

# AN ENHANCED PRIVACY PRESERVING SEARCH ENGINE FRAMEWORK USING GREEDY ALGORITHM

A.Smilien Rophie

PG Scholar, Department of Information Technology  
Francis Xavier Engineering College, Tirunelveli, Tamilnadu, India  
smilienrophie@gmail.com

Dr.A.Anitha

Professor, Department of Information Technology  
Francis Xavier Engineering College, Tirunelveli, Tamilnadu, India  
dr.aanitha@yahoo.com

**Abstract-** By using internet, the user collect all the information. Because of weak search quality the users get irrelevant information. By using Personalized Web Search (PWS), this paper achieves the high-quality search results. In the same manner, the user's privacy also retained. In the existing system, they never used PWS technique. In the proposed system, the relevance and the privacy can be attained by using greedy procedure. In the user profile, the user preferences can be stored in the form of hierarchical structure. The greedy algorithms DP and IL are used to raise the discriminating power and minimizes the information loss.

**Keywords** - Privacy, Personalization, Profile, Greedy algorithm

## I. INTRODUCTION

Web use mining is utilized to concentrate data in view of the client log [26]. Web Usage mining is the procedure of using data mining strategies to the disclosure of usage examples from Web information, focused on towards different applications. The usage information gathered at the distinctive sources will denote the navigation examples of various portions of the general Web traffic, differing from single-client, single-webpage searching conduct to multi-client, multi-website access designs [26]. The Methodologies of web usage mining are

- Information gathering
- Information preprocessing
- Information Discovery
- Design examination

User profiling is introduced as a method for furnishing the client with his/her interests [27]. A framework that is fused into the Internet Explorer and keeps up a dynamic client profile in a type of naturally built topic ontology.

## II. EXISTING SYSTEM

User profile can be generalized only offline, and it is used to personalize all query from a same substance abuser. The existing personalization does not even serve to develop the various search level for few queries [13]. The existing way does not capture the customization of concealment requirements. In the existing system, all the susceptible topics are observed using an complete metric called surprisal done on information theory. Iterative user fundamental interactions are required in many personalization search results [13]. Then the search results are refined with some metric such as average rank, rank scoring, etc., This is inaccessible for all runtime profiling, due to it affect too much exposure of privacy breach, and also requires processing time for profiling[13]. Therefore, these guessing metrics to reach the search quality without iterative aspect interaction of user.

### III. ENHANCED PWS MODEL

The main objective of web search system is user for searching the data in a personalized way. It also rises the quality of the web search. An enhanced privacy preserving PWS model is proposed as follows:

- User Profile.
- Hiding Privacy Details.
- User Profile Generalization.
- Search results.

The below diagram shows how the enhanced model is done in PWS.

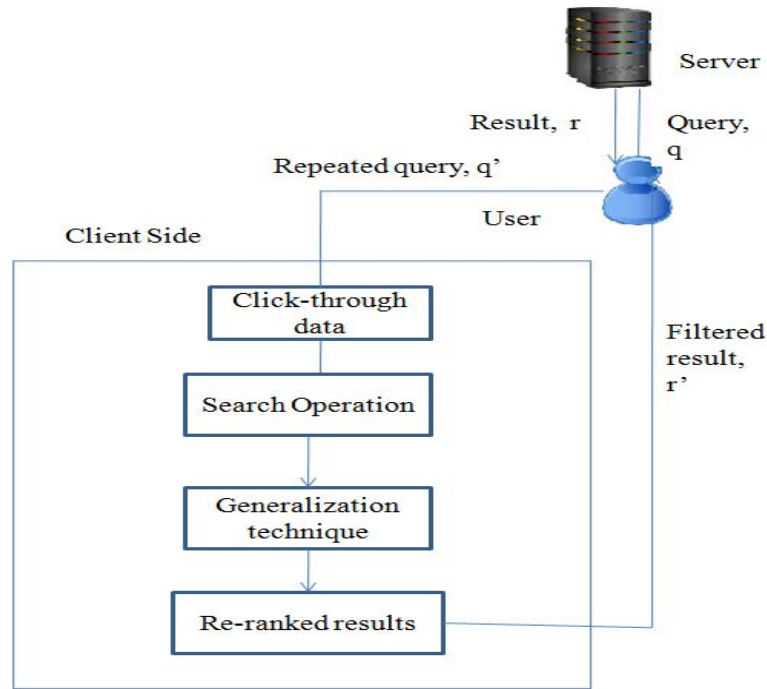


Figure 1. Architecture diagram

#### A. User Profile

The histories, documents, past preferences, etc., are stored in user profile [13]. The user profile adapts a hierarchical format. All profiles can be saved in a repository. The user profile gets the details of his/ her knowledge.

