

**International Journal of Innovative Research in Science,  
Engineering and Technology**

*(An ISO 3297: 2007 Certified Organization)*

Vol. 5, Issue 7, July 2016

# **Profile Creation & Privacy Protection in Personalized Web Search**

VaibhavMenkudale<sup>1</sup>, B. M. Patil<sup>2</sup>

P.G. Student, Dept. of CSIT, M B E Society's College of Engineering, Ambajogai, Maharashtra, India<sup>1</sup>

Professor, M B E Society's College of Engineering, Ambajogai, Maharashtra, India<sup>2</sup>

**ABSTRACT:** The web search engine is an important source for looking information on the web. An amount of information in search engine increases day by day. But in sometimes result obtained through search engine is irrelevant because it does not meet user's intentions. Personalized Web Search (PWS) overcomes this problem by providing better search results. PWS is an effective way to improve the search engine quality by providing customization in search results. PWS framework can adaptively generalize profiles by queries while respecting user specified privacy requirements. Extensive experiments demonstrate the effectiveness of our framework.

**KEYWORDS:** Privacy protection, personalized web search, utility, risk, profile

## **I. INTRODUCTION**

For ordinary people web search engine has long become the most essential portal for searching information on the web. However users might encounter failure when search engine reciprocates insignificant results that do not meet their actual requirements. Such insignificant result is largely due to the vast variety of users, context and background as well as the ambiguity of texts [1]. For individual user needs Personalized Web Search (PWS) is used that provides finer search results.

To figure actual purpose of the user behind the issued query, user information needs to be gathered and analysed. The solutions to PWS can normally be categorized into two types i) Click-Log-Based methods work on repeated queries created by the same user it is simple and uncomplicated, this strategy imposes bias to clicked pages in the user's query history. These methods have been indicated to perform consistently and exceptionally well. ii) Profile-Based methods are used for user profiling techniques to enhance the search experience with complex user-interest models. These techniques are effective for all sorts of queries but there are some possibilities of unstable performance under some situations.

Although there are some advantages and disadvantages for both types click-log-based and profile-based methods of PWS techniques. The profile based PWS has significantly more efficiency in enhancing the quality of web search results, with growing usage of personal and behaviour information to profile its users. The user's unwillingness to reveal their private information during search has become a main obstacle for the wide propagation of PWS. Privacy issues are growing from the insufficiency of protection for such data.

## **II. RELATED WORK**

In [2] authors used 12 days MSR query log to evaluate the five personalized search strategies that include two clicklog-based and profile-based a method of a large scale evaluation framework for personalized search. Advantage of the strategies evaluated by the author is search precision is evaluated by real user clicks recorded in query logs automatically. Main disadvantage in this method is some queries may be affected due to personalization. In [3] they develop an intelligent client side web search agent (UCAIR) which is a web browser plug-in that acts as a proxy for web search engines. They developed a technique for implicit user modelling in information retrieval and present a decision theoretic framework. Search precision over the popular search engine Google by search agent but it lacks user

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

modelling and also not adaptive to individual users. In [4] authors propose many approaches to adapting search results. It considers each user’s need for significant information without any user effort. It uses detailed analysis of user’s one day search history for achieving user preferences by user profile based on modified collaborative filtering. Disadvantage in this, every user need different information for their query. Therefore, with the different information need search result should be adapted to every user. In [5] methods used for mining contextual information from long term search history are statistical language modelling based methods.

In this PWS system is classified as Profile-Based Personalization and Privacy Protection. These are

- **Profile-Based Personalization:** PWS mainly focus on improving the search utility. The basic idea of these works is to tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal. In the remainder of this section, we review the previous solutions to PWS on two aspects, namely the representation of profiles, and the measure of the effectiveness of personalization.
- **Privacy Protection:** Generally there are two classes of privacy protection problems for PWS. One class includes those treat privacy as the identification of an individual. The other includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server [1].

