PROTECTING PRIVACY IN PERSONALIZED WEB SEARCH

Kaveri Kadasiddayya Hiremath[1], Kapilesh S. Swami[2], Vishal V. Bhanawase[3]
[1]M.E, Computer Science and Engineering,
MGM'S College of Engineering, Nanded, Maharashtra, India,
[2,3]Research Scholar, K L University, Vijayawada, Andhra Pradesh, India,

Abstract

With the increasing size of the Internet the users of search providers regularly demand search results that are accurate to their needs. Personalized Web Search (PWS) is one of the options available to users in order to return the search results based on their personal data provided to the search provider. It has proved that the quality of various search services on the Internet is improved effectively. While, the evidences show that users' unwillingness to disclose their private information during search is a major challenge for search services. We propose a PWS framework that can adaptively generalize profiles by queries aiming at providing better search results, which are tailored for individual user needs while respecting user specified privacy requirements. The runtime generalization aims to strike a balance between the utility of personalization and the privacy risk of exposing the generalized profile. We present a method, called GreedyIL that prefers generalization. The experimental results also show that GreedyIL significantly outperforms non personalization in terms of efficiency.

Keywords: Personalized web search (PWS), profile, Privacy protection.

I. INTRODUCTION

Now a days, people who are looking for the useful information, may require the Web Search Engine for net services. Many web computing systems are running real time database services. Their information change continuously and expand incrementally. So here, web data services have a major role and they significantly improve in monitoring and controlling the data propagation and information truthfulness. So, the web search engine has long become the most important gateway for ordinary people. Sometimes Search engines may return irrelevant results, those do not match user's real intentions. Such irrelevance is largely due to the ambiguity of texts as well as the enormous variety of users' contexts and backgrounds. Personalized web search (PWS) is one of the general categories of search techniques that aims to provide better search results, which are related to the individual user needs. To figure out the user intention behind the issued query, we have to collect and analyze the user information [1]. For PWS, Profile based methods are effective in improving the quality of web search.

The size of the Internet continues to grow. So, the users of search providers regularly demand search results that are accurate to their needs. Personalized Search is one of the options available to users in order to return the search results based on their personal data provided to the search provider. So, the concerns of privacy issues are raised however as users are typically uncomfortable to disclose personal information to an often faceless service provider on the Internet. This paper aims for personalized search without compromising the user privacy and it also discusses ways that privacy can be enriched so that users can become more comfortable for releasing their personal data in order to receive more accurate search results [15].

There are two contradicting effects need to be considered during the search process to protect user privacy in profile-based PWS. On the one hand, the search quality with the personalization utility of the user profile has to be improved and on the other hand, hide the privacy contents existing in the user profile are to be hidden to place the privacy risk under control. Some of the previous studies suggest that people are willing to compromise privacy if the search engine yields better search quality with personalization by supplying user profile. This can be achieved by personalization with the help of only a small (and less sensitive) portion of the user profile, namely a generalized profile, in an ideal case. Thus, it protects user privacy without compromising the personalized search quality. In general, there is a balance between two desirable features that is the search quality and the level of privacy protection achieved from generalization [14].

The solutions of Personalized Web Search can be categorized into two types, one is click-log-based methods and other profile-based methods. The click log based methods are based on simply imposing the bias to clicked pages in the user's query history and profile based method improves search quality experiences with complicated user-interest models that are generated from user profiling techniques. In improving the quality of web search, the profile-based PWS is proved to be more effective [1]. Now-a-days to profile its users, search with increasing usage of personal information, which can be gathered implicitly from click-through data, query history, browsing history, bookmarks, user documents, and so on.

Our main objectives are summarized as following: We propose a user customizable privacy-preserving personalized web search framework UPS, which generalizes profiles for each query as per the user-specified privacy requirements that provides better search results as per the individual user needs. For hierarchical user profile, the problem of privacy-preserving personalized search is expressed systematically as Risk Profile Generalization. Our extensive experiments have proved the efficiency and effectiveness of our UPS framework.

