

# Efficient Privacy on Personalized Web Search Using Web Transformation Technique in User Profile

V. Arunpandi

Department of Computer science & Engineering  
Angel College of Engineering and Technology  
Tirupur, India  
arunpandi077@gmail.com

S. Sundaramoorthy

Department of Computer science & Engineering  
Angel College of Engineering and Technology  
Tirupur, India  
sundaramoorthy87@gmail.com

**Abstract**—The time required for query processing over the internet is high due to the massively increasing amount of data on the server. Sometimes we may get irrelevant information as a result for a query. So we go for Personalized Web Search (PWS) to make the query processing good. In PWS, the query processing is done with the help of user profile. The user profile is created in two manners namely implicit and explicit. The implicit method creates the user profile from user's browser histories, email, document and etc., without any effort from the user. Through this method the profile created with some user's personal and secret information. Exposure of secret information on web leads to the privacy problem. In another way that the profile was created by explicit method. In this method the users requested to create their profile manually on the web. After profile creation the query processing is takes place. At each time a query is generated by a user that is combined with the personalized profile to generate a personalized query. Now the generalized query is send to the server. The server process the query then ranks the collected information. Finally the results are given to the client side and viewed to the user. The profile is updated in both ways at each time of query processing (automatically) and also by the manual update. To increase the privacy protection the profile details is reviewed at users own time. They can hide their secret information from the profile. Each profile updating process checks the newly added field information with the already hided field information. If any newly added field information matches with the personalized information then a notification is generated automatically to alert the user to personalize their profile.

**Keywords**—Privacy Protection; Personalized Web Search; Profile Based Web Search; Privacy Protection in PWS;

\*\*\*\*\*

## I. INTRODUCTION

Searching information for a query makes the process slow due to the large amount of data over the servers. The data are replicated (i.e. duplicated) and stored in more than one server. So the irrelevant information for queries is high. To eliminate multiple, irrelevant information's and to reduce the time consumption for query processing, a new searching method is introduced called profile based web searching. The profile based search is used to improve the quality of information search over the internet. In implicit and explicit methods the profile creation was done. Through this, query processing is effectively performed. In the user profile the user details are included as profile field information in implicit or explicit methods. High exposure of user information over the internet leads to the security problem. The third party may extract user's private information from the profile. Providing protection for the user profile is major task. This paper discusses the different profile creation methods and privacy problems in existing systems, provides the way to handle privacy problem in proposed system.

## II. RELATED WORK

Personalized web search is a way to increase the quality of the search. The profile creation has involves in searching user information and collecting the obtained information. The collected information contains web cache details and user desktop activity details. The using the collected information the profile was created. The desktop activity detail gives large amount of data for profile. At each time of query processing the profile was updated. Each selected links are ranked and stored as a profile detail under corresponding category. The personalization takes place in the type. First one the

personalization is done at client side. The query is combined with the profile to generate the generalize query. Now the query is processed at the server side and the results are presented at client side. Second one is server side personalization. In this the profile is kept at the server side. When the query is generated the personalization is takes place at the server side. The third one is client-server cooperative personalization. The profile is kept at client side and personalization is also takes place at client side. The existing system gives the personalized web search method and proposed system gives the way to increase the protection for user privacy details.

## III. EXISTING SYSTEM

The Personalize web search (PWS) introduces UPS method. That holds the user profile at client side. The query (Q) is generated at the user side and given to the online profiler. The profiler generates the generalized profile (G). The combination of both the generalized query (Q) and generalized profile (G) is send to the server. The server search for the result and provides the obtained result to the client side. The online profiler rearranges the received results (R').

The UPS framework personalizes the profile at each time before the runtime profiling to avoid the exposure of profile data. This method has some disadvantages. It requires the user interaction at each time profile personalization. Large profile leads to the time to generate generalize the profile and to generalize the query. Existing system does not support for the runtime profiling. Privacy Requirements are not customizable in this framework.

