

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

A UPS Framework for Providing Privacy Protection in Personalized Web Search

V. Sai kumar¹, P.N.V.S. Pavan Kumar²

PG Scholar, Dept. of CSE, G Pulla Reddy Engineering College, Kurnool, Andhra Pradesh, India¹

Assistant Professor, Dept. of CSE, G Pulla Reddy Engineering College, Kurnool, Andhra Pradesh, India²

ABSTRACT: Web search engines (e.g. Google, Yahoo etc.) are used to find the information among a huge amount of data within a less amount of time. Most of users prefer these search engines for getting information and the data is searched based on the keyword or query given by users over the internet. The data over the internet is growing dramatically and the users spend lot of time to get the information where the users are provided with irrelevant information. Even the users are uncomfortable in exposing private preference information to search engines. In this case, personalized web search has demonstrated its effectiveness in improving the quality of various search services over the internet and provide privacy protection in PWS applications that models based on the user interest as hierarchical user profiles. A framework called UPS that is used to generalize profiles by queries at the same time respecting user specified privacy requirements. Mainly, this paper provides effective search and privacy protection in personalized web search by runtime generalization that strike balance between search quality and privacy.

KEYWORDS: web search, Privacy protection, UPS framework

I. INTRODUCTION

Generally, web search engines are mostly used for getting information on web. Sometimes user will get some irrelevant results that do not match their real intentions. In some cases search engine provides same set of results without the query submitted. In the personalized web search system which gives output to the user based on the highly ranked pages, thus the personalized web search(PWS) aims to provide better search results based on the user query. So, for this user information has to be collected and analyzed so that perfect search results are required for the user behind the issued query and solution is provided for this problem by Personalized Web Search. Basically classified into two types, one is click-log-based methods and other is profile-based ones. The method click-log-based is very easy and straight forward, in this method the operations performed based upon the number of clicks made by the user on pages in query history. Even though in this method has been demonstrated to perform consistently and considerably well but it only work on repeated queries from the same user, which is a one of the disadvantage.

In profile based method improve the search experience with complicated user interest models generated from user profiling techniques. This method can be more effective for almost all sorts of queries and they are reported to be improper under some situations. Even though there are some reasons and considerations for both types of PWS techniques, but profile based personalized web search has proved in improving the quality of web search, with help of increasing usage of one's personal and behavioral information to profile its user which is collected through the history of query and browsing history, bookmarks. Based on the collected personal data one can easily get the entire scope of user personal data and protecting privacy issue that arises due to lack of protection for the data. Privacy concern has become the major barrier for wide use of personalized web search.

II. BACKGROUND

In order to protect privacy in profile based method, researchers have to consider two important things and contradicting issues during search process. And in the first issue they attempt to improve the search quality with the personalization utility of the user profile. In the second issue they need to hide the privacy contents existing in user

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

profile in order to control privacy risk of user profile. In situation like identical situation, significant gain can be obtained by personalization at the expense of only small portion of the user profile namely generalized profile. That is way user privacy can be protected without comprising the personalized search quality. In normal there is a compromise between the search quality and the level of privacy protection achieved from generalization.

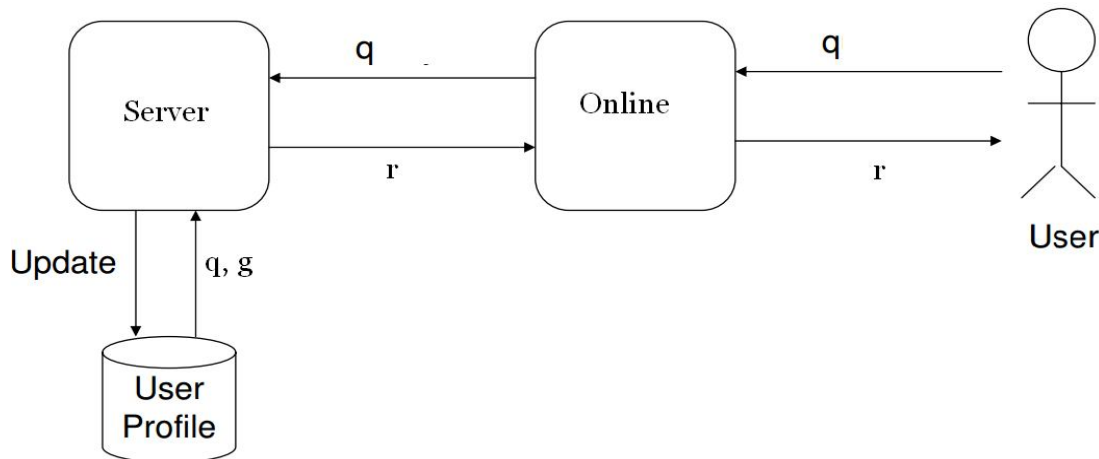


Fig1: Existing System Architecture

So far, the previous works of privacy preserving PWS are far from optimal. And the problems with the existing methods are explained in the following observations

