International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Enhanced Privacy in Personalized Web Search

Kalyani R. Kshirsagar[1], Prof. N.R.Wankhade[2]
[1]*Department of Computer Engineering, Late G.N. Sapkal College of engineering Nasik,kshirsagarkalyani2411@gmail.com*
[2]*Department of Computer Engineering, Late G.N.Sapkal College of engineering Nasik,nileshrw_2000@yahoo.com*

Abstract-Web search engines are very important portal for ordinary people that are looking for useful information on the web. Generic search engine cannot identify the users search goals behind the query, if users enter improper keyword, ambiguous keyword and lack of user's ability to express what they need exactly. Personalized web search overcome these above problem. Personalized web search (PWS) is ability that identifies different users needs who issue the same query for web searching for carry out data retrieval as a part of his/her interests. User hesitated to disclose their private preference information to search engines which has become major issue in personalized web search. Thus, a balance must be stuck between user's privacy and quality of search. In this paper we have surveyed user attitudes towards the privacy protection and here we have provided methodology for securing rich users profile while transferring the query to the server side using cryptography random4 algorithm. User profiles summarize user's specific interests into a hierarchical organization according to particular interests.

Keyword- personalized web search, Generic search, user profile, search behavior, Search quality

I. INTRODUCTION

The generic search engine has gained a lot of popularity and increases its importance for users seeking information on the web day by day. Since the contents available in web is very vast and ambiguous, users at times experience failure when an irrelevant result of user query is returned from the search engine. Therefore, in order to provide better search result a technique is used called Personalized Web search. In personalized web search, user information is collected and analyzed in order to find intention behind issued query fired by user. There are two general category personalized web search namely click log based and profile based. The click log based methods is straightforward method. They simply imposed bias to clicked pages in users query history. This method inconsistent but can be work only on repeated queries from the same user which is strong limitation on its applicability. Profile based method improves the search experience with complicated user interest model generated from user profiling technique. There are advantages and disadvantages in both type of PWS techniques, the profile based PWS has demonstrated as more effectiveness in improving the quality of web search. As recently increasing with the usage of personal and behavior information to the profile its users, which are usually collected in various ways such as browsing history, click through data and so on. Such implicitly collected personal information can be easily a gamut of user's private life. Privacy issues will be rising from lack of protection for such type of data. To overcome this issue preserving user's privacy in personalized web search is necessary [1].

II. LITERATURE REVIEW

J. Teevan et al[2] described the methodology based on personalized web search via automated analysis of interests and activities and also investigated the feasibility of personalized web search by using automatic construction of user profile by using ranking algorithm. Text based personalized web search algorithms perform significantly better than explicit relevance feedback. Such personalization algorithms can significantly improve current web search.J.Castelli-Roca et al [6] proposed that new mechanisms that introduced a high cost in terms of computation and communication. They presented a novel protocol specially designed to protect the user's privacy in front of web search profiling. Their system provides a distorted user profile to the web search engine.X.Shen et al [9] discussed IR evaluation methods for retrieving highly relevant documents that the issue of privacy preservation in personalized search. They showed that client side personalization has advantages over the existing server side personalized search in preserving privacy. They distinguished and done four levels of privacy preservation, and analyzed the various software architectures for personalized search. They also investigated the privacy protection of current search systems.

III. PROPOSED SYSTEM

We propose user customizable Privacy Preserving Search. This paper targets at bridging the conflict needs of personalization and privacy preservation and provide better solution where users can decide their own privacy settings based on a structured user profile. In this we are providing more security to User profile by using Random4 algorithm.Random4 algorithm encrypt the query given by user and then transfer to the server for further search. This facility provides more confidentiality to the user profile and his/her private information. User can choose his/her area of interest, its sub interest and a specific topic in which user interested in. User can inform the server about sensitivity level. Here we use 0-means less sensitivity, 1-means high sensitivity. The keyword with less sensitive will go to server for search. The Keyword with high sensitive will not go to server to avoid in the middle attack. When the attacker trying to see the user data, he cannot see the sensitive data (keywords) given by user. Sensitive keyword is not published here, to preserve privacy for user. We ensure that it is very a simple and effective technique for user a good suggestion and also promises for effective and relevant data retrieval. In addition to this, we are implementing the proposed framework as suggesting relevant web pages to the user and then use this strategy to make the web search more personalized. Feedback will be provided from client side for the each search result cane given to server to improve the search result performance and quality of search. More security will be provided to user profile on transferring query data from client to server using cryptography.

