

Privacy Protection in Personalized Web Search

Guided By SharmilaChopade

Uday D. More¹, Kavita T², Nanekar, Sana³, M. Sothe⁴
^{1,2,3,4}(Comp Dept, DYPIET,, Talegaon, Pune University, India))

Abstract- Personalized Web Search (PWS) has provided its effectiveness in improving the quality of various search services on the Internet. Personalized search is a promising way to improve the accuracy of web search, and has been attracting much attention now days. But effective, personalized search requires aggregating and collecting user information, which cause privacy infringement for many users; these infringements have become one of the main obstacles to deploying personalized search applications, and great challenge of how to do privacy preserving personalization. Study privacy protection in PWS applications that model user preference as hierarchical user profiles. Propose a PWS framework called UPS (User customizable Privacy-preserving Search) that can adaptively generalize profiles by queries while respecting user specified privacy requirements. Our runtime generalization has aims of keeping a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the user generalized profile.

Keyword-privacy protection, personalized web search, utility, risk, profile.

I. INTRODUCTION

The web search engine is an important portal for ordinary people who are looking for useful information on the web. Generally, users facing failure and get improper results when search engines return irrelevant results that do not meet their real intentions. A typical search engine provides a similar set of results without considering of who submitted the query. So, the requirement arises to have personalized web search system, which gives

Appropriate output to the user. Personalized Web Search (PWS) is a general category of search techniques which aims to provide better search results, according to users need. So, for this user information has to be collected and analyzed so that the perfect search results required for the user behind the issued query is given to the user. The solution for this is Personalized Web Search (PWS). It can generally be categorized into two types, first is click-log-based methods and the second is a profile-based ones. The click logs based methods are simple and straight forward. This method performs the search based upon clicking pages in the users query history. Also this method has been demonstrated to perform consistently and considerably good, it only works on repeated queries from the same user, which is a main limitation and restriction to certain applications. On other hand profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be proved more effective for almost all sorts of queries, but are reported to be improper under some situations. Although there are reasons and considerations for both types of PWS techniques, the profile-based personalized web search (PWS) has proved its more effectiveness in improving the quality of web search with the increasing usage of one personal and behavioral information to profile its users, which is generally gathered implicitly with the help of query history browsing history, click-through data, bookmarks, user documents and so

on. Such type of collected personal data can easily reveal an entire scope of user's private life. Protecting privacy issues arising from the lack of protection for such data, for example the AOL query logs scandal, not only raise panic among individual users, but also downs the data publisher's enthusiasm in offering personalized service. In fact, privacy protections have become the major barrier for a wide use of PWS services.

II. MOTIVATION

Researchers have to consider two main contradicting effects during the search process, for protecting user's privacy in profile-based PWS. On the one hand, they try to improve the quality of search with the personalization utility of the user profile. On the other hand, to place the privacy risk under control, they need to hide the privacy contents existing in the user profile. A few previous studies suggest that people are wishing to compromise privacy if the personalization is done by supplying user profile to the search engine yields better search quality. In an ideal, significant growth can be obtained by personalization at the expense of only a small, less-sensitive portion of the user profile, namely a generalized profile. Consider the example 1) Different users may use exactly the same query (e.g., "Washington") to search for different information (e.g., the Washington DC city in America or the George Washington first president of America), but existing search engines return the same results for these users. (2) Information needs of user's may change over time. The same user may use "Washington" as a Washington DC city in America and sometimes as the president of America. Existing search engines are unable to distinguish such cases. So it is clear that without knowing more user information and the search interest of a user it is impossible for a search engine to know which sense "Washington" refers to in a query. So in order to get results, must use more user information and personalize search results according to each individual user. Again, consider the query "Washington" to see how personalized search may help improve search accuracy. The intended meaning of "Washington " can often be easily determined by exploiting some naturally available information about a user. Any of the following additional information about the user could help determine the intended meaning of "Washington" in the query: (1) the user is a history student as opposed to a travel agent. (2) Before entering this query, the user had just bookmarked or viewed a web page with many words related to the Washington. Exploiting such user information to optimize the ranking of search results for a particular user is very appealing because it does not require any extra effort from the user. In general, personalized search is considered as one of the most promising techniques to break the limitation of current search engines and improve the quality of search results. Thus, without compromising the personalized search quality user privacy can be protected. In general, there is a tradeoff between the level of privacy protection and search quality which is achieved by generalization. Unfortunately, the previous works of privacy preserving PWS are far from optimal. Following observations shows the problems with the existing methods:

1. The existing profile-based PWS does not support runtime profiling.
2. The existing methods do not take into account the customization of privacy requirements.
3. When creating personalized search results, many personalization techniques require iterative user interaction.

III. LITERATURE SURVEY REVIEW

As per Z. Dou, R. Song, and J.-R. Wen, A large scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies(including two click log based and three profile based)using 12-days MSR query log.1.Search accuracy is evaluated by real user clicks recorded in query logs automatically. Personalization may lack effectiveness on some query.

