

HEB

# Privacy Protection recommended system Using Personalized Web Search

CASS


Ritu Jhalani \*

M.Tech Student, Suresh Gyan Vihar University

*Email ID- editorcassstudies@gmail.com***ABSTRACT:**

Web mining has been explored to a large scale and different techniques have been proposed for many applications that includes Web Search, Classification and Personalization etc. In this report, we focus the significance of studying the evolving nature of the Web personalization. Web personalization is the process of customizing the structure and content of Web site in order to fulfill the needs of target users, taking benefit of the information collected from the analysis of the user's navigational behaviour and usage pattern data in correlation with other information collected in the Web context, namely, structure, content and user profile data. Due to the volatile size of the Web, the domain of Web personalization has gained great hand both in the research and digitization field. To improve internet quality and ranking of a particular web page it's necessary for developer to know the user click or user's navigational behaviour. Web personalization is just like an approach that customizes the information or services provided by a web site to an end user so that he or she can get relevant information as search result. In this paper we are focusing on all the categories of Web personalization. More specifically, we introduce Web personalization system, emphasizing the Web usage mining module. Different tasks are associated to implement Web personalization. A review of methods that are implemented for personalization, challenges including technical & security issues that occur is also given, along with the overview of the most popular tools and applications.

**Key Words :** Web personalization, Web usage mining, user profiling, WWW.

Access this Article Online	Quick Response Code: 
Website: <a href="http://heb-nic.in/cass-studies">http://heb-nic.in/cass-studies</a>	
Received on 26/10/2018	
Accepted on 28/10/2018 © HEB All rights reserved	

## **I. INTRODUCTION**

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. New personalization technologies are becoming increasingly widespread, raising a multitude of privacy challenges. The Web had become more social, a place where people use their real identities and communicate with their family, friends,

Finally, the Web had become mobile, frequently accessed through smart phones, providing new information and possibilities that can be used for personalization. Personalization has the potential to amplify and complicate the Internet's inherent privacy risks and concerns. For example, personalized content in a social network system can reveal potentially embarrassing information directly to friends, family, and colleagues. Personalizing content according to the physical location of the user can reveal the location to unauthorized third-party entities. Examples of these types of personalization are readily apparent at many web services operating today in which users are facing a complicated privacy landscape.

Personalized search gains popularity as there is the demand for more relevant information. Research has indicated low success rates among major search engines in providing relevant results; in 52% of 20,000 queries, searchers did not find any relevant results within the documents that Google returned.

In our Existing System, Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs.

### **Disadvantages of Existing System**

- The existing profile-based PWS do not support runtime profiling.
- The existing methods do not take into account the customization of privacy requirements.

### **Research Gap**

To protect user privacy in profile-based PWS, searchers have to consider two contradicting effects during the search process. On the one hand, they attempt to improve the search quality with the personalization utility of the user profile. Thus, user privacy can be protected without compromising the personalized search quality. In general, there is a tradeoff between the search quality and the level of privacy protection achieved from generalization. Unfortunately, the previous works of privacy preserving PWS are far from optimal.

Based on the study conducted by reviewing the literature, we may conclude the following

1. Previous work presents many scalable solutions for users to automatically build user profiles and rich query log based on search.
2. These profiles arrange a user's interests into a hierarchical organization according to specific interests.
3. There is no solid model proposed to do effective web personalized search.

### **Research Objective**

1. **On this topic some information will be required and need to find some advanced algorithms to sustain privacy with effective web personalized search.**
2. **Need to hide the privacy contents existing in the user profile to place the privacy risk under control**
3. **Significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) portion of the user profile,**

## **II. LITERATURE REVIEW**

**In [12, 3], better search results can be achieved with privacy guarantee if personalization is only performed based on a less sensitive or less specific part of the user profile, namely a generalized profile. The main idea is to build a hierarchical user profile and not to expose the sensitive part of the profile to the search engine by acquiring the level of privacy requirement from the user.**