### III. IMPLEMENTATION

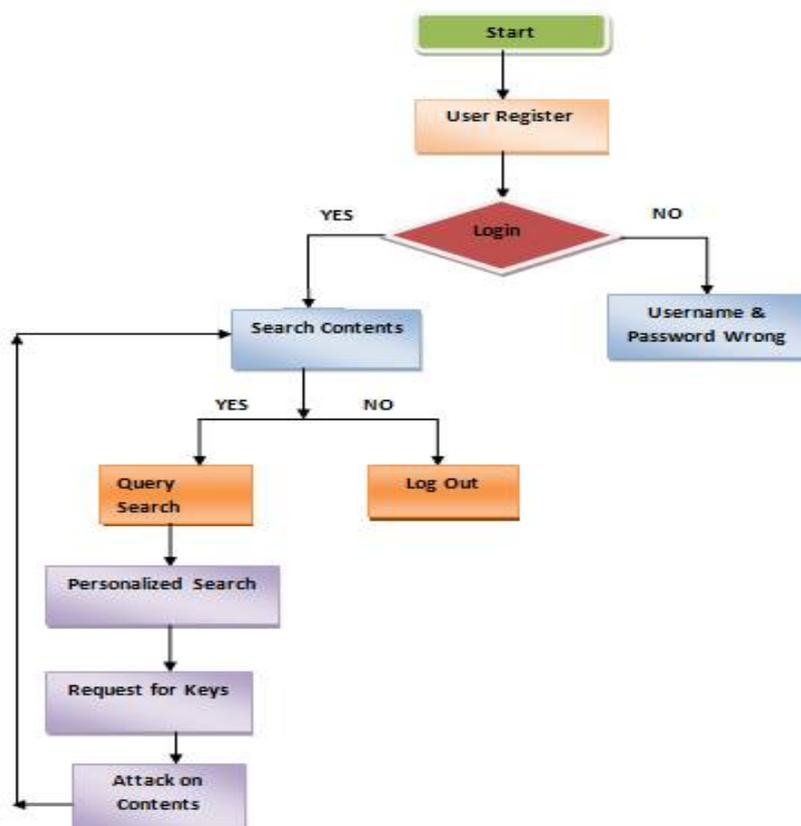


Fig. 1. Flow Chart for User

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

❖ **User**

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations such as view my details, query search, personalized search, personalized search comparisons, attack content details, request for general key, request for personalized key and logout. If user clicks on my details button, then the server will give response to the user with their tags such as user ID, name, mobile no, address, pin code and email ID.

- **Query Search**

In this module, the user can search query. Before searching any query, the user should request general key, then admin will provide a general key. Then enter general key, select field to search, enter key word and search, it will display all related contents with their tags. After searching a content rank will be increased.

- **Personalized Search**

In this module, the user can search contents. Before searching contents, the user should request personalized key, then admin will provide personalized key, then enter key and enter keyword, then user will get a related contents with their tags. After searching content the rank will be increased.

- **Personalized Search Comparison**

In this module, the user can view the comparison between greedy DP & greedy IL. After personalized searching, the greedy IL will be generated. If the user clicks on personalized search button, it will display all personalized search details with their tags such as user name, keyword used, date, time, using greedy DP and using greedy IL.

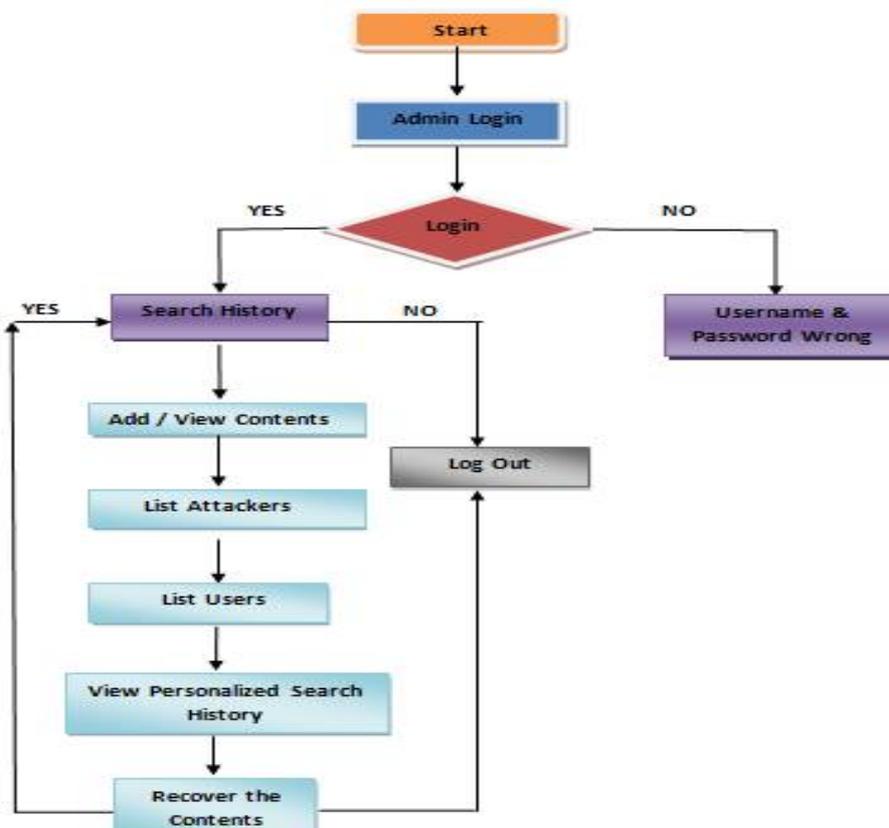


Fig. 2. Flow Chart Admin

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

- **Time delay Generation chart**

In this module, we can view the time delay Generation chart results. This chart shows the time delay by using greedy DP and time delay using greedy IP. After viewing or search the content, rank will be increased and also the time delay will be display, the time variation can be shown in this chart.

