A Novel Framework for User Customizable Privacy Preserving Search

Prof. G. Pradeep, E. Priyanga, M. Durgadevi, T. Kaviya, A. Priyanka

Department of Information Technology,

A.V.C College of Engineering,

Mannampandal.

Abstract :-The objective of the Personalized web search (PWS) is to provide an effective and efficient search results, which are tailor mode for individual user needs. we build user profiles based on user preference and these profiles are then used to re-rank the search results and rank the order of user-examined results.User privacy can be protected without affecting the personalized search quality. However, users are troubled, with exposing personal preference information to search engines has become a major limitation for profile based personalized web search.The Privacy-preserving personalized web search framework is called UPS framework which can generalize profiles for each query according to user-specific privacy requirements. .In general, there is a tradeoff between the search quality and the level of privacy protection achieved from generalization. Effective generalization algorithms namely GreedyDP and GreedyIL are used to support the runtime profiling. Experiments are conducted on real web search data show that the algorithms are effective in enhancing the stability of the search quality and avoids the unnecessary exposure of the user profile.

Keywords:Privacy,Rerank,Generalization,Profile

_____*****_____

I. 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 the failure when search engines return irrelevant results that does 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. The existing information retrieval systems are mostly keyword-based and retrieve the relevant documents or information by matching keywords. Personalized web search (PWS) is a general category of search techniques providing an effective and efficient search results, which are tailore mode for individual user needs. As the check, user information has to be collected and analyze 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 users, which provides 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 [1]. Although there are pros and cons for both types of PWS techniques, the profile-based PWS has demonstrated more effectiveness in improving the quality of web search recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history [2], [3], [4], data, for instance the AOL query logs scandal not only raise panic among individual users, but also dampen the data-publisher's enthusiasm in offering personalized service. In fact, privacy concerns have become the major barrier for browsing history [5], click-through data[1] bookmarks, user documents [2], and so forth. Unfortunately, such implicitly collected personal data can easily reveal a gamut of user's private life. Privacy issues rising from the lack of protection for such wide proliferation of PWS services. Here, we also provide the privacy for the users. Our work aims at providing protection against a typical model of privacy attack, namely eavesdropping.

II. RELATEDWORKS

In this section, the overview of accompanying works are discussed. We focus on the abstract of profile-based personalization and privacy protection in PWS system.

Z.Dou et al(2007) described that the majority of queries searched through search engines are short and ambiguous, and different users may have completely different information needs and goals under the same query. For example, a biologist may use query "mouse" to get information about rodents, while programmers may use the same query to find information about computer peripherals. When such a query is submitted to a search engine, it takes a moment for a user to choose which information he/she wishes to get. Personalized search is considered a solution to this problem since different search results based on preferences of users are provided. They developed a large-scale personalized search evaluation framework based on query logs.Further Z.Dou et al proposed two click-based personalized search strategies and three profile-based personalized search strategies. Personalization has different effectiveness on different queries, users, and search contexts. Finally they proved that click-based personalization strategies perform consistently and considerably well though they can only work on the repeated queries.The drawbacks of the work is personalization which may lack effectiveness

1517

on some queries, and there is no need for personalization on such queries.Different strategies may have variant effects on different queries.

