

Supporting Privacy Protection in Personalized Web Search

¹Pulligilla.Manoj Kumar, ² M.Sruthi

¹M.Tech (CS), Department of Computer Science & Engineering, Sri Indu Institute of Engineering & Technology, Sheriguda (v)Ibrahimpattanam(M), RR District,501510.

²Assistant Professor, Department of Computer Science & Engineering, Sri Indu Institute of Engineering & Technology, Sheriguda (v)Ibrahimpattanam(M), RR District,501510

¹manoj.pulligilla@gmail.com, ²msruthi1248@gmail.com

Abstract:-With expanding number of websites the Web clients are expanded with the monstrous measure of information accessible in the web which is given by the Web crawler (WC). The point of the WC is to give the pertinent output to the client with the client's conduct snap were they performed. WC gives the pertinent result in the interest of the client successive snap based strategy. From this technique no affirmation to the client protection furthermore no securities were giving to their information. Subsequently clients were anxious for their private data amid inquiry has turned into a noteworthy obstruction. They were numerous methods were proposed by specialists the vast majority of that in view of the server side, it has give less security. For minimizing the security hazard here we propose the customer side based procedure with the mix of Eager strategy to keep the client information that we connected in learning mining zone.

Keywords: Web Search Engine, personalized search, user query logs, content search and privacy preserving.

1. INTRODUCTION:

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. Therefore, in order to provide better search result a general category of search technique Personalized Web search is used. In personalized web search, user information is collected and analyzed in order to find intention behind issued query fired by user. There are two categories of PWS, namely click-log-based and profile-based. The click-log based methods are straightforward— they simply impose bias to clicked pages in the user's query history. This strategy has been performing well but it work on repeated queries from same user which is a strong limitation to its applicability. While profile-based methods improve the search experience generated from user profiling techniques. Profile-based methods can be potentially effective for almost all sorts of queries, but are reported to be unstable under some circumstances. There are both advantages and disadvantages for both type of PWS technique, profile based PWS is more effective for improving search result. The user profile is made from information ga-

thered.

Searching is one of the common factor to know the information from the internet. Internet is one of the service providers, which provide the search result to the user with the help of the Web search engine (WSE) [1]. It employ by storing information about many web pages. WSE is a tool which allows the web user for finding information from the World Wide Web. WSE is one of the software that searches for and identifies the content or item from the web engine or web server or web database with correspond keywords or character specified by the user and finding particular sites on the World Wide Web [2]. Data search and information retrieval on the Internet has located high demands on search engines. Many search engines like Google, Yahoo provide a relevant and irrelevant data to the user based on their search. To avoid the irrelevant data the technique called Personalized Web Search (PWS) were arise. Inferring user search goals is very important in improving search-engine relevance and personalized search [3, 4]. This is based on the user profiles based on the click through log and the feedback session [5]. These data were generated from the frequent query requested by the user, history of query, browsing, bookmarks and so on.

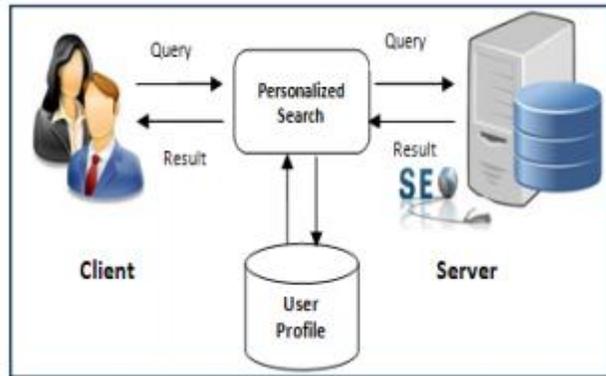


Fig 1.PSE Architecture

By these methods personal data were easily reveal. While many search engines take advantage of information about people in common, or regarding particular groups of people, personalized search based on a user profile that is unique to the individual person. Research systems that personalize search outcomes model their users in different ways. The Personalized Web Search provides a unique opportunity to consolidate and scrutinize the work from industrial labs on personalizing web search using user logged search behavior context. It presents a fully anonymized dataset, which has anonymized user id, queries based on the keywords, their terms of query, providing URLs, domain of URL and the user clicks. This dispute and the shared dataset will enable a whole new set of researchers to study the problem of personalizing web search experience. It decreases the likelihood of finding new information by biasing search results towards what the user has already found. By using these methods privacy of the user might be loss because of clicking the relevant search, frequently visited sites and providing their personal information like their name, address, etc. in this case their privacy might be leak. For this privacy issue, many existing work proposed a potential privacy problems in which a user may not be aware that their search results are personalized for them [6, 7].

