ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

PWS Applications that Model User Preferences as Hierarchical User Profiles

K.Laxman
M.Tech Student,
Department of CSE,
Global Institute of Engineering & Technology, Chilkur, RR District, Telangana.

Mr.Syed.Mazharuddin
Associate professor,
Department of CSE,
Global Institute of Engineering & Technology, Chilkur, RR District, Telangana.

Mrs. M.Jhansi Lakshmi
Associate professor & HOD,
Department of CSE,
Global Institute of Engineering & Technology, Chilkur, RR District, Telangana.

ABSTRACT:

Personalized web search has denoted its success in improving the grade of different search services on the internet. The proof reveal that user's disinclination to tell their personal information during search has becomes a major barricade for the wide build-up of pws.In this we study private safety in pws applications that representation user desire as hierarchical user profiles. Generalize profile by queries while reference user specified a private requirement using a pws framework ups. Two predictive metrics utility of personalization and the privacy risk are used for build – up of 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.

Index Terms: Privacy protection, personalized web search, utility, risk, profile.

INTRODUCTION:

The web search engine is widely used by the users for searching useful information on the web. But the amount of information on the web grows continuously so it becomes very difficult for web search engines to find information that satisfies user's individual needs. Due to the enormous variety of user's contexts and backgrounds, as well as the ambiguity of texts, search engines return irrelevant results that do not meet the users real intentions. For providing better search results a general category of search techniques, personalized web search (PWS) is used. To figure out the user intention behind the issued query, user information has to be collected and analyzed.

There are two types of solutions to the PWS :

1) Click-log-based method: This is a straightforward method. The click-log based methods uses clicked pages in the users query history. But it has strong limitation that it can only work on repeated queries from the same user .

2) Profile-based methods: Profile-based methods can be used effectively for almost all sorts of queries, but under some circumstances the results are unstable. It improves the search experience with complicated user-interest models generated from user profiling techniques.

There are pros and cons for both types of PWS techniques, but 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. It is usually gathered implicitly from query history, browsing history, click-through data, bookmarks, user documents, and so forth. Unfortunately, such implicitly collected personal data can easily disclose a span of user's private life. Privacy issues are raised from the lack of protection for such data, for instance the AOL query logs scandal , raise panic among individual users, and also dampen the data publishers enthusiasm in offering personalized service. So the privacy concerns have become the major barrier for wide proliferation of PWS services. Existing system have a privacy-preserving personalized web search framework UPS.

User specifies the privacy requirements and according to the requirements user profiles are generalized. The problem of privacy-preserving personalized search is formulated as δ-Risk Profile Generalization, by using two conflicting metrics, personalization utility and privacy risk, for hierarchical user profile. Two simple and effective generalization algorithms, GreedyDP and GreedyIL are developed, which support runtime profiling.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

GreedyDP tries to maximize the discriminating power (DP), and the GreedyIL attempts to minimize the information loss (IL). To enhance the stability of the search results and to avoid the unnecessary exposure of the profile an inexpensive mechanism is used for deciding whether to personalize a query in UPS. UPS allows customization of privacy needs; and it does not require iterative user interaction.To protect user privacy in profile-based PWS, researchers 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. They need to hide the privacy contents existing in the user profile to place the privacy risk under control. Significant gain can be obtained by personalization at the expense of only a small (and less-sensitive) portion of the user profile, namely a generalized 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.

LITERATURE REVIEW:

In this paper, author study this problem and provide some preliminary conclusions. It presents a largescale evaluation framework for personalized search based on query logs and then evaluates with the click and profile based strategies. By analyzing the results, author reveals that personalized search has significant improvement over common web search on some queries but it has little effect on other queries. Author also reveals that both long term and short-term contexts are very important in improving search performance for profile-based personalized search strategies. In this paper, author tries to investigate whether personalization is consistently effective under different situations. The profile-based personalized search strategies proposed in this paper are not as stable as the click-based ones. They could improve the search accuracy on some queries, but they also harm many queries. Since these strategies are far from optimal, author will continue his work to improve them in future. It also finds for profile-based methods, both long-term and short-term contexts are important in improving search performance. The appropriate combination of them can be more reliable than solely using either of them. From the author, they studied how to exploit implicit user modeling to intelligently personalize information retrieval and improve search accuracy.

Unlike most previous work, it emphasizes the use of immediate search context and implicit feedback information as well as eager updating of search results to maximally benefit a user. Author presented a decision-theoretic framework for optimizing interactive information retrieval based on eager user model updating, in which the system responds to every action of the user by choosing a system action to optimize a utility function. Author propose] 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 exploiting the viewed document summaries to immediately re-rank any documents that have not yet been seen by the user. Using these techniques, author develops a client side web search agent (UCAIR) on top of a popular search engine (Google) without any additional effort from the user. From the author have explored how to exploit implicit feedback information, including query history and click-through history within the same search session, to improve information retrieval performance.

