INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

BEST KEYWORD COVER SEARCH

CH DAKSHAYANI[1], N.NAVEEN[2]

[1]M.Tech Student, Sree Rama institute of technology and science Kuppenakuntla,Penuballi,Khammam,TS INDIA

[2]Asst Prof,CSE DeptSree Rama institute of technology and science Kuppenakuntla,Penuballi,Khammam,TS INDIA

ABSTRACT:

It is common that the objects in a spatial database (e.g., restaurants/hotels) are associated with keyword(s) to indicate their businesses/services/features. An interesting problem known as Closest Keywords search is to query objects, called keyword cover, which together cover a set of query keywords and have the minimum inter-objects distance. In recent years, we observe the increasing availability and importance of keyword rating in object evaluation for the better decision making. This motivates us to investigate a generic version of Closest Keywords search called Best Keyword Cover which considers inter-objects distance as well as the keyword rating of objects. The baseline algorithm is inspired by the methods of Closest Keywords search which is based on exhaustively combining objects from different query keywords to generate candidate keyword covers. When the number of query keywords increases, the performance of the baseline algorithm drops dramatically as a result of massive candidate keyword covers generated. To attack this

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924

drawback, this work proposes a much more scalable algorithm called keyword nearest neighbor expansion (keyword-NNE). Compared to the baseline algorithm, keyword-NNE algorithm significantly reduces the number of candidate keyword covers

Index Terms— Spatial database, point of interests, keywords, keyword rating, keyword cover

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

ISSN : 2454-9924

methods can be potentially effective for almost all sorts of queries, but are Reported to be unstable under some Circumstances.

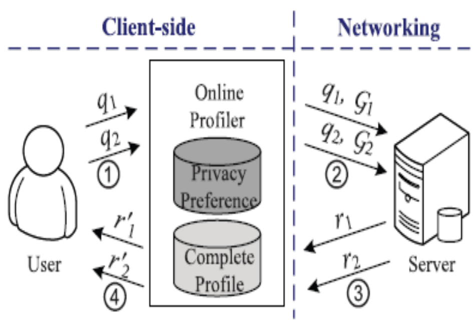


Fig. 1. System architecture of UPS.

Existing System:

In this section, we overview the related works. We focus on the literature of profile-based personalization and privacy protection in PWS system Previous works on profile-based PWS mainly focus on improving the search utility. The basic idea of these works is to tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal. In the remainder of this section, we review the previous solutions to PWS on two aspects, namely the representation of profiles, and the measure of the effectiveness of personalization The solutions in class two do not require third-party assistance or collaborations between social network entries. In these solutions, users only trust themselves and cannot tolerate the exposure of their complete profiles an anonymity server. In [12], Krause and Horvitz 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. Xu et al. [10] proposed a privacy protection solution for PWS based on

ISSN : 2454-9924

hierarchical profiles. Using a user-specified threshold, a generalized profile is obtained in effect as a rooted subtree of the complete profile. Unfortunately, this work does not address the query utility, which is crucial for the service quality of PWS. For comparison, our approach takes both the privacy requirement and the query utility into account.

Proposed System:

In this section, we first introduce the structure of user profile in UPS. Then, we define the customized privacy requirements on a user profile. Finally, we present the attack model and formulate the problem of privacy preserving profile generalization. For ease of presentation, Table 1 summarizes all the symbols used in this paper 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.

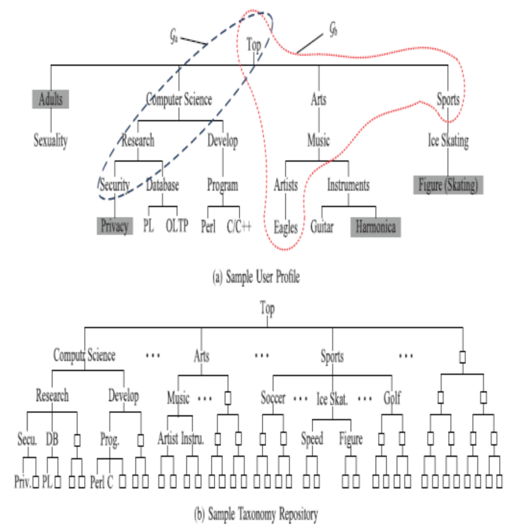


Fig. 2. Taxonomy-based user profile.

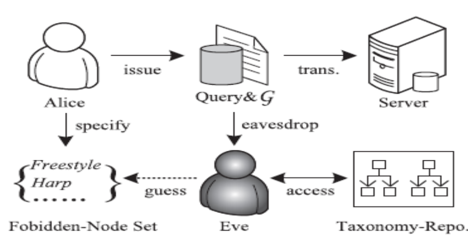


Fig. 3. Attack model of personalized web search.

UPS PROCEDURES

In this section, we present the procedures carried out for each user during two different execution

ISSN : 2454-9924

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.

1. offline profile construction,

2.offline privacy requirement customization,

3. online query-topic mapping, and

4. online generalization.

