

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

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

## Approach for Providing Privacy of Personalization in Web Search for Users

Shraddha Tare<sup>1</sup>, Sachin Chavan<sup>2</sup>

P.G. Student, Department of Computer Engineering, MGM's College of Engg. & Tech, Kamothe, Maharashtra, India<sup>1</sup>

Professor, Department of Computer Engineering, MGM's College of Engg. & Tech, Kamothe, Maharashtra, India<sup>2</sup>

**ABSTRACT:** Internet provide accuracy and more quality in search results with the use of PWS (Personalize Web Search), which gives user the personalization in searching on internet also personalization assures the correctness of search results. In the personalization search it will aggregates users sensitive information and use it for accuracy in search result. This is the main obstacle for growth of PWS and this is the main challenge for ensuring the privacy in personalization. Hierarchical user profile is the most used in PWS for search result accuracy. We propose framework for PWS called UPS (User customizable Privacy protecting Search). UPS can alter the generalize profile by queries with the consideration of users precise requirements. Our proposed system generally focuses on more accurately creating the generalization of user profile on bases of their interests and harmonizing between two predictive metrics that calculate the efficacy of personalization and the risk of illuminating users private and sensitive information.

**KEYWORDS:** Risk, Personalize web search, Privacy, Taxonomy, Web search, Sensitivity, Profile, Privacy protection.

### I. INTRODUCTION

For searching information on search engine is the most common thing now a days for people. Users might come across disappointment when results of search engine give in return inappropriate results that are not the actual user requirements. This is because of large number of variety users, background, context and as well as the uncertainty of texts. For each individual user need the personalization in search which gives them the accurate search results, for understanding the actual meaning of the users query user information needs to be collected and analysed.

The solution for this in PWS in categorized into two types one is Click-Log Based method and Profile Based method. In Click-Log Based method analyse the cookies and check the most clicked pages and links. It is very simple and uncomplicated. This strategy analyse the query of individual user from the users query history. This method is performing exceptionally well.

In the Profile based method user profiling techniques is use which enhancing the experience of search with the complex user interest models. These methods are cooperative for different queries, but also there are some chances of unsteady performance under some situations. Each technique wether it is Click-Log base or Profile base they have their own advantages and disadvantages [1].

For preserving the privacy of users in personalization main challenge is fill the gap between users variance need for personalization and also protection of users privacy. The solution on that is user can declare own terms of their regarding the privacy and in structural user profiling it can be possible. There are several benefits of this technique. On client side user hierarchical profile will build automatically also there is no need of every user to declare their personal interests in details. This algorithm is collected all the information of user interest and users hidden goal behind the query automatically, it is design in that way. The hierarchical profile is generate for each individual user in that profile users interests are arrange in hierarchical format most common interest are put on the higher level and less common interests are put at the lower level. The interests of user and the level of interest will decided by the using users browsing history , documents downloaded by the users, bookmarks, email, links click by the users, pages visited by users, etc. With the help of analysing all these, algorithm can decide the level of interest of topic whether the topic is more common or less common.

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

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

## II. LITERATURE SURVEY

Z.Dou, R.Song, and J.-R. Wen [2] introduced the 12 days MSR query log to establish the five personalized search strategies that include two click-log-based and profile-based a method of a large scale evaluation framework for personalized search. Benefit of the strategies evaluated by the author is search accuracy is evaluated by real user clicks recorded in query logs automatically. Main drawback in this method is some query may affect due to personalization. sX. Shen,B.Tan, and C.Zhai [3] 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 introduced technique for implicit user modelling in information retrieval and present decision theoretic framework. Search precision over the popular search engine Google by search agent but it lacking user modelling and also not useful to individual users. K. Sugiyama, K. Hatano, and M. Yoshikawa [4] propose many approaches to adapting search results. It considers each user's need for important information without any user effort. It uses complete analysis of user's one day search history for achieving user preferences by user profile based on customized collaborative filtering. Disadvantage in this, every user need different information for their query. Therefore, with the different information need search result should be personalized to every user. B.Tan,X.Shen, and C.Zhai [5] introduced methods for mining contextual information from long term search history are statistical language modelling based methods. The major advantage of methods is utilizing it for a more accurate estimate of query language model. Disadvantage is that, the problem in web search engine occurred is only one size available for all documents to return which is based on query and none for particular. Jordi Castellà-Roca, Alexandre Viejo, Jordi Herrera-Joancomartí [6] present a novel protocol called UUP (Useless User Profile) specially designed to protect user's privacy in front of web search profiling this system provide distorted user profile to web search engine. The advantage of the novel protocol is that it deals with user privacy with web search engine. Server side changes are not required for this scheme. Ashok S.O, Vikas Kumar, and Sanjay K Jena [7] present a new dimension of attributes which can be used to get more personalized search results. These attributes are based on the user actions after the user has opened the given link. The attributes contain actions like saving/printing/bookmarking the documents etc. Which are the clear indicators of user relevance.

