A Framework For User Customizable Privacy Preserving Search

1R. Padmanaban, 2Priscilla Valentine, 3A.Nivetha, 4A. Aruna

1Assistant Professor, Department of Computer Science, Christ College of Engineering and Technology, Pondicherry, India.
2Student, B. Tech, Department of Computer Science, Christ College of Engineering and Technology, Pondicherry, India.
3Student, B. Tech Department of Computer Science, Christ College of Engineering and Technology, Pondicherry, India.
4Student, B. Tech Department of Computer Science, Christ College of Engineering and Technology, Pondicherry, India.

ABSTRACT

The effective manner of retrieving the information is the Personalized web search. This web search improves the quality of search services. However the evidences of various result analysis and proof show those users dislike when search engine displays unnecessary results. Users are unwilling to express their personal information. Existing personalized web search does not work equally well for all users under different situations. Therefore, we introduce a framework called UPS user customizable privacy preserving search where the user privacy is maintained in hierarchical system by creating individual user profiles for all users. This UPS framework analysis the given query understands the exact intension of the user and displays only the relevant data to the user. Experiments on the search engine show that our framework will increase the efficiency and accuracy of search engines.

INTRODUCTION

Recently people are often using search engines to find all kinds of information on the internet. As the quantity of information is continuously growing in internet so it has become very tediously difficult for the search engines to find information that satisfies the user. The same user may search for the details for the dove as a type of bird and some other time wants to know about the cleaning product dove. In many cases users experience failure when search engines return unwanted results that do not meet the real intension of the user. Such unwanted amount of large information leads to vagueness of texts. Only for the repeated queries from the same user the existing system works well. So there emerges the need for the user in personalized web search system which gives relevant output to the user as highly ranked pages UPS framework provides better results by considering the needs of the individual user. For this the user data has to be collected and analyzed so that the perfect search results required for the user behind the issued query is to be given to the user. Online profilers are used to collect information from the user then convert the profile to generalized profile. Generalized profile contains the details about search history and search occurrences with the help of Greedy algorithm. This greedy algorithm improves the search results by creating a repository of Taxonomy for making quicker access. Thus unwanted search results are removed by the implementation of prune leaf function. Greedy IL algorithm uses heuristics techniques, which increases the effectiveness of the generalization on large set of results. One important discovery is that finding of prune-leaf operation decreases the discriminating power of the profile.

Personalized search are adapted according to an individual's interests by collecting information about the individual with some specific query provided. Previously used search engines find results based on key words. Personalized search engine, such as Google have becoming more and more complex with the goal to "understand exactly what the user meant and produce the result exactly what the user want. Thus the search engines uses greedy algorithms and produce results based on the number of links from the sites by the use of mathematical algorithms, the higher a rank a site has the topper it is kept on the page. Search engines have two degrees of expertise: the trivial expert and the deep expert. The trivial expert degree serves as a proof for whoever knows some specific information on a given event. The deep experts are having comprehensible knowledge that they have the unique capacity to deliver relevant information to each individual. The search engine will act as a trivial expert and simply locate that data if the person knows what he or she wants. But
search engines are also capable of deep expertise in that they rank results indicating that those near the top are more relevant to a user’s wants than those below.

Research systems that personalize search results model their users in various ways. But user supplied data can be hard to collect and keep up to date. Others have built implied user models based on content previously read by the user in web or their history of interaction with Web pages. There are several systems available for personalizing Web search results. One technique that Google’s are using to personalize search results for its users is by tracking log in time and enabling the search history of the user in his browser. Google believes that you like that particular page when you keep using the particular web site again and again. So when we search for data in any web site Google's personalized search algorithm gives the page energy and keeps moving it up with the help of their ranks. Google will personalize your results even if you are logged out and it keeps a log record for 180 days of that particular web browser has searched by linking the cookie in that browser with the previously searched results.

In order to make a better understanding of how personalized search results are being presented, is by making a comparative study on group of searches results of an authorized users. The researchers have found that 14.5% of results show variation in the results due to personalization; however this differs largely by search query and result ranking position. It considers the results with high ranks through personalization technique include companies and politics. An important point to note is that the top ranked URLs won’t to change based on personalization since with most are occurring at the lower ranks.

