ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 1, *January 2015*

Supporting privacy protection in personalized web search for knowledge mining

Mr. Stibu Stephen [1], Mr. A.Venugopal [2]

Scholar, Department of Computer Science, Sree Narayana Guru College, K.G.Chavadi, Coimbatore, Tamil Nadu[1]

Assistant Professor, Dept., of Computer Science, Sree Narayana Guru College, K.G.Chavadi, Coimbatore, Tamil Nadu[2]

Abstract: With increasing number of websites the Web users are increased with the massive amount of data available in the internet which is provided by the Web Search Engine (WSE). The aim of the WSE is to provide the relevant search result to the user with the behavior of the user click were they performed. WSE provide the relevant result on behalf of the user frequent click based method. From this method no assurance to the user privacy and also no securities were providing to their data. Hence users were afraid for their private information during search has become a major barrier. They were many techniques were proposed by researchers most of that based on the server side, it has provide less security. For minimizing the privacy risk here we propose the client side based technique with the combination of Greedy method to prevent the user data that we applied in Knowledge mining area.

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

1. INTRODUCTION

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.

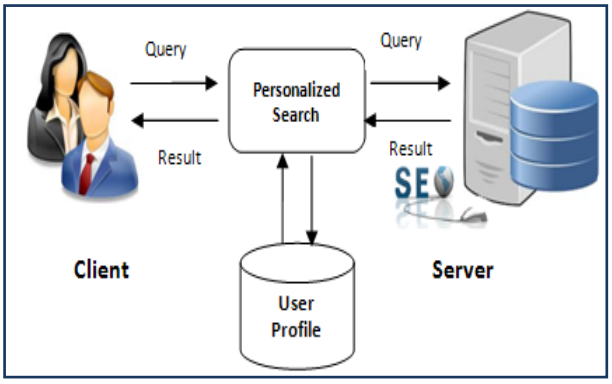

Fig 1: Personalized Search Engine Architecture

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].

It affords a host of services to people, and several of these services do not necessitate information to be grouped about a person to be customizable. While there is no warning of privacy assault with these services, the stability has been tipped to errand personalization over privacy, yet when it comes to search [8]. That approaches does not protect privacy issues rising from the lack of protection for the user data. To providing better privacy we propose a privacy preserving with the help of greedy method by providing the hybrid method of the discriminating power and prevent the information loss.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 1, *January 2015*

2. LITERATURE REVIEW

In [9] this paper, author study this problem and provide some preliminary conclusions. It presents a large-scale evaluation framework for personalized search based on query logs and then evaluates with the click and profile based strategies. By analyzing the results, author reveals that personalized search has significant improvement over common web search on some queries but it has little effect on other queries. Author also reveals that both long term and short-term contexts are very important in improving search performance for profile-based personalized search strategies. In this paper, author tries to investigate whether personalization is consistently effective under different situations. The profile-based personalized search strategies proposed in this paper are not as stable as the click-based ones. They could improve the search accuracy on some queries, but they also harm many queries. Since these strategies are far from optimal, author will continue his work to improve them in future [10]. It also finds 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.

From the author [11], they studied how to exploit implicit user modeling to intelligently personalize information retrieval and improve search accuracy. Unlike most previous work, it emphasizes the use of immediate search context and implicit feedback information as well as eager updating of search results to maximally benefit a user. Author presented a decision-theoretic framework for optimizing interactive information retrieval based on eager user model updating, in which the system responds to every action of the user by choosing a system action to optimize a utility function. Author propose [12] specific techniques to capture and exploit two types of implicit feedback information: (1) identifying related immediately preceding query and using the query and the corresponding search results to select appropriate terms to expand the current query, and (2) exploiting the viewed document summaries to immediately re-rank any documents that have not yet been seen by the user. Using these techniques, author develops a client side web search agent (UCAIR) on top of a popular search engine (Google) without any additional effort from the user.

From the [13] author have explored how to exploit implicit feedback information, including query history and click-through history within the same search session, to improve information retrieval performance. Using the KL-divergence retrieval model as the basis, author proposed and studied four statistical language models for context-sensitive information retrieval, i.e., FixInt, BayesInt, OnlineUp and BatchUp. It uses TREC AP Data to create a test set for evaluating implicit feedback models. The current work can be extended in several ways: First, it has only explored some very simple language models for incorporating implicit feedback information. It would be interesting to develop more sophisticated models to better exploit query history and click through history. For example, this may treat a clicked summary differently depending on whether the current query is a generalization or refinement of the previous query. Second, the proposed models can be implemented in any practical systems. It currently develops a client-side personalized search agent, which will incorporate some of the proposed algorithms. Author will also do a user study to evaluate effectiveness of these models in the real web search. Finally, author should further study a general retrieval framework for sequential decision making in interactive information retrieval and study how to optimize some of the parameters in the context-sensitive retrieval models.