2. RELATED WORK.

Previous works has focused on improving search result on profile- based PWS. Many representations for profile are available, some of them are term lists/vectors or bag of words to represent their profile while recent work create profile in hierarchical structure. The hierarchical representations are constructed with existing weighted topic hierarchy/graph, such as Wikipedia or the hierarchical profile is generated via term-frequency analysis

on the user data. UPS framework can adopt any hierarchical representation. Two classes of privacy protection problems for PWS is identified. One class treats privacy as identification of individual. Other considers data sensitivity as the privacy. Typical literature works in for class one try to solve the privacy problem on different levels, which includes the pseudoidentity, the group identity, no identity, and no personal information. The first level solution is proved to fragile and the third and fourth levels are impractical because of high cost in communication and cryptography. Therefore, the existing efforts focus on the second level. Online anonymity for PWS provides anonymity by generating a group profile of k users. Using this approach, the relation between the query and a single user is broken. The useless user profile (UUP) protocol shuffle queries among a group of users who issue them. As a result no entity can profile a certain individual. The shortcomings of class one solution are the high cost. In Class two solutions, users only trust themselves and don't tolerate the exposure of their complete profiles to 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. Privacy Enhancing personalized web search proposed a privacy protection solution for PWS based on hierarchical profiles. Using a user-specified threshold, a generalized profile is obtained in effect as a rooted sub tree of the complete profile. This paper provides personalized privacy protection in PWS. A person can specify the degree of privacy protection for her/his sensitive values by specifying "guarding nodes" in the taxonomy of the sensitive attribute. Thus, this paper allows user to customize privacy requirements in hierarchical user profiles

3. SYSTEM STUDY

Presented System

The existing profile-based Personalized Web Search does not support runtime profiling. User profile is generalized only once offline, and used to personalize all queries from a same user. Such "one profile fits all" strategy has drawbacks for the variety of queries. Also, the existing profile-based personalization does not even help to improve the search quality for some ad hoc queries. The existing methods do not take into account the customization of privacy requirements. In existing system, all the sensitive topics are detected using an absolute metric called surprisal based on the information theory which assumes that the interests with less user document sup-

port are more sensitive. However, this assumption can be doubted with a simple example: 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. Iterative user interactions are required in many personalization techniques for creating personalized search results. Search results are refined with some metrics such as rank scoring, average rank, and so on. This is infeasible for runtime profiling, since it poses too much risk of privacy breach, and also requires processing time for profiling. Therefore, we need predictive metrics to measure the search quality without iterative interaction of user.

Disadvantage: 1. All the sensitive topics are detected using an absolute metric called surprisal based on the information theory. 2. The existing profile-based PWS do not support runtime profiling. 3. The existing methods do not take into account the customization of privacy requirements. 4. Personalization techniques require iterative user interactions when creating personalized search results.

Proposed System

Indeed, the privacy concern is one of the major barriers in deploying serious personalized search applications, and how to attain personalized search through preserving users’ privacy. Here we propose a client side personalization which deals with the preserving privacy and envisions possible future strategies to fully protect user privacy. For privacy, we introduce our approach to digitalized multimedia content based on user profile information. For this, two main methods were developed: Automatic creation of user profiles based on our profile generator mechanism and on the other hand recommendation system based on the content to estimate the user interest based on our client side meta data. Fig 2: Proposed Architecture Above figure shows our proposed architecture which is built in the client side mechanism and here we protect the data from the server, so only we provide a privacy to the client user. Every query from the client user were provided by the separate requests to the server, this hides the frequent click through logs or content based mechanism, from this user can protect the data from the server. In the same case our mechanism maintains the online profiler about the user hence it hides the click logs and provides a safeguard to the user data. After that, online profiler query were processed in the manner of generalization process, it is used to meet the specific prerequisites to handle the user profile and it is based on the preprocessing the user

