

A Privacy Protection in Personalized Web Search for Knowledge Mining: A Survey

Miss. P.T. Sawant, Prof. Amrit Priyadarshi, Miss Monali D Jagtap
ME II Computer, Asst. Professor and PG Coordinator, ME II Computer
^{1,2,3}DGOI, FOE, Daund, Pune University (MH), India
poojasawant102@gmail.com, amritpriyadarshi@gmail.com, monali2512@gmail.com

Abstract –The web search engines (e.g. Google, Yahoo etc.) help the users to find required useful information on the World Wide Web (WWW). But it has become increasingly difficult to get the expected results from the web search engine because contents are available in web is very vast and ambiguous. Due to tremendous data opportunities in the internet, the privacy protection is very essential to preserve user search behaviors and their profiles. In this paper system present a novel protocol specially designed to protect the users' privacy in front of web search profiling. Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. System proposed two greedy algorithms namely GreedyDP and GreedyIL. These two algorithms are used for runtime generalization. The proposed protocol preserves the privacy of the individuals who deal with a web search engine. System provides a distorted user profile to the web search engine. It offers implementation details and computational and communication results that show that the proposed protocol improves the existing solutions in terms of query delay.

Keywords- Privacy Protection, personalized web search, Search Engine, user profile.

I. INTRODUCTION

The web search engine (WSE) is very important in web life. The web search engine has gained a lot of popularity and importance for users seeking information on the web. Since the contents available in web is very vast and dubiety, users at times experience failure when an irrelevant result of user query is returned from the search engine. Therefore, a user provides better search result a general category of search technique is Personalized Web search Personalization means the search engine can help users to filter the useful information for them by using user's interest. Search engine will pick the users' interest at the top of results, so it is very convenient for users to pick useful information. The solutions to Personalized Web Search (PWS) can generally be categorized into two types, 1) click-log-based methods and 2) profile based method. The click-log based methods are mild and straightforward. The click based method performs the search based on the clicked pages in the user's query history. Supposing this method has been show to perform consistently and considerably well, click based

1.1 Need of Personalization System

Personalization also known as customization, web search engines shows the results which are general and not adaptable or specifically individual users need. For a particular query fired to the search engine, different results are provided for different users. Search results are organized for every user considering one's interest, preferences and information needs. The need for personalization arises due

method only work on repeated queries from the same user, which is a strong inhibition and restricted for certain applications. In contrast, profile-based methods promote the search experience with complicated user interest models generated from user profiling techniques. Profile-based method can be demonstrated more effective for almost every sort of the queries, but are reported to be inopportune under some situations. In fact, privacy concerns have become the major barrier for wide use of PWS services. In this paper we will study how to provide privacy for the personalized web search applications that model the user preferences as hierarchal user profiles. We provide a framework called UPS (User Customizable Privacy-Preserving Search) which provides the search results by adapting to the user's information needs and also provides privacy according to the user specified privacy requirements which help the user to choose content and degree of detail of the profile information that is exposed to the search engine. An online prediction mechanism is provided for deciding whether to personalize the query and what to expose in the user profile at runtime.

to the two reasons because the different users have different interest and different background. For the same query, they have different information needs and goals. Secondly, User information needs may change over time. Users may have variety of requirements based on the time and circumstances. For example, on a web search engine user may use query "apple" to find information about apple computer peripheral when he/she wants to buy apple computer mouse and a computer user may submit same

query to find the information about the apple computer peripheral but the search engine also show the result about the fruit apple. Search engines can not to differentiate between such cases.