X. Shen, B. Tan, and C. Zhai, present a decision theoretic framework and develop techniques for implicit user modeling in information retrieval. They developed an intelligent client-side web search agent (UCAIR) that can perform eager implicit feedback.Searchagent can improve search accuracy over the popular Google search engine. They generally lack user modeling and are not adaptive to individual users.

K. Sugiyama, K. Hatano, and M. Yoshikawa, Propose several approaches to adapting search results according to each user's need for relevant information without any user effort, and then verify the effectiveness of our proposed approaches.

As per B. Tan, X. Shen, and C. Zhai, they proposed Statistical language modeling based methods to mine contextual information from long term search history. Exploit it for a more accurate estimate of query language model. The web search engines, suffer from the problem of documents to return is "one size fits all" the decision of which documents to return is based on query, without consideration of a particular user's preferences and search context.

A. Krause and E. Horvitz, User profile, descriptions of user interests, can be used by search engines to provide personalized search results. Many approaches to creating user profile collect user information through proxy servers (to capture browsing histories)or desktop bots (to capture activities on a personal computer). Both these techniques require participation of the user to install the proxy server or the both.

IV. EXISTING SYSTEM

The existing profile-based Personalized Web Search does not support runtime profiling. 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. One evidence reported in is that profile-based personalization may not even help to improve the search q utility for some ado queries, though exposing user profile to a server has put the user's privacy at risk. The existing methods do not take into account the customization of privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. For example, in, all the sensitive topics are detected using an absolute metric called surprise based on the information theory, assuming that the interests witless user document support are more sensitive. However, this assumption can be doubted with a simple counterexample: If a user has a large number of documents about "sex," the surprise of this topic may lead to a conclusion that "sex" is very general and not sensitive, despite the truth which is opposite. Unfortunately, few prior work can effectively address individual privacy needs during the generalization. Many personalization techniques require iterative user interactions when creating personalized search results. 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, need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

1. User fires a query 'q' online to the server.
2. Query is stored on server and it generates user's profile 'g', which is resided at the server side. Client does not have control over that profile.
3. Server gives online response 'r' to the client according to query.

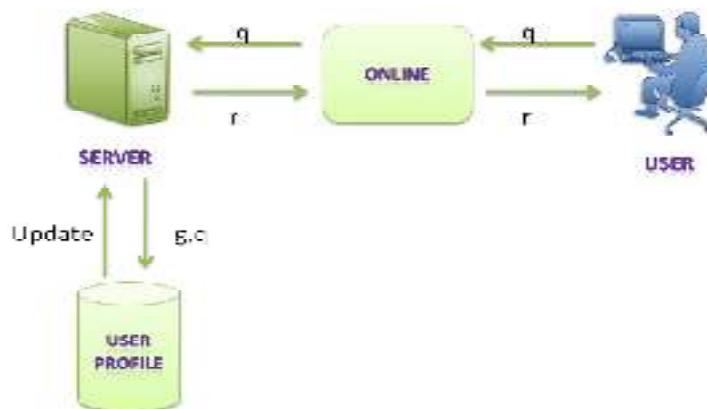


Fig 1: Existing system

• **Disadvantages :**

1. Users might experience failure when search engines return irrelevant results that do not meet their real intentions.
2. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts.
3. The existing profile-based PWS do not support runtime Profiling.
4. The existing methods do not take into account the customization of privacy requirements.
5. Many personalization techniques require iterative user interactions when creating personalized search results.
6. All the sensitive topics are detected using an absolute metric called surprisebased on the information theory.

V. PROPOSED SYSTEM

framework assumes that the queries do not contain any sensitive information, and aims at protecting the privacy in individual user profiles while retaining their usefulness. UPS Framework which generalized profiles for each query according to user specified privacy requirements. The problems of privacy preserving personalized search as Risk Profile Generalization, with its NP-Hardness proved. A Tradeoff between search quality and level of privacy protection achieved from generalization. Generalization algorithms are used namely Greedy and Greedily to find out an utilization of user search and improving performance.

