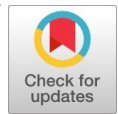# Extended model for Privacy Enhanced Personalized Web Search Ranking System

Nidhi Saxena, Bineet Gupta

Abstract: *Our existing society is totally dependent on web search to fulfill our daily requirements. Therefore millions of web pages are accessed every day. To fulfill user need number of websites and webpages are added .The growing size of web data results to the difficulty in attaining useful information with a minimum clicks. This results to the acquisition of personalization a major place in Web search. But the use of personalization breaches privacy in searching. Personalization with privacy is leading issue in current web environment.*

This paper aims at user satisfaction by using user identification based personalization approach in web search engine. Beside personalization the proposed model creates privacy during personalization. The proposed system will prove to be user friendly with less efforts and privacy concern.

Keywords : *Personalization, Privacy ,Ranking, User Search Log*

I. INTRODUCTION

WWW is a medium through which information is accessed in large scale. Search engine is a tool which is capable of acquiring useful information from huge source of data. Millions of web pages are retrieved every day.

User generally creates short query which sometimes does not provide required results due to variation of user interest. For example the search string "kingfisher". This short query does not differentiate between flight kingfisher and the bird kingfisher therefore results will be common having information of both and not be properly ranked based on user interest. This leads to the problem of Objectivity which signifies that different users keep different interest though the search string could be same. During such circumstances search engine needs full string for specific search. It affects searching time and ranking accuracy and reduces the efficiency of Search engine. In order to overcome such issues personalization is used in Search engine. Personalization creates user search log and store search results for further consideration. It ranks the search result based on personal interest. In personalization methodology user profile is created on the basis of previous search record. User interest is explored based on Web search log maintained.Profile based technique has limitation of unique identification. Single web search log is maintained at the server end or client for each system used for searching. Results are ranked based on the

Manuscript published on 30 August 2019.
*Correspondence Author(s)
Nidhi Saxena*, Information Technology, SRMGPC. Lucknow UP, India. Email: nidhi.shivansh@gmail.com
Bineet Gupta, Computer Applications, SRMU, Lucknow UP, India. Email:bkguptacs@gmail.com

search history contained. Such system lack user interest personally and delay the searching process. There also exists the problem of privacy with personalization. Social websites acquires user personal record from search engine in order to popularize there product or to give personalized results. This technique results to privacy issues. The developed model is an extension to the profiling technique used for personalization. The proposed work considers relevant Profile (Have individual search result on client), Quality Results with minimum hits along with Privacy consent.

II. BACK GROUND

Personalization researches were based on creating profile and keeping navigation track. [1], [2]. Based on Web server log and web contents user interest were analyzed and future demand were calculated [3]. Later on various techniques were introduced which includes Automatic Identification technique [4] study the user's interest from previous click events and use it as a factor personalization. UCAIR [5] system implement eager implicit feedback.

From 2010 onward the personalization was the main demand of search engine. Work was carried on implicit profile creation. FTPW [6] and other algorithms [7], [8] calculates time devoted on web page along with page visiting frequency. These techniques were effective for web page forecast. Recent research conducted in the field of personalization includes implicit techniques of profile making based on click events [9], [10], [11]. Mouse movement was analyzed for predicting user behavior [12],[13],[14]. Previous queries and results were considered for ranking [15]. User behavior becomes the basis for ranking judgment [16]. Previously the studies like Private Information Retrieval (PIR) [17], keeps the query private while retrieving user information. The system preserves the privacy of user personal information and permits general information to be accessed by others. Other works on privacy protect individual data entries by restricting data discloser [18],[19].For the purpose of measuring the efficiency of Ranking algorithm various methods were developed. A Mathematical Formula for the Search Engine Ranking SEREET (Search Engine Ranking Efficiency Evaluation Tool) [20] uses hit and miss concept during ranking which considers the position of web pages returned during searching. The algorithm mathematically assesses the efficiency of a search engine rank.

Algorithm [20]
Let there be m hits and n misses.
Let i, ,represent the position of a website name on the search output.

$W_i = m + n + 1 - i$, if the ith name is a hit.
$W_i = 0$, if the ith name is a miss.

Then the efficiency of ranking of search engine is given by-

$$E = \sum W_i * \frac{I = m + n}{2} \, \text{---------------------------} \, *100\% \qquad (1)$$
$$(m + n)*(m+n+1)$$

III. PROBLEM STATEMENT

There exist a significant difference between how well search engines might perform if they have to provide results to the individual, and how well they work if they have to satisfy everyone. This results to the need of individual interest based personalization.

