ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

An Advanced Approach for Privacy Protection in Web Based Personalized User Search

Shraddha Tare[1], Prof. Sachin Chavan[2]

PG Scholar, Dept. of Computer Engineering, MGM's College of Engineering & Technology, Kamothe, Navi Mumbai, India

Assistant Professor, Dept. of Computer Engineering, MGM's College of Engineering & Technology, Kamothe, Navi Mumbai, India

ABSTRACT: Internet provides an efficient way to search information on search engines and PWS (Personalized Web Search) bespeak of its potency that ameliorate the quality of search engines. Personalized search needs to gather user information and aggregation which causes the contravension for many users, this contravension have become the major hurdle for escalation of PWS and challenge for preserving privacy in personalization. We study privacy protection in PWS applications that model user preference as hierarchical user profiles in privacy protection in PWS application. We propose a framework called UPS (User customizable Privacy-preserving Search) that can modify generalize profiles by queries, while considering user specified requirements for privacy protection. Our runtime generalization mainly focuses on balancing between two predictive metrics that assess the utility of personalization and the privacy risk of revealing the user generalized profile.

KEYWORDS: Risk; Personalize web search; Privacy; Taxonomy; Servers; Web search; Sensitivity; Profile; Privacy protection; Utility

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 reciprocate 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, users information need to be gathered and analysed.

The solutions to PWS can normally be categorized into two types i) Click-Log-Based methods works on repeated query created by the same user it is simple and uncomplicated, this strategy impose 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 signify more efficiency in enhancing the quality of web search freshly, with growing usage of personal and behavior 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

II. RELATED WORK

In [2] authors used 12 days MSR query log to evaluate 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. 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 query may affect 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 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 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. The major advantage of methods is utilizing it for a more precise 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. In [6] authors 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.

III. PROPOSED ALGORITHM

Protecting the privacy in individual user profiles is the main aim. Our PWS framework called UPS ensures that queries do not contain any sensitive information of user. According to users specified privacy requirements generalization of profiles for each query done by UPS framework. GreedyDP and GreedyIL are the generalization algorithms to find out an exploitation of user search and performance improvisation.

A. *GreedyDP Algorithm*

In proposed model UPS the GreedyDP named as greedy utility based on predictive metrics for supporting online profiling. GreedyDP is a greedy algorithm works in a bottom up manner. Prune-leaf manner generates a candidate profile which are needed to be recomputing in GreedyDP is the major problem. The process of pruning leaf t from $\mathcal{G}1_i$ to obtain $\mathcal{G}1_{i+1}$ formally we denoted as $\mathcal{G}1_i \xrightarrow{-t} \mathcal{G}1_{i+1}$. With the finite-length transitive closure of prune-leaf the optimal Profile G1 * can be generated.

The GreedyDP works in bottom up manner which is starts from the $\mathcal{G}1_0$ up to the $\mathcal{G}1_i$ that is i[th] iteration of G1. In every i[th] iteration leaf topic selected by GreedyDP $t \in T_{G1i}(q)$ for pruning, with the current iteration $G1_{i+1}$ trying to achieve maximum utility of the output. We also retain Best Profile So Far during the iterations, which show that $G1_{i+1}$ have the great discriminating power while achieving the ∂-risk constraint. When the root topic is generalized by the profile the iterative process terminated. The final outcome (G1 *) of the algorithm is the Best Profile So Far. The main disadvantage of the GreedyDP algorithm is that it demands re-computation with privacy risk and discrimination power of all candidate profile which are produced by attempts of prune-leaf on all $t \in T_{G1i}(q)$. It required sufficient memory and it also increase the computational cost.