Related Works:

Personalized searching technique has been proposed many years before but still the personalized strategy is no clear weather personalization is consistently efficient on different queries for different users. We introduce a large-scale evaluation framework, which express that personalized search has gradual improvement over frequently used queries, and it has very little effect on other queries. It also reveals that both long term and short-term histories are very important for improving the search performance for profile based personalized search. (Z. Dou, 2011)

Proposed search algorithm will personalized the user by considering user’s interaction with a wide variety of information in present web search. It does not rely on the unrealistic assumption of the people while searching. It uses a method that covers all implicit information about the user’s interest, this technique re-rank the feedback given from the framework. Thus it is very useful to obtain information from the previous issued queries or other information. (F. Qiu,2006).

It provides a client side privacy protection framework called PWS, Personalized web search. PWS does not support run-time profiling and it personalizes all queries from the same user indiscriminately. It does not customize the privacy requirement. (X. Shen,2005)

Existing System:

The existing profile based Personalized Web Search does not support runtime profiling. The user profile is works typically well when only it is in offline mode and only at this phase it is used to personalize all queries from a same user indiscriminately. Thus this profile is fixed perfectly in all strategy and has drawbacks based on the different types of queries. One truth is that profile-based personalization may not help to improving the search quality for some ad hoc queries; though exhibiting the user profile to a server thereby put the user’s privacy at risk.

The existing methods do not customize the privacy requirements. Thus it makes some of the user privacy requirements to be overprotected while others are less protected. For example, all the sensitive topics of an individual are detected using the algorithm that the users interest with less document support are more sensitive which is proves based on the information theory. However, this belief can be cleared with a simple example: If a user has large number of documents about his personal details then unfortunately only few works can be effectively address individual privacy needs during the generalization.

They usually change the search results with some algorithms which require multiple user interactions, such as average rank, rank scoring, and so on. However it does not work well for runtime profiling because it will not only develop too much risk of personal data, but also take maximum time for executing and creating profiles. Thus, we need some relative information metrics to measure the breach risk search and quality after personalization, without affecting sequential user interaction.

In Personalized Web Search: They are focused on two types of personalized web searching techniques they are Click-log based and Profile based personalized web search (X. Xiao, 2006).

A. Click-Log-Based:

In this type of personalized web search techniques, personalization is been executed out based on the number of clicks made by the user. The searched information gets recorded by the clicks made in the query logs of the history in the web browser. In this method, the web pages which are repeatedly viewed by the user is
recorded in the past browser history for a particular query is been personalized and score is computed for a particular web page and web search results are produced according to that score. Click log based personalized web search approaches will perform consistent and considerably well when it is working on frequent queries (M. Spertta, 2005). Limitation of the previous click log based approach is that they have poor performance in multi user environments.

B. Profile-Based:

In this approach, search results are sorted based on personal interest of the user profile. Mainly there are two techniques are used for creating user profile. The first technique is one that is issuing words which are repeatedly used in documents based on this a large number of profiles are been created with less perfection and less efficiency in search results by this method. The second technique is that of using of the existing method of ontology such as DMOZ (Z. Dou, 2010). This approach eases formation and preservation of most of the problem and this technique has found that profile based personalized in web search has become unstable when users search is been history increased.

Proposed System:

This paper proposes User customizable Privacy-preserving Search called UPS which generalizes profile for each query of the user as per their privacy specification. It has developed two simple and efficient generalizing algorithms namely Greedy DP and Greedy IL which is supported in runtime profiling (13) where the discriminating power is been increased using Greedy DP. (M. Spertta, 2005). While Greedy IL minimizing the information loss (IL) (M. Spertta, 2005). It also suggests a proper method for the user’s query for deciding whether to personalize a query or not through online decision in the UPS framework. In this method the data are arranged based on the previously searched information in the hierarchical order. The issued query is processed before each runtime profiling with the result of the online decision through greedy algorithms. Thus the stability of the searched result is been increased with efficiency and security.