**[13] automatically builds a hierarchical user profile in the client side based on user specified privacy settings.**

**In Knowledge-basedScheme [10] a similar approach is proposed to generate distorted user queries from a semantic point of view in order to preserve the utility of user profiles. In addition, linguistic analysis techniques are used to properly interpret complex queries submitted by users and generate new semantically-related queries accordingly.**

**[8] proposed a different way to protect user privacy by embellishing the search queries with decoy terms that exhibit similar specificity spread as the genuine terms, but point to plausible alternative topics.**

**[14] concentrated only on anonymizing user profiles by clustering them into user groups by taking into account the semantic relationships between query terms while satisfying the privacy constraints.**

### **Problem Identification**

**A huge task of association rule hiding is to be carried out. One of the method of extracting the association rule is by reducing the support and confidence sensitive data, this method is largely followed by many researcher the most common way to hide the sensitive data is ISL and DSR, And this also followed by many researcher for hiding any specific rule various approaches were proposed such as classification and clustering. Many of them have applied the technique of converting the sensitive data in a way that we get the customized version of database. As already explain above that by reducing support and confidence we can extract the association rule but some modification is required to achieve our aim.**

### **Proposed Technique**

**Association rule mining provides useful methods in market basket database. To explain this let us consider an example: Let transition  $T = \{\text{Trans 1, Trans 2, Trans 3 ...}, \text{Trans } n\}$  be the set of items and  $D$  the basic information of transactions. Usually there are one or more items in  $I$  in a transaction  $T$ .  $A$  with an association rule,  $B$  having the form any transition which value define are in  $X \rightarrow Y$ , i.e.  $X$  define the occurrence of  $Y$  such that  $X \cap Y = \text{Null}$  and where  $X$  and  $Y$  are subsets of  $T$  which are non-empty. A set of items is known as item set and  $X$  (subset of  $I$ ) is known as antecedent.**

Now we get the liberty to hide any given association rule which was not possible earlier.

We are taking a suitable example, to explain the association rule: let us take a minimum support of 3 and minimum confidence is 7. Now we can calculate the association rule as follows. i.e.

support =30%

Confidence=70%

Support  $A \Rightarrow B = \frac{\text{Common item in any giving table}}{\text{Total no transaction in any table}}$

Total no transaction in any table

$Y \Rightarrow X = \frac{4}{6} * 100 = 66.66\% \sim 66\%$

Our main task of hiding certain sensitive information is done by privacy preserving web mining. By hiding the private information we can be assured that no one can discover our data through web mining. Our purpose is the modification of database so that no one should be able to discover it. Provided a transaction database and setup of sensitive items A, We can be assured of our privacy. As shown in the above example that if element X is sensitive then these rules [ $XY \Rightarrow Z$  (50%, 75%),  $XZ \Rightarrow Y$  (50%, 75%)] would be applied and our data is much secured now as it can't be discovered by any technique of web mining.

### III. Result Analysis of previous methods

By simple ISL algorithm it is difficult to hide D and X. By modifying transaction Tran2 from Y to YD, One can check this out but still it is complicated to hide the rule  $D \rightarrow X$  by ISL algorithm.

Table 1: Insert into database

<u>TranID</u>	<u>Products</u>	<u>Bit Map</u>
Tran1	XYD	1101
Tran2	Y	0100
Tran3	XZD	1011
Tran4	Y	0100
Tran5	XYD	1101

Table 2: (Hiding  $D \rightarrow X$  by ISL approach)

<u>TranID</u>	<u>Products</u>	<u>Bit Map</u>
---------------	-----------------	----------------

Tran1	XYD	1101
Tran2	YD	0101
Tran3	XZD	1011
Tran4	XY	1100
Tran5	XYD	1101

From the above explanation one can easily figure out that the rule  $D \rightarrow Y$  cannot be hide by ISL algorithm. As by modify Tran2 from Y to YD we cannot hide the rule but the rule  $D \rightarrow X$  will have confidence and support of 60% and 75% respectively.