Tan et al(2006) communicated that the existing retrieval systems, including the web search engines, suffer from the problem of "one size fits all": the decision of which documents to return is made based only on the query, without consideration of a particular user's preferences and search context. When a query (e.g., "python") is ambiguous, the search results are inevitably mixed in content (e.g., containing documents on the snake and on the programming language), which is certainly non optimal for the user, who is burdened by the need to sift through the mixed results. There are a wide variety of search contexts, from the user's background and interests, personal document collection (e.g., emails and saved web pages), to what activities the user is doing before submitting the query (e.g., reading an article on wildlife). Further the authors focused on the user's search history, which is often kept in log format and records what queries the user made in the past and what results he/she chose to view. Search history can be divided into short-term and long-term types. Short-term search history is limited to a single search session, which contains a (normally consecutive) sequence of searches with a coherent information need and usually spans a short period of time. Long-term search history is, in contrast, unlimited in time scope and may include all search activities in the past.Their work systematically studied that how to exploit a user's long-term search history to improve retrieval accuracy. Further they proposed mixture models to represent a user's information need and apply statistical language modeling techniques to discover relevant context from the search history, and exploit it to obtain improved estimates of the query model.Tan et al found that mined search history information, can substantially improve retrieval performance for both recurring and fresh queries, and works best when click through data is used with a discriminative weighting scheme for past searches. The issue related to the work shows that the web search engines, suffer from the problem of "one size fits all" the decision of which documents to return is made based only on the query, without consideration of a particular user's preferences and search context.When a query is ambiguous, the search results are inevitably mixed in content.

K.Sugiyama et al(2004) defined that the web search engines help users find useful information on the WWW. However, when the same query is submitted by different users, most search engines return the same results regardless of who submits the query. In general, each user has different information needs for his/her query. For example, for the query "Java," some users may be interested in documents dealing with the programming language, "Java," while other users may want documents related to "coffee." Therefore, Web search results should adapt to users with different information needs. Three types of Web search systems provide such information: (1) systems using relevance feedback, (2) systems in which users register their interest or demographic information, and (3) systems that recommend

information based on users' ratings. In these systems, users have to register personal information such as their interests, age, and so on, beforehand, or users have to provide feedback on relevant or irrelevant judgements with scaling ranges from 1(very bad) to 5 (very good), and so on.Further the authors proposed several approaches that can be used to adapt search results according to user's information need.Similarly compare the retrieval accuracy of the proposed approaches. Compared with our prior works, we scrutinize user's browsing history in a day closely and it allows each user to perform more fine-grained search by capturing changes of each user's preferences without any user effort. Such a method is not performed in typical search engines. Several experiments were conducted in order to verify the effectiveness of the approaches: (1) relevance feedback and implicit approaches, (2) user profiles based on purebrowsing history, and (3) user profiles based on the modified collaborative filtering.The issue related to the work was each user has different information needs for his/her query. Therefore, the search results should be adapted to users with different information needs.The discovery of patterns from usage data by itself is not sufficient for performing the personalization tasks.

X.Shen et al(2005) explained the major deficiency of existing retrieval systems is that they generally lack user modeling and are not adaptive to individual users. This inherent non-optimality is seen clearly in the following two cases: (1) Different users may use exactly the same query (e.g., "Java") to search for different information (e.g., the Java island in Indonesia or the Java programming language), but existing IR systems return the same results for these users. Without considering the actual user, it is impossible to know which sense "Java" refers to in a query. (2) A user's information needs may change over time. The same user may use "Java" sometimes to mean the Java island in Indonesia and some other times to mean the programming language. Without recognizing the search context, it would be again impossible to recognize the correct sense.Any of the following immediate feedback information about the user could potentially help determine the intended meaning of "Java" in the query: (1) The previous query submitted by the user is "hash table" (as opposed to, e.g., "travel Indonesia"). (2) In the search results, the user viewed a page where words such as "programming", "software", and "applet" occur many times. Further they proposed specific techniques to capture and exploit two types of implicit feedback information: (1) identifying related immediately preceding query and using the query and the corresponding search results to select appropriate terms to expand the current query, and (2) exploiting the viewed document summaries to immediately rerank any documents that have not yet been seen by the user. Using these techniques, we develop a client-side web search agent UCAIR (User-Centered Adaptive Information Retrieval) on top of a popular search engine (Google).The weakness of the work is 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.A major deficiency of existing retrieval systems is that they generally lack user modeling