3.1. Flow of Our Proposed System will be as Follows

1. Create Account in PWS system
2. User chooses his/her area of interest and makes privacy settings in offline mode.
3. Online profile is generated when user enter query by using greedyIL algorithm.
4. According to user privacy settings in sensitive topic the query transfer towards server in encrypted form by using random4 algorithm.
5. Result will be provided to the user in re-ranked form

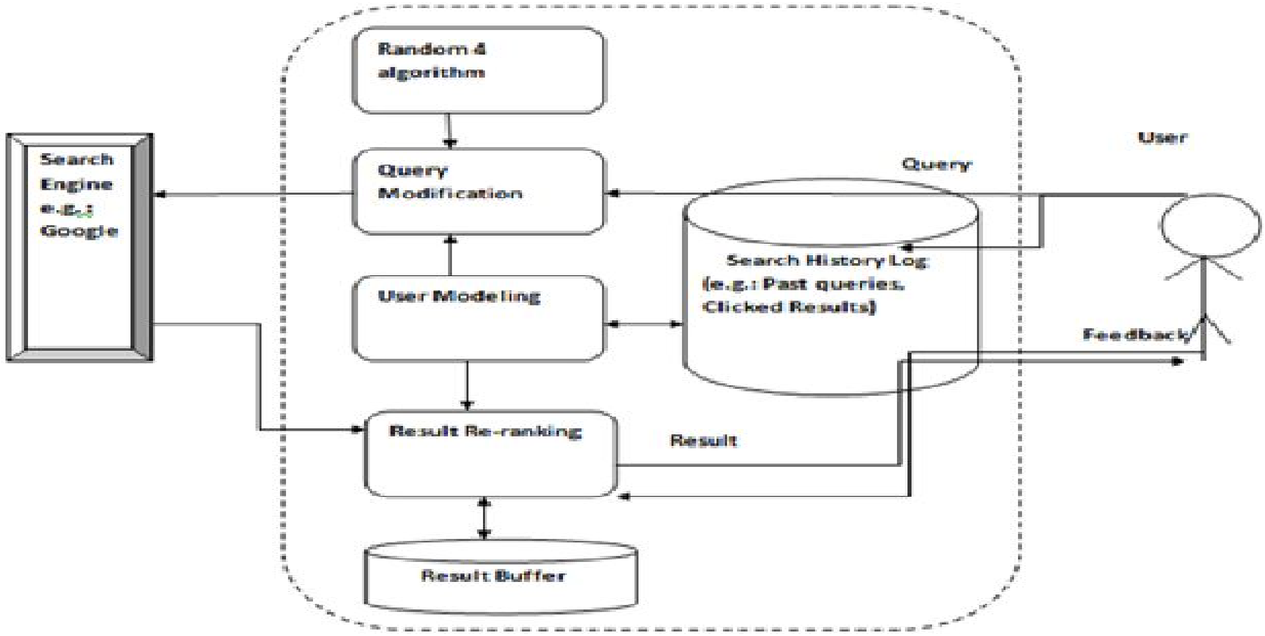


Fig. 1 Proposed Architecture

IV. ALGORITHMIC STRATEGY

4.1. Greedy Algorithm

We are using greedyIL algorithm for online profile generation when user fires a query. GreedyIL algorithm increases efficiency of profile and reduces the discriminating power of the profile. There is no any algorithm on the name of Greedy and GreedyIL. The both name has given by user for related to their task. Actually there is only greedy algorithm provide the solution for tree based input like(largest/shortest path on tree, finding max/min depth on the tree, etc) When we create profile construction, in that time the Greedy algorithm called as Greedy DP. When we find the sensitivity node/item, in that time the Greedy algorithm called as Greedy IL.

Steps of Greedy Algorithm

1) A item set (Category set), from which a solution is created
2) A selection function, which chooses the best item tube added to the solution
3) A feasibility function, which is used to determine if an item can be used to contribute to a solution
4) An objective function, which assigns a value to absolution, or a partial solution
5) A solution function, which will indicate when we've discovered a complete solution

4.2. Random4 Algorithm

The random 4 algorithm is based on randomization and issued to convert the input into a cipher text incorporating the concept of cryptographic salt. Random 4 algorithms are used for avoiding the attack from hackers in the client side.

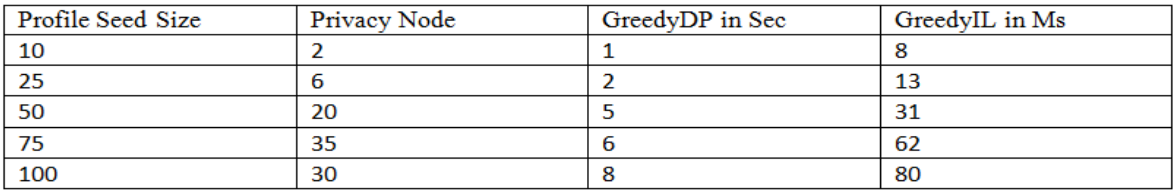
Steps for algorithm
1) Read input text
2) Loop through all characters in the input text
3) Convert every character into cipher text using cryptographic technique

4) Store the encrypted text

V.RESULT ANALYSIS

For the personalized search experiments, we measure the effectiveness of re-ranking in terms of Top-n Recall and Topn Precision. For example, at n = 100, the top 100 search results are included in the computation of recall and precision, whereas at n = 90, only the top 90 results are taken into consideration. Starting with the top one hundred results and going down to top ten search results, the values for n include n = {100, 90, 80, 70, ..., 10}. The Top-n Recall is computed by dividing the number of relevant documents that appear within the top n search results at each interval with the total number of relevant documents for the given concept.

