

A Survey on Privacy Assurance in Personalized Web Search

Abhilasha V. Biradar¹, K. B. Sadafale²

^{1,2}Department of Information Technology, Sinhgad College of Engineering Pune, India

Abstract: *Web search engine is a portal for ordinary people to access required information from the ,but generic web search provides the same results to different users for the same query. To overcome this drawback, personalized web search (PWS) develops which provides tailored results to the particular user based on his interest and preferences. For this purpose user information has to be collected and analyzed to find out the intention behind the query. However, the major barrier to the wider use of PWS is a user unwillingness to reveal their private information. Security mechanism should be provided by search engine so that users ensures for its privacy. For minimizing this privacy risk, different techniques will be discussed in this paper.*

Keywords: Web search engine, Personalization web search, user profile, privacy risk.

1. Introduction

Searching is one of the standard technique to find out the information from the internet with the guidance of the Web search engine (WSE). It is necessary for every retailer who is providing services to provide coherent and easy to use of information. WSE is one of the software that seeks for and identifies the content from the web engine or web database with matching keywords or character that is specified by the user and finds particular sites on the World Wide Web. There are many search engines like Google, Yahoo, who provides a significant and irrelevant data to the user based on their search. To avoid the irrelevant data, the method called Personalized Web Search (PWS) emerged. The Understanding of user intention behind query is essential in improving search engine relevance and personalized search. The user profiles, which means on click-through the log and the feedback session. By using the above methods, users' privacy might be a loss because of clicking the powerful search, commonly visited sites and providing their personal data like their name, DOB, etc. in this case their privacy might be the leak. Web personalization is the method of customizing a Web site to the need for each user or set of users, which takes advantage of the knowledge gained through the analysis of the user's navigational behavior.

Besides the personalized results, security is required in the personalized web search. Users are not interested in exposing their information during a web search. Becoming a significant concern in profiling the user in personalized web search. There should be a technique that recognizes profiles according to information given by the user. If the search engine knows more about the user, more concrete results will be gained by search provider. But users cannot believe in a search engine that information that is provided by the users not practiced illegally. Search engines can deliver accurate and exact data if users think search engine and provide more information. Hence, search engines should give security mechanism such that user will be assured of its privacy and its information should be kept secure. In personalized web search, user information is gathered and evaluated to discover the intention of issuing query fired by the user. The search is completed by providing queries to retrieval system in the form of a set of words. If remote

users fire the same query, the system will yield the same results without considering the user. But search results should be produced by considering the user in the equation so that specific users can get various search results for the same query by keeping the record of user's personal information and interests.

2. Related Work

Personalized web search (PWS) vary from existing web search, which returns same research results to all users for same queries, without considering user interests and information needs. PWS classifies into two types; one is click-log-based methods and other profile-based ones [1,2]. The click log-based methods rely on selecting the clicked pages in the user's past query history. The main shortcoming of this approach is which works on the regular set of queries by the same users only. Profile-based methods are more efficient in getting a better quality of web search with increasing use of personal and navigational information to profile its users, which is usually collected inherently from query history, surfing history, click-through data, bookmarks, etc. [1,2]. The main disadvantage of this method is that it needs the personal user data sends to the server, so this privacy problem makes the user discomfort. Personalized web search implements on both server side (in the search engine) or client side (in the user's computer or a personalization agent) [10]. For server-side personalization, user profiles are built, updated, and stored on the search engine side. For client-side personalization, user information is received and stored on the client side (on the user's computer or a personalization agent), normally by installing a client software or plug-in on a user's computer. Even though, having various advantages of personalized search, there is no large-scale use of personalized search services currently. Personalized web search has several difficulties that block its massive real world applications:

Privacy is a problem.

Users are dynamic.

Queries should not be divided in the same manner about personalization

In [2], the prototype of UPS is proposed, together with a greedy algorithm GreedyDP to help online profiling based on auspicious metrics of personalization convenience and privacy risk.

Krause and Horvitz in [4] employ analytical procedures to determine a probabilistic model, and then use this model to produce the near-optimal incomplete profile. They had preceded and investigate a distribution of secrecy in personalization, wherever people can opt to share personal data, in a standing or on-demand manner, in return for expected enhancements in the quality of an online service.

