International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 04 | Apr-2016 www.irjet.net p-ISSN: 2395-0072

Profile Based Personalization Using User Privacy Preserving Search

Miss. Snehal R. kawalkar [1], Prof. P. L. Ramteke [2]

[1]*Student, Department of computer science & Information Technology, H.V.P.M. College of Engineering & Technology, Amravati, India.*
[2]*Professor, Department of computer science & Information Technology, H.V.P.M. College of Engineering & Technology, Amravati, India*

Abstract - *Personalized web search is a promising way to improve search quality by customizing search results for people with individual information goals. However, users are uncomfortable with exposing private preference information to search engines. On the other hand, privacy is not absolute, and often can be compromised if there is a gain in service or profitability to the user. Thus, a balance must be struck between search quality and privacy protection. This paper models preference of users as hierarchical user profiles. It proposes a framework called UPS which generalizes profile at the same time maintaining privacy requirement specified by user. It has been found that UPS framework is one of the efficient techniques which guarantees the user privacy and retrieves the contents as per user requirement accurately. During execution query will be executed and then it will be maintaining the privacy depending the profile of the user. The information extraction is based on User Profile. An efficient user profile can improve the search engines performance by identifying the individual interest. GreedyIL algorithm used to improves the efficiency of the generalization.*

Key Words: Personalized web Search, Search engine, Profile, GreedyIL.

1. INTRODUCTION

It has turned into ever harder for users to discover data on the WWW that gratify their entity desires since information resources on the WWW continue to grow. Under these circumstances, Web search engines help users find useful data on the WWW. However, when the similar query is presented by dissimilar users, mainly search engines give again the similar results regardless of who submits the query. Commercial Web search engines are expected to process user queries under tight response time constraints while being able to operate under heavy querytraffic loads. Personalization is being used by most online service platforms (OSPs) such as search, advertising, shopping, etc.

The goal is to lure users by offering a better service experience customized to their individual interests. A popular trend is to employ profile based personalization, where OSPs build extensive profile for the user and personalize the content based on this profile. While OSPs definitely track rich user histories, they can infer a great deal more by mining this rare data. Casually talking, OSPs can decide user's interests and biases on different categories, which can then be used for personalization. Web search results should adapt to users with different data wants. In order to predict such information wants, there are numerous methods relate data mining techniques to extract usage patterns from Web logs. However, the discovery of patterns from usage data by itself is not sufficient for performing the personalization tasks.

In this paper propose a privacy-preserving personalized web search structure UPS, which can simplify profiles for every query according to user-specified privacy requirements. Relying on the description of two incompatible metrics, explicitly personalization usefulness and confidentiality risk, for hierarchical user profile prepare the difficulty of privacy-preserving personalized search as Risk Profile overview, with its NP-hardness proved. We develop effective simplification algorithms GreedyIL, to maintain run time profiling.

2. LITERATURE REVIEW

In this section, many profile representations are available in the literature to facilitate different personalization strategies.

M. Spertta and S. Gach, [1] systematically examined the issue of privacy preservation in personalized search. The four levels of privacy protection is distinguished, and analyze various software architectures for personalized search. This work showed that client-side personalization has advantages over the existing server-side personalized search services in preserving privacy, and envision possible future strategies to fully protect user privacy.

Y. Xu, K. Wang, G. Yang[4] proposed the notion of online anonymity to enable users to issue personalized queries to an un-trusted web service while with their anonymity preserved. The challenge for providing online anonymity is dealing with unknown and dynamic web users who can get online and offline at any time. Introduces the notion of online anonymity to ensure that each query entry in the query log cannot be linked to its sender and an algorithm that achieves online anonymity through the user pool is proposed. This approach can be extended to deal with personally identifying information that may be contained in the query. The method is also applicable to general web services where there is a need to anonymize the query, with or without personalization.

Krause and Horvitz [2] employ statistical techniques to learn a probabilistic model, and then use this model to generate the near-optimal partial profile. One main limitation in this work is that it builds the user profile as a finite set of attributes, and the probabilistic model is trained through predefined frequent queries. These assumptions are impractical in the context of PWS.

In J. Castelli-Roca, A. Viejo and J. Herrera[5] presents a novel protocol Useless User Profile (UUP) protocol, specially designed to protect the users' privacy in front of web search profiling. System provides a distorted user profile to the web search engine. Also offers implementation details, computational and communication results that show that the proposed protocol improves the existing solutions in terms of query delay. The protocol also provides an affordable overhead while offering privacy benefits to the users.

