International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.5., 2015
(Sept.-Oct.)

REVIEW ARTICLE

ISSN: 2321-7758

UPS EFFICIENT SEARCH ENGINE BASED ON WEB-SNIPPET HIERARCHICAL CLUSTERING

MS.MANISHA DESHMUKH, PROF. UMESH KULKARNI
Department of Computer Engineering, ARMIET, Department of Computer Engineering, VIT

MANISHA DESHMUKH

ABSTRACT
The quality of various search services on the Internet is effectively improved by using personalized web search (PWS). Personalized web search is a promising way to improve search quality by customizing search results for people with individual information goals. However, evidences show that user's reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. Privacy protection in PWS applications model user preferences as hierarchical user profiles. PWS framework called UPS can adaptively generalize profiles by queries while respecting user specified privacy requirements. Runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. Two greedy algorithms, namely Greedydp and GreedyIL, are used for runtime generalization. An online prediction mechanism for deciding whether personalizing a query is beneficial is provided. Extensive experiments demonstrate the effectiveness of the framework. The experimental results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.

Keywords - Greedydp and GreedyIL, PWS, UPS, hierarchical profiles, PWS

INTRODUCTION

The web search engine has long become the most important portal for ordinary people looking for useful information on the web. However, users might experience failure when search engines return irrelevant results that do not meet their real intentions.[5] Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query.[6]

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, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability.[5] In contrast, profile-based methods improve the search experience with complicated user-interest models 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. [8]

Although there are pros and cons for both types of PWS techniques, the profile-based PWS has demonstrated more effectiveness in improving the quality of web search recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query

The UPS 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. The two greedy algorithms were proposed, namely GreedyDP and GreedyIL, for the online generalization. [10]

Existing system was having some drawbacks like:

Controlling Session attacks

Capability to capture the series of queries

log file optimization and maintenance was not considered

Proposed work

The existing problems are addressed in our UPS. The framework works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements. The online phase handles queries as follows:

1. When a user issues a query qi on the client, the proxy generates a user profile in runtime in the light of query terms. The output of this step is a generalized user profile Gi satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.[10]
2. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search.
3. The search results are personalized with the profile and delivered back to the query proxy.
4. Finally, the proxy either presents the raw results to the user, or re ranks them with the complete user profile.

Create Admin Module

In this module to create admin login, search data details, display search details and user monitoring. Admin login page admin gave the username and password to enter the admin page. in search data details you have to give the product name, url, description and image. In search data display displays the stored search product and you have to edit all search product attributes. In user tracking to track the what type of product to be searched.