1. Present profile-based PWS does not support runtime profiling. And a user profile is typically generalized for only once offline and used to personalize all queries from a same user in discriminatingly. In such way that "one profile fits all" strategy certainly has drawbacks given the variety of queries. In some cases, it already proved that Profile-based personalization may not even help to improve the search quality for some ad hoc queries and even exposing the user profile to a server has put the user's privacy at risk. And for better results to make an online decision on:
 - a. whether to personalize the query and also to expose the user profile at runtime.
 - b. Till now no previous work has supported such feature.
2. And the present methods do not take into account the customization of privacy requirements. This makes some user privacy to be over protected while others are insufficiently protected. For example, in the case of all sensitive topics are detected using an absolute metric called surprised based on the information theory, by assuming that the less user interests document support are more sensitive.
3. And even many personalization techniques require iterative user interactions while creating personalized search results. And they usually refine the search results with some metrics which require multiple user interactions.
4. There are two classes of privacy protection problems for PWS. One class includes these treat privacy as the identification of an individual. And other includes the sensitivity of the data, particularly the user profiles while exposing to the PWS server.

III. RELATED WORK

Above drawbacks can be overcome by using UPS (User Customizable Privacy Preserving Search) framework. As given in Fig.2, UPS framework consists of untrust search engine and number of clients. Every client accessing the search service trusts no one but himself/ herself. And the key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself. The proxy maintains the complete user profile, in a hierarchy of nodes with semantics, and the user-specified privacy requirements represented as a set of sensitive nodes.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

Framework works in two phases for every user, namely the offline and online phase. In the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. And online phase handles queries as follows:

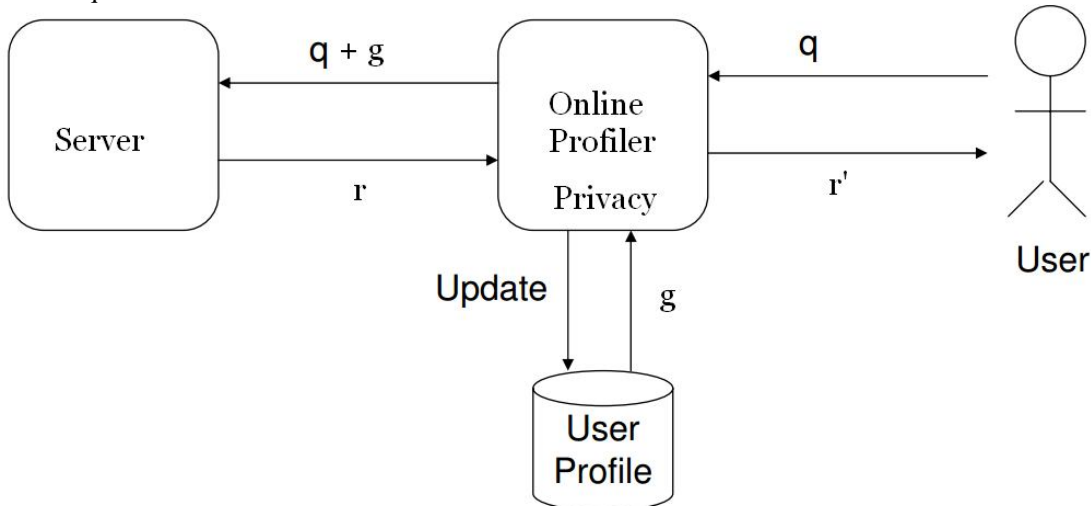


Fig2: System architecture of UPS framework

1. When user gives a query q_i to the client, the proxy generates a user profile in runtime in the light of query terms. And the output of this step is a generalized user profile G_i satisfying the privacy requirements. And generalization process is followed by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.
2. Both the query and the generalized user profile are sent together to the PWS server for personalized search.
3. And the search results are personalized with the profile and return back to the query proxy.
4. Lastly, the proxy may presents the raw results to the user or reranks them with the complete user profile.