In X. Xiao and Y. Tao[7] presented a new generalization framework based on the concept of personalized anonymity. This technique performs the minimum generalization for satisfying everybody's requirements, and thus, retains the largest amount of information from the microdata.

J. Teevan, S.T. Dumais, and D.J. Liebling, [8] examines variability in user intent using both explicit relevance judgments and large-scale log analysis of user behavior patterns. They characterize queries using a variety of features of the query, the results returned for the query, and people's interaction history with the query. Using these features, the authors build predictive models to identify queries that can benefit from personalization.

X. Shen, B. Tan, and C. Zhai,[3] Information retrieval systems (e.g., web search engines) are critical for overcoming information overload. A major deficiency of existing retrieval systems is that they generally lack user modeling and are not adaptive to individual users, resulting in inherently non-optimal retrieval performance.

In 2007, Z. Dou [9] proposed Average Precision metric, to measure the effectiveness of the personalization in UPS. Susan T. Dumais[10] introduces a search algorithm that considers user's prior interactions with a wide variety of content, to personalize their current web search. Rather than relying on the unrealistic assumption that people will precisely specify their intent when searching, it pursues techniques that leverage implicit information about the user's interests.

In 2013, S.Vanitha [13] proposed a technique extracts the web pages based on two methods. The information extraction is based on User Profile and Click through data. The main advantage of this system is that it is possible to extract personalized web pages. An efficient user profile can improve the search engines performance by identifying the individual interest.

Lidan Shou and Gang Chen et al [6], uses hierarchical user structure for modeling user interests. The system provides generalization of user profile with use of an online profiler at the client side.

3. PROPOSED SYSTEM

We propose a privacy-preserving personalized web search framework UPS, which can generalize profiles for each query according to user-specified privacy requirements. Relying on the definition of two conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as Risk Profile Generalization, with its NP-hardness proved. To prevent the information loss while performing runtime generalization, a greedy algorithm is used here.

We provide an inexpensive mechanism for the client to decide whether to personalize a query in UPS. This decision can be made before each runtime profiling to enhance the stability of the search results while avoid the unnecessary exposure of the profile.

Advantages:

1. It enhances the stability of the search quality.
2. It avoids the unnecessary exposure of the user profile.

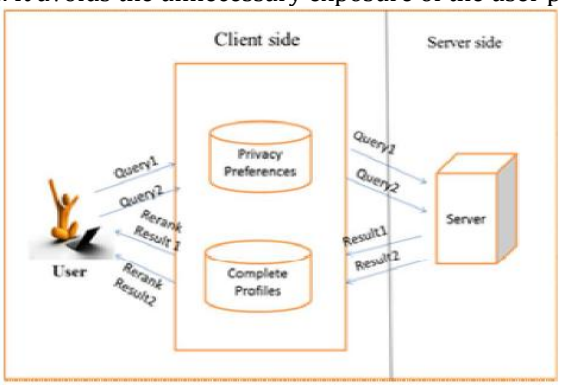


Fig-1: Proposed System Architecture

The Proposed system consists of four modules:

- Profile-Based Personalization
- Generating User Profile
- Online Decision
- Privacy Protection in PWS system

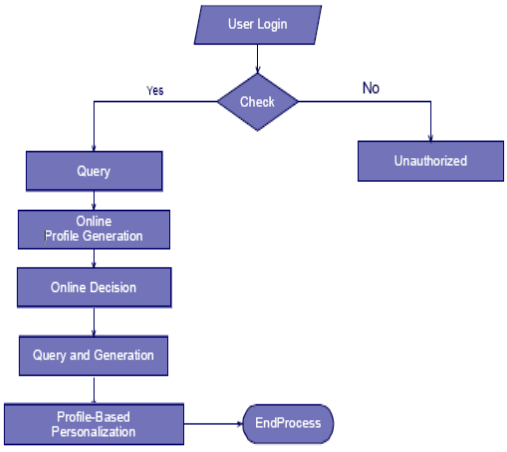
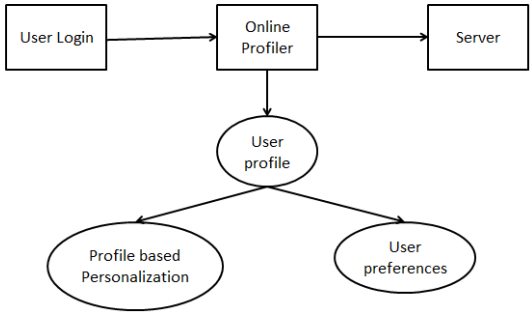


