Supporting Privacy Protection in Personalized Web Search

A.Arun[1] S.Suma[2]
[1]PG Scholar [2]Assistant Professor
[1,2]Department of Computer Science and Engineering
[1,2]Valliammai Engineering College, Chennai, India

Abstract— Personalized web search (PWS) has demonstrated that it is effective in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to discover their private information during search has become a major revetment for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as stratified user profiles. We propose a PWS framework called UPS that can adaptative generalize profiles by queries while respecting user-specified privacy requirements. Our runtime generalization aims at striking a balance between two prophetical metrics that valuate the utility of reification and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime induction. We also provide an online prevision mechanism for deciding whether personalizing a query is beneficial. Extensive experiments demonstrate the effectiveness of our framework. The simulation results also reveal that GreedyIL significantly outperforms GreedyDP in terms of efficiency.

Key words: Privacy Protection, Web Search

I. INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks.

A. Marketing / Retail

Data mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaigns such as direct mail, online marketing campaign etc. Through the results, marketers will have appropriate approach to sell profitable products to targeted customers. Data mining brings a lot of benefits to retail companies in the same way as marketing. Through market basket analysis, a store can have an appropriate production arrangement in a way that customers can buy frequent buying products together with pleasant. In addition, it also helps the retail companies offer certain discounts for particular products that will attract more customers.

B. Finance / Banking

Data mining gives financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner.

C. Manufacturing

By applying data mining in operational engineering data, manufacturers can detect faulty equipments and determine optimal control parameters. For example semi-conductor manufacturers has a challenge that even the conditions of manufacturing environments at different wafer production plants are similar, the quality of wafer are lot the same and some for unknown reasons even has defects. Data mining has been applying to determine the ranges of control parameters that lead to the production of golden wafer. Then those optimal control parameters are used to manufacture wafers with desired quality.

D. Governments

\Data mining helps government agency by digging and analyzing records of financial transaction to build patterns that can detect money laundering or criminal activities.

E. Law Enforcement

Data mining can aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviors.

- Large quantities of data: The volume of data so great it has to be analyzed by automated techniques
- Noisy, incomplete data: Imprecise data is the characteristic of all data collection.
- Complex data structure: conventional statistical analysis not possible
- Heterogeneous data stored in legacy systems.

II. RELATED WORK

Although personalized search has been proposed for many years and many personalization strategies have been investigated, it is still unclear whether personalization is consistently effective on different queries for different users, and under different search contexts. In this paper, we study this problem and get some preliminary conclusions. We present a large-scale evaluation framework for personalized search based on query logs, and then evaluate five personalized search strategies including two click-based and three profile-based ones using 12-day MSN query logs. By analyzing the results, we reveal that personalized search has significant improvement over common web search on some queries but it also has little effect on other queries. [1] 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. [2] Long-term search history contains rich information about a user's search preferences, which can be used as search context to improve retrieval performance. In this paper, we study statistical language modeling based methods to mine contextual information from long-term search history and exploit it for a more accurate estimate of the query language model. Experiments on real web search data show that the algorithms are effective in improving search accuracy for both fresh and recurring queries. The best performance is achieved when using click through data of past searches that are related to the current query. [3] Web search engines help users find useful information on the World Wide Web (WWW). However, when the same query is submitted by different users, typical search engines return the same result regardless of who submitted the query. Generally, each user has different information needs for his/her query. Therefore, the search result should be adapted to users with different information needs. In this paper, we first propose several approaches to adapting search results according to each user's need for relevant information without any user effort, and then verify the effectiveness of our proposed approaches. [4] One hundred users, one hundred needs. As more and more topics are being discussed on the web and our vocabulary remains relatively stable, it is increasingly difficult to let the search engine know what we want. Coping with ambiguous queries has long been an important part of the research on Information Retrieval, but still remains a challenging task. *Personalized search* has recently got significant attention in addressing this challenge in the web search community, based on the premise that a user's general preference may help the search engine disambiguate the true intention of a query. [5]

III. SYSTEM ARCHITECTURE

A. *Main Use of System:*

Increasing usage of personal and activity information to profile its users, this is usually collected implicitly from query history, browsing history, click-through data bookmarks, user record, and so forth.