The problems associated with personalization are –

- How to top rank the interest oriented web pages for the corresponding query made by the user

- How to preserve privacy while acquiring needful user personal details.

In all above approaches the user is recognized as a computer system and personalized search is according to the search made from the system. So it is still a problem to provide search result according to user interest on single system. The main challenge lies in ranking the webpages is how to consider each user as unique entity

Beside personalization there lies another challenge to provide privacy with personalization as lot of information get disclosed in order to provide personalization.

IV. PERSONALIZATION MODEL

Difficulties faced in existing web search systems lead to the requirement of smart client-side web search agent based on user identification.

The developed tool is an extension to profile based personalization technique. The System keeps track of user behavior while searching. It records and identifies each user separately for providing personalized results to each user.

The developed system considers privacy with personalization by creating each search history separately at client web log.

Each log consist of-

- Snap record of user on client. It is for the purpose of Face identification at the time of login. On the basis of
  Snap record separate log history is maintained. These Snaps will be taken at the time of first search by the user.
- String or Query for searching
- Searched or visited Web pages.

- Time spend on each visited page.

The System works in 2 steps-

Step 1- User identification is performed using face matching. Based on the snap record user is identified as anew user or old user.

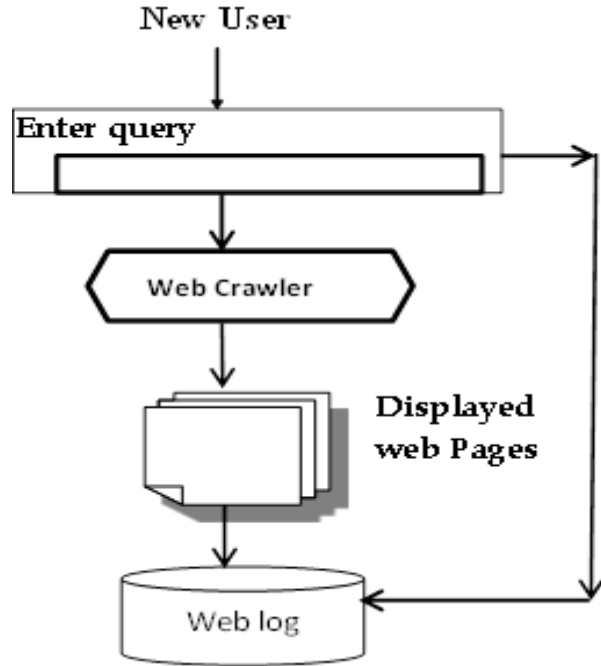Step 2- The system works separately for New and Old User

Fig.1. New user Search Model

For New user the system does not perform ranking. It records the user identity along with search record in the web log. This record helps in future when the same user delivers the same string for searching. The searched pages by new user get assembled in user web log along with time spend on each page.
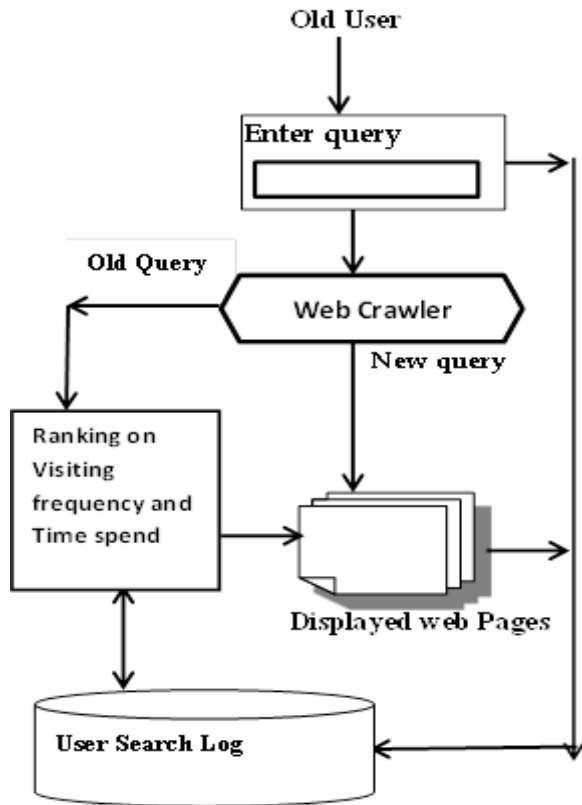
Fig.2. Old user Search Model

For the old user the keyword is tallied with the earlier searched strings. If match found the ranking is performed based on frequency and time spend on each page taken from web log.   But if the keyword is new then crawler will display the crawled results without ranking. The proposed ranking starts by the second time user search.