II. RELATED WORK

Privacy protection problems for PWS have two classes. One class includes those treat privacy as the identification of an individual. The other one includes those consider the data, particularly the user profiles, exposed to the PWS server. The typical works in the literature of protecting user identifications try to solve the privacy problem on different levels, as per the group identity, no personal information, the pseudo identity, and no identity. Solution to the first level is proved too fragile. Due to high cost in communication and cryptography, other levels that is the third and fourth levels are impractical. Thus, the existing efforts focus on the second level. The useless user profile (UUP) protocol is proposed in [4] of shuffle queries among a group of users who issue them. So, a certain individual cannot be profiled by any entity. They assume the existence of a trustworthy third-party anonymizer ind it is not readily available over the Internet at large. To provide a distorted user profile to the web search engine, Viejo and Castell-a-Roca [9] use legacy social networks instead of the third party. In this context, every user acts as his or her neighbour's search agency. They can decide to submit the query on behalf of who issued it, or they can forward it to other neighbors. The failure or faults of current solutions in class one is the high cost introduced due to the collaboration and communication.

The class two solutions do not require third-party assistance or collaborations between social network entries. Here, users only trust themselves and cannot tolerate the exposure of their complete profiles to an anonymity server.

Krause and Horvitz [5] employ techniques to learn a probabilistic model, and then using this model to generate the partial profile which is near-optimal. Limitation for this work is that the user profile is built as a finite set of attributes, and this model is trained through predefined frequent queries and these assumptions are not practical in the context of PWS.

Xu et al. [7] proposed a privacy protection solution for PWS based on hierarchical profiles. In effect, a generalized profile is obtained as a rooted subtree of the complete profile, using a user-specified threshold and unfortunately, this work does not address the query utility, which is crucial for the service quality of PWS. Xiao and Tao [8] proposed Privacy-Preserving Data Publishing (PPDP). For his/her sensitive values, a person can specify the degree of privacy protection by specifying "guarding nodes" in the taxonomy of the sensitive attribute.

To classify queries with the help of their their click entropy, Teevan et al. [6] collect a set of features of the query. While these works are helpful in terms of whether to personalize or not to and they assume the availability of massive user query logs and user feedback.

F Liu, C Yu and W Meng et. al. [16] proposes technique to learn user profiles from users' search histories. There are two types of profiles.

A user profile and a general profile which are learned from the search history of user and a category hierarchy, respectively. By combining these two profiles, map a user query into a set of categories. They represent the user's search intention and they serve as a context to disambiguate the words in the user's query. There are many profile learning and category mapping algorithms and a fusion algorithm are used and evaluated. Experimental results show that the proposed technique to personalize Web search is both effective and efficient.

Personalized web search (PWS) has proved its effectiveness in improving the quality of various search services on the Internet. While, the evidences show that users' unwillingness to disclose their private information during search has become a major challenge for search services. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. Also, we propose a PWS framework called UPS that can adaptively generalize profiles by queries which aims at providing better search results, which are related to the individual user needs while respecting user specified privacy requirements. The generalization aims to strike a balance between the utility of personalization and the risk of exposing the generalized profile. We present a method, called GreedyIL that prefers generalization. The experimental results also show that GreedyIL significantly outperforms non personalization in terms of efficiency.

III. METHOD

To meet the aim of our project, we used GreedyIL Algorithm that goes through various steps, explained as follows:

A. UPS procedures

The above problems are addressed in our UPS that is User customizable Privacy-preserving Search framework and the framework aims at protecting the privacy in individual user profiles while retaining their usefulness for PWS and assumes that no any sensitive information is contained in queries. User profile is constructed and then customized with the user-specified privacy requirements. The system works under two phases- online phase and offline phase. Offline phase collects user information and query while the online phase works as follows:

1. In first stage, when a user issues a query qi on the client, the proxy generates a user profile. It's output is a generalized user profile Gi that satisfies the privacy requirements.
2. Both the query and the generalized user profile are sent together to the PWS server further for personalized search.
3. The search results are personalized with the profile. Later, they are delivered back to the query proxy.
4. Finally, the proxy either sends the raw results to the user, or it re ranks them with the complete user profile. Fig. 1 displays the architecture of PWS system.