The framework permit users to qualify customized privacy demand via the hierarchical profiles. In addition, UPS also performed online generalization on user profiles to preserve the individualized privacy without compromising the search quality

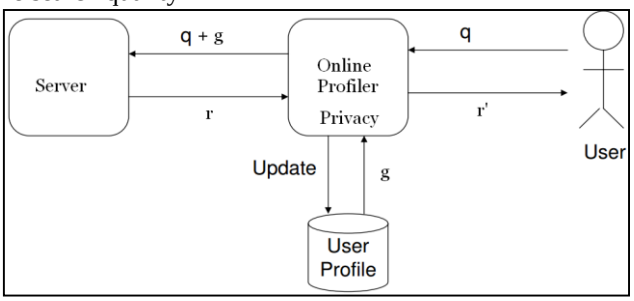

Fig. 1: Architecture of system

The discus about Figure: 1.1 proposes 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 inconsistent metrics, namely reification utility and privacy risk, for hierarchical user profile. We formulate the problem of privacy preserving personalized search as attempt Profile induction, with its N P-hardness proved. We develop two simple but effective generalization algorithms, GreedyDP and GreedyIL, to support runtime profiling. While the former tries to maximize the diacritic power DP, the latter stirring to minimize the information loss IL. By exploiting a number of heuristics, GreedyIL out performs GreedyDP probative. We provide an affordable 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 required exposure of the profile. Our extensive condition demonstrates the efficiency and effectiveness of our UPS framework..

IV. SYSTEM ANALYSIS

1) Profile-Based Personalization.
2) Privacy Protection in PWS System.
3) Generalizing User Profile.
4) Online Decision.

A. *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 apparatus that automatically creates user profiles correspond the user preferences, and a content based recommendation algorithm that estimates the user's interest in unknown content by matching her profile to metadata depiction of the assemblage. Both features are integrated into a personalization system.

B. *Privacy Protection in PWS System*

We propose a PWS framework called UPS that can conclude profiles in for each query according to user specific privacy requirements. Two predictive metrics are proposed to measure the privacy breach risk and the query utility for gradable user profile. We develop two simple but efficient generalization algorithms for user profiles allowing for query level customization using our proposed metrics. We also provide an online anticipation mechanism based on query utility for decisive whether to personalize a query in UPS. Extensive experiments demonstrate the efficiency and effectiveness of our framework.

C. *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 processes separate the user profile by taking the auspicate parent user profile into account. The process adds the inherited edibility to the properties of the local user profile. Thereafter the computing loads the data for the aspect and the background of the map according to the described selection in the user profile.

Additionally, using references enables caching and is helpful when expect an implementation in a devising

environment. The reference to the user profile can be used as an identifier for already processed user profiles. It allows personating the customization process once, but reusing the result doubled 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 scheme, which check after a specialized timeout or a specific event, if the user profile has not changed yet. Additionally, as the generalization process regard remote data services, which might be updated of times, the cached generalization results might become outdated. Thus selecting a precipice caching strategy requires careful analysis.

D. Online Decision

The profile-based personalizations modify little or even reduce the search quality, while existent 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 individualize a query. The basic idea is unequivocal. If a distinct query is identified during generalization, the entire unraced profiling will be decease and the query will be sent to the server without a user profile.

