ISSN:0975-9646

Mahesh Holkar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4709-4711

Supporting privacy Preserved in personalized web search Engine

Mahesh Holkar, Kunal Patil, Khemraj Kakade, Ajay Pakhale
[1]Computer Dept, *Sandip Institute of engineering & management, Nashik*

***Abstract*- Personalized web search (PWS) has representing for improving the quality of various search services on the internet. Since content in Internet is growing rapidly, the search provider users demand accurate, search results as per their need. But one option is available to users is PWS means personalize web search that is specially used for the personal data of user provided to the search provider. Even if evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. This paper models preference of users as hierarchical profiles of users. It proposes a framework which is known as UPS & it generalizes profile at the same time, it is used for the maintaining the privacy requirements by user[1]. We use the two Greedy algorithms namely Greedy-DP & Greedy-IL which are used for run-time generalization also we provide an online prediction mechanism for deciding whether personalizing a query is beneficial in this paper.**

Index Terms- UPS, PWS, generalization, and Proliferation

I. INTRODUCTION

The web search engine is the very important portal for Common people looking for useful data on the web. However, users usually experience failure and get unrelated results when search engines return irrelevant results that do not meet their real goal. Now a days modern technique personalize web search is used for in order to provide the better search result .In personalize web search, user information is gathered & analyze in order to find the goal behind issued query fired by user.

Click-log-based & profile based are two categorize of PWS. 1) The click-log based method is very straight-forward & simple, it imposes bias to clicked pages in the users query history. This method is performing good, But it works on the repeated query from the same users which is the strong limitation of its applicability[1]. 2) While profile-based method enhances the search quality using the profiling techniques. Profile-based methods are varying, but it is effective for the lots of queries. There are advantages & disadvantages for both type of PWS techniques, profile based PWS is more effective for improving the search results. The user profile is created with the help of whatever the user has searched up till now.

II. LITERATURE SURVEY

The existed personalized web search which are based on profile they do not support the runtime.

In this search method A user profile is typically generalized for only once offline, and used to personalize all queries from a same user randomly. Such "one profile fits all" strategy certainly has drawbacks given the variety of queries. One proof reported in is that profile-based personalization may not even help to improve the search quality for some temporary queries, though exposing user profile to a server has put the user's privacy at risk.[2]

The existing methods do not take into consideration the customization for privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. Many of personalization techniques require recursive user interactions when creating personalized search results. They usually filter the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on. This paradigm is, however, infeasible for runtime profiling, as it will not only pose too much risk of privacy violation, but also demand prohibitive processing time for profiling. Thus, we need predictive metrics to measure the search quality and violation risk after personalization, without incurring iterative user interaction.

Some major Drawbacks are given below-

1. It do not support run-time profiling
2. This method do not take into account customization of privacy requirements.
3. All the sensitive topics are detected.[2]

III. PROPOSED SYSTEM

We developing a system using privacy preserving personalized web search framework named as UPS, which will generalize profile for each query as per the user specified privacy requirements. Depending on the definitions of two conflicting metrics, namely personalization utility & privacy risk for generating hierarchical user profiles also we categorize the problem of privacy .We implementing two simple but efficient & effective generalization algorithms named as Greedy-DP & Greedy-IL To support runtime profiling .In those one tries to maximize the discriminating power (DP) & other i.e. latter attempts minimize the information loss (IL).[1]

We are trying to provide an cheaper mechanism for the client to decide whether to personalize a query in UPS .This decision is taken before each runtime profiling to improve the stability of the search results while avoid the unnecessary expose of user profile.

IV. SYSTEM ARCHITECTURE

A. No personalization

In web searching application, client-server architecture as given in the following fig, it's a common scenario where client i.e. web browser send the request in the form of query to the server i.e. search engine. The search engine analyze user interest as per the information provided in the

query, & finds its index structure relevant to that query & returns the list of results as per the ranking to the user to view. A search engine maintains user search information in the logs for various purposes such as personalization & anti-spam. Thus it's up to the search engine to maintain user search logs & not to automatically remove them. Typically there are three software architecture that used for basic client-server model of the web search to include personalized search.

In this section, we describe those three types of software architectures & analyze what are different privacy preserving levels can be achieved using different architectures[3].

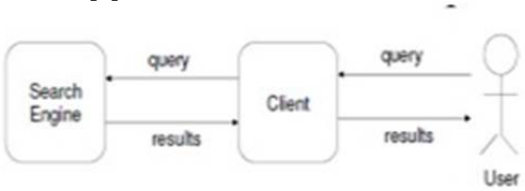

Fig 4.1. No Personalization

B. Server-side personalization

In the server side personalization shown in the above fig. the personal information which can be identified is stored on the search engine side. The Search engine also responsible for creating & updating user profiles using either users explicit input i.e. as per the user interest specified in the users query or information which is maintained by search engines in users logs implicitly i.e. query & search history. Both methods require the user to create an account to identify him.[3].