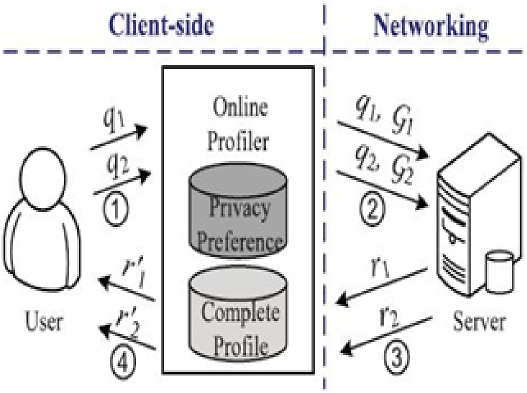

Fig. 1. Architecture of UPS

IV. User Profile Personalization

This paper introduces an approach to personalize digital multimedia content based on user profile information. So, two main mechanisms were developed: one, a profile generator which automatically creates user profiles that represent the user preferences, and second, a content-based recommendation algorithm which estimates the user's interest in unknown content by matching his profile to metadata descriptions of the content. Both of these features are integrated into a personalization system.

V. Generalization

User will register to the server, with his profile which is a representation of user interests. Subsequently, 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 re-ranks them with the complete user profile.

The generalization process has to meet specific prerequisites to handle the user profile. This is achieved by preprocessing the user profile. At first, the process initializes the user profile by taking the indicated parent user profile into account. The process adds the inherited properties to the properties of the local user profile. Thereafter the process loads the data for the foreground and the background of the map according to the described selection in the user profile.

VI. PRIVACY SEARCH

The profile-based personalization contributes little or even reduces the search quality, while exposure of the profile to a server would disturb the user's privacy. Here, privacy is maintained by user profile based web search, which maintain search status for user query response.

Any unauthorized user should be prevented from accessing the system. Password authentication can be introduced. To ensure the safety of the system, perform regular monitoring of the system so as to trace the proper working of the system.

VII. OUTCOME

Outcome of the proposed system is personalizing web search from the network for client query. It is search engine for personally.

VIII. ATTACK MODEL

There is a typical model of privacy attack, namely eavesdropping. The system aims at providing protection against it. To corrupt user's privacy, the eavesdropper Eve successfully obstruct the communication between user and the PWS-server via man-in-the middle attack. Consequently, whenever user issues a query q, Eve will capture the entire copy of q together with a runtime profile G. Based on G, Eve will attempt to touch the sensitive nodes or information of Alice by recovering the segments which are hidden from the original H and then computing a confidence for each recovered topic, with the help of background knowledge available in the taxonomy repository R which is public. This can be avoid by the customized privacy requirement and generalization.

IX. EXPERIMENTAL RESULTS

In this section, we empirically evaluate the proposed approaches. In first experiment, we study the scalability of GreedyIL algorithm in terms of response time. Second, we study the precision verses recall. Third, we study the computation performance. Fourth section reveals click through performance and fifth section includes concept relationship.

A. PWS Computation Time

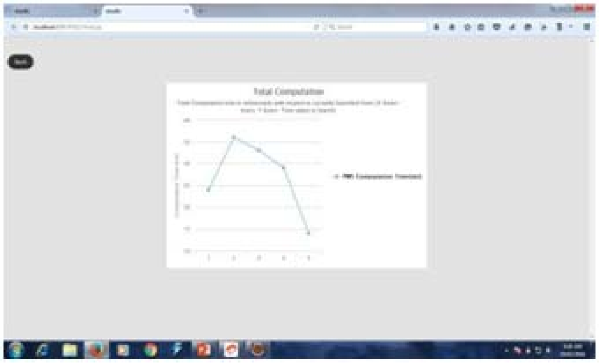

Fig. 2. PWS Computation Time

Fig. 1 illustrates the results of average response time for search results in miliseconds. Here, we study the scalability of GreedyIL algorithm in terms of response time.