$$sup_R\langle t \rangle = \sum_{t' \in C(t,R)} sup_R(t') \quad (1)$$

Based on the below formulae, the calculation of conditional probability as

$$Pr(t|s) = \frac{sup_R(t)}{sup_R(s)}, \quad t \in subtr(s, R) \quad (2)$$

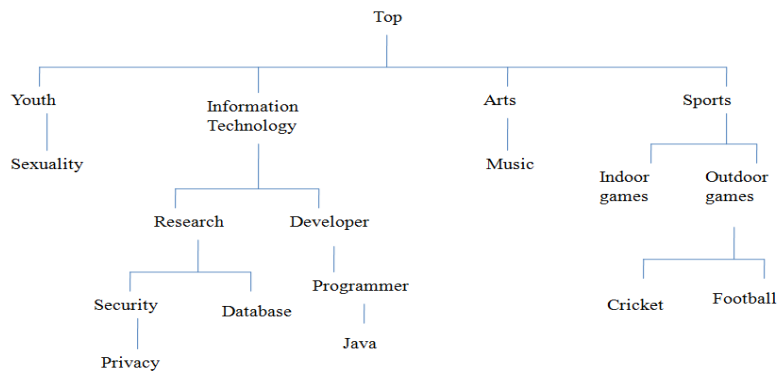


Figure 2. User Profile

### B. Hiding Privacy Details

A PWS framework that observe profiles for each query concordant to user specified requirements. The online prediction process based on query service program for judging either a query is personalized or not. The server return their results and put it in the PWS client. The updation takes place in the click-through database. By this way, the privacy details are put in a hidden node. In the user profile, the sensitivity nodes are defined as  $S = \{ \text{Youth, Privacy, cricket} \}$ . Some user may hesitate to show their personal information (e.g., Youth, Privacy)

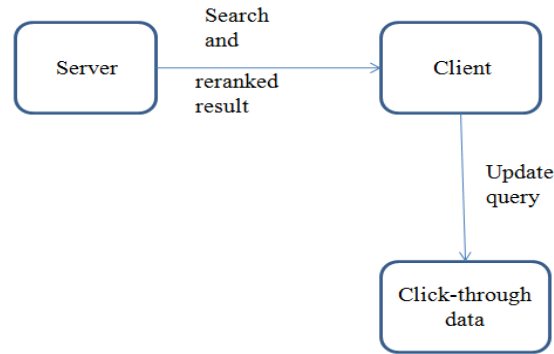


Figure 3. Privacy Protection

### C. Generalizing User Profile

The online generalization used to avoid unnecessary data and also eliminate noisy data.

### D. Attack Model

The eavesdropper Eve effectively interrupt the interaction between Alice and the server through few measures, like man-in-the middle attack, threatening the PWS server. From profile G, Eve would pressure to reach all the announcement nodes of Alice by giving back the sections invisible from the unique and calculating a confidence for every retrieved topic present in R.

Knowledge bounded: The background experience of the challenge is limited to the repository R.

Sessions bounded: The already captured information is never present for mapping identical target in a long time. This assumption seems powerful, but are justifiable in practice session.

### E. Greedy Algorithm

By using this greedy algorithm, the PWS could reach the search quality. The greedy algorithm maximizes discriminating power and minimizes the information loss.

#### 1) GreedyDP Algorithm

Greedy DP named as Greedy Utility to auxiliary online profiling based on some predictive metrics of privacy risk and personalization utility. The best profile can be maintained at each and every iteration.