## III. PROPOSED METHODOLOGY

Our proposed framework is UPS which assumes that the queries do not contain any sensitive information of users like account details, personal information etc., and main focus is preserving the privacy of individual user profiles while maintaining their worth. UPS Framework which generalized profiles for each query according to user precise privacy requirements. For generalization UPS uses two algorithms GreedyDP and GreedyIL [8]. Following are the models consist in a system design:

- i) Construction Of Profile
- ii) Privacy Requirements customization
- iii) Query-Topic Mapping
- iv) Generalization Of Profile

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

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

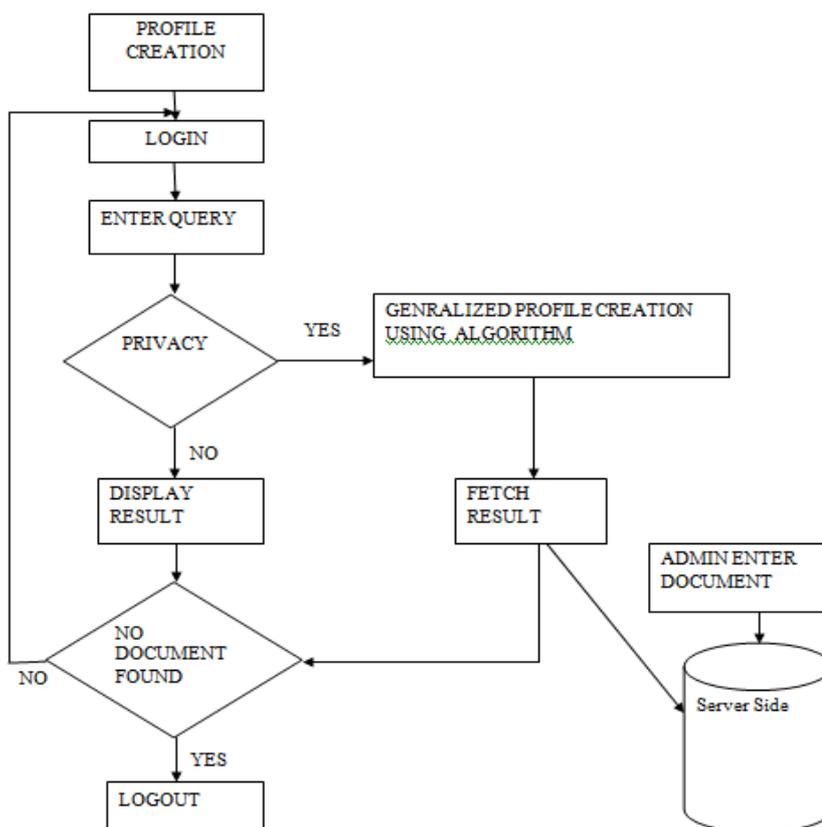


Fig. 1 System Architecture Of PWS System Using GreedyDP And GreedyIL Algorithms

We implement the GreedyDP algorithm in our proposed system UPS. GreedyDP also known as greedy utility which is based on predictive metrics for supporting online profiling [9]. Work of GreedyDP is in bottom up manner approach. Prune-leaf manner generates a candidate profile which are needed to be recomputing in GreedyDP is the major problem. GreedyIL algorithm is used for new and better profile generalization. GreedyIL uses heuristics based on profuse conclusions for improve the effectiveness of the generalization.

## IV. EXPERIMENTAL RESULTS

Results are based on the comparison of the existing system of GreedyDP and our proposed system GreedyIL on the bases of quality of search and the better privacy protection. Search quality refer to the acceptable search result on providing the query and more precisely generalize profile as per user's interest build up in their profile which give us the comparison result of GreedyDP and GreedyIL.

Results are based on the comparison of attributes. On the X-axis there are query sets used during the profile generalization. Query sets are referred by QS. QS1-Distinct, QS2-Medium, QS3-Ambiguous, QS4-Very ambiguous. Similarly on the Y-axis there are number of related URL's are plotted. These URL's are based on the users generalized profiles. Our results show that the improvement achieved by the GreedyIL is 13% more than the existing GreedyDP. On the Y-axis response time is plotted. Response time calculated by the comparison the time taken by the GreedyDP and GreedyIL algorithms. Improvement in response time of our proposed GreedyIL is 12% more than the GreedyDP.

