

Design and Implementation of User Identification Based Personalization

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

[#] Department of Information Technology, ^{#1}SRMGPC, ^{#3}JIT

^{#1}Tiwariganj, Faizabad Road, Lucknow, India

^{#3} Barabanki, India

¹nidhi.shivansh@gmail.com

³mohd.husain90@gmail.com

^{*} Department of Information Technology, D.S.M.N.R.U.

Lucknow, India

²drvinodini@gmail.com

Abstract— The use of Search Engine is now a days common among each and every society. On the other side the websites are available in bulk and a single search can give various different results. There still exist problem of fetching results based on user interest in order to save time and complexity while searching. The personalized search based on user unique identification can solve the current problem to large extent. In this paper we have taken a novel 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) used for our work and designed an algorithm. The experiment on web search shows that our search agent will prove more user friendly as it will make the searching fast and easy. Therefore the developed search engine is an enhancement in the field of web mining.

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

I. INTRODUCTION

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 effect.

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

With the increasing demand of personalization, number of 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.

Unfortunately, it was found that explicit collection of user feedback is not supported by all users. On the other hand the implicit methods 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 developed 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.

II. BACK GROUND

Personalization is the process 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 behaviour. Such information are stored in user profile, analysed 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] in 2005 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. A new form of probabilistic profile RLT profile was introduced [3] which can be used to express major entries of web search such as

users, websites, and requests. These profiles capture topic and reading level distributions, which then use in amalgamation with search record data to characterize and compare diverse entities. 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 [4]. There are personalization approaches based on explicitly creating user profile by collecting some information from user [5]. For profile developing some cases document present at user machine is used for personalization [6]. The assumption is that if a operator keeps a document on his/her machine there is a strong possibility 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 [7] –

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[8].When Web server logs and web contents were used for classifying user navigation patterns and predicting users' future requests[9]. It calculate interest degree of a webpage to a user. Two measures were introduced 'Frequency' and 'Duration'. Further primary goal of Automatic Identification [10] 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) [11] developed that can perform eager implicit feedback, e.g., query expansion based on previous queries and instant outcome reranking based on clickthrough information. Weighted Association Rules[12] assign a significant weight to each page built on the time spent by user on each page and visiting frequency of each page, taking in to account the degree of interest instead of binary weighting. STRank[13] 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 like FTPW algorithm [14] keep record of each page visited by the user and time devote on that page [15] [16] [17]. 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 [18]. The algorithm shows that as the number of parameter increases, the effectiveness also increases. A new ranking algorithm Ratio Rank [19] 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[20] considers link of the webpages. Here the relevancy of the webpages resumed is high, as the user behavior is also considered to rank the webpages.

III. USER IDENTIFICATION BASED PERSONALIZATION (UIBP)

We develop an intelligent client-side web search agent (UIBP) that can perform web search based on personalization. Experiments on web search show that our search instrument can improve search accuracy over the popular Google search engine.

Our developed model is an extension to the existing user profile development technique used in implicit personalization. Apart from creating history about page visited and time spend , the unique identification record is also maintained in search history . This record distinguish two user's search on the same machine. The results 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 along with algorithm. C section contains developed simulator.

The idea is taken in order to make two person 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.

