

NOVEL APPROACH to PERSONALIZATION with UNIQUE USER PROFILE

Nidhi Saxena^{#1}, Vinodini Katiyar^{#2}, Mohd. Husain^{#3}

[#] Department of Computer Science, ^{#1}SRMGPC, ^{#3}JIT
^{#1}Tiwariganj, Faizabad Road, Lucknow, India

^{#3} Barabanki, India

¹nidhi.shivansh@gmail.com

²mohd.husain90@gmail.com

*Department of Computer Science SRMU
Deva Road, Lucknow, India

²drvinodini@gmail.com

Abstract— In recent year, the massive influx of information onto WWW has facilitated user, not only retrieving data, but also ascertaining knowledge. However user usually suffers from the problem of information overload. Therefore personalization is a general need in web search. This paper focuses on personalization techniques and introduces an approach based on unique record entry taking voice as a parameter for recording data on client side web log history. In this we have taken ranking on the basis of user web search history taking frequency, time spend and click event as a parameter for personalization. The system proposed is evaluated using web search records on client web log. It is an attempt to improve the accuracy of personalization by keeping unique and more precise information of user in web search log.

Keywords— User Search History, Web Search Log, Ranking, Personalization, User Profile.

I. INTRODUCTION

The web pages are growing rapidly day by day. Thus it is the necessary to have efficient information retrieval tool which can retrieve the relevant information to the user from the vast data source. The requirement of user differs in context to the interest, preferences and information need. There came the necessity of personalization. Personalization is the practice of customizing the Web search experience of individual users. User information need differs on the basis of interest and requirements. For instance a zoologist user may use query “kingfisher” to find the information about the bird Kingfisher while on the other side someone may be searching for the flight Kingfisher may give the same query. Personalization differentiates between users. The main challenge is to identify the user and to read the mind of user. This paper aims to uniquely identify the user and to keep unique record in order to improve the performance of Ranking and to provide search based on interest.

II. PERSONALIZATION APPROACH

Personalization combines the features of Page Rank with that of user choice and interest. Personalization is based on individual observation. In web search personalization information about user from the profile or cookies is collected and re-rank to provide satisfactory result to user. On the basis of user interest user profile are generated. They may be implicitly or explicitly created. In both the cases user profile is generated. In explicit, profile is created on the basis of certain feedback given to search system. In implicit user profile is generated on the basis of past user search history. User Search Profile may be on the server side or Client side. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. RELATED WORK

Researches are conducted on dynamic personalization and re-ranking is done in order to merge it with original ranking [10].

Previous work on profile based Personalization focus on improving the search utility. Many profile based methods are available. Some techniques like Ratio Rank[1,2], iHITS [3] uses web structure in link and out link weight with visit count for personalization. Some algorithms were based on creating user search log history of sites visited, frequency and time spend on that side [4, 5, 6, 7, 8] and there by calculating the user interest. In order to remove unwanted pages (not of user interest) different fields are included in web log for better web history and to improve the performance of filtering and ranking.

There are personalization methods based on explicitly creating user profile by collecting some information from user [9]. In some cases document present at user machine is used for personalization [8]. The assumption is that if a user keeps a document on his/her machine there is a strong possibility that the user is interested in those documents. In all above work a client profile is maintained either on the

server or on the client end. In web log on server side search is recorded based on IP address [11].

IV. RESEARCH APPROACH

We propose a system which makes user identification unique based on which ranking is done. According to study IP address is taken as a unique identification for personalization but on single machine more than one user can make a search therefore web log on server based on IP address does not provide accurate personalization. In order to overcome it, our attempt is to use voice as recognition. The proposed system is considering voice for identification of user and taking it for personalization. For each new user, voice will be recorded on client machine. User profile on client machine will have a record according to voice and query made by user. For the purpose of personalization and ranking following parameters will be considered.

A. Voice Matching

For the purpose of Personalization the search result data will be filtered from client side profile using Voice matching. The result ranked by Search engine will be re-ranked based on personalization.