By DSR approach:

Table 3: (Hiding  $D \rightarrow X$  by DSR approach)

<u>TranID</u>	<u>Products</u>	<u>Bit Map</u>
Tran1	YD	0101
Tran2	Y	0100
Tran3	XZD	1011
Tran4	XY	1100
Tran5	XYD	1101

Now we can see that By this DSR approach, rule  $D \rightarrow X$  is hidden and now confidence and its support is 40% and 66%. This is less than ISL algorithm. But it has side effect that now the rule  $X \rightarrow D$  is also hidden.

#### IV. Result Analysis of Proposed Algorithm

Table 4

A Data Set

<u>TranID</u>	<u>Products</u>
Tran1	XYZ
Tran2	XYZ
Tran3	XYZ
Tran4	XY
Tran5	X
Tran6	XZ

Suppose MCT is 50%.

Now by calculating the confidence of the above transition. We get

$X \rightarrow Y$  (66.66%),

Confidence  $A \Rightarrow B = \frac{\text{Total Support in Number}(X \cup Y)}{\text{Total support in Number}(X)}$

Total support in Number(X)

$$X \Rightarrow Y = \frac{4}{6} * 100 = 66.66\% \sim 67\%$$

$$X \rightarrow Z \text{ (66.66\%)}$$

$$X \Rightarrow Z = \frac{4}{6} * 100 = 66.66\% \sim 67\%$$

$$Y \Rightarrow X = \frac{4}{4} * 100 = 100\%$$

$$Y \Rightarrow Z = \frac{3}{4} * 100 = 75\%$$

$$Z \Rightarrow X = \frac{4}{4} * 100 = 100\%$$

$$Z \Rightarrow Y = \frac{3}{4} * 100 = 75\%$$

If we want to hide item X then for this we take rule in which X is in RHS. i.e.  $Y \rightarrow X$  and  $Z \rightarrow X$ . and both of these have greater confidence. Now by taking the rule  $Y \rightarrow X$  and searching for transaction which supports both Y and X i.e.,  $Y = X = 1$ . And we found four such transition i.e. Tran1, Tran2, Tran3 and Tran4 with  $X = Y = 1$ . By Putting 0 for item X in the above four transactions. We get table 3 as the modified table for further working.

Table 5: After hiding  $Y \rightarrow X$

<u>TranID</u>	<u>bitmap(for XYZ)</u>
Tran1	011
Tran2	011
Tran3	011
Tran4	010
Tran5	100
Tran6	101

Now by calculating confidence of  $Y \rightarrow X$ , we get 0% which is less than the minimal confidence which implies that this rule is now hidden. By following the same procedure we take rule  $Z \rightarrow X$ , and find out for transactions in which  $X = Z = 1$ . We get only one such transaction Tran6 which has  $X = Z = 1$ . Now by updating transaction i.e. putting 0 instead of 1 in X. Here we get confidence of  $Z \rightarrow X$ , of 0% which is less than the minimum confidence. Which show that now this rule is also hidden. We have to now find out the rules in which X in Left hand Side.

**Table 6: After hiding Z->X**

<b>TranID</b>	<b>Bitmap ( for XYZ)</b>
<b>Tran1</b>	<b>011</b>
<b>Tran2</b>	<b>011</b>
<b>Tran3</b>	<b>011</b>
<b>Tran4</b>	<b>010</b>
<b>Tran5</b>	<b>100</b>
<b>Tran6</b>	<b>001</b>

We get such two rule in which A is on left hand side i.e.  $X \rightarrow Y$  and  $X \rightarrow Z$ . Both of them have confidence less than minimal confidence so there is no need to hide these rules. By modifying the above table we get the modified database i.e. by hiding item X Which is shown in the table 4. So it is clear that the hybrid algorithm unnecessarily scans the database. Because it scans the data base to find the same sensitive item X in LHS and it doesn't make any difference because item X is already hidden in the data base. Proposed algorithm 2 removes this problem of hybrid algorithm.