Using the KLdivergence retrieval model as the basis, author proposed and studied four statistical language models for contextsensitive information retrieval, i.e., FixInt, BayesInt, OnlineUp and BatchUp. It uses TREC AP Data to create a test set for evaluating implicit feedback models. The current work can be extended in several ways: First, it has only explored some very simple language models for incorporating implicit feedback information. It would be interesting to develop more sophisticated models to better exploit query history and click through history. For example, this may treat a clicked summary differently depending on whether the current query is a generalization or refinement of the previous query. Second, the proposed models can be implemented in any practical systems. It currently develops a client-side personalized search agent, which will incorporate some of the proposed algorithms. Author will also do a user study to evaluate effectiveness of these models in the real web search. Finally, author should further study a general retrieval framework for sequential decision making in interactive information retrieval and study how to optimize some of the parameters in the context-sensitive retrieval models. This paper was motivated by two emerging trends: web users want personalized services and web users want privacy. One challenge is that personal information must be made anonymous under the assumption that the participating parties,

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

including the web service, are not completely trusted, due to systematic collection of personal information in addition to queries. Another challenge is the online and dynamic nature of web users. Author proposed the notion of online anonymity to protect web users and proposed an approach to maintain online anonymity through time. This approach makes use of a third party called the user pool and it does not require the user pool to be trusted. The simulation study on real US demographics showed promising results: it is feasible to achieve personalization for reasonable privacy settings. From this approach they requires users to contribution the server full access to personal information on Internet, which break users' privacy.

In this paper, author inspects the possibility of accomplish a balance between users' privacy and search quality. First, an algorithm is provided to the user for collecting, abbreviation, and organizing their personal information into a hierarchical user profile, where general terms are ranked to higher levels than explicit terms. Through this profile, users control what section of their private information is uncovered to the server by adjusting the minDetail threshold. An additional privacy measure, expRatio, is proposed to approximation the amount of privacy is exposed with the specified minDetail value. Yet, this paper is an exploratory work on the two features: First, author deal with unstructured data such as personal documents, for which it is still an open problem on how to define privacy. Secondly, author try to bridge the conflict needs of personalization and privacy protection by breaking the premise on privacy as an absolute standard. Also, author believe that an enhanced balance between privacy protection and search quality can be achieved if web search are personalized by allowing for only revealing those information associated to a specific query. It performs less protection for the user data and they were no assured for the user data and their profile information's.

In this paper the author studied the existing generalization methods are insufficient because they cannot assurance privacy protection in all cases, and frequently acquire redundant information loss by performing too much generalization. In this paper, author proposes the idea of personalized secrecy, and develops a new generalization structure that takes into account customized privacy necessities. This technique successfully avoid privacy intrusion even in scenarios where the existing approaches fail, and results in generalized tables that permit accurate aggregate analysis.

This work lays down a solid theoretical foundation for developing substitute generalization strategies. For instance, the greedy algorithm presented in this paper is not optimal, in the sense that it does not necessarily achieve the lowest information loss. Discovering the optimal solution is a demanding problem. As another example, in performance, the recipients of the published data are often specialized users (e.g. scientists), who may explicitly specify the analytical tasks (such as association rule mining) required. This information may be utilized to free a table that is highly efficient for those tasks, without breaching the privacy constraints formulated by data owners.

EXISTING SYSTEM:

Profile based PWS:

» 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.
» Profile-based personalization may not even help to improve the search quality for some ad hoc queries, though exposing user profile to a server has put the user's privacy at risk.
» A better approach is to make an online decision on whether to personalize the query and what to expose in the user profile at runtime.

Customization of privacy requirements:

» This considers, all the sensitive topics are detected using an absolute metric called surprisal based on the information theory, assuming that the interests with less user document support are more sensitive.

Iterative user interactions:

» They usually refine the search results with some metrics which require multiple user interactions, such as rank scoring, average rank, and so on.
» This paradigm is, however, infeasible for runtime profiling, as it will not only pose too much risk of privacy breach, but also demand prohibitive processing time for profiling.
» Thus, we need predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Disadvantages:

» The existing profile-based PWS do not support runtime profiling.
» The existing methods do not take into account the customization of privacy requirements.
» Many personalization techniques require iterative user interactions when creating personalized search results.

