ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Privacy Protection Methodology in Personalized Web Search

Kenche Vamshi Krishna
M.Tech Student
Auroras Scientific Technological and Research Academy.

Archana Reddy
Guide
Auroras Scientific Technological and Research Academy.

Abstract:
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 ambiguous, users at times experience failure when an irrelevant result of user query is returned from the search engine. Web search engines (e.g. Google, Yahoo, Microsoft Live Search, etc.) are widely used to find certain data among a huge amount of information in a minimal amount of time. These useful tools also pose a privacy threat to the users. Web search engines profile their users on the basis of past searches submitted by them however, effective personalized search requires collecting and aggregating user information, which often raises serious concerns of privacy infringement for many users. Indeed, these concerns have become one of the main barriers for deploying personalized search applications, and how to do privacy-preserving personalization is a great challenge. This paper models preference of users as hierarchical user profiles. It proposes a framework called UPS which generalizes profile at the same time maintaining privacy requirement specified by user. Two greedy algorithms namely GreedyDP and GreedyIL are used for runtime generalization. Also, an online prediction mechanism to decide whether to personalize a query or not is provided in this paper.

Keywords: ***Privacy Protection, profile, personalized web search, risk, Search Engines***

Introduction:

Personalized search refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided. Pitkow et al. describe two general approaches to personalizing search results, one involving modifying the user's query and the other re-ranking search results.

Google introduced personalized search in 2004 and it was implemented in 2005 to Google search. Google has personalized search set up for not just those who have a Google account but everyone as well. There is not very much information on how exactly Google personalizes their searches, however, it is believed that they use user language, location, and web history.

Early search engines, like Yahoo! and AltaVista, found results based only on key words. Personalized search, as pioneered by Google, has become far more complex with the goal to "understand exactly what you mean and give you exactly what you want." Using mathematical algorithms, search engines are now able to return results based on the number of links to an from sites; the more links a site has, the higher it is placed on the page. Search engines have two degrees of expertise: the shallow expert and the deep expert. An expert from the shallowest degree serves as a witness who knows some specific information on a given event. A deep expert, on the other hand, has comprehensible knowledge that gives it the capacity to deliver unique information that is relevant to each individual inquirer. If a person knows what he or she wants than the search engine will act as a shallow expert and simply locate that information. But search engines are also capable of deep expertise in that they rank results indicating that those near the top are more relevant to a user's wants than those below.

While many search engines take advantage of information about people in general, or about specific groups of people, personalized search depends on a

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

user profile that is unique to the individual. Research systems that personalize search results model their users in different ways. Some rely on users explicitly specifying their interests or on demographic/cognitive characteristics. But user supplied information can be hard to collect and keep up to date. Others have built implicit user models based on content the user has read or their history of interaction with Web pages.

Benefits:

One of the most critical benefits personalized search has is to improve the quality of decisions consumers make. The internet has made the transaction cost of obtaining information significantly lower than ever. However, human's capability of processing information has not expanded much. When facing overwhelming amount of information, consumers need a sophisticated tool to help them make high quality decisions. Two studies examined the effects of personalized screening and ordering tools, and the results show positive correlation between personalized search and the quality of consumers' decisions.

The first study was conducted by Kristin Diehl from University of South Carolina. Her research discovered that reducing search cost led to lower quality choices. The reason behind this discovery was that 'consumers make worse choices because lower search costs cause them to consider inferior options.' It also showed that if consumers have a specific goal in mind, they would further their search, resulting in an even worse decision. The study by Gerald Haubl from University of Alberta and Benedict G.C. Dellaert from Maastricht University mainly focused on recommendation systems. Both studies concluded that a personalized search and recommendation system significantly improved consumers' decision quality and reduced the number of products inspected.

Personalized search gains popularity because of the demand for more relevant information. Research has indicated low success rates among major search engines in providing relevant results; in 52% of 20,000 queries, searchers did not find any relevant results within the documents that Google returned. Personalized search can improve search quality significantly and there are mainly two ways to achieve this goal.

The first model available is based on the users' historical searches and search locations. People are probably familiar with this model since they often find the results reflecting their current location and previous searches.

There is another way to personalize search results. In Bracha Shapira and Boaz Zabar's "Personalized Search: Integrating Collaboration and Social Networks", Shapira and Zabar focused on a model that utilizes a recommendation system.[36] This model shows results of other users who have searched for similar keywords. The authors examined keyword search, the recommendation system, and the recommendation system with social network working separately and compares the results in terms of search quality. The results show that a personalized search engine with the recommendation system produces better quality results than the standard search engine, and that the recommendation system with social network even improves more.