- **Attack content**

In this module, user can attack contents, and then user should enter content URL to attack, then user will get all information about content, then user can add malicious data and click on attack button. After attacking successful, the attacker details will send to admin.

- ❖ **Admin**

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as add contents, view all contents, list all searching history, list ranking of images, list of all personalized search, attacker details, recover contents, list of all user and logout.

- **Add contents**

In this module, the admin can add n-number of contents. If the admin want to add a new content, then admin will enter a URL, domain, title, description, uses, related images of the particular content ,then submit and that data will stored in data base. If admin want view to the newly added content, then click on view contents button, it will display the all contents & with their tags, the initially rank will be zero.

- **List of users**

In this module, the Admin can view list of all users. Here all register users are stored with the details such as user ID, user name, E mail ID, mobile no, Location, date of birth, address, pin code, general key and personalized key.

- **View list all searching history**

This is controlled by admin; the admin can view the all searching history. If admin clicks on search history button, then the server will display the all searching history with their tags such as user name, key word used, field searched, time & date.

- **Attacker details**

In this module, the admin can view the attacker details. If admin clicks on attacker details button, the admin will get attacker information with their tags such as attacker name, attacked content URL and attacked content ID. After attacking content, the admin will recover the content.

## IV. EXPERIMENTAL RESULT

User Name	Keyword Used	Date & Time	Using Greedy DP	Using Greedy IL
pankaj	CS	24/07/2016 11:31:52	87	5
pankaj	www.google.com	24/07/2016 11:31:24	81	8
pankaj	www.car.com	24/07/2016 11:10:20	9	0
pankaj	Research	24/07/2016 11:12:42	6	0
manish	ARTS	24/07/2016 11:17:24	6	0
manish	www.google.com	24/07/2016 11:30:15	10	null
manish	cs	24/07/2016 11:30:18	11	null
manish	yahoo	24/07/2016 11:30:22	7	null
pankaj	yahoo	24/07/2016 11:31:16	3	null
pankaj	car	24/07/2016 11:31:39	31	null
pankaj	arts	24/07/2016 11:31:48	11	null
vaibhav	cs	27/07/2016 19:59:59	48	null
vaibhav	www.google.com	27/07/2016 21:41:27	0	0
amit	www.google.com	27/07/2016 21:49:26	60	2
jayant	www.google.com	27/07/2016 22:05:13	16	3
jayant	cs	27/07/2016 22:06:23	21	0
jayant	yahoo	27/07/2016 21:50:51	0	null
jayant	www.car.com	27/07/2016 21:51:11	23	null
jayant	SPORTS	27/07/2016 21:59:35	16	0
jayant	ARTS	27/07/2016 22:03:20	0	1
jayant	development	27/07/2016 21:56:52	0	null