1.2 Personalization System Approach

When applied to search, personalization would involve the following steps: 1) to collect and represent information about the user in order to understand the user's interests. 2) Use this information to either filter the results returned from the initial retrieval process, or precisely include this information into the search process itself to select personalized results. In this approach web search personalization systems use gathered information about user from profiles, cookies and to conduct and revise the search to maximize the user satisfaction. The user profiles are created which specifies the user's interests, preferences and information needs to better personalize the search results. There are two ways to generate user profiles- explicit and implicit user profiling. In the explicit approach users create their profiles manually by providing some kind of feedback to a search system. In implicit user profiling, the user profile is created from users past behavior, such as by determining the documents they chose for viewing, the duration of time spent viewing a document or page browsing or scrolling actions. This is being done in the background automatically by the search system. Personalization of web search can be done at either server side or client side. Many problems arise on personalizing the web at server side like server should

II. LITERATURE REVIEW

This paper focuses on the literature of profile-based personalized and privacy protection in PWS system.

A. Privacy enhancing personalized web search

[4] In this paper the personalized web search is a promising way to improve search quality by customizing search results for people with individual information goals. However, users are very difficult with exposing private preference information to search engines. The privacy is not absolute, and often can be compromised if there is a gain in service or profitability to the user. Thus, the balance must be struck between search quality and privacy protection. This paper presents an extensible way for users to automatically build rich user profiles. These profiles recap a user's interests into a hierarchical organization according to specific interests. Two parameters for specifying privacy need are proposed to help the user to choose the content and degree of detail of the profile information that is exposed to the search engine. The privacy enhancing personalized experiments showed that the user profile improved search quality when compared to standard MSN rankings. Most importantly, the results

maintain all the search history for each and every user. It also has to search the history of a particular user when a user submits any ambiguous query. The performance of the server gets down when many users submits the query at the same time. Therefore, most of the techniques employ client side approach as all the search histories and queries are maintained at the client system making the faster way to access the user profile.

1.3 Background of personalized Search

There are generally two categories of personalized web search first is click based and second one is profile based.

1) Click Log Based Method –

Click-log method performs the search based upon clicked pages in the user's query history. This is simple and straight forward. The click based method can only work on repeated queries from the same user. This is the strong limitation and restricted for certain applications, although this method has been demonstrated to perform consistently and considerably well.

2) Profile Based Method –

Profile-based methods improve the search experience by generating user profiling techniques with complicated user-interest models. Profile-based methods provide more effective for almost all sorts of queries, but it is to be improper under some situations. There are reasons and consideration for both types of personalized web search techniques.

verified our hypothesis that a significant improvement on search quality can be achieved by only sharing some higher-level user profile information, which is probably less sensitive than detailed personal information.

B. To personalize or not to personalize: Modelling Queries with Variation User Internet

[12] In this paper author examine there is a lot of variation across queries in the benefits that can be achieved through personalization. For some queries, each one who issues the query is looking for the same thing. For other queries, different people need very different results even though they express their need in the same way. They consider the variability in user intent using both explicit relevance judgments and large-scale log analysis of user behavior patterns. Although variation in user behavior is correlated with variation in explicit relevance judgments the same query, there are many another factors, such as result entropy, result quality, and task that can also affect the variation in behavior. They characterize queries using various features of the query, the results returned for the query, and the people's interaction history with the query.

Using these features author build predictive models to identify queries that can benefit from personalization.

C. Anonymizing User Profiles for Personalized Web Search

[10] In this paper authors proposes a novel bundling technique that clusters user profiles into groups by taking into account the semantic relationships between the terms while satisfying the privacy constraint. Problem of grouping user profiles (represented as a weighted term list) are studied, so that user privacy is adequately protected while the grouped profiles are still effective in enabling personalized web search. Anonymization goal is to avert linking attacks that associate a user with an individual term in the anonymized user profile set. They proposes a Bayes-optimal privacy notion to bind the prior and posterior probability of associating a user with an individual term in the anonymized user profile set. In this paper author evaluate the approach through a set of preliminary experiments using real data demonstrating its feasibility and effectiveness.

D. Personalized Search via Automated Analysis of Interested and Activities

[1] In this paper author systematically examined the issue of privacy preservation in personalized search. The four levels of privacy protection are distinguished, and analysed various software architectures for personalized search. This work showed that the client-side personalization has beneficial over the existing server-side personalized search services in preserving privacy, and envision probable future strategies to fully protect user privacy. They formulate and study search algorithms that consider a user's prior interactions with a wide variety of content to personalize that user's current Web search. Rather than relying on the improbable assumption that people will precisely specify their intent when searching, they pursue techniques that influence implicit information about the user's interests. This information are used for the re-rank Web search results within a relevance feedback framework. In this paper system explore rich models of user interests, built from both search related information, such as previously issued queries and already visited web pages, and another information about the user such as the documents and email the user has read and created.