### a) Profile Based PWS

A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminately. Such “one profile fits all” strategy certainly has drawbacks given the variety of queries. Profile-based personalization may not even help to improve the search quality for some ad hoc queries, though exposing user profile to a server has put the user’s privacy at risk. A better approach is to make an online decision on whether to personalize the query and what to expose in the user profile at runtime.

b) Customization of Privacy Requirements

This considers, all the sensitive topics are detected using an absolute metric called surprisal based on the information theory, assuming that the interests with less user document support are more sensitive.

c) Iterative User Interaction

They usually refine 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 breach, but also demand prohibitive processing time for profiling. Thus, we need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

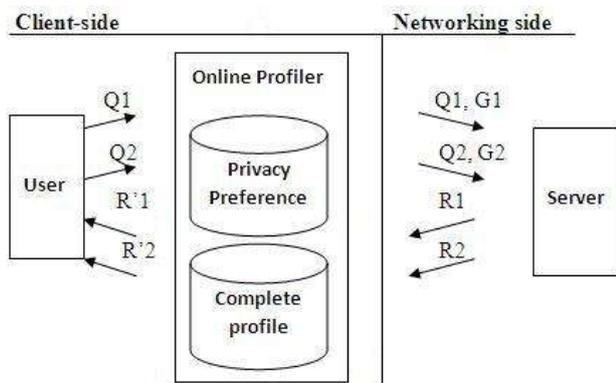


Fig.1 System Architecture of UPS

IV. PROPOSED SYSTEM

The amount of data over the web is increasing day by day. A personalized web search is technique to search a data over the internet. An important requirement for building personalized web search is to build user profiles that represent the users' interests. There are two representations commonly used for user profiles. One is using frequently occurring words in user documents. This creates large profiles where profile terms have low precision and have insufficient context to determine the user interests. The other is using user’s interaction to create the profile explicitly. In both ways privacy-preserving personalization is a major problem. This project provides a better technique to protect the user profile.

The proposed system uses user interaction for explicit profile creation in a privacy-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified privacy requirements. Relying on the 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. We provide an inexpensive mechanism for the client to decide whether to personalize a query in UPS. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile. General and specified profile details are analyzed for relation among it. Founded relations are displayed as notification to the users. Our extensive experiments demonstrate the efficiency and effectiveness of our UPS framework.

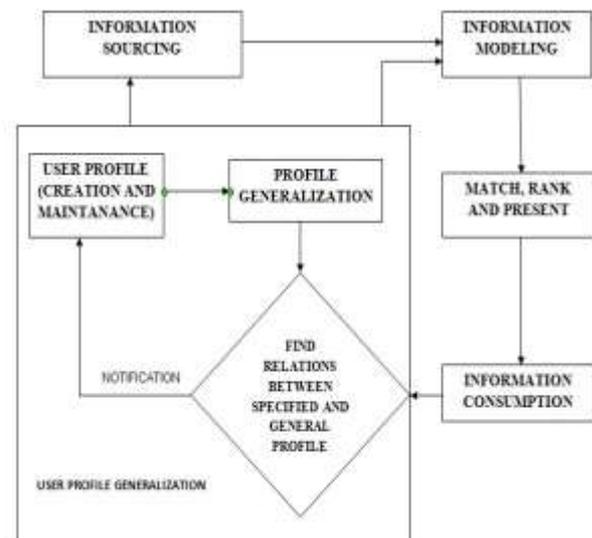


Fig.1 Profile Generalization in UPS

UPS provides a way to generalize the user profile. The user explicitly gives the details and implicitly some details from browse history, documents and click-based are taken for profile creation. The general and specified details are classified by the user interaction. After this the general and specified fields are analysed to find the relations among these fields. If any available then a notification is generated and displayed to user. The profile may be generalized again.

The explicit profile creation and the customized generalization of the profile lead to improved query searching with improved user profile protection. The major merits of the proposed works are

- Increasing usage of personal and behaviour information to profile its users, this is usually gathered implicitly from query history, browsing history, click-through data bookmarks, user documents, and so forth.
- The framework allowed users to specify customized privacy requirements via the hierarchical profiles.
- UPS also performs online generalization on user profiles to protect the personal privacy without compromising the search quality.
- The profile modification is done in both explicit and implicit manner.

a) User Profile Modeling

A user profile is created explicitly. In the explicit manner the user need to interact with the system to create the user profile manually. The user need to input some own details (eg: Name, age, designation, area of interest, etc..). After filling the form the profile creation is completed by clicking the submit option. Now the profile can be displayed to the user. Further the profile can be modified by explicitly or implicitly. To update the user profile, we need to classify the web pages accessed by a user into particular category. This process is taken place when browsing process is started after profile creation. Each links are categorized and arranged in hierarchical profile order for easy understanding and easy of query matching. The query matching process uses the profile to generalize the profile along with the query. The user can protect their details by hiding the profile details manually.