Fig-2: System flow diagram

3.1 Profile-Based Personalization



Personalization is the process of presenting the right information to the right user at the correct instant. In order to study on a user, systems must gather personal data, investigate it, and accumulate the consequences of the analysis in a user profile. Data can be composed from users in two traditions: unambiguously, for instance ask for comment such as preferences or ratings; or perfectly, for instance detect user behaviors such as the time spent reading an on-line document. The accessible profile-based PWS do not hold runtime profiling. A user profile is usually inclusive for only one time offline, and utilized to personalize all query from a similar user indiscriminatingly. Such "one profile fits all" strategy certainly has drawbacks given the variety of queries. The existing methods do not take into account the customization of privacy requirements. This possibly creates several user privacy to be overprotected while others insufficiently protected. For example, all the sensitive topics are detected using an absolute metric called surprised based on the information theory, supposing that the interests with less user document support are more sensitive. Many personalization techniques require iterative user interactions when creating personalized search results. They typically process the search results with some metrics which require multiple user communications, such as average rank, rank scoring, and so on.

We propose a method to personalize the search results of a user based on their profile information. The two main mechanisms used for this purpose are: a profile generator that is used to create the profile of the user based on the inputs and preferences given by the user and an algorithm that ranks the search results based on the preferences and interests of the user.

3.2 Generalizing User Profiles

Generalizing the profile of the user is the most important step in the entire framework. Since our proposal focuses mainly on the privacy requirements of the user, the generalized profile of the user has a very important role to play in the entire process. The process of generalization is done by using a parent profile and an inherited profile. The parent profile is the original profile of the user that contains all the details of the user. The inherited profile contains the profile with the necessary privacy requirements of the user. The profile of the user is stored in the form of a graph that has a hierarchical structure. There are root nodes and child nodes for specific nodes. If a parent node is in privacy protection, all this children nodes are automatically privacy protected by the reverse need not be true. The privacy requirements specified by the user are indicated in the graph using a marking system. The inherited profile has all the attributes of the parent profile with additional attributes that are not present in the parent profile. If there is an update in the user profile, the change is propagated in the generalization process and the graph is modified accordingly. This process of generalization occurs when the user inputs a search query and the complete profile of the user should not be shown to the server. Here, the generalized profile with the privacy requirements of the user is only shown to the server. Hence, the privacy of the user is maintained.

3.3. Online Decision

The personalization of web search leads to lot of unwanted information of the user being shown to the server, with not much improvement in the quality of the search results. This puts the privacy of the user in risk. To make sure that the user gets relevant and efficient results even for distinct queries, we propose an online decision method. Here, it is decided if a query should be personalized or not. If the query is distinct, i.e., it is very different and not much related to the preferences of the user, the profile that is created during runtime (the generalized profile) of the user is discarded and the query is sent to the server without a user profile.

3.4 Privacy Protection in PWS system

Different users have different requirements of privacy protection. While some users may not want anyone else to know or hold any of their personal information, others may be willing to share some personal information for better search results or services. Thus the level of privacy protection may need to be tuned for Different users to accommodate Different preferences for the tradeoff of personalization and privacy protection.

Generally there are two classes of privacy protection problems for PWS. One class includes those treat privacy as the identification of an individual. The other includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server. Typical works in the literature of protecting user identifications (class one) try to solve the privacy problem on different levels, including the pseudoidentity, the group identity, no identity, and no personal information. Solution to the first level is proved to fragile. The third and fourth levels are impractical due to high cost in communication and cryptography. Therefore, the existing efforts focus on the second level.

A PWS framework called UPS that generalize profiles for each query according to user-specified privacy requirements. Mainly, two predictive metrics are proposed to evaluate the privacy breach risk and the query utility for hierarchical user profile. But effective generalization algorithms are used for user profiles allowing the query-level customization by using proposed metrics. And online prediction mechanism based on query utility for deciding whether to personalize a query in UPS.

