International Journal of Engineering Research and General Science Volume 3, Issue 2, March-April, 2015
ISSN 2091-2730

Personalization of Web Search With Protected Privacy

S.S DIVYA, R.RUBINI,P.EZHIL

Final year, Information Technology,KarpagaVinayaga College Engineering and Technology, Kanchipuram [D.t]

Final year, Information Technology,KarpagaVinayaga College Engineering and Technology, Kanchipuram [D.t]

Assistant Professor, Information Technology,KarpagaVinayaga College Engineeringand Technology, Kanchipuram [D.t]

ABSTRACT

Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting userspecified privacy requirements. Our 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. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework. The experimental results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.

Keywords—Privacy protection, personalized web search, utility, risk, profile,UPS,GreedyDP,GreedyIL.

1.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. 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. 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 [1], it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. 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

2.RELATED WORKS

Z. Dou, R. Song, and J.-R. Wen [1]We present a large- scale evaluation framework for personalized searchbased on query logs and then evaluate five personalized search algorithms (including two click-based ones and three topical-interest-based ones) using 12-day query logs of Windows Live search. By analyzing the results, we reveal that personalized Web search does not work equally well under various situations.

Teevan, S.T. Dumais, and E. Horvitz [2]We formulate and study search algorithms that consider a user's prior interactions with a wide variety of content to personalize that user's current Web search. Rather than relying on the unrealistic assumption that people will precisely specify their intent when searching, we pursue techniques that leverage implicit information about the user's interests. This information is used to re-rank Web search results within a relevance feedback framework. We explore rich models of user interests,

built from both search-related information, such as previously issued queries and previously visited Web pages, and other information about the user such as documents and email the user has read and created.

M. Spertta and S. Gach [3]Explore the use of a less-invasive means of gathering user information for personalized search. In particular, we build user profiles based on activity at the search site itself and study the use of these profiles to provide personalized search results. By implementing a wrapper around the Google search engine, we were able to collect information about individual user search activities.

A. Pretschner and S. Gauch [4]The user profiles are structured as a concept hierarchy of 4,400 nodes. These are populated by `watching over a user's shoulder' while he is surfing. No explicit feedback is necessary. The profiles are shown to converge and to reflect the actual interests quite well. One possible deployment of the profiles is investigated: re-ranking and filtering search results. Increases in performance are moderate but noticeable and show that fully automatic creation of large hierarchical user profiles is possible[X. Shen, B. Tan, and C. Zhai 5]we systematically examine the issue of privacy preservation in personalized search. We distinguish and define four levels of privacy protection, and analyze various software architectures for personalized search. We show 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.

The existing profile-based PWS do not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Such “one profile fits all” strategy certainly has drawbacks given the variety of queries. A better approach is to make an online decision on whether to personalize the query (by exposing the profile) andWhat to expose in the user profile at runtime. Previous works on profile-based PWS mainly focus on improving the search utility and do not take into account the customization of privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. Many personalization techniques require iterative user interactions when creating personalized search results.

4.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. We design two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). 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.

MODULES

4.1 User Profile and Semantic Data Building

4.2 Rdf For User Uploaded Data.

4.3 Search over Indexed Data and Offline Profiling.

4.4 PSWS with UPS Framework.

4.1 User Profile and Semantic Data Building:

Consistent with many previous works in personalized web services, each user profile in UPS adopts a hierarchical structure. Moreover, our profile is constructed based on the availability of a public accessible taxonomy, denoted as R, which satisfies the following assumption. User profile is constructed based on the sample taxonomy repository.

The Resource Description Framework (RDF) is constructed for semantic data on a Relational Database containing Structured as well as unstructured data. A Schema is identified for the relational database and a RDF representing the schema of the database is constructed through model provided by the jena api. The Model contains all the information's about the data linkages in the schema. In this process the schema can also be altered based on admin requirement so that the search process can be effective.

4.2 Rdf for User Uploaded Data.

The RDF is also generated by mining the text contents uploaded by the users in blogs and the contents of the file are analyzed and the meta contents are manipulated. The meta contents are the key for search process so that the file can be rendered on demand. The Text mining process analyses the text word by word and also picks up the literal meaning behind the group of words that constitute the sentence. The Words are analyzed in WordNet api so that the related terms can be found for use in the meta content in generation of RDF. Generally RDF runs in the web services of Servers in all over the world to provide the schematic datas that the server holds in db to the distribution in the web to access it. So the user uploaded content will also be analyzed in realtime servers in their own natural language processing strategies and the results are obtained in a RDF format so that it can be understood by other Servers.

4.3 Search over Indexed Data and Offline Profiling

Similar data's are grouped together that relate to the same resource. The data level process is subjected to structure level processing by indexing the semantic data elements. Multiple RDFs are grouped and structured together to form a master RDF data that holds all the semantic information's of a Server that support reasoning in any formats of query processing. The Different resources are interlinked with high degree of relational factors by the predicates in the triples. The Query processing is handled directly in the RDF file by iterating the triples forming a discrete relation with the Service query and the URI representing the location of the resource is returned. As this process is handled in webservices in real time servers. Hence the structure-oriented approach to RDF data management where data partitioning and query processing make use of structure patterns generated by the RDF. 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. UPS consists of a nontrusty search engine server and a number of clients. Each client (user) accessing the search service trusts no one but himself/ herself. The key component for privacy protection is an online profiler implemented as a search proxy running on the client machine itself. The proxy maintains both the complete user profile, in a hierarchy of nodes with semantics, and the user-specified (customized) privacy requirements represented as a set of sensitive-nodes. In this section, we present the procedures carried out for each user during two different execution phases, namely the offline and online phases. Generally, the offline phase constructs the original user profile and then performs privacy requirement customization according to user-specified topic sensitivity. The subsequent online phase finds the optimal _-Risk Generalization solution in the search space determined by the customized user profile. Specifically, each user has to undertake the following procedures in our solution:

4.3.1. Offline profile construction,

4.3.2. Privacy requirement customization,

4.3.1. Offline-Profile Construction. The first step of the offline processing is to build the original user profile in a topic hierarchy H that reveals user interests.