B. Visiting Frequency

It is the number of times a page is visited by user. It is calculated through web Search log entry. The page visited maximum have the highest frequency. It is assumed that page with highest frequency is of interest to user.

C. Time Spend

It is the time spent on each visited page. It is assumed that more we spend time on a page the higher is the interest. The record is maintained in the web history.

D. Click Event

Sometimes user opens a page but has no interest. This will show high but inappropriate results. Therefore new parameter click event on page is taken. Beside this, mouse movement is also documented in order to judge the interest of user on particular page.

The system first search the user voice in the web search log on client side. If the voice matches, the query is searched. If the profile contains similar search then ranking is done based on parameters specified else a new record with the search will be added to the web search log.

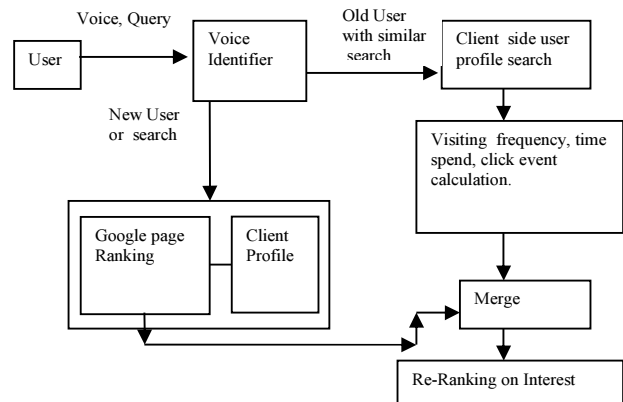


Fig.1. System Data Flow Diagram

On the basis of System Data Flow the proposed algorithm is shown below-

If New user (Voice matching)

Then

Google Page Rank (R_j)

Add data in client web log

Return (R_j)

Else

If query match not present in web log

Then

Add data in client web log

Google Page Rank (R_j)

Return (R_j)

Else

Calculate max.Visiting Frequency (MF)

Calculate max. Time spend (MT)

Calculate max. Click event (MC)

Ranking on interest ($R_i = MF+MT+MC$)

Google Page Rank (R_j)

$R_j = R_j + R_i$

Return (R_j)

V. EVALUATION METHOD

For the purpose of evaluation two web search logs are compared. The first one is the Web search engine standard transaction log and other is the proposed client side web log having some static data.

Uid	Date	Time	Search url
ce00160c04c41580 87704275d69fbedc	25/Ap r/2014	04:08:5 0	http://www.flykingfisher.com
38f04d74e6511375 87e9ba3f4f1af315	25/Ap r/2014	04:08:5 0	http://www.kingfisherworld.com
ce00160c04c41580 87704275d69fbedc	15/Ap r/2014	06:08:1 5	http://a-z-nimals.com/animals/kingfish
38f04d74e6511375 87e9ba3f4f1af315	25/Ap r/2014	07:08:1 5	http://en.wikipedia.org/wiki/Mouse

TABLE 1. WEB SEARCH ENGINE STANDARD TRANSACTION LOG

Table1 showing the data recorded on standard transaction log when search is made from any machine. IP address of the client's computer is user identification. This is sometimes also an anonymous user code address assigned by the search engine server, which is our example in Table 1. Personalization in this way can only on IP address and there remain a problem that more then one user may be using the same machine therefore it is not user specific personalization.