Two factors are taken for ranking-

- Frequency of page visit- Number of times the user visit the page is counted. This record is kept in a database along with URL searched.

- Time spend on each page- Sometimes the user may not be interested in the page he searched therefore the system keep record of time spend on each page in order to prove the relevance of the web page visited.

The results will reflect user interest in ranked pages as if user has earlier visited same pages number of time and has spent maximum time in reading those pages signifies that he is interested in those web pages. The result will also include other web pages but they will be lower ranked. Similar search made by the user will make the algorithm more accurate each time and relevant results will be displayed.

V. PRIVACY CONTROLLED PERSONALIZED SEARCH

Other than personalization the system aims to provide privacy enhanced personalized search. After the purpose of personalization in web search the user information search pattern are disclosed publicly. This lead to the unwanted messages, mails and calls of companies who wanted to grab customers. These activities are generally performed by social websites for their popularity. For the purpose of privacy it becomes necessary to restrict personalization in search.

The proposed work is an enhancement in personalization technique which offers privacy along with personalization.

This Search engine provides privacy in the following manner-

- The System maintains user account on client end and algorithm works from client therefore all information are confined and not disclosed on web.

- User is identified before search which restricts unauthorized access.

- Each user has separate profile and cannot be accesses by other user.

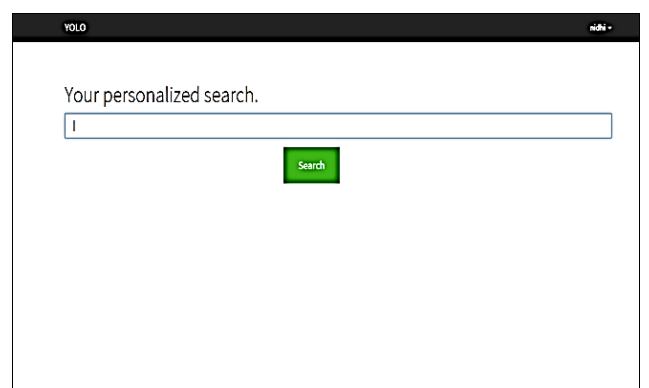## VI. IMPLEMENTATION AND SIMULATION

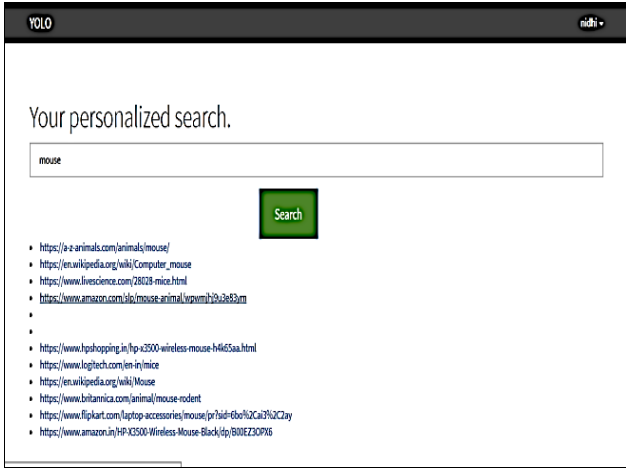**Fig.3. Login Page**

Fig.4. Home Page
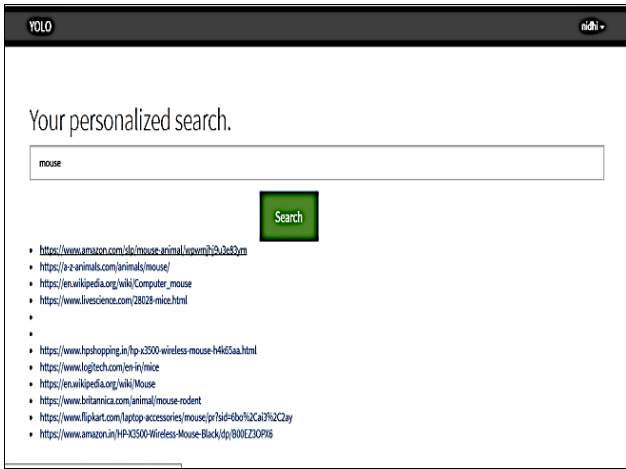
Fig.5. Ranked Result for Query

Fig.6. Varied result for same query next time.

