ENHANCED SECURITY IN PERSONALISED WEB SEARCH

Leya Zacharias[1], Neema George[2]

Abstract— **Web search engine (WSE) provides massive amount of information .The aim of the WSE is to provide the relevant search result to the right user with the behavior of the queries they have given .Here the user privacy is under risk and also no securities were providing to their data. Personalized web search is done by adopting the Meta search approach which applies on one of the commercial search engine yahoo to execute a definite search. In order to derive personal favorite, the client is responsible for receiving the user's requests, submitting the requests to the web server, displaying the returned results and collecting profile details and search history. The server which is responsible for managing important tasks such as forwarding the requests to a commercial search engine, preparation and reranking of explore results before they returned to the client. The user's privacy is preserved by storing the user profiles for particular user on client's side. The server-client model is adopted in which user queries are forwarded to a server for making the training and reranking quickly. Privacy is preserved for both click-log based method and profile-based method. User's positive and negative feedback to improve the privacy is also considered. Here a working prototype of the clients on the yahoo platform is implemented**

Index Terms*— *Personalized web search, Search engine, Profile based method, click log based method, Greedy algorithm.

I. INTRODUCTION

Ordinary people look useful information from the web search engines. Web is now a day's common tool to everyone for searching their interesting topics and getting useful information .But it is not always necessary that they get the appropriate result that they are searching for**.** Search engine will never consider about who is submitting the query and what they want. Personalized web search (PWS) is a general category of search techniques which aims on providing better search results, according to individual user needs. Search-engine relevance and personalized Web Search can be improved by inferring user search goals, which is based on the click through log and the feedback session[12, 3].The User information collected and analyzed, so that the perfect search result can be given to the user according to the query.WSE [8] is the software that searches and identifies for the content or item from the web engine or web server or web database with correspond keywords or character specified by the user and finding particular sites on World Wide Web[13]. Data search and information retrieval has located high demands on search engines. Different search engines like Google, Yahoo provides relevant as well as irrelevant data to the user based on their search. Personalized Web Search (PWS) is used to avoid the irrelevant data arises which can be done based on the user profiles based on the click through log and the feedback session. From the frequently asked queries requested by the user, history of query, browsing, bookmarks the data were generated.

Fig 1: Personalized search engine Architecture

There are generally two categories, first are click-log-based methods and second is profile-based ones. Click-log method Performs the search based upon clicked pages in the user's query history. This is simple and straight forward. It can only work on repeated queries from the same user, which is a strong limitation and restricted for certain applications, although this method has been demonstrated to perform consistently and considerably well. But profile-based methods improve the search by generating user profiling techniques with complicated user-interest models. Profile-based methods provide more effective for almost all sorts of queries, but it is to be improper under some situations. There are reasons and consideration for both types of PWS techniques. The profile-based PWS has proved its more effectiveness in improving the accuracy of web search recently, with increasing usage of one's personal and behavioral information to profile its users, which is usually gathered implicitly with the help of query history, browsing

Manuscript received March, 2015.

Leya Zacharias*, CSE, M G University Kottayam, India, 9496736469.*

Neema George*, CSE, M G University Ettumanoor, India, 9400281133.*

histories, click-through data, bookmarks, user documents and so on. But, such type of collected personal data can easily reveal a threat of user's private life.

This paper proposes a privacy-preserving personalized web search structure UPS. According to user-specified privacy requirements every query can simplify profiles. Based on two incompatible metrics, explicitly personalization usefulness and confidentiality risk, for hierarchical user profile prepare the difficulty of privacy-preserving personalized search as risk profile overview, with its NP-hardness proved. Two simple but effective simplification algorithms, Greedy DP and Greedy IL will provide run time profiling. While the previous tries to maximize the discriminating power (DP), the other tries to minimize the information loss (IL).

II. RELATED WORK

This paper [4] web user want personalized services and web users want privacy was the two emerging trends motivated. The main challenge is that personal information must be anonymous under the assumption that the participating parties, including the web services, are not completely trusted, due to systematic collection of personal information in addition to queries. Another challenge is the online and dynamic nature of web users. This paper proposed the notion of online anonymity to protect web users and proposed an approach to maintain online anonymity through time. Third party user pool is used but it does not require the user pool must be trusted .The simulation study show promising results by which personalization achieve for reasonable privacy settings.

To improve information retrieval performance, the [3] authors have explored how to exploit implicit feedback information including query history and click-through history within the same search session. With the help of using the KL Divergence Retrieval model as the basis, author proposed and studied four statistical language models for context sensitive information retrieval. They are Fixit, BayesInt, OnlineUp and Batch Up. With the help of Trec AP Data to create a test set for evaluating implicit feedback models. The current process can be extended in different ways. First, for incorporating implicit feedback information it has only explored some very simple language models to better exploit query history and click through history it would be interesting to develop more sophisticated models.
As an example, this may treat a clicked summery differently depending on whether the current query is a generalization or refinement of the previous query. The next step is that the decided methods can be implemented in any practical systems currently it is developed as a client-side personalized search agent, who will incorporate some of the proposed algorithms. To evaluate effectiveness of these models in the real web search and to strengthen the same author will also do a user study. As a result, author should further study a general retrieval framework for sequential decision making in interaction information retrieval. He also conducts several studies on how to optimize some of the parameters in the context –sensitive retrieval models.