E. Preserving User Privacy in Web Search Engine

[7] In this paper author propose a new protocol designed to protect the privacy of the users from a web search engine that tries to profile them. In the preserving user privacy system uses social networks to provide a distorted user

profile to the web search engine. To address this privacy threat, current solutions have new mechanisms that introduce a high cost in terms of computation and communication. In this paper author present a new protocol specially designed to protect the users' privacy in front of web search profiling. They provide a distorted user profile to the web search engine. They offer implementation details and computational and communication results that show that the proposed protocol improves the existing solutions in terms of query delay.

F. Personalized Privacy Preservation

[3] A new generalization framework based on the concept of personalized anonymity. This framework performs the minimum generalization for satisfying everybody's requirements, and thus, retains the biggest amount of information from the microdata. They carry out a carefully theoretical study that leads to valuable insight into the behavior of alternative solutions. In particular, system analysis mathematically reveals the circumstances where the last work fails to protect privacy, and establishes the perfection of the proposed solutions.

G. A Large-Scale Evaluation and Analysis of Personalized Search Strategies

[6] In this paper author present a large-scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies (including two click-based and three profile-based ones) using 12-day MSN query logs. Here the consequences are examined and it is exposed that personalized search has important development over general web search on a number of query, but it also has tiny outcome on some additional question. In addition, it also demonstrates that uncomplicated click-based personalization approach performs constantly and significantly well, even as profile based ones is unbalanced in this research. Also discloses that both long-term and short-term contexts are very significant in humanizing search performance for profile-based modified search strategy. The profile-based personalized search strategies proposed in this paper are not as stable as the click-based ones. They could improve the search quality on some queries, but they also harm many queries. They also find for profile-based methods, both long-term and short-term contexts are important in improving search performance. The appropriate combination of them can be more reliable than solely using either of them.

H. Online Anonymity for Personalized Web Services

[8] In this paper system receives personalized web services; the user has to provide personal information and preferences, in addition to the query itself, to the web service. However, accurate personal information could identify the sender of sensitive queries, thus compromise user privacy. In this paper system proposes the notion of online anonymity to enable users to issue personalized queries to an untrusted web service while with their anonymity preserved. The challenge for providing online anonymity is dealing with unknown and dynamic web users who can get online and offline at any time.

I. Privacy Protection in Personalized Search

[5] In privacy protection, analytically observe the concern of privacy preservation in personalized search. Here discriminate and describe four levels of privacy protection, and analyze numerous software architectures for personalized search. It shows that client-side personalization

has advantages over the existing server-side personalized search services in preserving privacy in this situation; personalized web search cannot be done at the individual user level, but is possible at the group level. This may reduce the efficiency of personalization because a group's information need explanation is used to model an individual user's information need. In this paper, author systematically examines the issue of privacy preservation in personalized search. They differ and define the four levels of privacy protection, and analyze different software architectures for personalized search. But some of the drawbacks which occurs by this techniques are-

1. It does not fully protect user privacy.
2. They were not discussed different levels of privacy protection provided by search engines depending on a user's preference for the tradeoff between the privacy concern and the improved search service quality.