Fig- 1

Search Feed Module

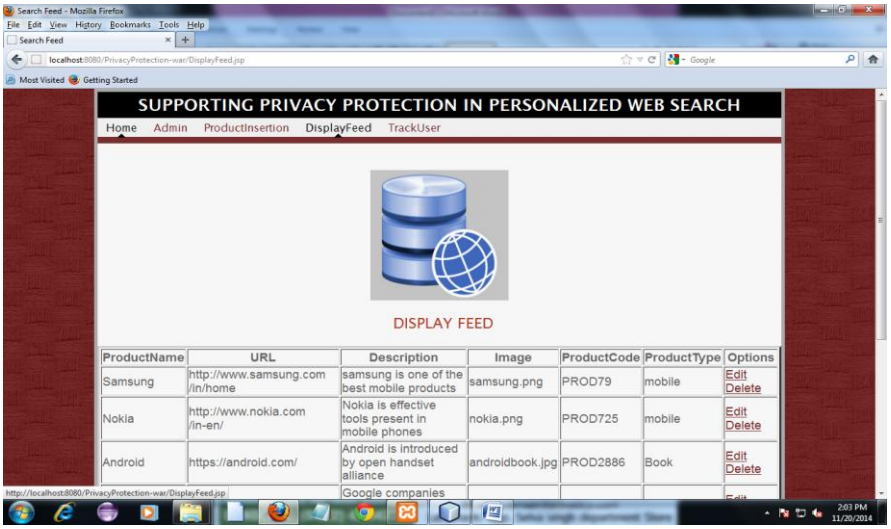

Fig- 2

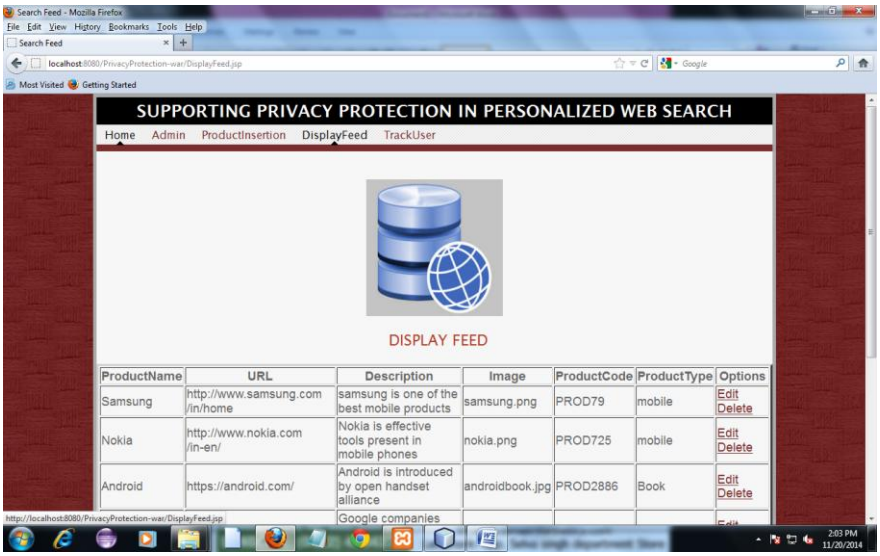

Fig-3

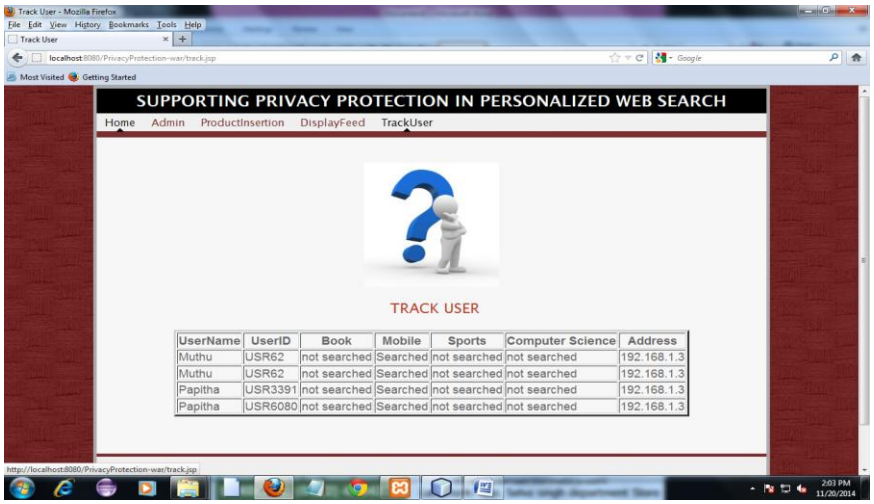

Fig-4

In search data details you have to give the product name, url, description and image. In search data display displays the stored search product and you have to edit and delete all search product attributes.

Search Feed Module In search data details you have to give the product name, url, description and image. In search data display displays the stored search product and you have to edit and delete all search product attributes.SQL injection is a code-injection technique, in which malicious SQL statements are inserted into entry fields for execution. For detecting SQL injection String Matching Algorithm is used.

User Tracking

In User Track module analyze the user what type of products to be searched. Display the user User id, list of product categories and ip address of user to find out the user's location. detecting brute force attack an attempt process is used. The user can login for three times, but if the user fails to attempt login for three times then IP address is blocked indicating non-genuine user and add it to blacklist.

Software/hardware Requirements:

- Operating System: Independent of Operating System
- Application Libraries: Java and J2EE
- Language: J2EE and Java
- Front End: Net Beans
- Database :Mysql
- Processor: Pentium IV. (& onwards).
- Memory (RAM):1GB RAM (32 bit) or 2 GB (64 bit)
- Hard disk: 40GB
- Internet access

RESULTS

UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality. Cover following advantages and efficient search technique.

Controlling Session attacks

Capability to capture the series of queries

log file optimization and maintenance was not considered

CONCLUSION

In this project Personalized web search (PWS) is used to improve the quality of various search services on the Internet. Privacy preserved PWS methods are used to protect the disclosure of personal information in search process. User customizable Privacy-preserving Search (UPS) framework is used to support privacy in search process. The UPS scheme is enhanced with attack resistant methods. Personalization utility is high in the personalized web search scheme. The system reduces the generalization risk levels. The system increases the attack control rate. Priority based user profile construction process is supported by the system

REFERENCES

[1]. 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.

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

[5]. 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.

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

[7]. 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.

[8]. 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.

[9]. J. Pitkow, H. Schu¨ tze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel, "Personalized Search," Comm. ACM, vol. 45, no. 9, pp. 50-55, 2002.

[10]. 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.

[11]. K. Hafner, Researchers Yearn to Use AOL Logs, but They Hesitate, New York Times, Aug. 2006.

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

[13]. J.S. Breese, D. Heckerman, and C.M. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI), pp. 43-52, 1998.

[14]. P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschu¨ tter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.

[15]. A. Pretschner and S. Gauch, "Ontology-Based Personalized Search and Browsing," Proc. IEEE 11th Int'l Conf. Tools with Artificial Intelligence (ICTAI '99), 1999.