Xu et al. in [7] proposed a privacy protection solution for PWS based on hierarchical profiles. Using a user specified threshold a generalized profile achieves in effect as a rooted subtree of the complete profile. These profiles summarize a user's interest in a hierarchical organization according to particular interests. Two parameters for specifying privacy requirements are proposed to help the user to choose the content and the degree of detail of the profile data that disclosed to the search engine. Drawback: The main disadvantage of this strategy was that this work does not address the query utility, which is essential for the service quality of PWS.

In this paper [9] the author studied the existing generalization methods are insufficient because they cannot assure privacy protection in all cases, and frequently acquire redundant information loss by performing too much generalization. In this paper, the author suggests the idea of personalized secrecy and develops a new generalized formation that takes into account customized privacy necessities.

Both [3] and [5] contribute online anonymity on user profiles by generating a group profile of k users. Using this method, the linkage between the query and a single user is broken.

In [6], the useless user profile (UUP) protocol is aimed to rearrange queries among a group of users who issue them. As a result, any entity cannot profile a certain individual. These works imply the existence of a trustworthy third-party anonymizer, which is not readily available over the Internet at large.

3. Literature Survey

By above survey, we compare some mechanisms used for the privacy protection in personalization web search.

Table 1: Comparison of privacy protection methods

Title	Author	Conference/ Journal	Mechanism	Limitation
1] Supporting Privacy Protection in Personalized Web Search	L.Shou, H.Bai, K.chen, G.Chen	IEEE-2014	Allows for customization of secrecy demands. Online decision to whether personalize or not.	User's identity to server is exposed. Reduction in personalization effectiveness.
2] Efficient Privacy Protection in Personalized Web Search	G. Chen, H. Bai, L. Shou, K. Chen, & Y. Gao	ACM-2011	Greedy Utility algorithm for generalizing user queries in PWS. Identification requires for server.	It requires more computational cost & recomputation of all queries.
3] Anonymizing User Profiles for Personalized Web Search	Y.Zhu, L.Xiong, C.Verdery	WWW Conference-2010	No identity, privacy protection using Bayes-optimal privacy notion. Online anonymity on user profiles by generating a anonymous id using anonymous proxy server.	Personalized Search results are not useful. It is not portable for all users.
4] A Utility-Theoretic Approach to Privacy in Online Services	A. Krause E. Horvitz	AI Research – 2010	Probabilistic model used for predefined frequent queries. Suggested to compromise privacy, if there is better search outcomes.	For distinct queries, it is not applicable. Privacy can't be negotiated, counting both metrics is better.
5] Online Anonymity for Personalized Web Services	Y. Xu, K. Wang, G. Yang & A.W.-C. Fu	ACM-2009	Anonymizer algorithm of the user profiles by producing foreign identification. No personal information.	lag in real-time response due to higher communication cost. Higher Encryption/Decryption cost.

4. Conclusion

This paper covers issues like need of personalized web search, how personalized web search can be implemented,

what are challenges in it and security issue of it. This paper gives a survey report of different ways to maintain privacy in a personalized web context. It also tells about the drawbacks of the existing privacy issues. The future scope of

this paper will be to overcome the existing system drawbacks and design a framework to maintain a complete privacy of the users so that they can work without any fear of working in personalized web environment.

References

- [1] Lidan Shou, He Bai, Ke Chen, and Gang Chen, "Supporting Privacy Protection In Personalized Web Search", IEEE transactions on knowledge and data engineering vol:26 no:2 year 2014.
- [2] 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.
- [3] 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.
- [4] A. Krause and E. Horvitz, "A Utility-Theoretic Approach to Privacy in Online Services", Journal of Artificial Intelligence Research 39 (2010) 633-662.
- [5] Xu, Yabo, et al. "Online anonymity for personalized web services." *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2009.
- [6] J. Castelli'-Roca, A. Viejo, and J. Herrera-Joancomarti', "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.
- [7] 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.
- [8] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.
- [9] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc.ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.
- [10] "Personalized Web Search", W. M. P. VAN DER AALST Eindhoven University of Technology, Eindhoven.