4.3.2. Privacy Requirement Customization. This procedure first requests the user to specify a sensitive-node set, and the respective sensitivity value for each topic.

4.4. PSWS with UPS Framework.

The online phase handles queries as follows:

When a user issues a query 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 satisfying the privacy requirements. Subsequently, the query and the generalized user profile are sent together to the PWS server for personalized search. The search results are personalized with the profile and delivered back to the query proxy. Finally, the proxy either presents the raw results to the user, or reranks them with the complete user profile. As the sensitivity values explicitly indicate the user's privacy concerns, the most straightforward privacy preserving method is to remove subtrees rooted at all sensitive-nodes whose sensitivity values are greater than a threshold. Such method is referred to as forbidding.

4.4.1 Online query-topic mapping

4.4.2 Online generalization.

4.4.1. Query-topic Mapping:

The purposes of query-topic mapping are

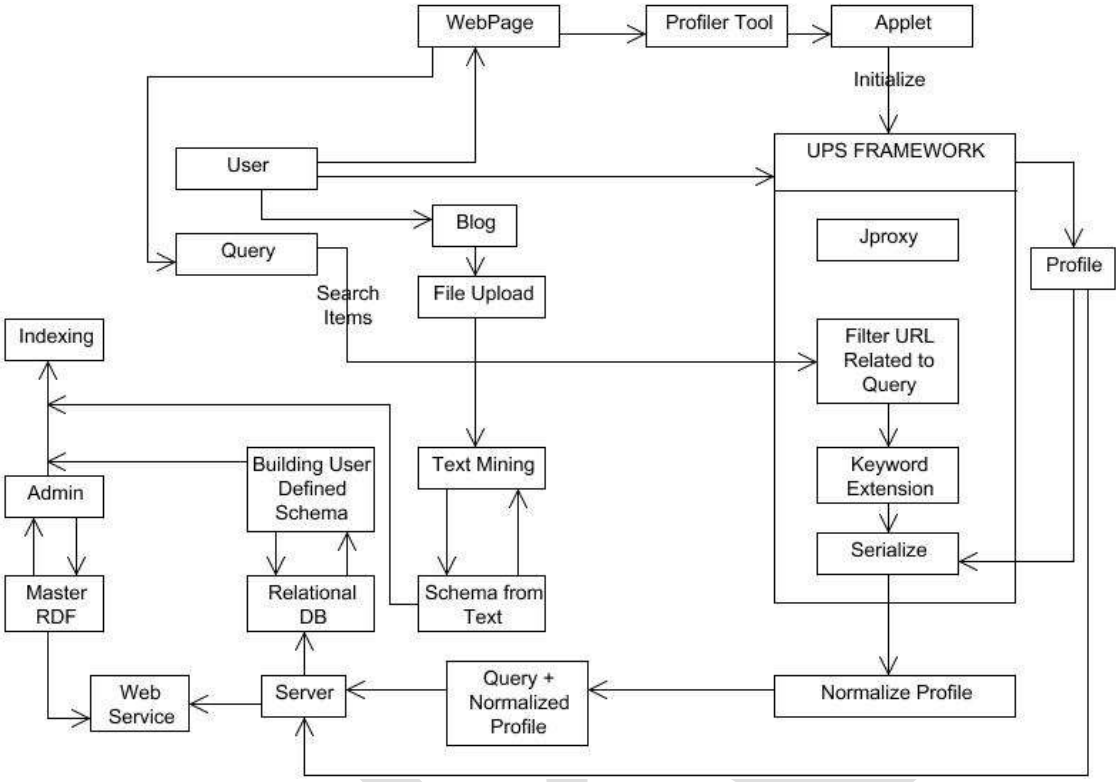
1) To compute a rooted subtree of H, which is called a seed profile, so that all topics relevant to q are contained in it; and

2) To obtain the preference values between q and all topics in H

4.4.2. Profile Generalization:

This procedure generalizes the seed profile G0 in a cost-based iterative manner relying on the privacy and utility metrics. In addition, this procedure computes the discriminating power for online decision on whether personalization should be employed.

5.ARCHITECTURE DIAGRAM



This paper presented a client-side privacy protection framework called UPS forpersonalized web search. UPS could potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The 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. We proposed two greedy algorithms, namely Greedy DP and Greedy IL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution. For future work, we will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentiality, and so on), or capability to capture a series of queries (relaxing the second constraint of the adversary in Section 3.3) from the victim. We will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

6.CONCLUSIONS

This paper presented a client-side privacy protection framework called UPS for personalized web search. UPS could potentially be adopted by any PWS that captures user profiles in a hierarchical taxonomy. The 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. We proposed two greedy algorithms, namely Greedy DP and Greedy IL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results also confirmed the effectiveness and efficiency of our solution. For future work, we will try to resist adversaries with broader background knowledge, such as richer relationship among topics (e.g., exclusiveness, sequentiality, and so on), or capability to capture a series of queries (relaxing the second constraint of the adversary in Section 3.3) from the victim. We will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

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

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