and are not adaptive to individual users Y.xu et al(2007) expressed the personalized search is a promising way to improve search quality by customizing search results for people with different information goals. Many recent research efforts have focused on this area. Most of them could be categorized into two general approaches: Re-ranking query results returned by search engines locally using personal information; or sending personal information and queries together to the search engine. A good personalization algorithm relies on rich user profiles and web corpus. This approach has privacy issues on exposing personal information to a public server. It usually requires users to grant the server full access to their personal and behavior information on the Internet.the authors targets at bridging the conflict needs of personalization and privacy protection, and provides a solution where users decide their own privacy settings based on a structured user profile. This benefits the user in the following ways: Offers a scalable way to automatically build a hierarchical user profile on the client side. It's not realistic to require that every user to specify their personal interests explicitly and clearly. Thus, an algorithm is implemented to automatically collect personal information that indicates an implicit goal or intent. The user profile is built hierarchically so that the higher-level interests are more general, and the lower-level interests are more specific. In this approach, a rich pool of profile sources is explored including browsing histories, emails and personal documents. Offers an easy way to protect and measure privacy. With a hierarchical user profile, the exposure of private information is controlled using two parameters. minDetail determines which part of user profile is protected. Interests in the user profile that does not satisfy minDetail are either too specific or uncommon, are considered private and hidden from the server. expRatio measures how much private information is exposed or protected for a specified MinDetail. The difficulty in the work is personal data are mostly unstructured, for which it is hard to measure privacy.Privacy is not absolute.

III. PROPOSEDWORK

Privacy-preserving personalized web search framework is UPS(User customizable Privacy-preserving Search).The framework aims to protecting the privacy in individual user profiles .UPS framework which can generalize profiles for each query according to user-specified privacy requirements.As shown in figure1 the problem of privacy-preserving personalized search as Risk Profile Generalization, with its NP-hardness proved.The framework works in two phases, namely the offline and online phase, for each user.Effective generalization algorithms, to support runtime profiling GreedyDP and GreedyIL.While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). 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.
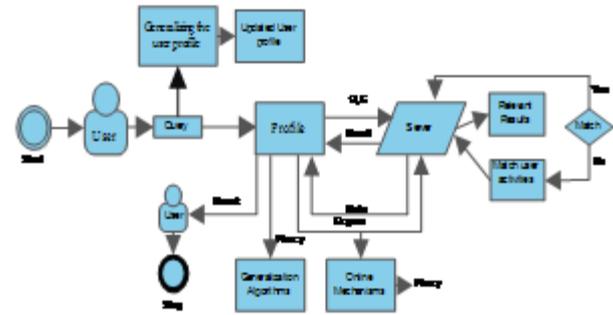

Figure1.Proposed Architecture of UPS

When a client issues an query on the server, the intermediary creates a client profile in runtime in the light of question terms. The output of this step is a generalized client profile Gi fulfilling the security required. The belief methodology is guided by considering two unsure measurements, in particular the personalization utility and the security risk, both characterized for client profiles.Subsequently, the question and the compute up user profile are sent together to the PWS server for customized search.The query items are customized with the profile and conveyed back to the query intermediary.They exhibits either the rough results to the user, or reranks them with the complete client profile.

IV. RESULTS AND DISCUSSION

An extensive experiments illustrate the efficiency and effectiveness of the UPS framework.In addition, UPS performes online generalization on user profiles to assure the personal privacy after compromising the search quality. The proposed greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization.GreedyDP performs,during iteration a best profile so far is maintained satisfying the risk constraints.The iteration stops when the root topic is reached.GreedyIL decrease the computational cost.GreedyIL states to terminate the iteration when risk is satisfied or when there is a single leaf left. The proposed work accomplishes knowledge, such as richer relationship among capacity (e.g., exclusiveness, sequentiality, and so on),or adequacy to better alternation of queries from the target. We will search added developed adjustment to build the user profile, and bigger metrics to indicate the performance (especially the utility) of UPS.