B. *GreedyIL Algorithm*

For a new profile generation GreedyIL algorithm is propounded. GreedyIL uses heuristics based on abundant conclusions for ameliorate the effectiveness of the generalization. One important outcome is that any prune-leaf operation decreases discriminating power of the profile. DP bespeak monotonicity by prune leaf, GreedyLP further decreases this measure with heuristic. The larger the isolation threshold, the minimum iterations the algorithm needs.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

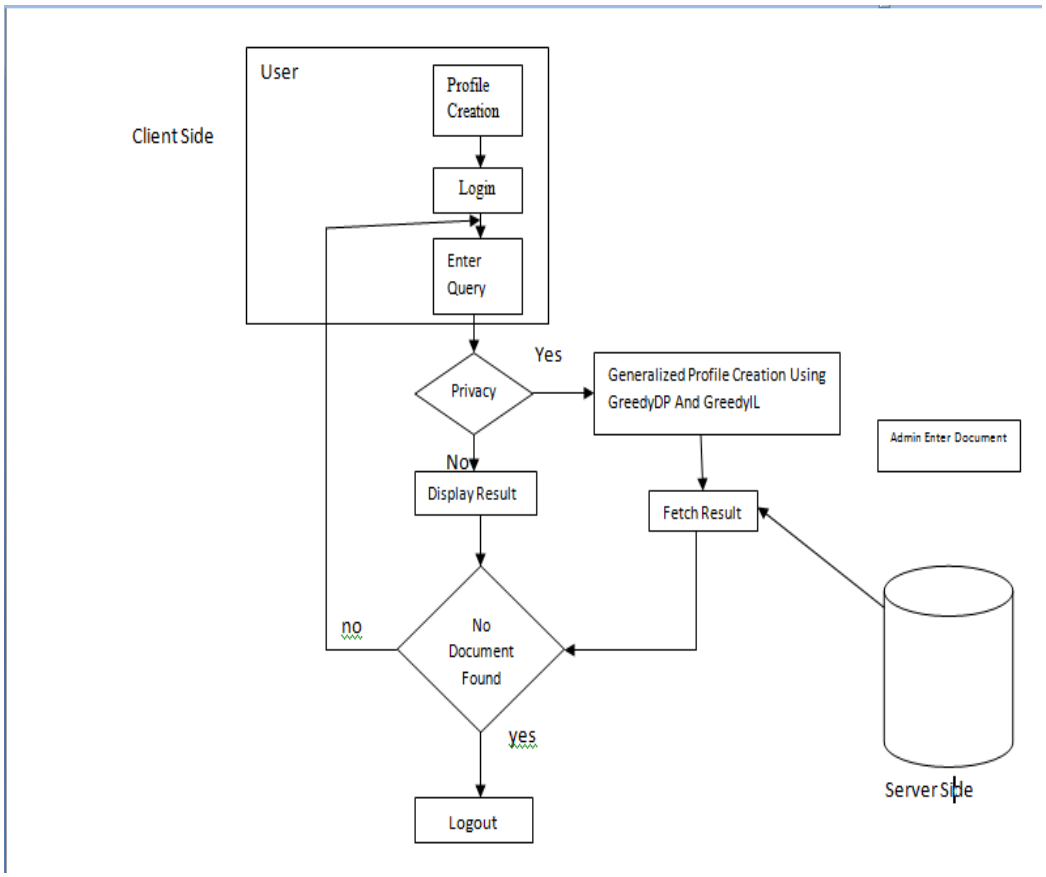

Fig1.System Architecure

IV. PROBLEM ANALYSIS

1. Construction Of Profile
2. Privacy Requirements customization
3. Query-Topic Mapping
4. Generalization Of Profile

1. *Construction Of Profile*
   The first step of processing is to create an original profile and each user profile embraces a hierarchical structure. Profile construction is based on the accessibility of public accessible taxonomy. User profiles reveals the user interest, we presume that preference of the users are represented in the set of plain text documents.

2. *Privacy Requirements Customization*
   This procedure requires sensitive-node set. These sensitive-node sets are the sensitive topics defined by the users. User defined sensitive-nodes that introduce the privacy risk to the user. The main approach of privacy protection of PWS is to keep privacy issues under control. Following are the two steps for privacy requirement customization:

a) Appeal the user to identify the sensitive topic and respective sensitivity value for each topic.
b) Calculate the cost layer of the profile by computing cost value of each node.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

3. *Query-Topic Mapping*
   Query-topic mapping used to calculate rooted sub-tree. Query q compute rooted sub-tree H which is called a seed profile. So that all topics related to q are restricted in it and obtain the preference value between q and all topics in H.