Here, [2] proposed that internet is one of the most important sources of knowledge in the present time. Internet renders a huge volume of data and information which grows and spreads dramatically every day after day. Online search engine tools are widely used to fetch and collect specific data among that information. However, these useful tools also represent a privacy threat for the users. The web search engines profile them by sorting and analyzing all the searchers that they have previously submitted. Current solutions propose new mechanisms that introduce a high cost in terms of computation and communication to address those privacy threats. The topic proposes a new scheme designed to protect the privacy of the users from an internet search engine that tries to profile them .The proposed system uses social networks to provide a distorted user profile to the web search engine. Without making any change in the server side the proposed protocol submits standard queries to their web search engine.

From this approach [5, 6] for gathering the required data and information from the web they require users to contribute the server full access to personal data and information. This in turn results in break user's privacy and search quality. An algorithm is provided to the user for collecting, abbreviation, and organizing their personal information into a hierarchical user profile. This is mostly a general terms are ranked to higher levels than explicit terms. With the help of this profile, users control what section of their private information is uncovered to the server by adjusting the minimum detail threshold. ExpRatio an additional privacy measure is proposed to approximation the amount of privacy is exposed with the specified minDetail value. Yet, this paper is an exploratory work on the two features: First, author deal with unstructured data such as personal documents, for which it is still an open problem on how to define privacy. Secondly, by breaking the premise on privacy as an absolute standard author try to bridge the conflict needs of personalization and privacy protection. Also, author believe that an enhanced balance between privacy protection and search quality can be achieved if internet search engines are personalized by allowing for only reveling those information associated to a specific query, It performs less protection and safety for the user data and they were no assured for the user data and their profile information's.

III. PROPOSED APPROACH

A personalized web search framework by the name UPS is formed to meet the user-specified privacy [1]. This can be used as a generalized profile for each query. With the help of definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as Risk Profile Generalization, with its NP-hardness proved.

Two simple but effective generalization algorithms where developed namely, GreedyDP and GreedyIL to support runtime profiling. GreedyDP tries to maximize the discriminating power (DP), and GreedyIL attempts to reduce the information loss (IL), GreedyIL outperforms GreedyDP significantly by exploiting a number of heuristics [1].

To personalize a query in UPS we provide an inexpensive mechanism for the client. This decision can be made before each runtime profiling to enable the stability of search results while avoid the un-necessary exposure of the profile.

A. Greedy Algorithm

A greedy algorithm is defined as a mathematical process that recursively constructs a set Recursion of objects from the smallest possible constituent parts. It is an approach to problem solving. Here the solution to a particular problem depends on solutions to smaller instances of the same problem.

Greedy algorithm look for simple, easy-to-implement solutions to complex, multi-step problems by deciding which next step will provide the most obvious benefit. As the optional solution to each smaller instance will provide an immediate output such algorithms are called greedy. Greedy algorithm doesn't consider the larger problem as a whole. In Greedy Algorithm, once a decision has been made, it is never reconsidered. Solutions to smaller instances of the problem can be straight forward and easy to understand are some of the advantages of using greedy algorithm. And the disadvantage is that it is entirely possible that the most optimal short-term solutions may lead to the worst long-term outcome. Greedy algorithms are often used in packets with the fewest number of hopes machine learning, business intelligence (BI), artificial intelligence (AI) and programming.

- Query Generalization using GreedyDP.

Anonymization methods are used to provide privacy for user query values. User query values are formed for privacy preservation. For the query generalization process Greedy discriminating power [1] (GreedyDP) algorithm is used. Generalized query values are updated in the user search history environment.

- Query Generalization using GreedyIL.

Query values are generalized with information lose factors. Greedy Information Loss (GreedyIL) algorithm is normally used for the generalization process[1]. Generalized query keywords are used in the search optimization process. Data usage is considered in the generalization process.

B. Personalized Search Process.

In the personalized search process Privacy Preserved web search is performed. Query optimization is used to improve the query keywords. Generalized keywords are used in the query optimization process. For the query optimization process Query weight values are used.
With the help of using heuristics based on several findings the GreedyIL [1] algorithm improves the efficiency of the generalization. Prune-leaf operation reduces the discriminating power of the profile is one of the most important findings. In other words it is explained that, the DP displays monotonicity by prune-leaf. The advantages of making the above runtime decision are, it improves the stability of the search quality and it avoids the unnecessary exposure of the user profile.