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

(A High Impact Factor & UGC Approved Journal)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 8, August 2017

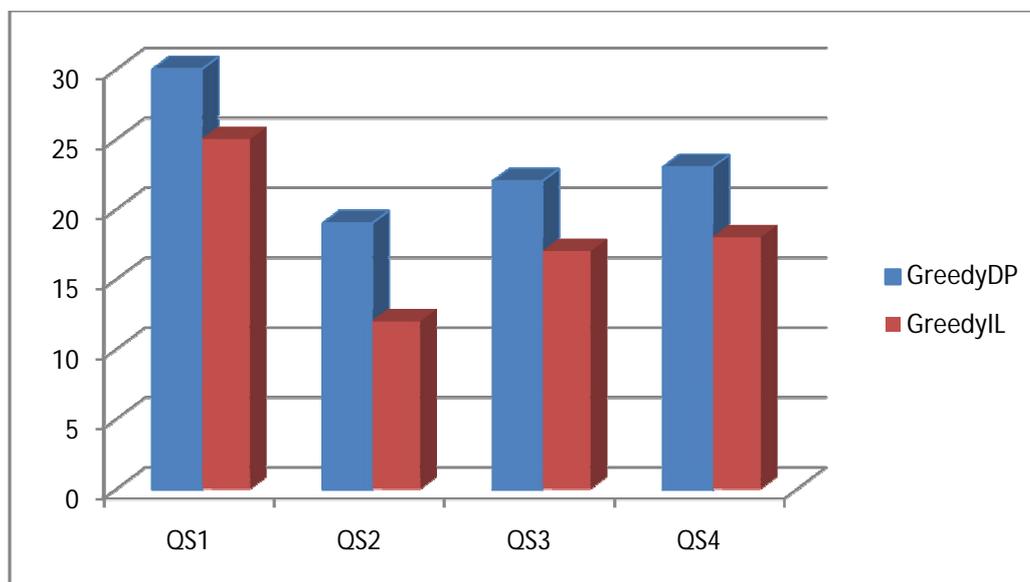


Fig2. Performance Comparison Graph Of GreedyDP and GreedyIL Algorithms On The Bases Of Response Time

The graph shows the comparison of the GreedyDP and GreedyIL on the bases on response time. The blue bars shows the response time of GreedyDP algorithm and the Red bars show the response time of GreedyIL algorithm. The bar charts shows the GreedyIL required less response time than the GreedyDP.

### V. CONCLUSION

This paper presented client side privacy protection with our UPS (User customizable Privacy protecting Search) framework. Captured user profile by any PWS in a hierarchical taxonomy can potentially support UPS. Our framework allows users to specify their requirement for privacy via the hierarchical profiles. UPS preserve the sensitive information of users with improving the search result accuracy. The results also generate the accuracy in search results as per the users query. The response time the UPS framework is also less than the existing PWS. For future work, we try to reduce the adversaries with the border background knowledge and more accurate metrics prediction for the better performance of UPS.

### REFERENCES

- [1] Lidan Shou, He Bai, Ke Chen, Gang Chen, "Supporting Privacy Protection In Personalized Web Search", IEEE Transaction On Knowledge And Data Engineering, Vol:26 No:2,pp.201-216, 2014
- [2] Z.Dou, R.Song, J.-R. Wen,"A LargeScale Evaluation and Analysis of Personalized Search Strategies", International World Wide Web Conference Committee, Proc. Int'l Conf ,No. 8, pp.581-590, 2007
- [3] s.X. Shen,B.Tan, C.Zhai,"Implicit User Modeling for Personalized Search", Information And Knowledge management (CIKM), pro 14<sup>th</sup> ACM Int'l Conf, No. ACM 1-59593-140-6/05/0010, 2005
- [4] K. Sugiyama, K. Hatano, M. Yoshikawa, "Personalizing Adaptive Web Search Based on User Profile Constructed without any effort from Users", Proc. 13th Int'l Conf., No. ACM 1-58113-844-X/04/0005, pp. 675-684, 2004
- [5] B.Tan,X.Shen, C.Zhai, "Mining Long-Term Search History to Improve Search Accuracy", Knowledge Discovery and Data Mining (KDD), Proc.ACM SIGKDD Int'l Conf., No. ACM 1-59593-339-5/06/0008 .pp. 718-723, 2006
- [6] Jordi Castellà-Roca, Alexandre Viejo, Jordi Herrera-Joancomartí, "Preserving user's privacy in web search engines", Computer Communications 32, Vol.32, pp.1541-1551,2009
- [7] Ashok S.O, Vikas Kumar, Sanjay K Jena, "An Extended Approach to Personalize the Web Search by Measuring the User Relevance" International Conference on Semantic Web and Web Services, No. SWWS'10, pp.152-158, 2010
- [8] J. Conrath, "Semantic Similarity based on Corpus Statistics and Lexical Taxonomy", Research Computational Linguistics (ROCLING X), Proc. Int'l Conf.pp. 19-33, 1997



ISSN(Online): 2319-8753  
ISSN (Print): 2347-6710

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

*(A High Impact Factor & UGC Approved Journal)*

Website: [www.ijirset.com](http://www.ijirset.com)

**Vol. 6, Issue 8, August 2017**

- [9] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "Ups: Efficient Privacy Protection, in Personalized Web Search", ACM SIGIR Conf. Research And Development In Information, Proc. 34th Int'l, pp. 615-624, 2011
- [10] D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, "Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies", Web Search and Data Mining (WSDM) Proc.Int'l Conf., pp. 139-148, 2008
- [11] Andreas Krause, Eric Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services", Journal of Artificial Intelligence Research, vol. 39, pp.633-662,2010