PROPOSED SYSTEM:

» To propose UPS (User customizable Privacy-preserving Search) framework, which is a privacy-preserving personalized web search framework, which can generalize profiles for each query according to user-specified privacy requirements.
» To develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. GreedyDP tries to maximize the discriminating power (DP), GreedyIL attempts to minimize the information loss (IL).
» The framework assumes that the queries do not contain any sensitive information, and aims at protecting the privacy in individual user profiles while retaining their usefulness for PWS.
» 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.
» During the offline phase, a hierarchical user profile is constructed and customized with the user-specified privacy requirements.
» The online phase handles queries as When a user issues a query qi 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 Gi satisfying the privacy requirements. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles.
» 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.
»
»

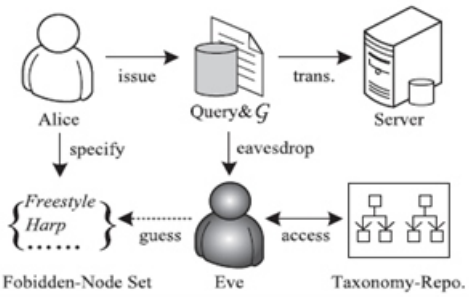

Fig. . Attack model of personalized web search.

Advantages:

» UPS provides runtime profiling, which in effect optimizes the personalization utility while respecting user's privacy requirements;
» Allows for customization of privacy needs; and
» Does not require iterative user interaction.
» Provides an inexpensive mechanism for the client to decide whether to personalize a query in UPS.

IMPLEMENTATION:

1.Profile-Based Personalization:

This paper introduces an approach to personalize digital multimedia content based on user profile information. For this, two main mechanisms were developed: a profile generator that automatically creates user profiles representing the user preferences, and a content-based recommendation algorithm that estimates the user's interest in unknown content by matching her profile to metadata descriptions of the content. Both features are integrated into a personalization system.

2.Privacy Protection in PWS System:

We propose a PWS framework called UPS that can generalize profiles in for each query according to user-specified privacy requirements. Two predictive metrics are proposed to evaluate the privacy breach risk and the query utility for hierarchical user profile. We develop two simple but effective generalization algorithms for user profiles allowing for query-level customization using our proposed metrics.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

We also provide an online prediction mechanism based on query utility for deciding whether to personalize a query in UPS. Extensive experiments demonstrate the efficiency and effectiveness of our framework.

3. Generalizing User Profile:

The generalization process has to meet specific prerequisites to handle the user profile. This is achieved by preprocessing the user profile. At first, the process initializes the user profile by taking the indicated parent user profile into account. The process adds the inherited properties to the properties of the local user profile. Thereafter the process loads the data for the foreground and the background of the map according to the described selection in the user profile. Additionally, using references enables caching and is helpful when considering an implementation in a production environment. The reference to the user profile can be used as an identifier for already processed user profiles. It allows performing the customization process once, but reusing the result multiple times. However, it has to be made sure, that an update of the user profile is also propagated to the generalization process. This requires specific update strategies, which check after a specific timeout or a specific event, if the user profile has not changed yet. Additionally, as the generalization process involves remote data services, which might be updated frequently, the cached generalization results might become outdated. Thus selecting a specific caching strategy requires careful analysis.

4.Online Decision:

The profile-based personalization contributes little or even reduces the search quality, while exposing the profile to a server would for sure risk the user's privacy. To address this problem, we develop an online mechanism to decide whether to personalize a query. The basic idea is straightforward. if a distinct query is identified during generalization, the entire runtime profiling will be aborted and the query will be sent to the server without a user profile.

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 GreedyDP and GreedyIL, 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.

FUTURE WORK:

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) 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] B. 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] 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.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[6] 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.

[7] X. Shen, B. Tan, and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.

[8] F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.

[9] 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.

[10] Y. Xu, K. Wang, B. Zhang, and Z. Chen, "Privacy-Enhancing Personalized Web Search," Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 591-600, 2007.

[11] K. Hafner, Researchers Yearn to Use AOL Logs, but They Hesitate, New York Times, Aug. 2006.

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

[13] J.S. Breese, D. Heckerman, and C.M. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI), pp. 43-52, 1998.

[14] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschu¨ tter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), 2005.

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

[16] E. Gabrilovich and S. Markovich, "Overcoming the Brittleness Bottleneck Using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge," Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI), 2006.

[17] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia," HP Labs, 2008.

[18] K. Ja¨rvelin and J. Keka¨la¨inen, "IR Evaluation Methods for Retrieving Highly Relevant Documents," Proc. 23rd Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR), pp. 41-48, 2000.

[19] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley Longman, 1999.

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

AUTHOR DETAILS:

Author 1:

K.Laxman,M.Tech,Department of CSE,Global Institute of Engineering and Technology ,Chilkur,RR District,Telangana

Author 2:

Mr.Syed.Mazharuddin,Associate professor, Department of CSE,Global Institute of Engineering and Technology
Chilkur,RR District,Telangana

Author 3:

Mrs. M.Jhansi Lakshmi,Associate professor,HOD of CSE,Global Institute of Engineering and Technology,Chilkur,RRDistrict,Telangana