III. COMPARISON STUDY

Sr. No.	Paper Name	Protocol Used	Feature	Problem
1	Preserving User Privacy in web Search Engines	Useless User Profile(UUP)	Protect user privacy in front of web search engines	Execution of query may be delayed, Collision occurs
2	Privacy Protection in Personalized Search	Intelligent Client Side Web Search Agent (UCAIR)	It improve the accuracy of web search	It does not fully protect user Privacy
3	A large scale evaluation & analysis of Personalized search strategies.	Click Based and Profile Based	Improve the search accuracy on same query	Only work on repeated query
4	Online Anonymity for personalized web services	Online Anonymity	Dealing with unknown & dynamic web user who can get online & offline any time	It include the third party means the user pool & it is not completely a trusted
5	Using social Network to distort user profiles generated by web search engine	Social Network	Preserves the privacy of the individuals who deal with a web search engine	High cost due to collaboration & communication
6	Ontology based personalized search & browsing	Filtering & Re-Ranking	Allow for the Fully automatic creation of large structured user Profile	Doesn't support integration of the system into web browser.
7	Anonymizing user profiles for personalized web search	Bayes-Optimal Privacy Notion	Reduced the time delay	Can't provide the privacy to location based query
8	Implicit User Modeling for personalized Search	Two click based and three profile-based ones	It is consistently effective on different queries for different users, and under different search Contexts.	It harms search accuracy under some situations.

IV. PROPOSED ENHANCEMENT

Web search engines are widely used to find data from huge amount of information in a minimal amount of time. However, these tools also pose a privacy threat to the users, web search engines profiles their users by storing and analyzing past searches submitted by them. UPS framework

seems to be more effective out of the methods discussed. For privacy protection, an online profiler is designed in this system, which can adaptively generalize profiles by queries while respecting user specified privacy requirements. The online profiler is at the client side where the complete user profile is stored along with the specified sensitive topics.

Runtime generalization aims at providing search efficiency along with privacy protection of user profiles.

V. CONCLUSION

The proposed approach personalized search is a promising way to improve search quality. Proposed system provide fast and relevant search that is personalized using User Profile. Based on User Profile the system keeps on updating user profile and an enhanced user profile is created. This paper

ACKNOWLEDGMENT

I express great many thanks to Prof. Amrit Priyadarshi and Department Staff for their great effort of supervising and leading me to accomplish this fine work. They were a great source of support and encouragement. To my friends and family, for their warm, kind encourages and loves. To every person who gave me something too light along my pathway. I thanks for believing in me.

REFERANCES

- [1] 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.
- [2] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," *Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI)*, 2005.
- [3] X. Xiao and Y. Tao, "Personalized Privacy Preservation," *Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD)*, 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," *ACM SIGIR Forum*, vol. 41, no. 1, pp.4-17, 2007.
- [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] J. Castelli-Roca and A. Vijeo and J. Herrera-Joancomarti, "Preserving user's privacy in web search engines," *Computer Comm.vol.32,Vol.32,no.13/14*, pp 1541-1551, 2009.
- [8] Y. Xu, K. Wang, G. Yang, and A.W.C Fu, "Online anonymity for personalized web services" *Proc.18th ACM conformation and knowledge management (CIKM)*, pp 1497-1500, 2009.
- [9] Vijeo and J. Castella-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," *Computer Networks*, vol. 54, no. 9, pp. 1343-1357, 2010.
- [10] 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.

presents the different approaches that have been implemented for personalizing web search. This paper mainly compares the various existing techniques in user search goals. The time complexity of proposed approach is going to reduce as compared with other protocol and by using this new protocol we are going to increase the efficiency of user search goal as well as to fulfill user information need by providing a secure personalization web search.

- [11] 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.
- [12] 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)*, pp. 163-170, 2008.

Authors



Miss Sawant Pooja T. received her B.E. degree in Computer Engineering from University of Pune in 2014. She is currently working toward the M.E. Degree in Computer Engineering from University of Pune. Her research interests lies in Data Mining, Text data mining and personalized information retrieval.



Prof. Amrit Priyadarshi received his M.Tech (CSE), PHD Persuing. Currently he is working as Assistant Professor and PG Co-ordinator in Department of Computer Engineering of DGOI, FOE, Swami-Chincholi, Daund, Pune.



Miss Jagtap Monali D. received her B.E. degree in Computer Engineering from University of Pune in 2014. She is currently working toward the M.E. Degree in Computer Engineering from University of Pune. Her research interests lies in Data Mining and Natural Language Processing.