Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 04, pp.15038-15041, April, 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

ENHANCED SECURITY PRESERVING IN PERSONALIZED WEB SEARCH

Ms. Shirisha, K., Sarika, P. and *Sushmitha, M.

1-1-18/36, TRT 120, Jawahar Nagar, RTC X Roads, Hyderabad

ARTICLE INFO

Article History:
Received 20th January, 2015
Received in revised form
02nd February, 2015
Accepted 20th March, 2015
Published online 30th April, 2015

Key words:

Personalized web search(pws),
GreedyIL,
GreedyDP.

ABSTRACT

Personalized web search (PWS) has improved various search services on internet. Now a days, as the reluctance of the users has been increased to hide their private information while searching. This has become the major problem for the wide proliferation of PWS. Here, we study how to protect PWS applications, so that user preferences can model as hierarchical user profiles. In this we are proposing a PWS framework known as UPS which can generalize user profiles by using queries with some privacy requirements. During run time generalization, it aims a balance between two predictive metrics which evaluate the use of personalization and privacy risk ij exposing their generalized profile. In run time generalization, we are presenting two greedy algorithms, Greedy DP and GreedyIL. Moreover, we are using an online prediction mechanism to decide whether personalizing a query is beneficial or not. This results the GreedyIL significantly outperforms GreedyDP in terms of efficiency.

Copyright © 2015 Shirisha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and *reproduction in any medium, provided the original work is properly cited.*

INTRODUCTION

1.1 Objective of the project

In today's world, the web search engine has become most important in our day to day life for browsing required information from the web. Sometimes, user may get irrelevant data than what required. The reason behind this failure is various number of user's contexts and ambiguity between texts. Personalized web search (PWS) is basically a type of search techniques which gives the better search results based on individual user needs. PWS collects user information and analyzes to know the intention of the user behind the query. The solutions for PWS is categorized into two types:

1. Click-log-based methods
2. Profile-based ones.

Click-based methods are very simple and straight forward, they give the clicked pages in the user's query history. Though this is performing consistently and considerably, as it works on repeated queries from the same user has become a strong limitation. Profile-based methods improve the search techniques with some complicated user-interest models given from user profiling techniques. This profile-based methods can be effective for almost all types of queries, but it is unstable for some queries under some circumstances. Even though, there are some disadvantages for both of the types of PWS, the profile-based PWS is more effective in improving the quality of web search with the increased usage of personal information to profile, which are gathered implicitly from query history, click through data bookmarks user documents. But, such collected data can easily reveal the private information of the user. Therefore, such privacy issues were raising, the AOL query logs scandal not only make users panic, but also dampen the data-publisher's enthusiasm in giving the personalized service. In fact, the privacy issues have become the major problem for the PWS services. To protect the user's information using profile-based PWS, researchers will consider two effects during search process. First one, they improve the quality of the search using personalization of the user profile. Second one, they have to hide the privacy contents in the user profile to avoid privacy risk. Some previous studies suggest that users are compromising privacy if the personalization by giving user profile to search engine yields better search quality. In ideal case, the gain can be obtained by personalization at only a small portion of the user profile, like a generalized profile. Therefore, privacy of the user can be protected without compromising the personalized search quality. Generally, there will be a tradeoff between the quality of search and the level of privacy protection achieved.

1.2 Existing system

The run time profiling was not supported by the existing profile-based Personalized Web search. The user profile was generalized only once and used to personalize all the queries from the same user. Obviously, due to this "one profile fits all" will lead to many drawbacks from different types of queries.

Corresponding author: Sushmitha, M.
1-1-18/36, TRT 120, Jawahar Nagar, RTC X Roads, Hyderabad

Example for this drawback was the search quality for some ad hoc queries was not improved by the profile-based personalization, even though the user profile was exposed to the server which puts the user into the privacy risk. The customization of privacy requirements does not taken into account in the existing system. Due to this, the user's information may not be protected from others. Consider an example, the sensitive data was detected using an absolute metric called surprisal based on the information theory, by assuming that the lesser document support are more sensitive. But this assumption was not sure. For example, if the user information contain more about "sex" then the surprisal will be that "sex" is not a sensitive data and it is a general one. But few priorities can say that the privacy is needed during generalization.

Most of the personalization techniques need iterative user interactions while doing personalized search results. Usually they do the search results based on some metrics for multiple user interactions, like scoring, average rank. However, this is infeasible for profiling at runtime, as it not only lead to privacy risk but also demand prohibitive processing time for profiling. Therefore, we need some measures to know the quality of search and risk after personalization, without any iterative user interaction.