B. Computation Performance

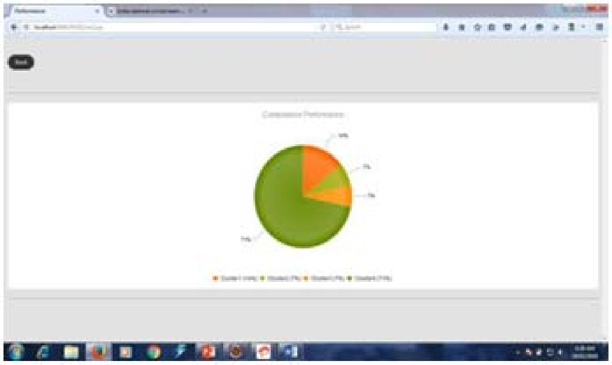

Fig. 3. Computation Performance

The concepts are divided into four clusters according to the decreasing click count. Fig. 3, shows the percentage of concepts retrieved online.

C. Seeds click through

The click count of links are displayed in fig. 4 for a user profile which reveals the user interests and detects whether user has reached the related information.

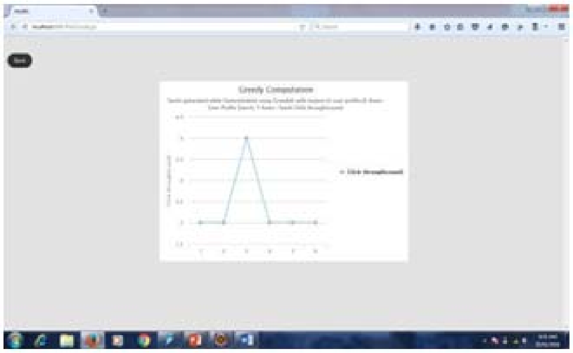

Fig. 4. Seeds click through

D. Concept Relationship

Fig, 5 shows that how many percentage of concepts is extracted for the particular percentage of queries.

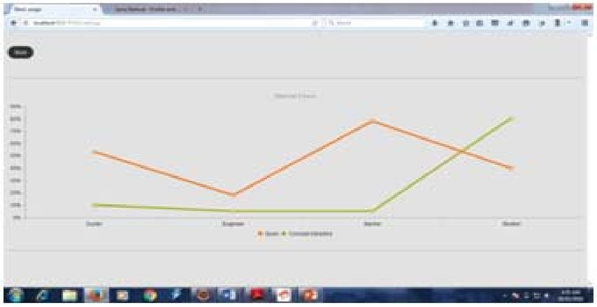

Fig. 5. Concept Relationship

X. CONCLUSION

Personalized web search (PWS) is one of the active research field which is related to the retrieval of the relevant web page results based on the user preferences and interest. This paper focuses mainly on the personalization process in various stages and there are various techniques and algorithms contained in each stage that have been discussed. The proposed survey will help the researchers for developing a better solution for Personalized Web search technique. A client-side privacy protection framework has been presented for PWS. UPS could potentially be adopted by PWS that captures user profiles in a hierarchical taxonomy. On user profiles, UPS can perform online generalization to protect the user privacy without compromising the search quality. An interesting direction for future work can be to use more advances extraction capabilities.

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, February 2014.

[2] Dmitri V. Kalashnikov, Zhaoqi (Stella) Chen, Sharad Mehrotra, “Web People Search Via Connection Analysis,” IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.11, February 2008.

[3] Natthakan Iam-On, Tossapon Boongoen, Simon Garrett, Chris Price, “A Link-Based Cluster Ensamble Approach for Categorical Data Clustering,” IEEE Transactions on Knowledge and Data Engineering, Vol.24, No.3, March 2012.

[4] 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.

[5] Krause and E. Horvitz, “A Utility-Theoretic Approach to Privacy in Online Services,” J. Artificial Intelligence Research, vol. 39, pp. 633-662, 2010.

[6] 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.

[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. Xiao and Y. Tao, “Personalized Privacy Preservation,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2006.