ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

A Survey on Privacy Protection in Personalize Web Search Mechanism

Gunjan K Naigaonkar, Payal Dhavale, Tejaswi Bhosale, Neha Kumari, Neha Rani

Asst. Professor, Dept. of CS, G.H Raisoni College of Engineering and Management, Wagholi, Savitribai Phule Pune University, Pune, India

Student, Dept. of CS, G.H Raisoni College of Engineering and Management, Wagholi, Savitribai Phule Pune University, Pune, India

Student, Dept. of CS, G.H Raisoni College of Engineering and Management, Wagholi, Savitribai Phule Pune University, Pune, India

Student, Dept. of CS, G.H Raisoni College of Engineering and Management, Wagholi, Savitribai Phule Pune University, Pune, India

Student, Dept. of CS, G.H Raisoni College of Engineering and , Wagholi, Savitribai Phule Pune University, Pune, India

ABSTRACT: As the size of the Internet continues to grow the users of search providers continually demand search results that are accurate to their needs. Personalized Search is one of the options available to users in order to sculpt search results returned to them based on their personal data provided to the search provider. This raises concerns of privacy issues however as users are typically uncomfortable revealing personal information to an often faceless service provider on the Internet. This paper aims to deal with the privacy issues surrounding personalized search and discusses ways that privacy can be enriched so that users can become more comfortable with the release of their personal data in order to receive more accurate search results. Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified privacy requirements.

KEYWORDS: Privacy protection, personalized web search, utility, risk, profile

I. INTRODUCTION

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query. The solutions to PWS can generally be categorized into two types

1. Click-log-based methods and
2. Profile-based methods

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

Click-log-based methods

1. The click-log based methods are straightforward they simply impose bias to clicked pages in the user's query history.
2. It can only work on repeated queries from the same user, which is a strong limitation confining its applicability.
3. **Profile-based methods**

Profile-based methods can be potentially effective for almost all sorts of queries, but are reported to be unstable under some circumstances.

Improve the search experience with complicated user-interest models generated from user profiling techniques.

PWS has demonstrated more effectiveness in improving the quality of web search recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history, browsing history, click-through data bookmarks, user documents and so forth.

II. RELATED WORK

The existing profile-based Personalized Web Search does not support runtime profiling [1]. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly [3]. 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 quality for some ad hoc 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 surprisal based on the information theory, assuming that the interests with less user document support are more sensitive[1]. However, this assumption can be doubted with a simple counterexample: If a user has a large number of documents about "sex," the surprisal of this topic may lead to a conclusion that "sex" is very general and not sensitive, despite the truth which is opposite. Unfortunately, little prior work can effectively address individual privacy needs during the generalization.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

III. PROPOSED SYSTEM

To propose UPS (User customizable Privacy-preserving Search) framework, which is a privacy-preserving personalized web search framework, which can generalize profiles for each query according to user-specified privacy requirements?To develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. GreedyDP tries to maximize the discriminating power (DP), GreedyIL attempts to minimize the information loss (IL). The 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 for PWS.UPS consists of a nontrusty search engine server and a number of clients. Each client (user) accessing the search service trusts no one but himself herself. The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself. The proxy maintains both the complete user profile, in a hierarchy of nodes with semantics, and the user-specified (customized)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

privacy requirements represented as a set of sensitive-nodes.During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. The online phase handles queries as When a user issues a query qi on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile Gi satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.The query and the generalized user profile are sent together to the PWS server for personalized search.The search results are personalized with the profile and delivered back to the query proxy. Finally, the proxy either presents the raw results to the user, or reranks them with the complete user profile.

IV. PRIVACY PROTECTION MECHANISM IN PWS SYSTEM

Typical works in the literature of protecting user identifications (class one) try to solve the privacy problem on different levels, including the pseudoidentity, the group identity, no identity, and no personal information. Solution to the first level is proved to fragile. The third and fourth levels are impractical due to high cost in communication and cryptography. Therefore, the existing efforts focus on the second level. The useless user profile (UUP) protocol is proposed to shuffle queries among a group of users who issue them. As a result any entity cannot profile a certain individual. These works assume the existence of a trustworthy third-party Anonymized, which is not readily available over the Internet at large.

Viejo and Castell-a-Roca use legacy social networks instead of the third party to provide a distorted user profile to the web search engine. In the scheme, every user acts as a search agency of his or her neighbors. They can decide to submit the query on behalf of who issued it, or forward it to other neighbors. The shortcomings of current solutions in class one is the high cost introduced due to the collaboration and communication. The solutions in class two do not require third-party assistance or collaborations between social network entries.

In these solutions, users only trust themselves and cannot tolerate the exposure of their complete profiles an anonymity server. Krause and Horvitz employ statistical techniques to learn a probabilistic model, and then use this model to generate the near-optimal partial profile. Limitation in this work is that it builds the user profile as a finite set of attributes, and the probabilistic model is trained through predefined frequent queries. These assumptions are impractical in the context of PWS.

V. FLOW OF PROJECT

It consists of two main modules which are further divided into sub modules as follows:

I] Creation of user profile considering user's positive and negative preferences:

1) Dealing with Clickthrough data from user (Concept Extraction).
2) Creation of Concept-Relationship Graph.
3) Creation of user Profile

II] Applying Clustering algorithm on created profile:

1) Apply Agglomerative clustering algorithm on created profile.
2) Use of precision and recall to measure the performance.

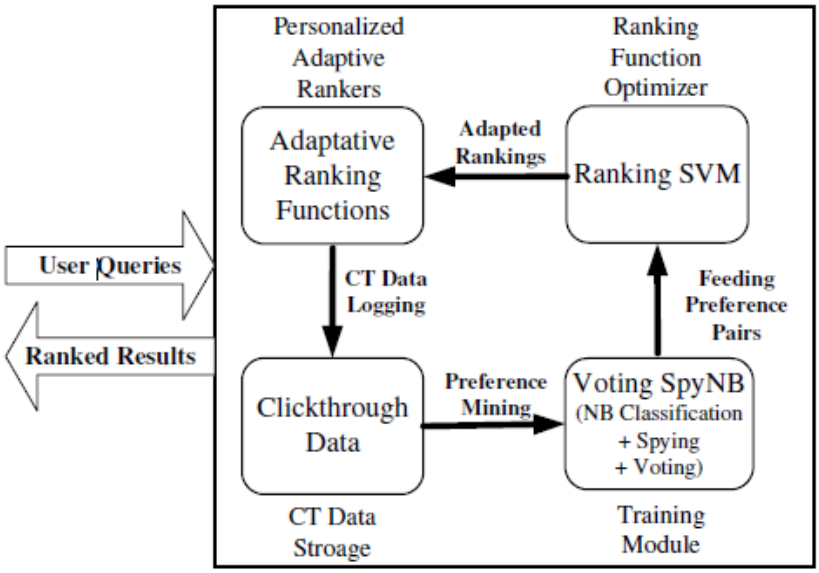
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

VI. SYSTEM ARCHITECTURES



VII.CONCLUSION

Here we conclude 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. In addition, UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality. We proposed two greedy algorithms, namely Greedy DP and Greedy IL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution.

REFERENCES

[1] 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.
[2] 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.
[3] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
[4] 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.
[5] 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.