V. RESULT

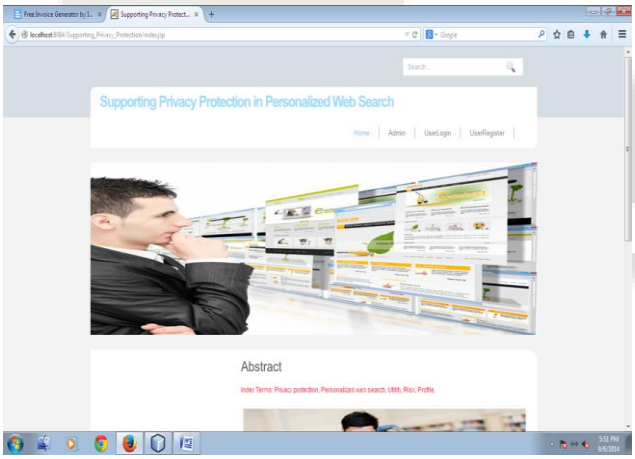

Fig. 2: Home Pages

The discuss about Figure: 1.2 Home pages for supporting privacy in personalized web search. It placed in Admin and your logging, user Registration.

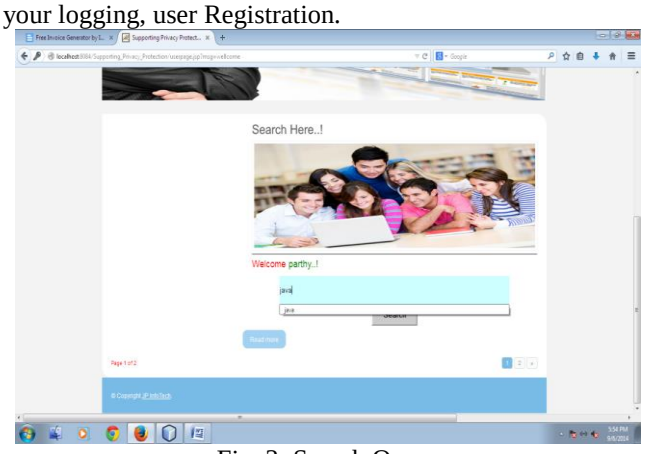

Fig. 3: Search Query

The discuss about Figure: 1.3 Search Query for particular user search in web page. It show the different content based on new words.

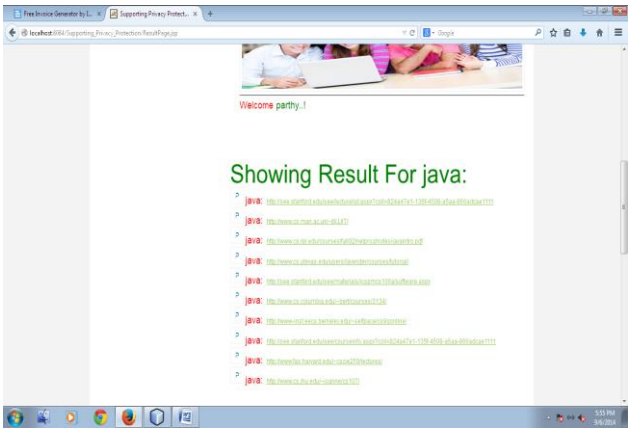

Fig. 4: Search Result

The discuss about Figure: 1.4 Search Result for user search in web page. Different words search in web page and showing result for exact result.

VI. CONCLUSION

This paper presented a client-side privacy insulation framework called UPS for individualized web search. UPS could potentially be adopted by any PWS that represent user profiles in a hierarchical taxonomy. The framework allowed users to specify construct privacy obligation via the hierarchical profiles. In addition, UPS also improvise online generalization on user profiles to preserve the personal privacy without conciliatory the search quality. We suggest two greedy algorithms, namely GreedyDP and GreedyIL, for the online generalization. Our experimental results expose that UPS could achieve quality search results while prolong user's customized privacy involve. The results also confirmed the effectiveness and efficiency of our solution.

VII. FUTURE WORK

We will try to resist individual with broader background knowledge, such as richer partnership among topics e.g., exclusiveness, sequentially, and so on, or capability to capture a series of queries from the victim. We will also seek more temporal 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.
[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] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia," HP Labs, 2008.
[14] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison Wesley Longman, 1999.
[15] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.