#### 2) GreedyIL Algorithm

The working of GreedyIL algorithm are as follows. Comparing generalized profile with user profile. If the user profile is greater than the generalized profile, then the pruning operation takes place. In GreedyIL, the priority queue could be maintained. The user candidate operator in the priority queue like  $op = (t, IL(t, G_i))$ , where the leaf topic  $t$  to be pruned by the  $op$ . This flow continues till delta value is reached. Two cases in GreedyIL are as given below. All the values can be loaded in Q.

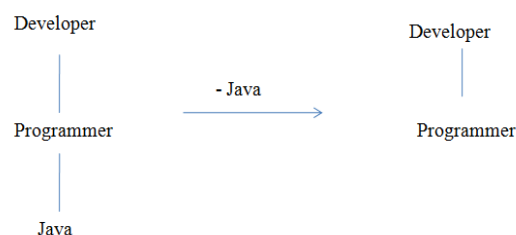


Figure 4. : leaf node has no siblings (Case 1)

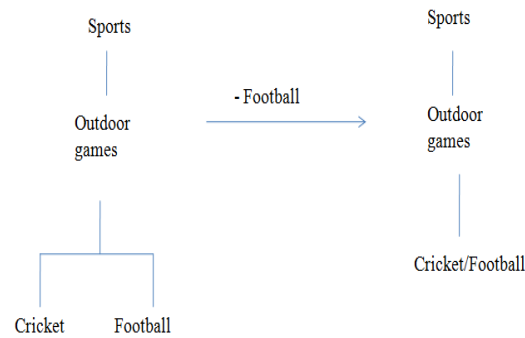


Figure 5. Leaf node with siblings (Case 2)

#### IV. RESULTS

All the user has a separate login. By using their own account and search any information. Here the user type a word as “Apple”. Apple denotes both fruit and iphone. During the first time all the data gets retrieved as follows:

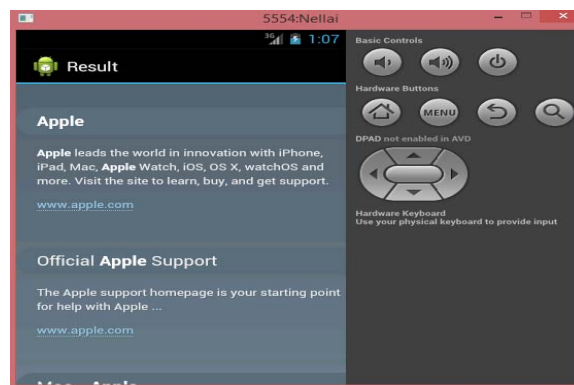


Figure 6. First Search Result of User Query

The user shown their interest as Apple support. Then the corresponding page will be opened.

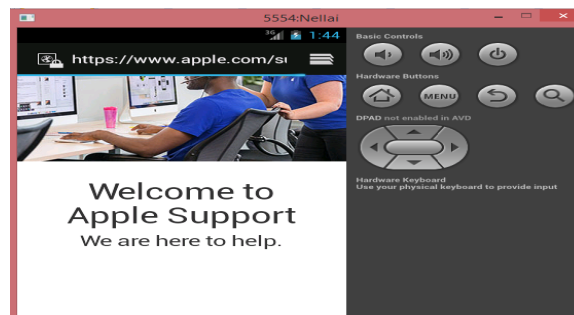


Figure 7. User interests

At the second time, directly the query goes to their user interests and retrieve only necessary information.

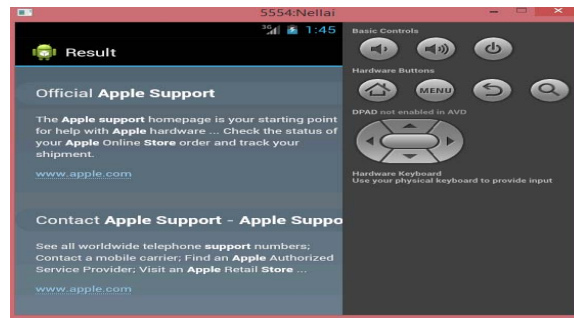


Figure 8. Next Relevant Search Result

## V. CONCLUSION

All the people uses only web search engine. So it retrieve all kinds of information. No filtering process is done at the internet. But by using this enhanced search technique using greedy algorithm, it filters all the data . So the people saves a bunch of time. The greedy algorithm increases the search quality.