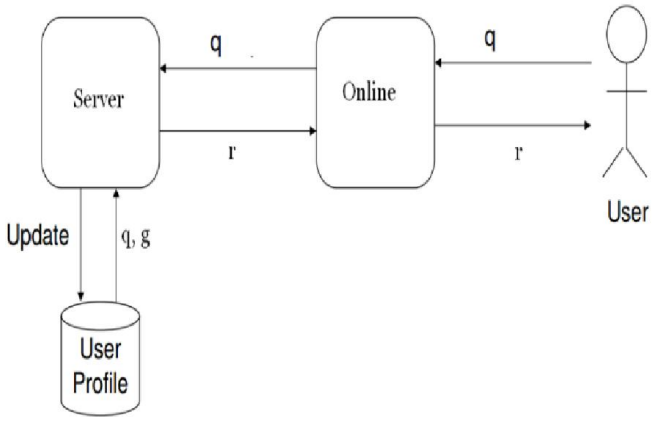


Fig 1.2.1. Architecture of existing system

1.3 Proposed system

Here, we are proposing a privacy-preserving personalized web search framework which is UPS, can generalize the profiles for every query according to the user requirements. It is based on the two conflicts for hierarchical user profile, namely privacy risk and personalization utility. Now we are solving the problem related to the privacy during searching process as Risk Profile Generalization, with its NP-hardness proved. During runtime profiling, we are developing two simple generalization algorithms, namely GreedyDP and GreedyIL.

If anyone tries to maximize the discriminating power (DP), leads to minimize the information loss (IL). By giving many number of heuristics, GreedyIL outperforms GreedyDP. To decide client whether to personalize a query in UPS, we are providing an inexpensive mechanism. Before each runtime profiling this decision can be made to enhance the stability of search results to avoid the unnecessary exposure of profile.

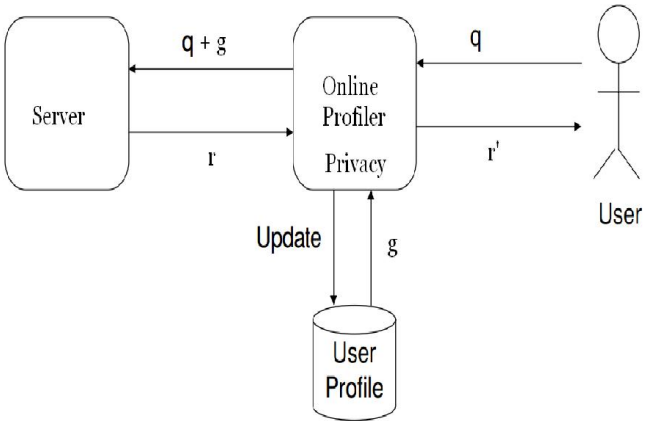


Fig 1.3.1. Architecture of proposed system

2. Modules

2.1. Profile-based personalization

In this module we use: A profile generator that automatically creates user profiles representing the user preferences. A content-based recommendated algorithm that estimates the user's interest in unknown content. The user profile is created in a hieracial structure based on availability of public accessible taxonomy. For example, in figure 2.a, we observe that owner of this profile is mainly interested in students of education, so the major fragments are formulated around this area of interest. The user is interested only in students then only that particular structure is taken for consideration.ie the path college – education-students is considered.

2.2. Privacy protection in pws

We propose a PWS framework called UPS(user customizable privacy preserving search) that generalizes profiles. We develop two simple generalization algorithms- Query level customization and Online prediction mechanism. A person can specify the degree of privacy preserving for his query by providing "guarding nodes" in the taxonomy for sensitive attributes.

2.3. Generalising user profile

The generalization process has some prerequisites to preprocess the user profile. First, initializes the user profile and add inherited properties of the local user profile. The process of generalization is obtained by the following algorithms. They are:

1. Brute force method
2. GreedyDL
3. GreedyIP

Brute force method

The brute force method generalizes all the sub rooted trees of the user profile to generate optimized generalized profile. The privacy preserving attributes are protected and the most frequently used sub tree is chosen and selected as the result.