#### b) Enhanced User Profile

Profile enhancement is the process of improving the amount profile data by updating in implicit or explicit method. An Enhanced User Profile improves the User Profile by using the Domain Knowledge. For preparing the Enhanced User Profile we have considered each URL of the User Profile, match it with Domain Knowledge URLs and add most relevant URLs to the Enhanced User Profile. The profile details are classified as specific and general. The data which are very common and occurred many times that are known as general fields. The data which are occurs rarely and user preferred to hide are called specific. Mostly the specific data are hided from the profile to make privacy protection better. Enhancement of profile is also known as profile updating.

#### c) Web Personalization

Web personalization is the process of "tailoring" web pages to the needs of specific users using the information of the users' navigational behavior and profile data. Automatically synthesizes index pages which contain links to pages pertaining to particular topics based on the co-occurrence frequency of pages in user traversals, to facilitate user navigation. Each URL for each query is categorized into specific or general terms. According to this the URL is added to the profile. When already hidden field is related to newly added field and that is not hidden, then the system automatically generates a notification that is displayed to the user to personalize the profile manually. If personalization is not taken place here the attacker may find the hidden field by using this new field. Web personalization minimizes the query size with the profile details that transmitted through the internet.

#### d) Web Transformation

Web transformation involves changing the structure of a website to facilitate the navigation for a large set of users instead of personalizing pages for individual users. An approach to reorganize web pages so as to provide users with their desired information in fewer clicks. Placing a same URL in more than one category makes the large profile. This leads to time complexity to match a query with the profile. Instead of placing a same query in different places we just link the categories. Each category can have more than one sub category and URL. For example cricket is main category. India, Pakistan, Sri Lanka are the teams that are classified as sub categories and that are kept under the cricket category.

The URLs are arranged under these categories by the specification.

#### e) Out Degree Threshold

Web pages can be generally classified into two categories: index pages and content pages. An index page is designed to help users better navigate and could include many links, while a content page contains information users are interested in and should not have many links. Thus, the out-degree threshold for a page is highly dependent on the purpose of the page and the website. Typically, the out degree threshold for index pages should be larger than that for content pages. Because the index pages are added to the profile. When the threshold is high the profile's weight is also high. The large profile makes the system slow and low performance. The automated notification is generated and displayed to the user to ask the user to minimize the profile weight by deleting some old and useless profile details. The profile weight is minimized or kept in normal size by the above action and this maintains the system performance in constant level.

## V. CONCLUSION

This paper presents a new way to protect the user privacy over the web, while using personalized web search. Time required for query processing is high due to the large amount of data over the internet. Because of duplication of same information, the ambiguous result increases. To avoid the waste of time and to reduce the irrelevant data, a profile based web search is introduced. This requires the user details to create the user profile in implicit or explicit manner. The UPS framework requires the user interaction at each time of profile personalization. The PWS has some disadvantages like iterative user interaction, it does not support for the runtime profiling. The proposed system provides some techniques to overcome these problems. The proposed system supports for runtime profiling. When the profile is exceeds the limit then the system generates the notification to reduce the profile weight.

## REFERENCES

- [1] Lidan shou, He Bai, Ke Chen and Gang Chen, "Supporting Privacy Protection in Personalized Web Search", IEEE Feb 2014.
- [2] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschutter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIRConf. Research and Development Information Retrieval (SIGIR), 2005.
- [3] A. Kobsa, "Privacy enhanced personalization", CACM vol. 50, no. 8, August 2007.
- [4] Y. Xu, B. Zhang, Z. Chen and K. Wang, "Privacy enhancing-personalized web search", 16<sup>th</sup> Int'l conf. world wide web (WWW) pp. 591-600, 2007.
- [5] Krishnan Ramanathan, Julien Giraudi, Ajay Gupta "Creating hierarchical user profiles using Wikipedia," inHPL-2008-127.
- [6] A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services," J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.
- [7] Xuehua Shen, Bin Tan, ChengXiang Zhai, "Privacy Protection in Personalized Search", in ACM SIGIR Forum Vol.41 No.1 June 2007.
- [8] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 449-456, 2005.

- [9] R.Kumar and A.Sharan, "Personalized web search using browsing history and domain knowledge", in *ieee* Feb. 2014.
- [10] Nicolaas Matthijs and Filip Radlinski, "Personalizing Web Search using Long Term Browsing History", *WSDM'11*, February 9–12, 2011, Hong Kong, China.
- [11] Yun Zhu, Li Xiong and Christopher Verdery, "Anonymizing User Profiles for Personalized Web Search", *WWW 2010*, April 26–30, 2010, Raleigh, North Carolina, USA.
- [12] Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without Any Effort from Users" in *WWW2007*.
- [13] 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.