4. *Generalization Of Profile*
   The procedure of generalization of profile will generalize the seed profile $G1_0$ in cost based iterative manner relaying on the privacy and utility metrics. This procedure also calculates the discrimination power for online result on whether personalization should be employed or not.

V. EXPERIMENT RESULTS AND ANALYSIS

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 in blue colour shows the result of GreedyDP algorithm and the bars in red colour shows the result of GreedyIL algorithm. On the X-axis there are number of query sets Q1-Distinct, Q2-Medium, Q3-Ambiguous, Q4-Very ambiguous are plotted. On the Y-axis there are number of related URL's are plotted. This related URL's are based on the user's profile. The improvement achieve by the GreedyIL is 13% more than the existing GreedyDP.

The response time comparison of the proposed system GreedyIL and the GreedyDP based on respose time taken by the query set. The number of query sets on X-axis and the average time is plotted on Y-axis. The response time varies with the respect to the user's profile. Improvement in response time of GreedyIL is 12% more than the existing GreedyDP.

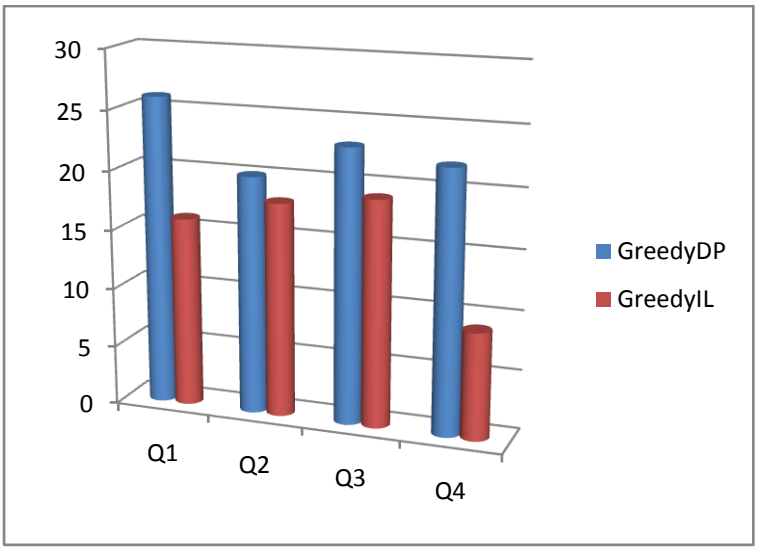

Fig1.Performance Comparison Based On Response Time

The comparison based on the scalability of varying profile size between existing GreedyDP and the proposed GreedyIL. The GreedyIL achieves the 11% of the enhancement in scalability than the GreedyDP.

VI. CONCLUSION AND FUTURE WORK

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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

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. Lidan Shou, 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.
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. Jordi Castellà-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. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, 'Preserving User's Privacy in Web Search Engines', Computer Comm., vol. 32, no. 13/14, 2009.

BIOGRAPHY

Shraddha Tare is a Master of Engineering (ME) student in the Computer Engineering Department, MGM's College of Engineering and Technology, Mumbai University. She received Bachelors of Computer Engineering (BE) degree in 2013 from Pillai's Institute of Information Technology, Engineering & Media and Research, Navi Mumbai, Mumbai University, MS, India. Her research interests are Data Mining, Computer Networks, Web Mining etc.

Sachin Chavan is an Asst. Professor in the Computer Engineering Department, MGM's College Of Engineering and Technology, Mumbai University. He received Master of Computer Engineering (ME) degree in 2010 from Mumbai University, Mumbai, MS, India. His Research Interests are Computer Networks, Data Mining, Web Mining etc.