TABLE 2. CLIENT MACHINE PROPOSED WEB SEARCH PROFILE

Voice address	Query	Search_url	Time Spend	Click Event
Pass1.mp3	Kingfisher	http://www.flykingfisher.com	10	15
		http://www.kingfisherworld.com	8	10
	mouse	http://en.wikipedia.org/wiki/Mouse...	12	8
Pass2.mp3	mouse	https://www.apple.com/in/magicmouse	11	9
		http://en.wikipedia.org/wiki/Mouse	12	9
		http://www.ncbi.nlm.nih.gov/genome	8	15
	Kingfisher	http://a-z-Nimals.com/animals/kingfish	9	10
		http://www.birdlife.org/datazone	8	5

Table2 is keeping voice record as a unique identification on client side. Web search history is maintained on the basis of Voice. Two different users based on voice Pass1 and Pass2 is shown. Search is carried by these users on two words “King fisher” and “mouse” having two different meaning. Voice Pass1.mp3 was interested in Kingfisher flight and computer mouse while other Pass2.mp3 was interested in kingfisher bird and mouse animal. Both works on same client machine but their unique records are

maintained in file by their voices. On the basis Voice their search records are distinguished .Ranking will become easier by having unique identification. Relevancy will be judged by taking other parameters stored in Web search profile.

VI. CONCLUSION

It is evident that in order to personalize a search, user profiling is essential. However, all the methods that are used for this purpose have a drawback. User was identified as a single client machine and the web log is maintained on the basis of IP address. This creates a problem of interest conflict. Our proposed system is an attempt to remove this conflict by creating search profile on each machine with information related to each user making search on it. For the user identification voice is recorded and recognized each time the user make search and the web log is filtered based on identification. In order to give more specific result the filtered web pages are ranked on the basis of visiting frequency, time spend and click event.

In future we will be implementing the proposed technique by creating the respective web log on client machine for storing search data by different users on which searching and Ranking can be done efficiently thereby improving the performance of Ranking of Web Search Engine.

The proposed approach has a wide future scope as there is a potential of using such profile in different areas, such as personalizing advertisements, suggesting news articles to read or perhaps even interesting social networking groups to join.

REFERENCES

- [1] Ranveer Singh , Dilip Kumar Sharma, “Ratio Rank: Enhancing the Impact of Inlinks and Outlinks”, in *Proc. of IEEE International Advance Computing Conference*, 2013.
- [2] Ranveer Singh ,Dilip Kumar Sharma,,”Enhanced Ratio Rank Enhancing Impact of Inlinks and Outlinks”, in *Proc. of IEEE Conference on Information and Communication Technologies*,2013.
- [3] Zhuang Ziming “iHITS: Extending HITS for Personal Interests Profiling”, in *Proc. of 19th International Conference on Advanced Information Networking and Applications*, vol. 2, pp. 747-751, 2005.
- [4] R Farsati , M.R. Meybode, A .Neiat, “ Web Page personalization based on Weighted Association Rules”, in *Proc. of International Conference on Electronic Computing Technology*, 2009.
- [5] Haibin Liu, Vlado Keselj, “Combined Mining of Web Server logs and web contents for classifying user navigation patterns and predicting user’s future requests”, in *Proc. of Elsevier Data and Knowledge Engineering*,,2007.
- [6] R. Khanchan, M. Punithavalli, “An Efficient Web Page Prediction based on Access Time Length and Frequency”, in *Proc. of IEEE 3rd International Conference on Electronics Computer Technology*

- (ICECT), 2011.
- [7] Nicolaas Matthijs, Filip Radlinski, “Personalizing Web Search using Long Term Browsing History”, in *Proc. of ACM WSDM’11*, pp.9–12, 2011.
- [8] Robin Singh Bhadoria, Deepak Sain, Rahul Moriwal, “ Data Mining Algorithm for personalizing user’s profile on Web” , *IJCTEE*, Vol.1, Issue 2, 2011.
- [9] C. Srinvas, “Explicit User Profiles for Semantic web search using XML”, *IJERA*, vol. 4, pp. 234-241, 2012.
- [10] Zhongming Ma, Gautam Pant, Sheng R. Olivia Liu, “Interest Based Personalization Search”, *ACM*, vol. 25, no. 1, 2007.
- [11] J. Jansen Bernard, “Search log analysis: What it is, what's been done, how to do it”, *Library & Information Science Research 28, Elsevier*, pp.407–432, 2006.