	1.Accuracy	2.Efficiency	3.Privacy
Existing	0	5	45
Proposed	5	50	95

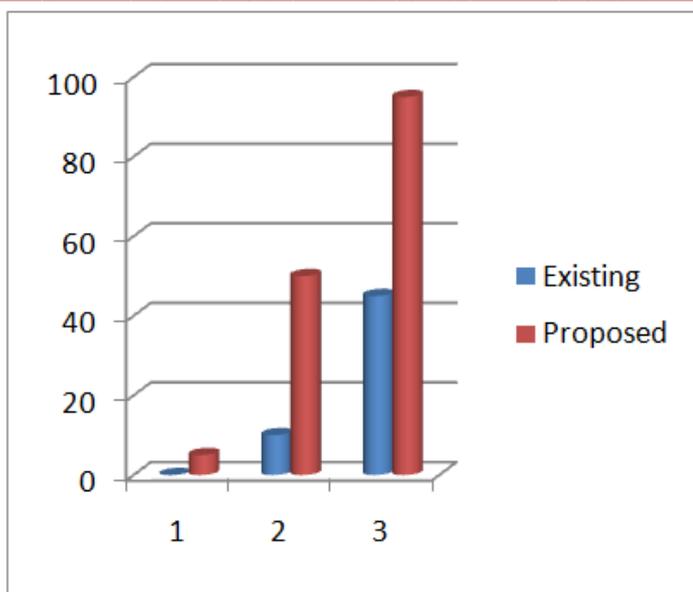Table1 Content of Performance Measures

Figure2.Performance measure of UPS

Figure3.Output of UPS Framework

V.   CONCLUSION

In this paper,we presented a client side protection assurance system called UPS for customized web search. UPS could potentially be received by any PWS that catches client profiles in a hierarchical scientific categorization. The system permitted clients to determine modified protection prerequisites through the various leveled profiles. Also, UPS additionally performed online opinion on client profiles to secure the individual protection without trading off the search quality. The proposed two Greedy algorithm, specifically GreedyDP and GreedyIL, for the online generalization improves the search efficiency the trial results that uncoveres the UPS which accomplishes quality indexed lists while protecting client's security prerequisites. The result confirmes the adequacy and proficiency of the work. In future served attempts to be made to prevent matches with more extensive foundation information, for example, improving relationship among subjects (e.g., exclusiveness,, sequentiality, and so on), or ability to catch a progression of inquiries from the victimized person. Further we look for more modern system to fabricate the client profile, and better measurements to predict the execution (especially the utility) of UPS.

REFERENCE

[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] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without any Effort from Users," Proc. 13th Int'l Conf. World Wide Web (WWW), 2004.

[3] X. Shen, B. Tan, and C. Zhai, "Implicit User Modeling for Personalized Search," Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM), 2005.

[4] Tan, X. Shen, and C. Zhai, "Mining Long-Term Search History to Improve Search Accuracy," Proc. ACM SIGKDD Int'l Conf. Knowledge discovery and data mining(kdd),2006

[5] 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.

AUTHOR(S) BIOGRAPHY

[1] **Pradeep Gurunathan** received his Master Degree in 1998 and obtained Master of Technology in Information Technology in Manonmaniam Sundaranar University in 2004. Currently he is working as a professor in the Department of Information Technology, A.V.C College of Engineering, Mannampandal, Mayiladuthurai, Tamilnadu, India. His Area of interest includes web services, service oriented architecture and internet technologies.

[2] **E.Priyanga** doing her bachelor degree in information technology at AVC College of Engineering, Mannampandal, Mayiladuthurai Tamilnadu, India during the year 2011-2015.

[3] **M.Durgadevi** doing her bachelor degree in information technology at AVC College ofEngineering, Mannampandal, Mayiladuthurai Tamilnadu, India during the year 2011-2015.

[4] **T.Kaviya** doing her bachelor degree in information technology at AVC College of Engineering, Mannampandal, Mayiladuthurai Tamilnadu, India during the year 2011-2015.

[5] **A.Priyanka** doing her bachelor degree in information technology at AVC College of Engineering, Mannampandal, Mayiladuthurai Tamilnadu, India during the year 2011-2015.