A. 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.

B. Frame work of UIBP

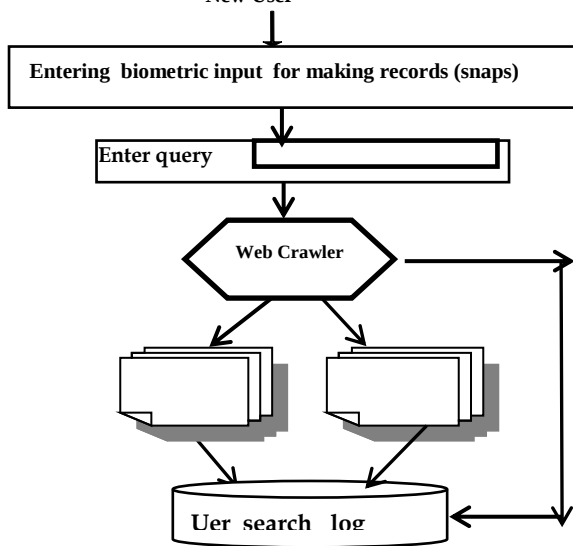


Fig 1. New User Search

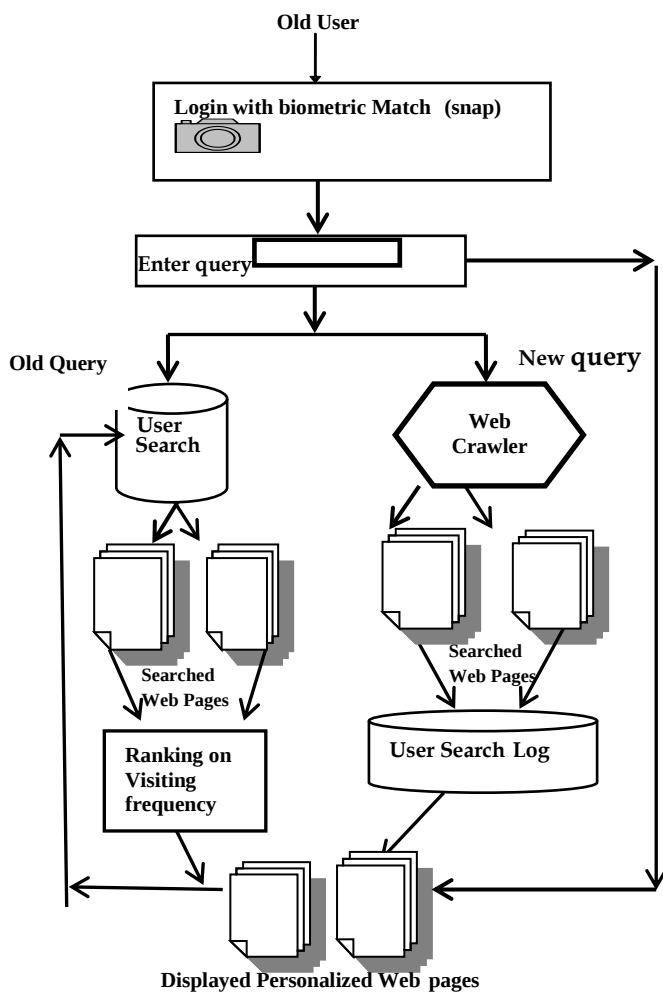


Fig 2. Old User Search

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 recognise 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 onces 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.

Algorithm for Search Model-

If New user (Face matching)

Then

Crawled (Page Ranked)Results (R_j)

Add data in client web log

Return (R_j)

Else

If query match not present in web log

Then

Crawled (Pagr Ranked)Results (R_j)

Add data in client web log

Return (R_j)

Else

Calculate max.Visiting Freq.(MF)

Calculate max. Time spend (MT)

Ranking on interest (R_i = MF+MT)

Crawled (Page Ranked) Results (R_j)

R_j= R_j+ R_i

Return (R_j)

