

In our search engine, we are giving each users a personal search space by providing individual search history that to in a customized form. Each user will have a user login account, according to which they will get a separate search space. All his further search will be affected by this personal search history. If a user visits a link regularly and most recently then this link will get displayed to him at the top of the search results.

4.2 PERSONALIZATION RESULTS

In Google , personalization is not included that will affect the users search in future. Like for a query “mouse” the link <http://www.webopedia.com> is coming as 7th link. If a user regularly visits this link then it should come at the top of the search results. But on testing it on google, if the user regularly visits this link then also it is presented as the 7th result only.

But in our search engine as user visits a link, his click events on a link and time of visits will increase which affects the rank of the results.

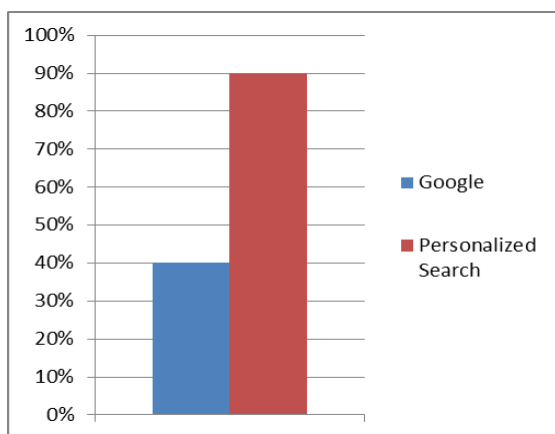


Fig 3: Comparison :Google with UIBP

The given graph illustrates that our personalized search engine results high in efficiency. Hence, our search engine rates high when compared on the discussed parameters-

1. Speed
2. Ranking
3. Efficiency

Table 1: Search Engine Technological Comparison

	Google	Yahoo!	UIBP
Speed	Very fast	Fast	Very fast as lengthy query not required
Ranking	Page Rank, hypertext Matching analysis	Keywords, click popularity	Based on individual search log
Efficiency (based on personalization)	Low	High	Very High Due to user identification, frequency and time spend data

5. CONCLUSION

This paper aims to develop a general framework to make use of the content and graph information effectively by leveraging information retrieval, machine learning, and knowledge discovery techniques for real-world applications, especially query log analysis and expertise retrieval.

A novel framework is proposed for modelling , which intends to discover better user importance and there by ranked the search results The intuition behind this model is common clicks and greater time spend URLs are of greater value than rarely visited pages and less time spend URLs .Further here the emphasis is on individual interest by maintaining separate search log for each user on single machine.

Based on this intuition, the given model presents the face recognition based search engine which create separate log by recognizing the faces who made search earlier from particular system. In future, this study may develop strong approaches for implementation to improve the ranking process.

6. REFERENCES

- [1] Tanudjaja, F. and Mui,L. 2002. Persona: A Contextualized and Personalized Web Search. In IEEE Proceedings of the 35th Annual Hawaii International Conference on System Sciences, pp. 1232 – 1240.
- [2] Zhou,D., Lawless,S. and Wade,V. 2012. Improving Search via Personalized Query Expansion using Social Media. Springer ,Vol. 5 No.3, pp 218-242.
- [3] Abel,F., Gao,Q., Houben ,G.J and Tao,K. 2011. Semantic
- [4] Enrichment of Twitter Posts for User Profile Construction on the Social Web, Springer, Vol. Part II, pp 375-389.

- [5] Srinvas,C.. 2012. Explicit User Profiles for Semantic web search using XML. *IJERA* ,Vol. 2 No. 6,pp 234-241.
- [6] Bhadoria,R.S. , Sain,D. and Moriwal, R. 2011. Data Mining Algorithm for personalizing user's profile on Web. *IJCTEE* , Vol. 1 No.2.
- [7] Bennett,P.N. , White, R. W. , Chu, W. and Dumais, S.T , Bailey,P. ,Borisyyuk, F. and Cui, X.. 2012. Modeling the Impact of Short- and Long-Term Behavior on Search Personalization. In *ACM Proceedings of the SIGIR Conference on Research and development in information retrieval*.
- [8] Speretta, M., Gauch,S.. 2005. Personalized Search Based on User Search Histories . *IEEE/WIC/ACM. In Proceedings of the International Conference on Web Intelligence*.
- [9] Liu,H.,Keselj,V.2007 Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users future requests.*Data & Knowledge engineering Elsevier*, Vol. 61 No. 2, pp 304– 330 .
- [10] Qiu,F and Cho,J..2006 .Automatic Identification of User Interest For Personalized Search' in *ACM. In Proceedings of the 15th International conference on World Wide Web*.
- [11] Shen,X. ,Tan,B. and Zhai,C. 2005. Implicit User Modeling for Personalized Search.. In *ACM Proceedings of 14th international conference on Information and knowledge management*.
- [12] Forsati, R., Meybodi, M.R. and Neiat, A.G. 2009. Web Page Personalization based on Weighted Association Rules .In *IEEE Proceedings of the International conference on Electronic Computer Technology*.
- [13] Guo,H. Chen,Q. Wang,X., Wang, Z. and Wu,Y.2009.STRank: A SiteRank Algorithm using Semantic Relevance and Time Frequency. In *IEEE Proceedings. of the International Conference on Systems, Man, and Cybernetics San Antonio, TX,USA* .
- [14] Liu,H. and Keselj,V.2007. Combine Mining of Web Server logs and web contents for classifying user Navigation patterns and predicting user's futurerequests, *Data and Knowledge engineering ACM Vol. 61 No. 2, pp 304-330*.
- [15] Khanchana,R. and Punithavalli, M.2011. An Efficient Web Page Prediction based on Access Time Length and Frequency. In *IEEE Proceedings of the 3rd International Conference on Electronics Computer Technology*.
- [16] [15] Matthijs, N. and Radlinski ,F. 2011. Personalizing Web Search using Long Term Browsing History. In *ACM Processings. Of the 4th International Conference on Web Search and Data mining* .
- [17] Agarwal,R., Arya, K.V. and Shekhar, S.2011. An Efficient Weighted Algorithm for Web Information Retrieval System.. In *IEEE Proceedings of the International Conference on CICN* .
- [18] Singh,R. and Sharma,D.K..2012. Ratio Rank :Enhancing Impact of Inlinks and Outlinks. In *IEEE Proceedings of the International Advance Computing Conference*.
- [19] Singh,R. and Sharma,D.K..2013. Enhanced Ratio Rank Enhancing Impact of Inlinks and Outlinks. In *IEEE Proceedings of the International Advance Computing Conference* .