## VI. REFERENCES

- [1] Aruna V., Palanivel K., "A Privacy-Preserving Personalized Web Search(PWS) Framework UPS," Global Journal of Advanced Engineering Technologies, Vol. 3, 2014.
- [2] Castella Roca J., Viejo A. and Herrera Joancomarti J., "Preserving User's Privacy in Web Search Engines," Computer Comm., Vol. 32, No. 13-14, pp. 1541-1551, 2009.
- [3] Chirita P.A., Nejdl W., Paiu R. and Kohlschutter C., "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.
- [4] Dou Z., Song R. and Wen J.R., "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.
- [5] Madhu G. and Govardhan A. and Rajinikanth T.V., "Intelligent Semantic Web Search Engines: A Brief Survey," International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1, 2011.
- [6] Priyanka Z. and Todmal S.R., "Privacy Protection Using CPS Framework in Personalized Web Search," International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 2, pp. 181-184, 2014.
- [7] Qiu F. and Cho J., "Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.
- [8] Radhika M. and Vijaya Chamundeeswari V., "Privacy Protection In Personalized Web Search Using Generalized Profile," ARPN Journal of Engineering and Applied Sciences, Vol. 10, No. 7, pp. 4-17, 2015.
- [9] Ramya V. and Gowthami S., "Enhance Privacy Search In Web Search Engine Using Greedy Algorithm," International Journal of Scientific Research Engineering & Technology (IJSRET), 2014.
- [10] Sai Kumar V., Pavan Kumar P.N.V.S., "A UPS Framework for Providing Privacy Protection in Personalized Web Search" International Journal of Innovative Research in Science, Engineering and Technology Vol. 4, 2015.
- [11] Shen X., Tan B. and Zhai C., "Implicit User Modeling for Personalized Search," Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM), 2005.
- [12] Shen X., Tan B. and Zhai C., "Privacy Protection in Personalized Search," SIGIR Forum, Vol. 41, No. 1, pp. 4-17, 2007.
- [13] Shou L., Bai H., Chen K., and Chen G., "Supporting Privacy Protection in Personalized Web Search," IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, 2014.
- [14] Srivastava J., Cooley R., Deshpande M. and Tan P.N., "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," SIGKDD Explorations, Vol. 1, pp. 12-17, 2006.
- [15] Teevan J., Dumais S.T. and Horvitz E., "Personalizing Search via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.
- [16] Teevan V., Dumais S.T. and Liebling D.J., "To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 163-170, 2008.
- [17] Viejo A. and Castella-Roca J., "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, Vol. 54, No. 9, pp. 1343-1357, 2010.
- [18] Xiao X. and Tao Y., "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.
- [19] Xu Y., Wang K., Zhang B. and Chen Z., "Privacy-Enhancing Personalized Web Search," Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 591-600, 2007.
- [20] Xu Y., Wang K., Yang G. and Fu A.W.C., "Online Anonymity for Personalized Web Services," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.
- [21] Yuan X., Song R., Wen J.R. and Dou Z., "Evaluating the Effectiveness of Personalized Web Search," IEEE Transactions On Knowledge And Data Engineering, Vol. 21, No. 8, pp. 1178-1190, 2009.
- [22] Zhu Y., Xiong L. and Verdery C., "Anonymizing User Profiles for Personalized Web Search," Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.
- [23] Jaideep S., Cooley R., Deshpande M. and Tan P.N., "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," SIGKDD Explorations, Vol. 1, pp.12-17, 2006.
- [24] J Vellingiri, S.Chenthur Pandian, "A Survey on Web Usage Mining," Global Journal of Computer Science and Technology, Vol. 11, Issue 4, March 2011.
- [25] Priyanka C. Ghegade, Vinod Wadane, "A Survey of Personalized Web Search in Current Techniques," International Journal of Computer Science and Information Technologies, Vol. 5 (6), 2014.
- [26] P. Menaka, A. Prathimadevi, "A Survey on Web Mining and Its Techniques," International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, September 2015.
- [27] Miha Grčar, Dunja Mladenič, Marko Grobelnik, "User Profiling for Interest-focused Browsing History".