C. Developed Simulator

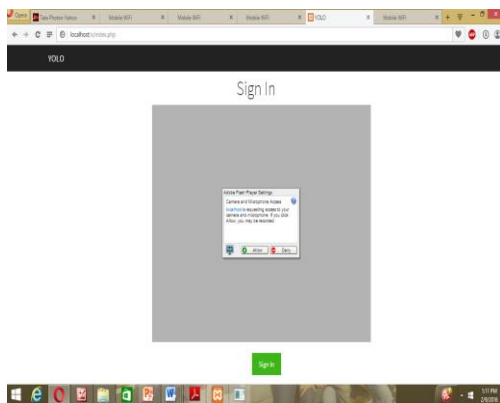


Fig. 3 User Login

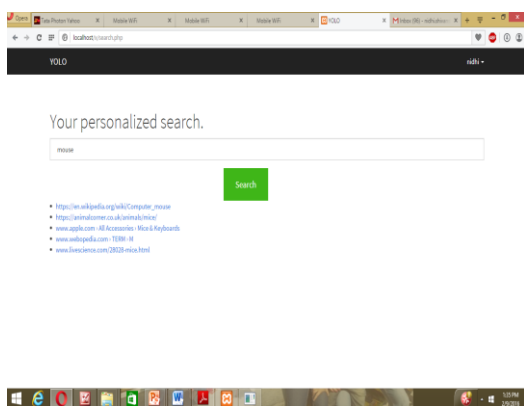


Fig. 4. Query on Search Engine

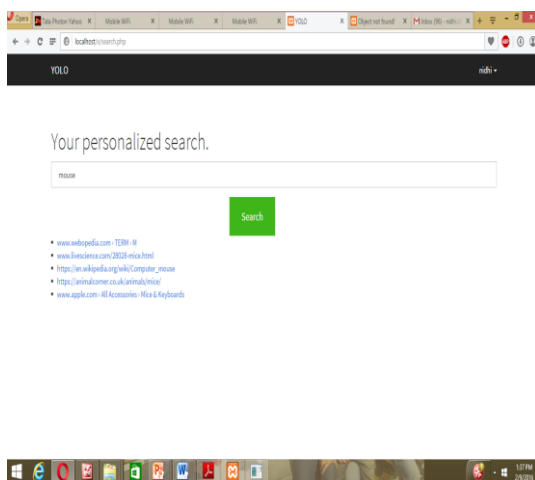


Fig. 5 Ranking Results

IV. RESULTS AND EVALUATION

User criteria	Searched	Time spend
User1	1 st search B->D->G	B=2min, D=3min, G=2min
	2 nd search G->D	G=2min, D=3.5min
User2	1 st search A->C	A=2min, C=1.5min
	2 nd search C->F->A	C=1min, A=2.5min, F=1min
User3	1 st search E->H	E=3min->H=2min
	2 nd Search H	H=1min

The performance of UIBP can be evaluated by comparing the performance of FTPW algorithm [14] and proposed algorithm which differ in number of parameters considered for experimentation.

FTPW is considering Visiting Frequency and time spend for ranking while the proposed algorithm is combining each user personal interest based on unique identification along with visiting frequency and time spend

The performance of algorithm is assessed using N cross Validation method, based on this method accuracy is calculated using the total number of correctly classified objects verses the total sample produced to classify.

The experimental setup uses three users on one machine and weights are plotted against various parameters for the query “Kingfisher”. As it is a generalized name which has three meaning- Beer, Bird or Airlines.

User Interest

User 1	Bird
User2	Beer
User3	Airlines

We consider 8 displayed webpages when we search by the keyword “kingfisher” denoted by A, B, C, D, E, F, G, H.

A, C, F-> associated to beer

B, D, G-> associated to bird

E, H-> associated to Airlines

Now a comparison is made between FTPW and User Identification based proposed algorithm (UIBP)

TABLE 1. USER SEARCH AND TIME SPEND

After ranking by FTPW results will be –

TABLE 2. FTPW RESULTS

Users	Ranked Result
User1	D->A->G->H->C->E->B>F
User2	D->A->G->H->C->E->B>F
User3	D->A->G->H->C->E->B>F

The ranked result will be based on visiting frequency and time spend. It will remain similar or all user making search from one IP address.

Now after ranking by User Identification based personalization results will be displayed -

TABLE 3. UIBP RESULTS

Users	Ranked Result
User1	D->G->B->A->C->E->F->H
User2	A->C->F->B->D->E->G->H
User3	H->E->A->B->C->D->F->G

If we compare the results of two algorithm FTPW is giving some required ranked result according to User1 only. For 8 instances A,B,C,D,E,F,G,H

TABLE 4. COMPARISON FTPW AND UIBP

Users	FTPW	UIBP
User1	3	8
User2	1	8
User3	1	8

VI. FUTURE SCOPE

Although a substantial number of promising achievements on Web mining and its applications have been presented in this paper, there are still numerous open issues that need to be further explored in future work.

In future, we can include the explicit activities of user along with implicit searching behavior in order to have more improved search results. Our future work will involve designing experiments that will allow us to observe user profiles over time to guarantee the incremental updates to the interest scores correctly reflect changes in user interest both explicitly and implicitly.

REFERENCES

- [1] Francisco Tanudjaja and Lik Mui, "Persona: A Contextualized and Personalized Web Search" in *Proc. of IEEE Hawaii International Conference on System Sciences*, 2002..
- [2] Dong Zhou, Séamus Lawless, Vincent Wade, "Improving Search via Personalized Query Expansion using Social Media", *Springer*, Vol. 15, Issue 3, pp 218-242, 2012.
- [3] Jin Young Kim Kevyn Collins-Thompson Paul N. Bennett, Susan T. Dumais, "Characterizing Web Content, User Interests, and Search Behavior by Reading Level and Topic", in *Proc. of fifth ACM International Conference on web search and Data mining*, 2012.
- [4] Fabian Abel, Qi Gao, Geert-Jan Houben, Ke Tao, "Semantic Enrichment of Twitter Posts for User Profile Construction on the