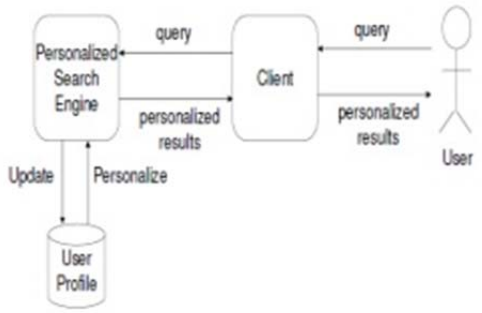

Fig 4.2. Server-side Personalization

But the latter methods does not requires any additional effort from the user and contains more description of user information need .The advantage of this architecture is that the search engine can use all of its resources (e.g., document index, common search patterns) in its personalization algorithm. Also, the client software generally requires no changes. This architecture is used by some general search engines such as Google Personalized.[7] In the current scenarios search engines required user to specify his/her interest in the form of query before information from logs is collected & used for personalization. If the users grants permission, the search engine will maintain all the personalized information which is already available on the server side. As per the users view it even does not have privacy protection of the level 1.

C. Client-side personalization

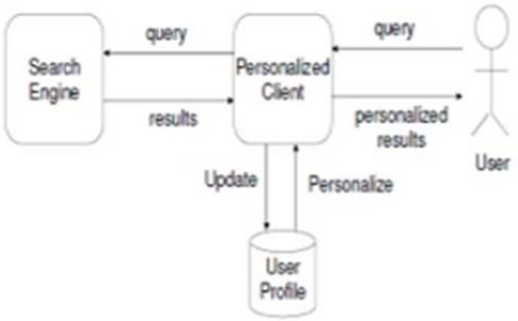

Fig 4.3. Client-side Personalization

In this architecture as shown in above fig, user sends the query to the search engine & search engine returns the relevant result to the particular user as per the common web search scenarios. In client side personalization there is a role of client side search agent, which does the query expansion & modifies the original query & send to the search engine. The client side personalized search agent also re-arranges or re-ranks the results of the query as per the users liking or interest after receiving results from the search engines. In this client side personalization architecture users search behavior is identified ,that means of what type of results user actually wants & to retrieve those results what query he/her fires on the search engine along with his contextual activities such as browsing history. The sensitive contextual information is generally not a major concern since it is strictly stored and used on the client side. Another benefit is that the overhead in computation and storage for personalization can be distributed among the clients. The main drawback of this architecture is that client side algorithm does not have knowledge of the server side (e.g., Page Rank score of a result document ,average ranking)[5].

D. Client-server co-operative personalization

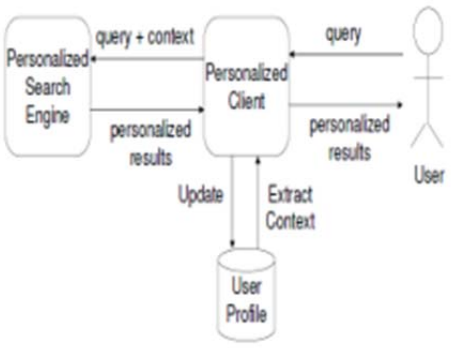

Fig 4.4. Client- Server co-operative Personalization

In the Client-Server Cooperative Personalization as shown in the above fig, there is the combination of both client side personalization & server side personalization. In the previous two architecture either only server or client was participating in the personalization, but here we are proposing such a architecture in which both client & server will participate in personalization. Here when user fires the query to the search engine at that time that users context information is extracted from the particular user profile

which is available on the client side. After extracting users contextual information will be added to the users actual query & sent to the search engine i.e. server.[6] After receiving query along-with contextual information search engine i.e. server will perform personalization. Here server side personalization is nothing but comparing search results with the particular user profile. If search engine founds any new results other than contextual information that new results will be updated to particular users profile & all the results along with new results & users previous context available in the particular user profile are will be sent to the that particular user.

The main advantage of this architecture is that every time user will get results that are relevant to his /her interest only.

V. CONCLUSION

The remarkable development of information on the Web has forced new challenges for the construction of effective search engines. The proposed work provides information on user customizable privacy preserving search framework-UPS for Personalized Web Search. UPS could potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The framework allowed users to specify customized privacy requirements via the hierarchical profiles. Another important conclusion we revealed in this proposed work is that personalization does not work equally well under various situations. The click entropy is used to measure variation in information needs of users under a query. Experimental results showed that personalized Web search yields significant improvements over generic Web search for queries with a high click entropy. For the queries with low click entropy, personalization methods performed similarly or even worse than generic search. As personalized search had different effectiveness for different kinds of queries, we argued that queries should not be handled in the same manner with regard to personalization. The proposed click entropy can be used as a simple measurement on whether a query should be personalized. For future work, we try to resist adversaries with border background knowledge including exclusiveness, sequentially and so on or the capability to capture a series of queries from the victim.

REFERENCES

[1]. Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 591-600, 2007.

[2]. 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.

[3]. 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.

[4]. A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services," J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.

[5]. Y. Xu, K. Wang, G. Yang, and A.W.-C. Fu, "Online Anonymity for Personalized Web Services," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.

[6]. 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.

[7]. J. Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.

[8]. 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),pp. 139-148, 2008.

[9] 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, pp. 615-624, 2011.

[10] A. Viejo and J. Castell_a-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.

[11] F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.

AUTHORS

First Author – Mahesh Holkar: Student of BE in Computer Engineering in *Sandip Institute of engineering & management, Nashik*

Second Author– Kunal Patil : Student of BE in Computer Engineering in *Sandip Institute of engineering & management, Nashik*

Third Author –Khemraj Kakade: Student of BE in Computer Engineering in *Sandip Institute of engineering & management, Nashik*

Fourth Author – Ajay Pakhale: Student of BE in Computer Engineering in *Sandip Institute of engineering & management, Nashik*