This paper [14] was motivated by two emerging trends: web users want personalized services and web users want privacy. One challenge is that personal information must be made anonymous under the assumption that the participating parties, including the web service, are not completely trusted, due to systematic collection of personal information in addition to queries. Another challenge is the online and dynamic nature of web users. Author proposed the notion of online anonymity to protect web users and proposed an approach to maintain online anonymity through time. This approach makes use of a third party called the user pool and it does not require the user pool to be trusted. The simulation study on real US demographics showed promising results: it is feasible to achieve personalization for reasonable privacy settings.

From this approach [15, 16] they requires users to contribution the server full access to personal information on Internet, which break users' privacy. In this paper, author inspects the possibility of accomplish a balance between users' privacy and search quality. First, an algorithm is provided to the user for collecting, abbreviation, and organizing their personal information into a hierarchical user profile, where general terms are ranked to higher levels than explicit terms. Through this profile, users control what section of their private information is uncovered to the server by adjusting the minDetail threshold. An additional privacy measure, expRatio, is proposed to approximation the amount of privacy is exposed with the specified minDetail value. Yet, this paper is an exploratory work on the two features: First, author deal with unstructured data such as personal documents, for which it is still an open problem on how to define privacy. Secondly, author try to bridge the conflict needs of personalization and privacy protection by breaking the premise on privacy as an absolute standard. Also, author believe that an enhanced balance between privacy protection and search quality can be achieved if web search are personalized by allowing for only revealing those information associated to a specific query. It performs less protection for the user data and they were no assured for the user data and their profile information's. In this paper [17] the author studied the existing generalization methods are insufficient because they cannot assurance privacy protection in all cases, and frequently acquire redundant information loss by performing too much generalization. In this paper, author proposes the idea of personalized secrecy, and develops a new generalization structure that takes into account customized privacy necessities. This technique successfully avoid privacy intrusion even in scenarios

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 1, *January 2015*

where the existing approaches fail, and results in generalized tables that permit accurate aggregate analysis. This work [18] lays down a solid theoretical foundation for developing substitute generalization strategies. For instance, the greedy algorithm presented in this paper is not optimal, in the sense that it does not necessarily achieve the lowest information loss.

Discovering the optimal solution is a demanding problem. As another example, in performance, the recipients of the published data are often specialized users (e.g. scientists), who may explicitly specify the analytical tasks (such as association rule mining) required. This information may be utilized to free a table that is highly efficient for those tasks, without breaching the privacy constraints formulated by data owners.

3. PROBLEM DEFINITION

Most of the existing works concentrate on server-side personalized search services in preserving privacy, it provide a less security to the user. To provide a security to the user from the profile-based PWS from the client side, many researchers have to deem two challenging effects during the search process of the user, (i) To increase the search quality by user profile and (ii) hide the privacy content to place the privacy risk under control.

In many studies tells that user suggestions and their click based method is the helpful way to provide a personalized search and at the same time they have trouble with the loss of their privacy under their providing contents.

Profile based method is an ideal case for providing the relevant search [18, 19]. Under this they were many drawbacks, it does not support on the runtime profiling, it can be based on the online and offline generalization, insufficiently protection of the data and require more iteration for obtaining relevant search.

4. PROPOSED SYSTEM

Indeed, the privacy concern is one of the major barriers in deploying serious personalized search applications, and how to attain personalized search though preserving users' privacy. Here we propose a client side personalization which deals with the preserving privacy and envision 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 estimates the user interest based on our client side meta data.

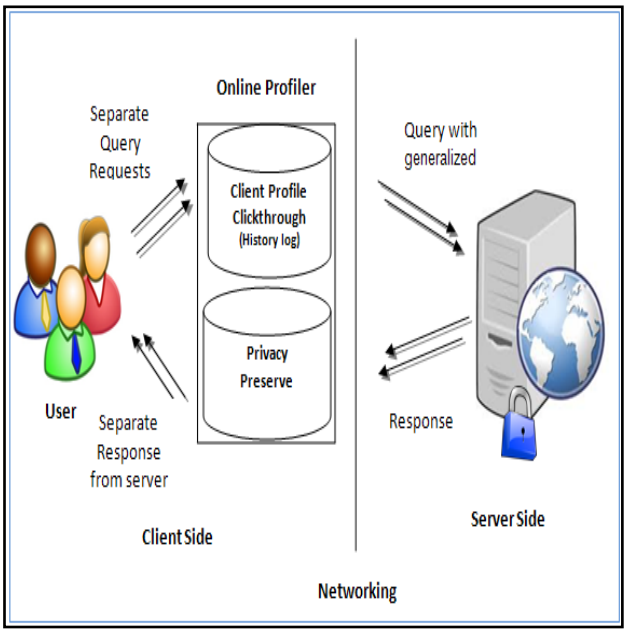

Fig 2: Proposed Architecture

Above figure shows our proposed architecture which is builds in the client side mechanism and here we protect the data from the server, so only we provides 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' information require and provide a 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.

DOI 10.17148/IJARCCE.2015.4115

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 1, *January 2015*

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, 2007

[17]. Xiao, Xiaokui, and Yufei Tao. "Personalized privacy preservation", *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. ACM, -2006.

[18]. Shou, Lidan, et al. "Supporting Privacy Protection in Personalized Web Search." (2012): 1-1.

[19]. 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.

DOI 10.17148/IJARCCE.2015.4115