4. GREEDY ALGORITHM

A **greedy algorithm** is a mathematical process that recursively constructs a set Recursion of objects from the smallest possible constituent parts. is an approach to problem solving in which the solution to a particular problem depends on solutions to smaller instances of the same problem.

Greedy algorithms look for simple, easy-to-implement solutions to complex, multi-step problems by deciding which next step will provide the most obvious benefit. Such algorithms are called greedy because while the optimal solution to each smaller instance will provide an immediate output, the algorithm doesn't consider the larger problem as a whole. Once a decision has been made, it is never reconsidered.

The advantage to using a greedy algorithm is that solutions to smaller instances of the problem can be straightforward and easy to understand. The disadvantage is that it is entirely possible that the most optimal short-term solutions may lead to the worst long-term outcome.
Greedy algorithms are often used in packets with the fewest number of hops machine learning, business intelligence (BI), artificial intelligence (AI) and programming. ad hoc mobile networking to efficiently route and the shortest delay possible.

The Greedy IL Algorithm

The Greedy IL algorithm is used in the implementation concept to provide generalization of queries by which the generalized user profile is generated. The Greedy algorithm enhance the efficiency of generalization using heuristics on several findings. Another important findings is that prune leaf operation reduces the discriminating power of the profile.
If G′ is a profile obtained by applying a prune-leaf operation on G, then $DP(q,\mathcal{G}) \geq DP(q,\mathcal{G}')$ Considering operation $\mathcal{G}_i \xrightarrow{-t} \mathcal{G}_{i+1}$ in the ith iteration, maximizing $DP(q,\mathcal{G}_{i+1})$ is equivalent to minimizing the incurred information loss, which is defined as $DP(q,\mathcal{G}_i) - DP(q,\mathcal{G}_{i+1})$

Symbol	Description
DP	Discriminating power
Gi	Generalized profile
q	Query

```
GET PROFILE (Greedy IL)
INPUT Query Q, Privacy Threshold δ
OUTPUT Profile P
1.Get Query Q.
2.Privacy Threshold δ  [0-1]
3.Split query into words QW.
4.Find Seed Profile G. (Any category level contains the category name inside the query words).
5.CID ={}.
6.For i=1 to QW.Count
  a. Find Category Ids which contains QW[i] in their category names.
  b.Add Category Ids to CID
7.Next
8.If CID.Count>0
9.Create Profile P.
10.For j=1 to CID.Count
  a.While RiskFactor(Q,CID[j]) > δ
            i.P.CategoryId= CID[j]
            ii.P.CategoryName  = Category Name of CID[j]
    a.End While
11.Next
12.Return P
```

6. CONCLUSIONS

A client side personalized web search that caters to the privacy needs of the user was proposed in this paper. Any personalized web search can adapt to this framework. This is helpful and advantageous because efficient search results are provided to the user based on their preferences in their profile. Also, their privacy needs are maintained and they can also be customized according to the varying needs of the user. We use GreedyIL algorithm which improves the efficiency of the generalization using heuristics based on numerous answers.

ACKNOWLEDGEMENT

I would like to thanks my Guide Prof. P. L. Ramteke and Principal Dr. A. B. Marathe, who provided me constructive and positive feedback during the preparation of this paper.

REFERENCES

[1] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," *Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI)*, 2005.

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

[3] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," *ACM* SIGIR Forum, vol. 41, no. 1, pp.4-17, 2007.

[4] Y. Xu, K. Wang, G. Yang, and A.W.C Fu, "Online anonymity for personalized web services" *Proc.18th ACM conformation and knowledge management (CIKM),* pp 1497-1500, 2009.

[5] Viejo and J. Castella-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.

[6] Lidan Shou, He Bai, Ke Chen, and Gang Chen "Supporting privacy protection in personalized web search" *IEEE Transactions on knowledge and data engineering,* Vol. 26, No. 2, February 2014.

[7] X. Xiao and Y. Tao, "Personalized Privacy Preservation," *Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD*), 2006.

[8] J. Teevan, S.T. Dumais, and D.J. Liebling, "To Personalize or Not to 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] S.Vanitha "A Personalized Web Search based on user profile and user clicks" International Journal of Latest Research in Science and Technology ISSN (Online):2278-5299 Volume 2, Issue 5: Page No.78-82,September-October 2013