The comparison table is as follows:

**Table 14: Algorithm Comparison**

<b>Algorithms</b>	<b>No. of Rules Eliminate</b>	<b>No. of Database Scans</b>
<b>Hybrid</b>	<b>6</b>	<b>6</b>
<b>Proposed Algorithm 2</b>	<b>6</b>	<b>2</b>

## **V. Conclusion& Future Work**

In order to protect sensitive rule from being displayed, we presented two fundamental approaches.

The first approach was totally dependent on the generation of the association rule with the help of taking the table1 and also hiding the item set which was generated by me. The second things is reducing the importance of the large amount of data it work on the under the user defined threshold, so that here are not any rules can be generated from the taken items of tables.

Although data mining is useful for us as it can provide patterns and relationships, but is not able to define the value or significance of these patterns. As it does not tell the user about the sensitivity of the patterns.

## **VI. References**

- [1]. Anu, Sharma, T. Al-khwaldeh Ali, and Singh Aarti. "9. Anu Sharma, Ali T. Al-khwaldeh, Aarti SingABSEP3S- An Agent Based Security Engine for Privacy Preserving in Personalized Search." *International Journal of computing Academics Research*, 2016: 170-176.
- [2]. Cutrell, E., D.C. Robbins, S.T. Dumais, and R. Sarin. "Fast, flexible filtering with Phlat- Personal Search and Organization made easy." *SIGCHI*. 2006.
- [3]. Dumais, S. T., E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. "Stuff I've seen: a system for personal information retrieval and re-use." *SIGIR*. 2003. 72-79.
- [4]. Jaison, G.V., and C.M. Varghese. "Privacy Protection in Personalized Web Search using Homomorphic Encryption." *Internation Journal of Scientific and Research Publications*, 2015.
- [5]. Lidan, Shou, Bai He, Chen Ke, and Chen Gang. "Supporting Privacy Protection in Personalized Web Search." *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING VOL:26 NO:2*, 2014: 1-15.
- [6]. Manali, Wadnerkar, and D.R. Ingle. "Supporting Privacy Protection in Personalized Web Searching and Browsing." *International Journal of computer Science and Information Technologies* 6, no. 4 (2015): 4086-4093.
- [7]. S. Porna Sai, s. Udhaya, R Suganya, K.S.Sangeetha. "Supporting privacy protection in personalized web search A survey." *Indian Journal of Innovations and Developments* 3, no. 3 (2014): 45-49.
- [8]. Saraswathi, Punagin, and Arya Arti. "A Novel query obfuscation scheme with user controlled Privacy and personalization." *International Journal of Computer Applications*, 2017.
- [9]. Sharvari, V. Malthankar, and Kolte Shilpa. "Client side Privacy Protection Using Personalized Web Search." *7th International Conference on Communication, computing and Virtualization 2016*. 2016.
- [10]. Su, Jessica, Ansh Shukla, Sharad Goel, and Arvind Nayayan. "De-anonymizing web browsing Data with Social Networks." *Internatioal World Wide Web Conference committee*. 2017.
- [11]. Xuehua, Shen, Tan Bin, and Zhai ChengXiang. "Privacy Protection in Personalized Search." *Special Interest Group on Information Retrieval Forum* 41, no. 1 (2007): 4-17.
- [12]. L. Shou, H. Bai, K. Chen, and G. Chen. Supporting privacy protection in personalized web search. *Knowledge and Data Engineering, IEEE Transactions on*, 26(2):453–467, 2014.
- [13]. Y. Xu, K. Wang, B. Zhang, and Z. Chen. Privacy-enhancing personalized web search. In *Proceedings of the 16<sup>th</sup>international conference on World Wide Web*, pages 591–600. ACM, 2007.
- [14]. Y. Zhu, L. Xiong, and C. Verdery. Anonymizing user profiles for personalized web search. In *Proceedings of the 19<sup>th</sup>international conference on World wide web*, pages 1225–1226. ACM, 2010.