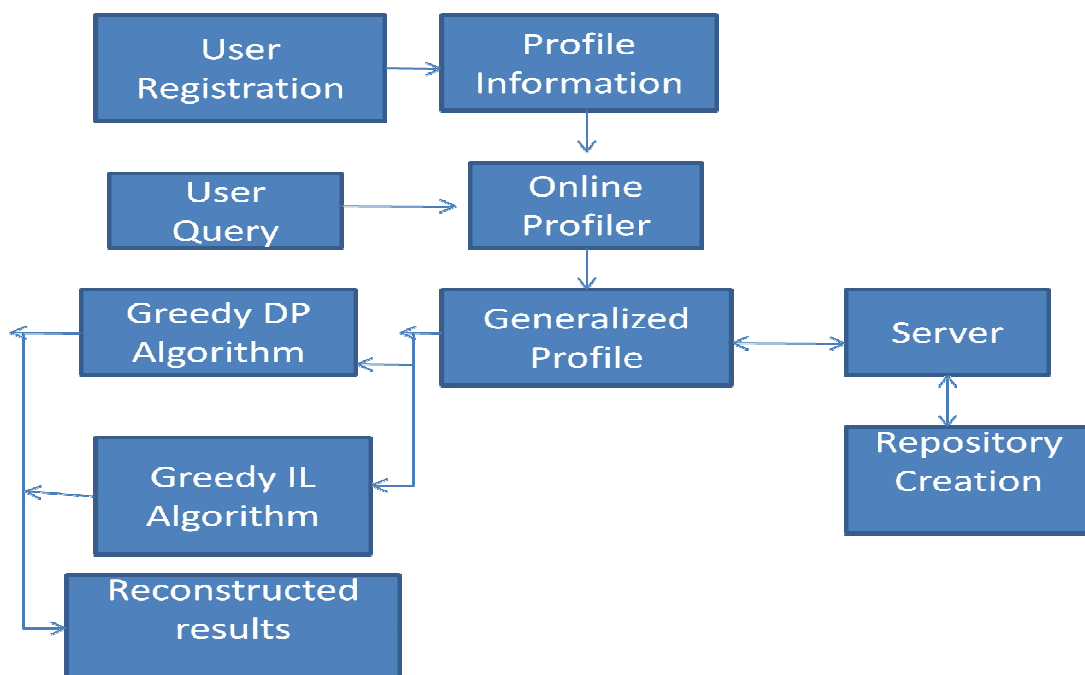


Fig3. Framework Workflow

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

A. Modules Description

1. *Profile-Based Personalization*: An approach to personalize digital multimedia content based on user profile information. Mainly these two main mechanisms were developed for this: the profile generator that automatically creates user profiles representing the user preferences, and the content-based recommendation algorithm that estimates the user's interest in unknown content by matching the profile to metadata descriptions of the content. These features are integrated into a personalization system.

2. *Privacy Protection in PWS System*: A PWS framework called UPS that generalize profiles for each query according to user-specified privacy requirements. Mainly, two predictive metrics are proposed to evaluate the privacy breach risk and the query utility for hierarchical user profile. But effective generalization algorithms are used for user profiles allowing the query-level customization by using proposed metrics. And online prediction mechanism based on query utility for deciding whether to personalize a query in UPS.

3. *Generalizing User Profile*: Generalization process has to meet specific prerequisites to handle the user profile. And it can be achieved by pre-processing the user profile. First, the process initializes the user profile by taking the indicated parent user profile into account. And the following process adds the inherited properties to the properties of the local user profile. Therefore, after the process loads the data for the foreground and the background of the map according to the described selection in the user profile.

In addition, by using references enables caching and is useful during the implementation in a production environment. And these references are used to the user profile that are used as an identifier for already processed user profile and allows performing the customization process once, but reusing the result multiple times. An update of the user profile is also propagated to the generalization process. It requires specific update strategies, that checks after a specific timeout or a specific event, when the user profile has not changed yet. As the generalization process involves remote data services which might be updated frequently and the cached generalization results might become outdated. So selecting a specific caching strategy requires careful analysis.

4. *Online Decision*: The profile-based personalization contributes little or even reduces the search quality, while exposing the profile to a server would for sure risk the user's privacy. In order to address this problem, an online mechanism is developed to decide whether to personalize a query. The basic idea is straightforward. if a distinct query is identified during generalization, the entire runtime profiling will be aborted and the query will be sent to the server without a user profile.

IV.CONCLUSION

A client-side privacy protection framework called UPS for personalized web search. It can be potentially adopted by any PWS that captures user profiles in a hierarchical taxonomy. And framework allows users to specify customized privacy requirements via the hierarchical profiles. It also performs online generalization on user profiles to protect the personal privacy without compromising the search quality. Two greedy algorithms, namely Greedy DP and Greedy IL are used for the online generalization.

REFERENCES

- [1]. 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.
- [2]. X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.
- [3]. 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.
- [4]. 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.
- [5]. X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.
- [6]. 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.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

BIOGRAPHY



V. Sai Kumar is a PG Scholar in Computer Science and Engineering at G Pulla Reddy Engineering College, Kurnool. My research area is Data Mining.



P.N.V.S. Pavan Kumar is working as Asst. Professor, Computer Science Department at G Pulla Reddy Engineering College, Kurnool. He has ten years of teaching experience. His research areas are Image Processing and DWH.