

# Privacy Protection Based on User Search History in Data Mining

I. Preethi and P. Sri Priya

**Abstract**--- As per the increasing number of websites the users of web are increased with the huge amount of data available in the internet which is provided by the Personalized Search Engine (PSE). The main aim is that it helps users to pick up the useful information as quickly for them of their own interest, which are been stored in the database. PSE provides the relevant result based on user frequent click method. When a query is given, the provided result by the search engine is based on ODP (Open Directory Project) operation as filtered. There are many techniques that were proposed by the server-side but it provided less security. For reducing the risk that has been obtained in the privacy here we propose the client based technique by merging Greedy method to the user data prevention that is been applied in Data mining area.

**Keywords**--- Content Search and Privacy Protection, Personalized Web Search, ODP, Profile

---

## I. INTRODUCTION

The result is been obtained we re-order Data Mining, the discovery of new and interesting patterns in large data sets, in an exploding field. Recently there has been a realization that data mining has an impact on security. As the thirst of the knowledge increases the access to the internet users, uses the search engines are more important in life. The problem occurs when the users' takes time in search engine. The only solution to the problem is Personalized web Search (PSE).

The PSE helps the users to access the information based on profile that is unique to the individual person through log

and the feedback session. These data are provided by frequent query requested by the user, history of query, browsing and so on. As the effectiveness of profile-based PWS it as improved the quality of web search with higher usage of personal information by the users.

The obtained result from the search engine is been filtered and categorized using ODP operations, which helps to find out the user interest also. Once the pages for the users of next session. For re-ordering we are using Greedy method.

## II. LITERATURE REVIEW

The personalized web search presents a fully anonymized dataset, which consists of user-id, queries based on the keywords, providing URLs, domain of URL and the user-click. This shared data will enable a new set of researchers to face the problem of personalizing web search experience. By this method the researcher's finds a problem because of clicking the relevant search frequently visited sites and providing information like their name, address, etc. For the privacy issue existing work was proposed a privacy problem in which the search results are personalized for them from the paper [1 , 2].

It provides a main service to people and several of those services that do not necessitate information to be grouped about a person to be customizable. While this does not provide warning of privacy assault with these services, when it comes to search from the paper [3]. This paper lacks from privacy issue to user data. To provide a better privacy, we propose privacy with the greedy method by providing the hybrid method and prevent the information loss.

In [4] this paper, it presents a large-scale evaluation for

---

*I. Preethi, Research Scholar, Vels University, Chennai.  
P. Sri Priya, Assistant Professor, School of Computing Sciences, Vels University, Chennai.*

web search on query logs and evaluates with the click and profile based strategies. By the result, author reveals a better improvement on queries but also reveals the both long term and short term contexts in improving search performance for profile-based personalized search strategies. The profile personalized search strategies proposed in this paper was not stable in click-based ones.

From the paper [5], they found that the profile-based methods, uses both the long-term and short-term contents are important in improving search performance. The combination of both them can be more reliable than using any them.

In [6] this papers, the author merged two trends: personalized services for the web users and privacy. The challenge was that personal information was not completely trusted as including the web services. Another challenge is the online and dynamic nature of web users. Author proposed of an online anonymity to protect web users and to maintain time. This simulation study results in feasible to achieve personalization for reasonable privacy settings.

From this paper [7], the author inspects to provide a balance between user's privacy and search quality. First he deals with unstructured data such as personal documents, which is still an open problem second, he try to bridge the conflict needs of personalization and privacy protection. But still it performs less protection for the user data and they were no assurance of user data and information of their profile. Where the existing approaches fails and results in generalized tables that permits accurate aggregate analysis.

From this paper [8], it lays down in developing substitute generalization strategies. The greedy algorithm used in this paper is not optimal if does not achieve necessarily the lowest information loss. Finding the optimal solution is a demanding problem. The information that is been utilized to free the table, that is highly efficient for a task, the privacy constraints formulated by data owners.

### III. EXISTING SYSTEM

The existing personalized web search on profile based was concentrated on server-side as the results of the search engines are common to all the users and it provides a less security to the user.

In many studies the click based method proved that the personalized search was very helpful but the privacy protection was poor as the contents as been lost.

The profile based method also provided better personalized relevant searches but the drawback is that it does not support runtime profiling and used to personalize all queries from a same user, as it is based on online and offline generalization.

### IV. PROPOSED SYSTEM

In concern with the privacy protection in the server side is less effective. Here we propose a personalization web search from the client side where it deals with preserving privacy for users. We used two methods to develop a digitalized content based on user profile information.

Creation of user profiles automatically based on generating profile mechanism and the other method is based on the system recommendation to estimate the content of user interest on client side. In the client side mechanism we protect the data from the server. The performance here is not only the user search but also the background activities like hiding the frequent click through logs or content based mechanism from which the user can protect the data from the server and personal information (e.g. Browser bookmarks, e-mails so on...)that could be added in the user profile for a richer structural model for personalization.

In this proposed system we use greedy method to increase the discriminating power and decrease the information loss. This method is used for generalizing the query, where it allows the data to be protected and addresses the privacy problems. In this it aims for individual user profiles for privacy protection.

## V. CONCLUSION

This paper presented the privacy protection in the client-side for personalized web search. In previous work we have seen that the relevant information has been provided by the user click logs, the privacy of the user might be loss in this case and the profile based method has drawbacks that runtime profiling has not been supported, it requires more iteration to search for the relevant data and it has insufficient protection based on online and offline generalization. To improve the better quality of the search engine in the client-side, for this issue we propose the Greedy method to prevent the user data and provide relevant search based on user interests. In future work this method can be implemented in mobile application.

## REFERENCE

- [1] 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
- [2] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in PersonalizedSearch," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.
- [3] 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.
- [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] J. Teevan, S.T. Dumais, and D.J. Liebling, "To Personalize or Notto 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.
- [6] Y. Xu, K. Wang, G. Yang, and A.W.-C.Fu, "Online Anonymity for Personalized Web Services", Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.
- [7] A. Viejo and J. Castell\_a-Roca, "Using Social Networks to DistortUsers' Profiles Generated by Web Search Engines", Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.By Web Search Engines,"
- [8] Shou, Lidanetal, "Supporting privacy protection in personalized web search" (2012):1-1