IV. PERFORMANCE EVALUATION

Experimental evaluation is done on the Macromedia Dreamweaver 8 with jsp support .Application server used here is Tomcat server. Mysql is used for backend application. Dreamweaver is a professional HTML editor for designing, coding and developing websites, webpage's and web applications. To run jsp pages, we need to run application server that supports java server pages so in order to support this tom cat server application is used. Experimental result shows that GreedyIL performs better than GreedyDP in terms of privacy and efficiency .Here both click-log based and profile based method is protected from unauthorized attack. So, any one with background knowledge can't access the user's private information. And also user can give feedback which will definitely improve privacy protection. Here minimum amount of information loss and maximum personalized search is considered.

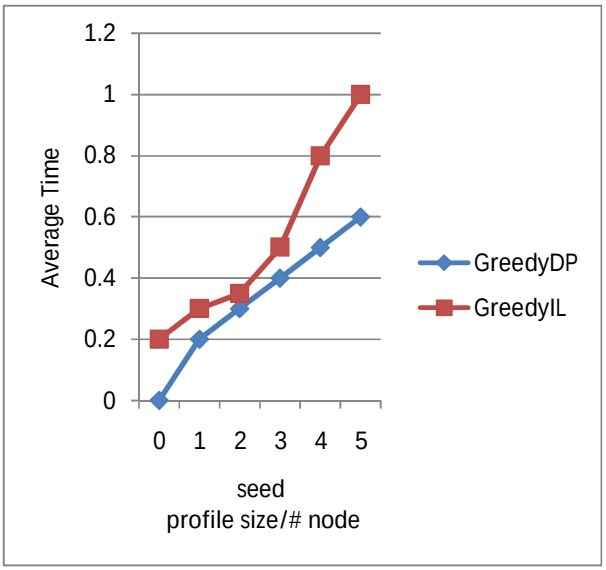

Fig 2 shows the comparison graph of GreedyDP and GreedyLP algorithm.

V. CONCLUSION

The day to day development of new information on the web has forced search engine to construct an effective one. Because of increasing number of user service engine must provide the relevant search result based on the user performance and their behavior. This work provides Ups framework for Personalized Web Search. This paper proposes a client-server model in which privacy is done for both click-log based method and profile-based method .Here keyword based search is done on the yahoo platform. Based on category and sub category given on the client side, a generalized profile will create for the user for the search done, and also with the help of the history and bookmark done by

the user. Based upon all the user will get most accurate result, which is personalized and also privacy is preserved. Also feedback from user is also included to improve privacy. Here user can either give a positive or negative feedback regarding the search or user can ask queries regarding the personalized search so that more security is provided.

ACKNOWLEDGMENT

The first author would like to thank her guide who has always been a motivation for carryout the paperwork. The first author would also like to thank all those people, who guided and supported. Without their valuable guidance and support, task would not have been possible and also likes to thank colleagues who have been a source of inspiration and motivation.

REFERENCES

[1] Lidan Shou , He Bai, Ke chen , and Gang Chen , "Supporting Privacy Protection in Personalized Web Search" vol. 26, no.2 ,Feb 2014.
[2] 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 .
[3] Shen, Xuehua,Bin Tan and Cheng Xiang Zhai."Context-sensitive information retrieval using implicit feedback"Pproceedings of the 28th annual international ACMSIGIR conference on research and development in information retrieval.ACM,2009.
[4] Xu,Yabo, et al. "Online anonymity for personalized web services" Proceedings of the 18th ACM conference on Information and knowledge management. ACM 2005.
[5] A. Viejo and J .Castell a-Roca," Using Social Networks to Distort User's Profiles Generated by Web Search Engines." Computer Networks ,vol 54 no. 9 pp.1343-1357,2010.
[6] Xu, Yabo ,et al." Privacy-enhancing personalized web search ." Proceedings of the 16th international conference on World Wide Web ,ACM ,2007.
[7] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.val (SIGIR), pp. 449-456, 2005.
[8] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW),2004.
[9] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating HierarchicalUser Profiles Using Wikipedia," HP Labs, 2008.
[10] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc.ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.
[11] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing User Profiles for Personalized Web Search," Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.
[12] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of Personalized Search Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.
[13] J.Teevan, S.T. Dumais and E. Horvitz, "Personalizing Search via Automated Analysis of Interest and Activities," Proc. 28th Ann.Int'1 ACM SIGIR Conf. Reasearch and Development in Information Retrieval (SIGIR),pp. 449-456, 2009.
[14] X. Xiao and Y. Tao, "Privacy Protection in Personalized Search," SIGIR Forum ,vol 41, no. 1pp. 4-17, 2007.

Leya Zacharias, Department of Computer Science & Engineering, Mangalam college of Engineering, Ettumanoor, Kerala, India.

Neema George, Assistant Professor, Department of Computer Science and Engineering, Mangalam college of Engineering, Ettumanoor, Kerala, India.