EXISTING SYSTEM:

The solutions to PWS can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based methods are straightforward— they simply impose bias to clicked pages in the user's query history. Although this strategy has been demonstrated to perform consistently and considerably well [1], it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. In contrast, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based methods can be poten-tially effective for almost all sorts of queries, but arereported to be unstable under some circumstances.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

DISADVANTAGES OF EXISTING SYSTEM:

- The existing profile-based PWS do not support runtime profiling.
- The existing methods do not take into account the customization of privacy requirements.
- Many personalization techniques require iterative user interactions when creating personalized search results.
- Generally there are two classes of privacy protection problems for PWS. One class includes those treat privacy as the identification of an individual, as described. The other includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server.

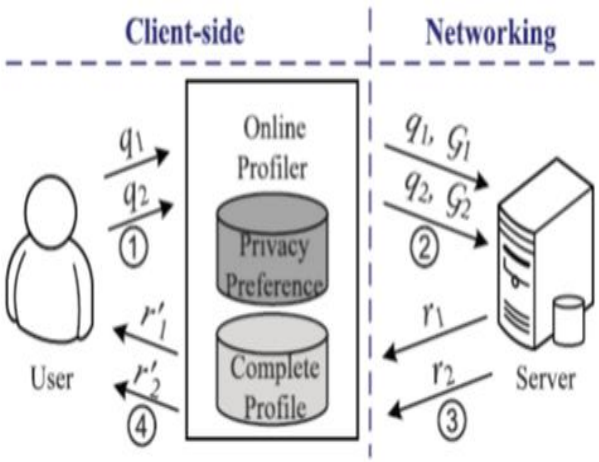
PROPOSED SYSTEM:

- We propose a privacy-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified privacy requirements.
- Relying on the definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as #-Risk Profile Generalization, with its N P-hardness proved.
- We develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). By exploiting a number of heuristics, GreedyIL out performs GreedyDP significantly.
- We provide an inexpensive mechanism for the client to decide whether to personalize a query in UPS. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile.
- Our extensive experiments demonstrate the efficiency and effectiveness of our UPS framework.

ADVANTAGES OF PROPOSED SYSTEM:

- ❖ Increasing usage of personal and behaviour information to profile its users, which is usually gathered implicitly from query history, browsing history, click-through data bookmarks, user documents, and so forth.
- ❖ 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.

SYSTEM ARCHITECTURE:



MODULES:

1. Profile-Based Personalization
2. Generalizing User Profile
3. Online Decision
4. Privacy Protection in PWS System

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Conclusion:

Disadvantage of search personalization is that internet companies such as Google are gathering and potentially selling your internet interests and histories to other companies. This raises a privacy issue. The issue is if people are content with companies gather and selling their internet information without their consent or knowledge. Many web users are unaware of the use of search personalization and even fewer have knowledge that user data is a valuable commodity for internet companies. A client side privacy protection framework called UPS i.e User customizable Privacy preserving Search is presented in the paper. Any PWS can adapt UPS for creating user profile in hierarchical taxonomy. UPS allows user to specify the privacy requirement and thus the personal information of user profile is kept private without compromising the search quality. UPS framework implements two greedy algorithms for this purpose, namely GreedyDP and GreedyIL.

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] Z. Dou, R. Song, and J.-R. Wen, "A Large-Scale Evaluation and Analysis of PersonalizedSearch Strategies," Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590, 2007.

[3] J. Teevan, S.T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis ofInterests and Activities," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Developmentin Information Retrieval (SIGIR), pp. 449-456, 2005.

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

[5] B. Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve SearchAccuracy," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD),2006.

[6] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User ProfileConstructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW),2004.

[7] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search," Proc. 14thACM Int'l Conf. Information and Knowledge Management (CIKM), 2005.

[8] X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using ImplicitFeedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development InformationRetrieval (SIGIR), 2005.

[9] F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search," Proc.15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.

[10] P. Agouris, J. Carswell, and A. Stefanidis, ``An environment for contentbased image retrieval from large spatial databases," ISPRS J. Photogram. Remote Sens., vol. 54, no. 4, pp. 263_272, 1999.

[11] M. Atallah and K. Frikken, ``Securely outsourcing linear algebra computations," in Proc. 5th ASIACCS, 2010, pp.48_59.

[12] M. Atallah and J. Li, ``Secure outsourcing of sequence comparisons," Int. J. Inf. Security, vol. 4, no. 4, pp.277_287, 2005.