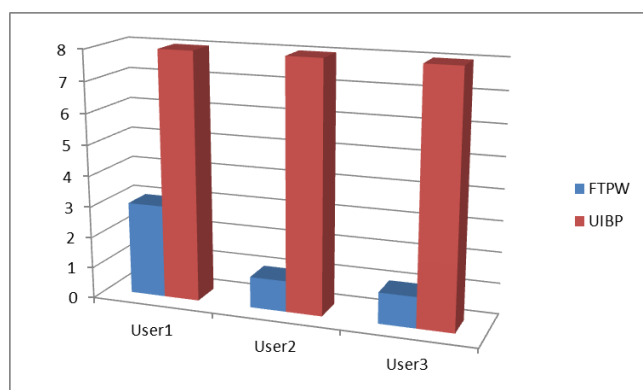


Fig. 6 Relative Ranking Accuracy

The above graph illustrate the performance of our algorithm as compared to FTPW. The accuracy of our algorithm is 100% in ranking FTPW is not so accurate as here the ranking is not individual based but system based.

V. 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

- Social Web”, *Springer*, Vol. Part II, pp 375-389,2011.
- [5] C. Srinvas,”Explicit User Profiles for Semantic web search using XML”, *IJERA* ,Vol. 2, Issue 6,pp 234-241 ,2012.
- [6] Robin Singh Bhadoria , Deepak Sain, Rahul Moriwal, “Data Mining Algorithm for personalizing user’s profile on Web”, *IJCTEE* , Vol. 1 , Issue 2, 2011.
- [7] Paul N. Bennett , Ryen W. White , Wei Chu , Susan T. Dumais , Peter Bailey , Fedor Borisyuk , and Xiaoyuan Cui , “Modeling the Impact of Short- and Long-Term Behavior on Search Personalization”, in *proc. of ACM SIGIR Conference on Research and development in information retrieval*,2012.
- [8] Mirco Speretta, Susan Gauch, Personalized Search Based on User Search Histories” in *proc. of IEEE/WIC/ACM International Conference on Web Intelligence* , 2005
- [9] Haibin Liu, Vlado Keselj ,” 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,Issue 2, pp 304 – 330 , 2007.
- [10] Feng Qiu, Junghoo Cho, “Automatic Identification of User Interest For Personalized Search”, in *proc. of the ACM 15th International conference on World Wide Web*, 2006.
- [11] Xuehua Shen, Bin Tan, ChengXiang Zhai,” Implicit User Modeling for Personalized Search.”, in *proc. of 14th ACM international conference on Information and knowledge management*, 2005.
- [12] R. Forsati, M. R. Meybodi, A. Ghari Neiat, “Web Page Personalization based on Weighted Association Rules”, in *proc. of IEEE International Conference on Electronic Computer Technology*, 2009.
- [13] Hongzhi Guo, Qingcai Chen, Xiaolong Wang, Zhiyong Wang, Yonghui Wu,” STRank: A SiteRank Algorithm using Semantic Relevance and Time Frequency”, in *proc. of IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA* , 2009.
- [14] Rohit Agarwal, Karm Veer AVer Arya, Shashi Shekhar,” An Architectural Framework for Web Information Retrieval based on User’s Navigational Pattern”, in *proc. of IEEE 5th International Conference on Information Systems*, 2010
- [15] Haibin Liu, Vlado Keselj,”Combined Mining of Web Server logs and web contents for classifying user navigation patterns and predicting user’s future requests”, *Data and Knowledge engineering ACM* ,Vol. 61 ,Issue 2, pp 304-330,2007.
- [16] R.Khanchana, 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*, 2011.
- [17] Nicolaas Matthijs, Filip Radlinski ,” Personalizing Web Search using Long Term Browsing History” , in *proc. of 4th ACM International Conference on Web Search and Data mining* , 2011.
- [18] Rohit Agarwal , K.V. Arya , Shashi Shekhar ,” An Efficient Weighted Algorithm for Web Information Retrieval System.”, in *proc. of IEEE International Conference on ICIN* ,2011.
- [19] Ranveer Singh , Dilip Kumar Sharma ,” Ratio Rank :Enhancing Impact of Inlinks and Outlinks”, in *proc. of IEEE International Advance Computing Conference*, 2012.
- [20] Ranveer Singh , Dilip Kumar Sharma ,” Enhanced Ratio Rank :Enhancing Impact of Inlinks and Outlinks”, in *proc. of IEEE International Advance Computing Conference*, 2013.