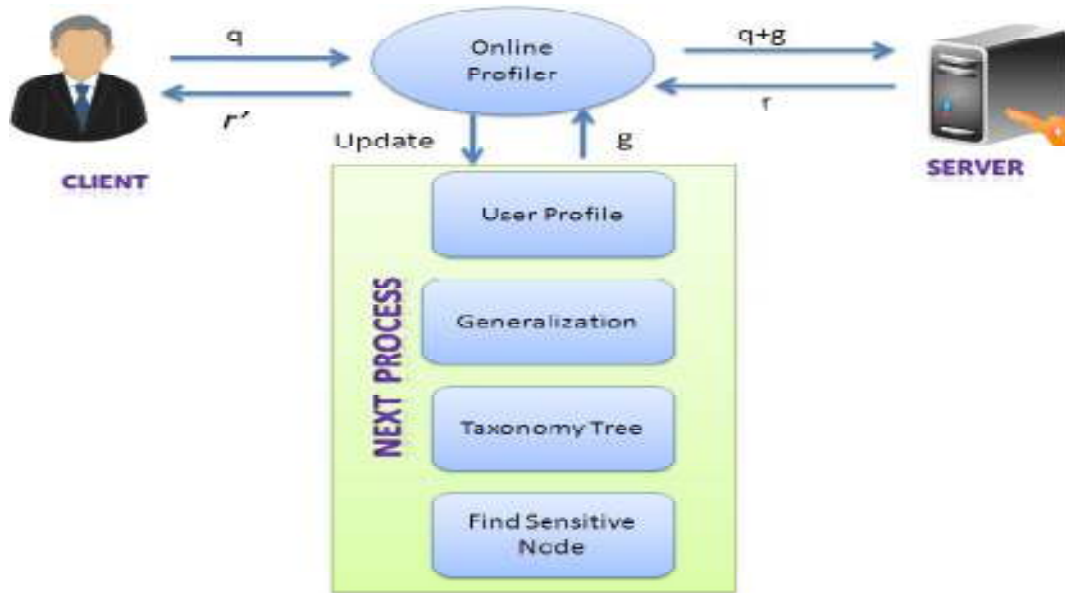


Fig2: Proposed system

1. User fire a query 'q' through a proxy refer as an on-line profiler to the server.
2. Then generalized profile is created by a proxy and both generalized profile and query are passed to the server.
3. Server gives response 'r' back to the proxy, then it decides either to re-ranked the search or provide as it is result to client as per the query.

• **Advantages :**

1. It enhances the stability of the search quality.
2. It avoids the unnecessary exposure of the user profile.
3. Provides runtime profiling, which in effect optimizes the personalization utility while respecting user's privacy requirements.

4. It allows for customization of privacy needs.
5. Does not require iterative user interaction.

VI. ATTACK MODEL

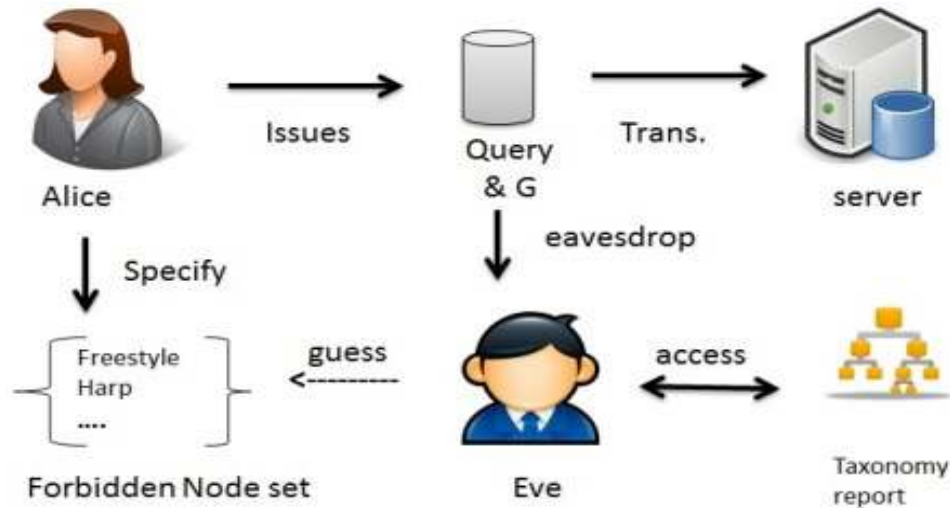


Fig 3. Attack model of personalized web search

Main aims at providing protection against a typical model of privacy attack, namely eavesdropping. As to corrupt Alice's privacy, the eavesdropper Eve successfully intercepts the communication between Alice and the PWS-server via some measures, such as man-in-the middle attack, invading the server, and so on. Consequently, whenever Alice issues a query q , the entire copy of together with a runtime profile G will be captured by Eve. Based on G , Eve will attempt to touch the sensitive nodes of Alice by recovering the segments. Web search. Hidden from the original and computing a confidence for each recovered topic, relying on the background knowledge in the publicly available taxonomy repository R .

VII. ACKNOWLEDGEMENT

The authors would like thanks to the publishers, researchers for making their resources available and teachers for their guidance. We also thank the college authority for providing the required infrastructure support. Finally, we would like to extend heartfelt gratitude to friends, family members.

VIII. FUTURE WORK AND CONCLUSION

This paper presented a client-side privacy protection framework called 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. For future work, we will try to resist adversaries with broader background knowledge, such as richer relationships among topics (e.g., exclusiveness, sequentiality, and so on), or capability to capture a series of queries from the Victim. We will also seek more sophisticated methods to build the user profile, and better metrics predict the performance (especially the utility) of the UPS.

IX. REFERENCES

- [1] 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, 2011.
- [2] A. Viejo and J. Castella-Roca, \Using Social Networks to Distort Users' Profile Generated by Web Search Engines "Computervices, "J. Arti_cial Intelligence Research, vol. 39, 2010.
- [3] 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), 2009.
- [4] J. Castell-Roca, A. Viejo, and J. Herrera-Joancomart_\Preserving User's Privacy in Web Search Engines, "Computer Comm., vol. 32, 2009.
- [5] J. Teevan, S.T. Dumais, and D.J. Liebling, \To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent, "Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2008.
- [6] 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.
- [7] 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.
- [8] X. Shen, B. Tan, and C. Zhai \Implicit User Modeling for Personalized Search,"Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM),2005.