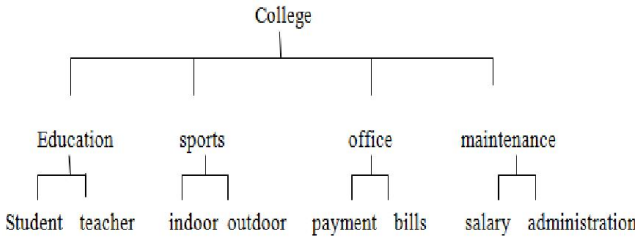


Fig 2a. User profile generation

GreedyDP Algorithm

If the complexity of the profile is more then the greedy method is used and one such algorithm used is greedy DL algorithm. We generate the profile as follows:

i. Firstly we prune the leaf node ie removing the leaf node from the tree.
ii. We repeat this step until we get a finite length of the tree.
iii. We follow bottom up approach and generate number of trees by eliminating a leaf node and select the best possible profile as optimal solution

This algorithm is mainly used for enhancing the discriminating power among the attributes so that we can easily prune the leaf nodes.

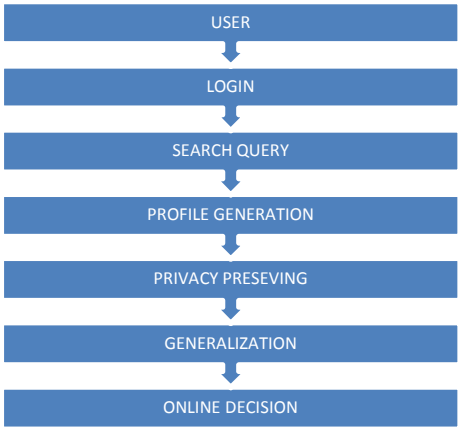
GreedyIL algorithm

The greedyIL algorithm is used to enhance the efficiency of the generalization. The algorithm is used to minimize the information loss that may be caused during the dividing process.

2.4. Online decision

We develop an online mechanism to decide whether to personalize a query or not. The basic idea is that if a distinct query is identified during generalization, the entire runtime profiling will be aborted and the query will be sent to the server without a user profile.

3. Data flow design



4. SCREEN SHOTS

Registration

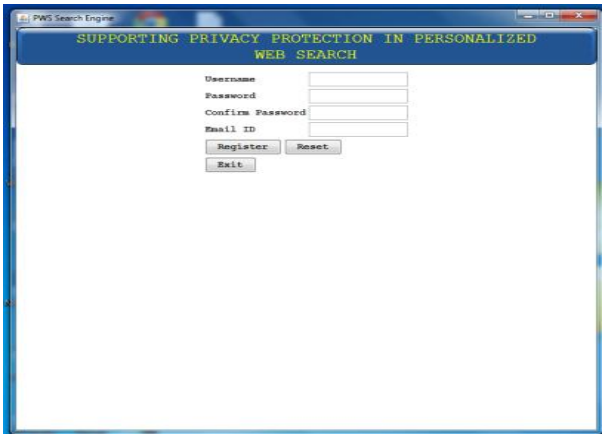


Fig 4.1. Registration screen

Login Page

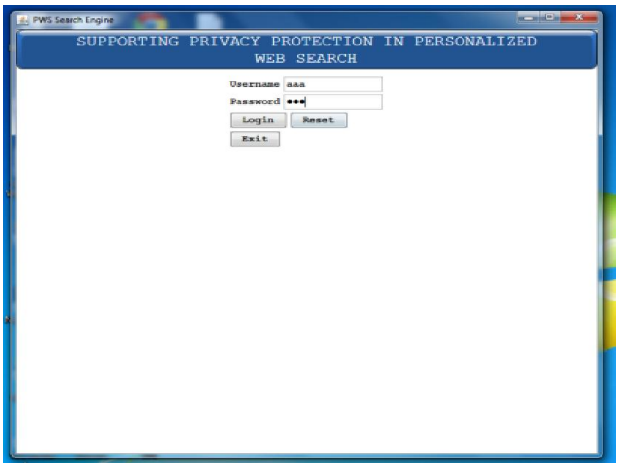


Fig 4.2. Login page screen

After successful login the user can able to give the queries according to his requirement.