Table1.Scalability of Varying Profile Seeds

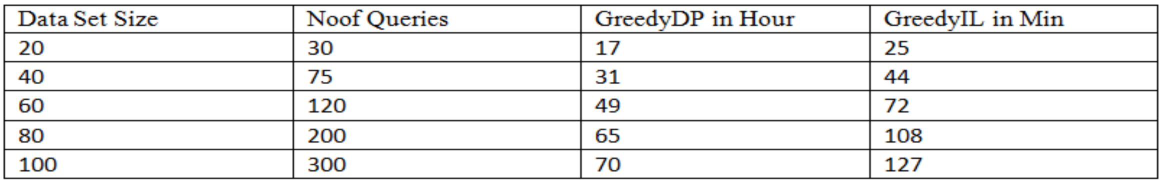


Profile Seed Size	Privacy Node	GreedyDP in Sec	GreedyIL in Ms
10	2	1	8
25	6	2	13
50	20	5	31
75	35	6	62
100	30	8	80

As of March 2015, the Google Open Directory contained more than 870,000 concepts. For experimental purposes, we use a branching factor of four with a depth of six levels in the hierarchy. Our experimental data set contained 100 concepts in the hierarchy and a total of 2000 documents that were indexed under various concepts.

The indexed documents were pre-processed and divided into three separate sets including a training set, a test set, and a profile set. For all of the data sets, we kept track of which concepts these documents were originally indexed under in the hierarchy. The training set was utilized for the representation of the reference, the profile set was used for spreading activation, and the test set was utilized as the document collection for searching. The test set documents that were originally indexed under a specific concept and all of its sub concepts were treated as signal documents for that concept whereas all other test set documents were treated as noise. In order to create an index for the signal and noise documents, a tf.idf weight was computed for each term in the document collection using the global dictionary of the reference ontology. The profile set consisted of 800 documents, which were treated as a representation of specific user interest for a given concept to simulate ontological user profiles. As we performed the automated experiments for each concept/query, only the profile documents that were originally indexed under that specific concept.

Table2.Scalability of Varying Data Sets



Data Set Size	Noof Queries	GreedyDP in Hour	GreedyIL in Min
20	30	17	25
40	75	31	44
60	120	49	72
80	200	65	108
100	300	70	127

CONCLUSION

This paper presented a client and server side privacy preservation framework called UPS for personalized web search as well as we used random4 algorithm for query encryption while transferring to the server side. This facility can protect user profile from attacks such as man in middle attack and provide privacy to each individual user of this system. The framework allowed users to specify customized profile for search with privacy requirements such as sensitive items via the hierarchical pre-build taxonomy profiles. In addition, UPS also performed online generalization on user profiles to protect the personal privacy i.e. sensitive item in user profile without compromising search quality. We proposed two greedy algorithms, namely GreedyDPand GreedyIL, for the online generalization and use to analyze user customized profile for the specific search query. In the proposed, we added a new concept called feedback on result which is going to play a major role on result re-ranking. It allows user to give feedback on query result to the server that will definitely improves the search quality of personalized web search. Our experimental results revealed that UP could achieve quality search results while preserving users customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution.

REFERENCES

[1] Linda Shou, He Bai, Ke Chen and Gang Chen,Supporting privacyprotection inpersonalized web search,, IEEE transaction on knowledgeand dataengineering vol:26 No:2 year 2014.

[2] J. Teevan, S.T. Dumais, and E. Horvitz, Personalizing Search viaAutomated Analysis of Interests and Activities,Proc. 28th Ann. IntlACM SIGIR Conf. Re-search and Development in Information
Retrieval (SIGIR), pp. 449-456, 2005.

[3] K. Sugiyama, K. Hatano, and M. Yoshikawa Adaptive Web SearchBased on User Profile Constructed without any Effort.Proc. 13thIntl Conf. World Wide Web(WWW),2004

[4] M. Spertta and S. Gach, Personalizing Search Based on User SearchHistories, Proc. IEEE/WIC/ACM Intl Conf. Web Intelligence (WI),2005.

[5] X. Shen, B. Tan, and C. Zhai Privacy Protection in PersonalizedSearch"SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.

[6] A. Viejo and J. Castell-a-Roca,Using Social Networks to DistortUsersProfiles Generated by Web Search Engines, ", Computer Networks, vol.54, no. 9, pp. 1343-1357, 2010.

[7] Y. Xu, K. Wang, B. Zhang, and Z. Chen Privacy-Enhancing PersonalizedWeb Search", Proc. 16th Intl Conf. (WWW), pp. 591-600, 2007..

[8] Y. Zhu, L. Xiong, and C. Verdery,Anonymizing User Profiles forPersonalized Web Search,", Proc. 19th Intl Conf. World Wide Web(WWW), pp. 1225-1226

[9] Shen, B. Tan, and C. Zhai ,Implicit User Modeling for PersonalizedSearch, , Proc. 14th ACM Intl Conf. Information andKnowledge Management,2005.10

[10] Arlein R.M., Jai B., Jakobsson M., Monrose F., ReiterM.K,"Privacypreservingglobal customization", In: 2nd ACM conference on electroniccommerce, pp. 176184.ACM Press, Minneapolis (2000).