profiles. Our architecture, not only the user’s search performance but also their background activities (e.g., viewed before) and personal information (e.g., emails, browser bookmarks) could be included into the user profile, permitting for the structure of a much richer user model for personalization. The sensitive contextual information is usually not a main aspect since it is strictly stored and used on the client side. A user’s personal information including user queries and click logs history resides on the user’s personal computer, and is exploited to better suppose the user’s information requirements and provide relevant search results. Our proposed algorithm uses the greedy method based on the discriminating power and information loss protection to inherit the relations. Here it uses the inherited method to generalize the query. It allows performing the customization process to protect the data and use the User customizable Privacy-preserving Search framework addressed the privacy problems. This aims at protecting the privacy in individual user profiles.

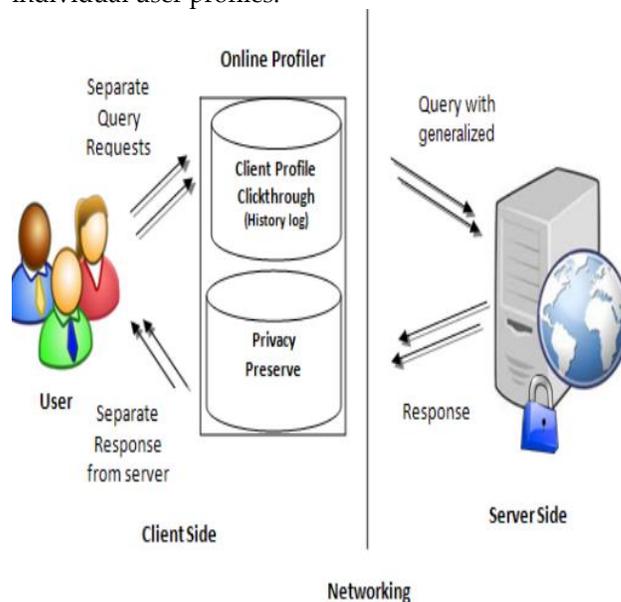


Fig 2. System Architecture

5. CONCLUSION & FUTURE WORK

Web users were increases because of available of information’s from the web browser based on the search engine. With the increasing number of user service engine must provide the relevant search result based on their behavior or based on the user performance. Providing relevant result to the user is based on their click logs, query histories, bookmarks, by this privacy of the user might be loss. For providing relevant search by using

these approaches the privacy of the user may loss. Most existing system provides a major barrier to the private information during user search. That approaches does not protect privacy issues and rising information loss for the user data. For this issue this paper proposes client based architecture based on the greedy algorithm to prevent the user data and provide the relevant search result to the user in future it can include this work in mobile application.

REFERENCES

- [1]. 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.
- [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]. 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.
- [5]. X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.
- [6]. 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.
- [7]. X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.
- [8]. 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.
- [9]. Dou, Zhicheng, Ruihua Song, and Ji-Rong Wen. "A large-scale evaluation and analysis of personalized search strategies", Proceedings of the 16th international conference on World Wide Web. ACM, 2007.
- [10]. 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.
- [11]. Shen, Xuehua, Bin Tan, and Cheng Xiang Zhai. "Implicit user modeling for personalized search." Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005.
- [12]. T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay, "Accurately Interpreting Clickthrough Data as Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05), pp. 154-161, 2005.
- [13]. Shen, Xuehua, Bin Tan, and Cheng Xiang Zhai. "Context-sensitive information retrieval using implicit feedback." Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2005.
- [14]. Xu, Yabo, et al. "Online anonymity for personalized web services." Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009.
- [15]. A. Viejo and J. Castell_a-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.
- [16]. Xu, Yabo, et al. "Privacy-enhancing personalized web search." Proceedings of the 16th international conference on World Wide Web. ACM,