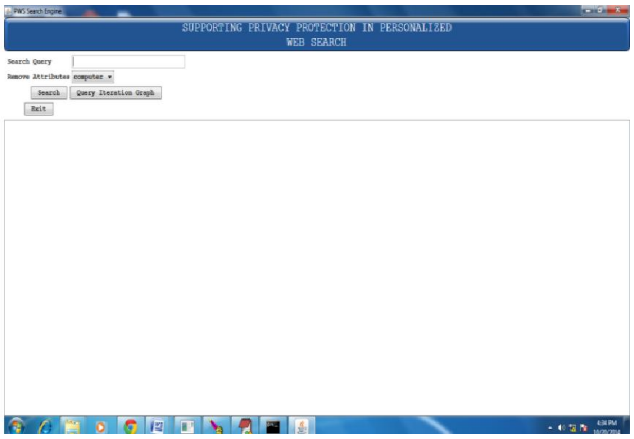


Fig 4.3. Search query screen

Online decision

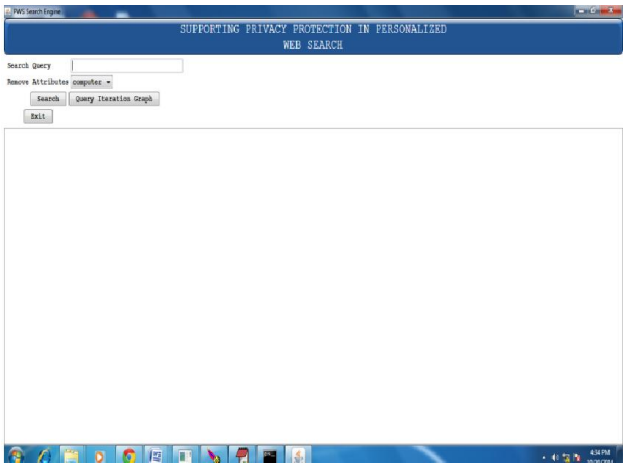


Fig 4.4. Online decision screen

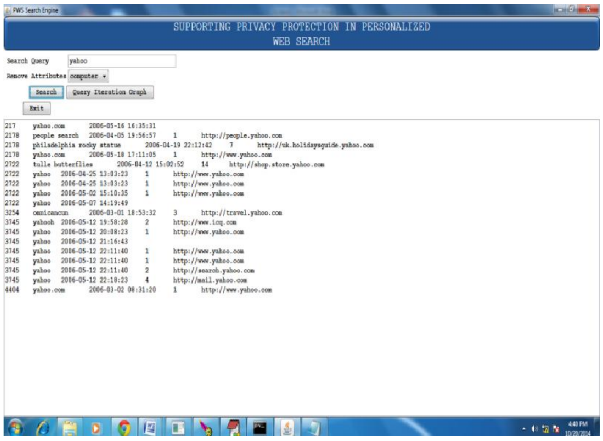


Fig 4.5. Search results screen

5. Conclusion and future work

We proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol in terms of Privacy and efficiency. One of the main ingredients in our proposed protocol is a novel secure multi-party protocol for computing the union (or intersection) of private subsets that each of the interacting players holds. Another ingredient is a protocol that tests the inclusion of an element held by one player in a subset held by another. Those protocols exploit the Fact that the underlying problem is of interest only when the number of players is greater than two. One research problem that this study suggests was described above; namely, to devise an efficient protocol for inequality verifications that uses the existence of a semi honest third party.

Such a protocol might enable to further improve upon the communication and computational costs of the second and third stages of the protocol of, as described in Sections above. Other research problems that this study suggests is the implementation of the techniques presented here to the problem of distributed association rule mining in the vertical setting the problem of mining generalized association rules and the problem of subgroup discovery in horizontally partitioned data.

6. REFERENCES

Adomavicius, G. and Tuzhilin, 2005. “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” IEEE Transaction on Knowledge and Data Engineering,vol. 17, no. 6, pp. 734–749.

Belkin N. J. and Croft, W. B. 1992. “Information filtering and information retrieval: Two sides of the same coin?” Communications of the ACM, vol. 35, no. 12, pp. 29–38,.

Chau, M. and Chen, H. 2008. “A machine learning approach to web page filtering using content and structure analysis,” Decision Support Systems, vol. 44, no. 2, pp. 482–494,.

Denning, P.J. 1982. “Electronic junk,” Communications of the ACM, vol. 25, no. 3, pp. 163–165,.

Foltz, P.W. and Dumais, S. T. 1992. “Personalized information delivery: An analysis of information filtering methods,” Communications of the ACM, vol. 35, no. 12, pp. 51–60,.

Jacobs, P.S. and Rau, L. F. 1990. “Scisor: Extracting information from online news,” Communications of the ACM, vol. 33, no. 11, pp. 88–97,.

Mooney, R. J. and Roy, L. 2000. “Content-based book recommending using learning for text categorization,” in Proceedings of the Fifth ACM Conference on Digital Libraries. New York: ACM Press, pp. 195–204.

Pollock, S. 1988. “A rule-based message filtering system,” ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232–254.

Sebastiani, F. 2002. “Machine learning in automated text categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1–47,.

Vanetti, M., Binaghi, E., Carminati, B., Carullo, M. and Ferrari, E. 2010. “Content-based filtering in on-line social networks,” in Proceedings of ECML/PKDD Workshop on Privacy and Security issues in Data Mining and Machine Learning (PSDML 2010).

****