1. Offline Phase:

In offline phase, a hierarchical user profile (y.xu, 2009) is build and modifies accordingly with the user-specified privacy supplies.

2. Online Phase:

In online phase, the proxy server creates a user profile on runtime when the user issues a query q to the client. By considering two important points in mind, namely the personalization application and the privacy risk which are defined for user profiles are directed by generalizing process. For most personalized search, the generalized user profile and the query are sent to together to the server their search results are been personalized with the profile and bring back to the online decisional. Lastly, the decisional re-ranks the results with the complete user.

![Fig. 3: 1UPS architecture](image)

Online Decision:

The profile-based personalization at the server side contributes little or even decreases the search quality, and by exposing the profile to the server leads to sure risk of the privacy in their data. In order to eradicate this problem we introduce an online mechanism which decides whether to personalize a query or not. If a distinct query is identified during generalizing, it will be sent to the server without profiling in the hierarchical structure of the user profile.
Greedy Algorithm:
A greedy algorithm follows the problem solving heuristic search techniques by making the local best choice at each stage of their searching process with the hope of finding global best results. Greedy algorithms are considered easy for execution and they are the simple approach to decide next step that provide better results. Yet the greedy process does not produce a best solution in many cases, still the greedy heuristic process obtains locally best results that roughly produce global optimal results in a considerable time.

3.2.1 GreedyDP Algorithm:
It works in a bottom up manner. Starting from the leaf node, for every loop it chooses leaf node topic for pruning thus it maximize and increases the usage of output to the users. At each iterational loop, a best user profile is maintained, thus it satisfies the Risk constraint. The iteration ends its search at the root node when the required data is reached. The optimal result is gained by selecting the best result. GreedyDp algorithms steadily increase the computational cost and memory requirement.

3.1.2 GreedyIL algorithm:
GreedyIL algorithm improves generalization efficiency. GreedyIL (Y. Zhu, 2010) maintains priority queue in descending order for each candidate prune leaf operators. This decreases the computational cost. GreedyIL (6),(9),(10) states to end of the iteration when Risk percentage is been satisfied or when there is a single leaf left. Since, there is less computational cost compared to GreedyDp (Lidan Shou, 2014) GreedyIL (Y. Xu,2009) outperforms GreedyDp (Lidan Shou, 2014.).

Performance Evaluation:
Security and privacy are the commonly used parameters in privacy protection techniques. The existing algorithm GreedyDp is been compared with the proposed approach GreedyIL. The system is been evaluated with the following parameters.

RESULTS AND DISCUSSION

4.1.1 Accuracy:
Accuracy = number of correct results/number of all returned results.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Number Of users</th>
<th>Existing Accuracy</th>
<th>proposed Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>185</td>
<td>65%</td>
<td>92%</td>
</tr>
<tr>
<td>Stock market</td>
<td>205</td>
<td>63%</td>
<td>96%</td>
</tr>
<tr>
<td>Sports</td>
<td>122</td>
<td>67%</td>
<td>93%</td>
</tr>
</tbody>
</table>

Fig. 4.1.1.1: table of accuracy of existing over proposed

In proposed approach accuracy level is quite higher when compared with the existing system methods.

Fig. 4.1.1.2: graph of accuracy of existing over proposed

4.1.2 Security:
It represents the complete study of the parameters and makes a comparison of the security level between the proposed and the existing system.
Thus the privacy of the user is been protected using the GreedyIL algorithm in our proposed approach.

**Conclusion:**

This work is used in client-side privacy protection framework called UPS for personalized web search. UPS could possibly be accepted by any PWS that uses user profiles in a hierarchical manner. The framework allowed the users to maintain their data confidentially without compromising the search quality by maintain the hierarchical profiles for individual users. We proposed two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization for the user. Our experimental results revealed that UPS can produce better quality search results while preserving user's customized privacy requirements. This system will produce the better and efficient results to our solution, UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search qualities.

**REFERENCES**


T’Zastelli-roca.A.vieja And J. Heera Joan Cmarti, 2009. ”Preserving user’s privacy in Web search engine”.