Fig. 3. Experimental Search Results

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

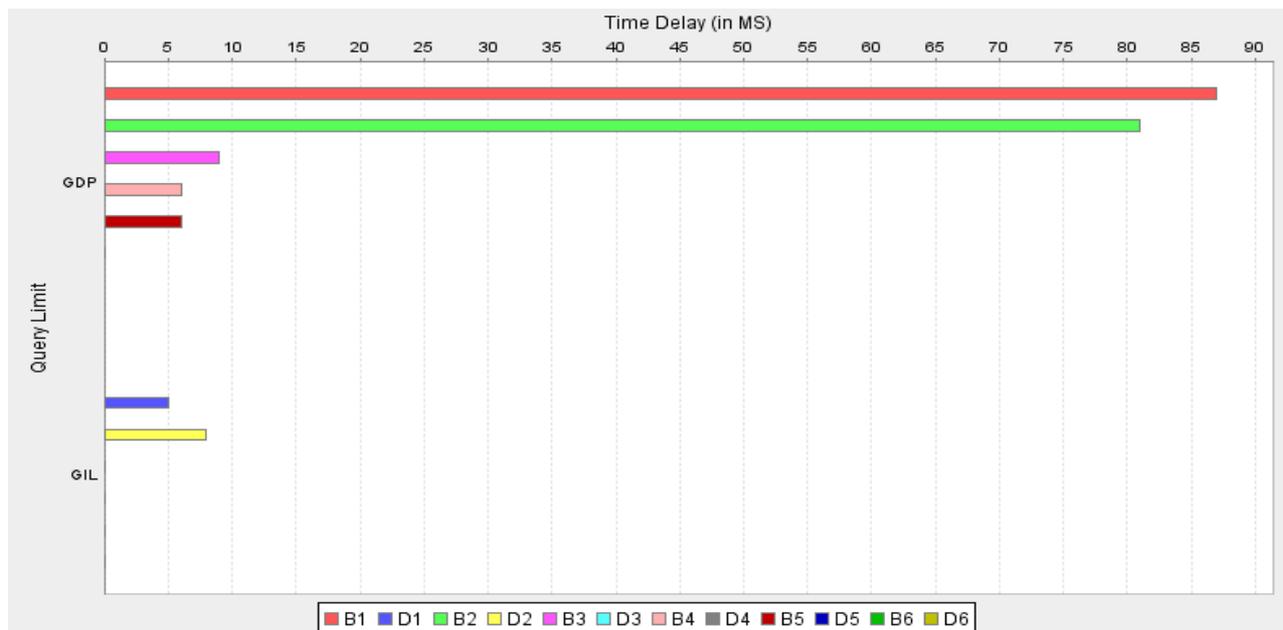


Fig. 4. Graphical Representation of Experimental Result

Search quality mention to the admissible search result on providing the query and the generalize profile as per user's interests constructed in their profile which gives the comparison of the proposed system of GreedyIL and the existing system of GreedyDP based on the quality of search. The bars B1, B2, B3, B4, and B5, B6 indicates the GDP values, in same way D2, D5 represents GIL values for different queries. These GDP & GIL values depend upon search time required. GIL values depend upon search results derived from search history. If the search result are not present in history then it will derived from whole database. So above figure 4 shows that, these search results are faster than the history database as compared with the search results of whole database. Also figure 3 shows the experimental values of GreedyIL and GreedyDP search with date and time, keyword for user.

## V. CONCLUSION

This paper presented a framework which is called as UPS for client side privacy protection in PWS (Personalize Web Search). Any PWS that captures user profiles in a hierarchical taxonomy can potentially espouse UPS. Framework allows users to specify personalized privacy requirements via the hierarchical profiles. In addition, UPS also provides protection to the personal privacy by performing online generation of user profiles without compromising the search quality. The results also ensure the effectiveness and accuracy of our solution. For future work, we will try to withstand adversaries with border background knowledge and better metrics predict the performance of the UPS.

## REFERENCES

- [1] LidanShou, He Bai, Ke Chen, and Gang Chen, 'Supporting Privacy Protection In Personalized Web Search', IEEE Transaction On Knowledge And Data Engineering, Vol:26 No:2,2014.
- [2] Z.Dou, R.Song, and J.-R. Wen, ' A LargeScale Evaluation and Analysis of Personalized Search Strategies', Proc. Int'l Conf,2007.
- [3] sX. Shen,B.Tan, and C.Zhai, ' Implicit User Modeling for Personalized Search', pro 14th ACM Int'l Conf.Information And Knowledge management (CIKM),2005.
- [4] K. Sugiyama, K. Hatano, and M. Yoshikawa, 'Personalizing Adaptive Web Search Based on User Profile Constructed without any effort from Users', Proc. 13th Int'l Conf., 2004.

## International Journal of Innovative Research in Science, Engineering and Technology

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 5, Issue 7, July 2016**

- [5] B.Tan,X.Shen, and C.Zhai, 'Mining Long-Term Search History to Improve Search Accuracy', Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2006.
- [6] JordiCastellà-Roca a, Alexandre Viejo a,b, \*, Jordi Herrera-Joancomartí,c,d, 'Preserving user's privacy in web search engines', Computer Communications 32, 2009.
- [7] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, 'Ups: Efficient Privacy Protection, in Personalized Web Search', Proc. 34th Int'l ACM SIGIR Conf. Research And Development In Information.2011.
- [8] J. Conrath, 'Semantic Similarity based on Corpus Statistics and Lexical Taxonomy', Proc. Int'l Conf. Research Computational Linguistics (ROCLING X), 1997.
- [9] D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, 'Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies', Proc. Int'l Conf. Web Search and Data Mining (WSDM), 2008. 10. J. Castellà-Roca, A. Viejo, and J. Herrera-Joancomartí, 'Preserving User's Privacy in Web Search Engines', Computer Comm., vol. 32, no. 13/14, 2009. Mr. Rajesh H. Davda1, Mr. Noor Mohammed, " Text Detection, Removal and Region Filling Using Image Inpainting", International Journal of Futuristic Science Engineering and Technology, vol. 1 Issue 2, ISSN 2320 – 4486, 2013
- [10] UdayModha, Preeti Dave, " Image Inpainting-Automatic Detection and Removal of Text From Images", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622 Vol. 2, Issue 2, 2012
- [11] Muthukumar S, Dr.Krishnan .N, Pasupathi.P, Deepa. S, "Analysis of Image Inpainting Techniques with Exemplar, Poisson, Successive Elimination and 8 Pixel Neighborhood Methods", International Journal of Computer Applications (0975 – 8887), Volume 9, No.11, 2010