Simulation shows the variation in the result using User Identification approach for personalization in search system. Figure 3 is a login page which takes user snaps for identifying user and redirect user to his session Home page in Figure 4. Figure 5 shows that user typed his query "mouse" and web pages related to query are displayed. Figure 6 shows how the results vary for same query on next time. User visited 4th URL for some time in first search therefore it ranked up on the second time. This system works separately for each user.

VII. RESULT AND EVALUATION

Performance of the developed system is evaluated by accessing its efficiency using SEREET [20].

Based on SEREET algorithm the efficiency of the developed system increases for each user from the second search when it keep records of earlier searches in its user search log on client machine. According to SEREET [20] for any user top ranked pages will always be hits and lower ranked pages will be miss. Therefore suppose on certain query 6 web pages were displayed. Out of which hits and miss are as under-

1. h1
2. h2
3. h3
4. h4
5. m1
6. m2

$W1 = m+n+1-1 = 4+2+1-1 = 6$
$W2 = 4+2+1-2 = 5$
$W3 = 4+2+1-3 = 4$
$W4 = 4+2+1-4 = 3$
$W5 = 0$
$W6 = 0$

Then
$\quad$ I= m+ n

$$E = \sum Wi * \frac{2}{(m+n)*(m+n+1)} *100\%$$

$$= (6+5+4+3) * \frac{2}{(4+2)*(4+2+1)} *100\%$$

$$=18* (2/ (6*7)) *100\%$$

$$= 85.71\%$$

On applying SEREET on given sample webpages we can see that efficiency is evaluated as 85% when out of 6 web pages displayed 2 are irrelevant for user. This efficiency increases with the increase in relevant pages and with each next search with same keyword. Therefore it provides relevant and accurate results
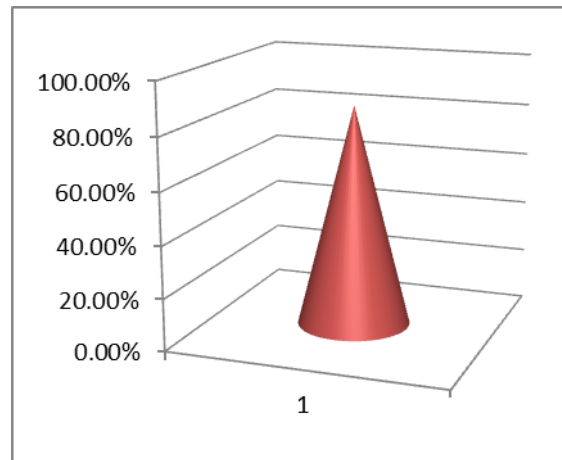

Fig.7. Ranking Efficiency

VIII. CONCLUSION AND FUTURE SCOPE

This personalized model is an extension to personalization technique used in search engine. It presents a novel approach by uniquely identifying each user and personalizes their search results. It aims at providing relevant results based on user interest. With the use of user friendly approach and privacy enhancement feature it is providing search engine a new direction. The developed system for searching and ranking is examined and its performance is evaluated. The simulation performed and test conducted illustrate the behavior and performance of the developed framework.

It can be judged that if this extension will be included with any search system the results of personalization will me more accurate than earlier. The system provides a broad benefit for all users who want interest based results along with privacy concern. This paper is an experimental work on small scale using client machine. There is a necessity to adopt such features along with an existing search engine while maintaining web log on large scale. The system can act as an extension with search engine which can provide refined personalized search with privacy concern.