GENERALIZATION TECHNIQUES

In this section, we first introduce the two critical metrics for our generalization problem. Then, we present our method of online decision on personalization. Finally, we propose the generalization algorithms.

Metric of Utility

The purpose of the utility metric is to predict the search quality (in revealing the user's intention) of the query q on a generalized profile G. The reason for not measuring the search quality directly is because search quality depends largely on the implementation of PWS search engine, which is hard to predict. In addition, it is too expensive to solicit user feedback on search results. Alternatively, we transform the utility prediction problem to the estimation of the discriminating power of a given query q on a profile G under the following assumption.

EXPERIMENTAL RESULTS

In this section, we present the experimental results of UPS. We

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924

conduct four experiments on UPS. In the first experiment, we study the detailed results of the metrics in each iteration of the proposed algorithms. Second, we look at the effectiveness of the proposed query-topic mapping. Third, we study the scalability of the proposed algorithms in terms of response time. In the fourth experiment, we study the effectiveness of clarity prediction and the search quality of UPS.

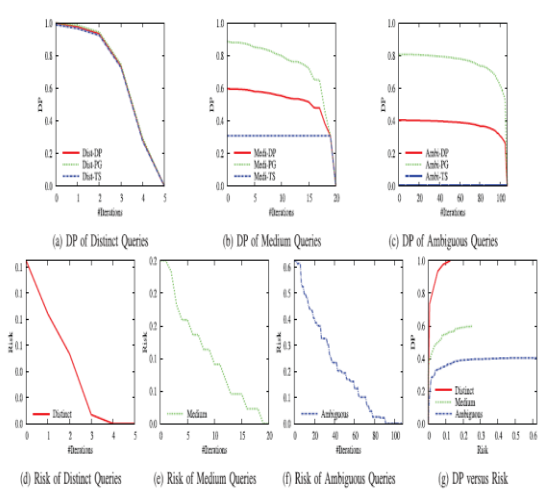


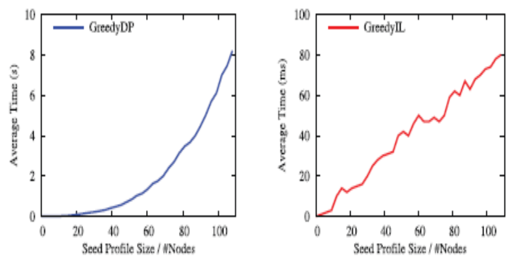
Fig. 5. Results of Distinct/Medium/Ambiguous queries during each iteration in GreedyDP/GreedyIL. All results are obtained from the same profile.

Scalability of Generalization Algorithms

We study the scalability of the proposed algorithms by varying 1) the seed profile size (i.e., number of nodes), and 2) the data set size (i.e., number of queries). For each possible seed profile size (ranging from 1 to 108), we randomly choose 100 queries from the AOL query log, and take their respective RðqÞ as their seed profiles. All leaf nodes in a same seed profile are given equal user preference. These queries are then processed using the GreedyDP and GreedyIL algorithms. For fair comparison, we set the privacy threshold _ ¼ 0 for GreedyIL to make it always run the same number of iterations as GreedyDP does. Fig. 7 shows the average response time of the two algorithms while varying the seed profile size. It can be seen that the cost of GreedyDP grows exponentially, and exceeds 8 seconds when the profile contains more than

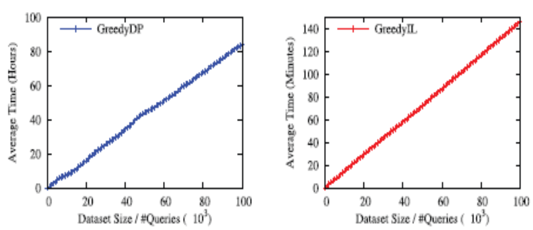
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924

100 nodes. However, GreedyIL displays near-linear scalability, and Significantly outperforms GreedyDP.



(a) Results of GreedyDP (b) Results of GreedyIL

Fig. 7. Scalability by varying profile size.



(a) Results of GreedyDP (b) Results of GreedyIL

Fig. 8. Scalability by varying data set size.

CONCLUSION

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 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

ISSN : 2454-9924

metrics to predict the performance (especially the utility) of UPS.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China (No. 61170034), and China Key Technology R&D Program (No. 2011BAG05B04.

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

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924

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.

ISSN : 2454-9924

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.

CH DAKSHAYANI is an M.Tech Department of Computer Science & Engineering, Sreerama Institute of Technology & science, Penuballi Mandal, Khammam, Kotha Kuppenkuntla

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES
ISSN : 2454-9924

Mr. N.Naveen is an efficient teacher, received M.Tech from JNTU Hyderabad is working as an Assistant Professor in Department of C.S.E, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, AP,India. He has published many papers in both National & International Journals. His area of Interest includes Data Communications & Networks, Database Management Systems, Computer Organization, C Programming and other advances in Computer Applications