REFERENCES

1. P. N. Bennett, R. W. White , W.Chu, S. T. Dumais, P. Bailey, F. Borisyuk and X. Cui, "Modeling the impact of short- and long-term behavior on search personalization," in *Proc. ACM SIGIR Conference on Research and development in information retrieval*, 2012.
2. M. Speretta, S. Gauch, "Personalized search based on user search histories," in *Proc. International Conference on Web Intelligence IEEE/WIC/ACM*, 2005.
3. H. Liu ,V. 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, no. 2, 2007, pp. 304– 330.
4. F. Qiu , J. Cho, "Automatic identification of user interest for personalized search," in *Proc. ACM 15th International conference on World Wide Web*, 2006.
5. X. Shen , B. Tan, C. Zhai, "Implicit user modeling for personalized search," in *Proc. ACM 14th international conference on Information and knowledge management*, 2005.
6. R. Agarwal, K. V. Arya, S. Shekhar, "An architectural framework for web information retrieval based on user's navigational pattern," in *Proc. IEEE 5th International Conference on Information Systems*, 2010.
7. R. Khanchana , M. Punithavalli, "An efficient web page prediction based on access time length and frequency," in *Proc. IEEE 3rd International Conference on Electronics Computer Technology*, 2011.
8. N. Matthijs, F. Radlinski, "Personalizing web search using long term browsing history," in *Proc. ACM 4th International Conference on Web Search and Data mining*, 2011.
9. A. Borisov, M. Wardenaar, I. Markov, M. de Rijke, "A click sequence model for web search," in *Proc. SIGR' 18th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2018, pp. 45-54.
10. A. Chuklin, I. Markov, M.de Rijke , "An introduction to click Models for web search," in *Proc. SIGIR 2015 38th International ACM Conference on Research and Development in Information Retieval*, 2015, pp. 1113-1115.
11. A. Chuklin , M. de Rijke "Incorporating clicks ,attention and satisfaction into a search engine result page evaluation model," in *Proc. 25th ACM International Conference on Information and Knowledge Management*, 2016, pp. 175-184.
12. Y. Chen, Y. Liu, S. Ma, M. Zhang, "User satisfaction prediction with mouse movement information in heterogeneous environment," *IEEE Transaction on Knowledge and Engineering*, 2017.
13. Y. Liu, Y. Chen, J. Tang, J. Sun, M. Zhang, S. Ma and X. Zhu "Different users, different opinions: predicting search satisfaction with mouse movement information," in *Proc. 38th International ACM Conference on Research and Development in Information Retieval*, 2015, pp: 493-502.
14. I. Aapakis, L. A. Leiva, "Predicting user engagement with direct displays using mouse cursor information," in *Proc. 39th International ACM SIGIR conference on Research and Development in Information Retieval*, 2016, pp. 599-608.
15. . Moshfeghi, K. Velinov, P. Triantafillou, "Improving search results with prior similar queries" in *Proc. 25th ACM International Conference on Information and Management*, 2016, pp.1985-1988.
16.]F. Cai, S. Wang, M.de Rijke, "Behavior based personalization in web search,"†*Journal of the Association for Information Science and Technology Wiley* , vol. 68, no. 4, 2016.
17. W. Gasarch," A survey on private information retrieval," *The bulletin of the European Association for Theoretical Computer Science (EATCS)*, vol. 82 , 2004, pp. 72—107.
18. R. Agrawal, R. Skriant, " Privacy preserving data mining," *ACM SIGMOD Conference on Management of Data (SIGMOD), Dallas, Texas*, May 2000.
19. A. Evfimievski, J. Gehrke and R. Srikant, " Limiting privacy breaches in privacy preserving data mining," *ACM SIGMOD/PODS , San Diego, CA*, 2003.
20. Tapia M, Alhalabi W S, Kubat M. ,"A mathematical formula for the search engine ranking efficiency evaluation Tool S.E.R.E.E.T", *International journal of computer science & Emerging Technologies* ,vol. 1, December 2010.

AUTHORS PROFILE

Nidhi Saxena is a research scholar in Computer Science from Shri Ramswaroop Memorial University Lucknow Deva Road UP, India. She has a teaching experience of about 10 years in Shri Ramswaroop Group of Professional Colleges Lucknow UP, India as an Assistant Professor in B. Tech Information Technology Department. Her area of research are Data mining, Web mining, Information retrieval and Security. She has published research papers in various journals and international conferences.

Dr. Bineet Kumar Gupta is working as an Associate Professor and Head, Department of Computer Application, Institute of Technology, Shri Ramswaroop Memorial University Lucknow Deva Road UP, India. Dr. Gupta has about 10 years of Higher education teaching and research experience at National and International platforms. His area of research are E- learning, Information Retrieval and Security, Software Engineering, Human Computer Interaction, ICT etc. He has published more than 30 research papers in various journals, conferences, and symposium of repute.