International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015
ISSN 2091-2730

Personalizing search based on user search histories

Ayush Patni , Arathy Ajaya Kumar , Mr G Senthil Kumar

SRM University

ayushpatni11@gmail.com , Contact:8939186462

Abstract— Various searches on the internet is done using personalized web search(PWS).But however , failures might be experienced when irrelevant results are returned that do not meet the requirements of the users. Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. However, evidences show that user's confidential information during search has known to be public due to proliferation of PWS. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified privacy requirements.

Keywords-Java Runtime Environment , Personalised Web Search, User Personalised Search, privacy , generalization , Query , User Interface

Introduction

Personalized search[1] refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided. It introduces potential privacy problems in which a user may not be aware that their search results are personalized for them, and wonder why the things that they are interested in have become so relevant.An interesting point about personalization that often gets overlooked is the privacy vs personalization battle. The overall goal of the data mining process[4]is to extract information from a data set and transform it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of lage quantities of data to extract previously unknown interesting patterns such as groups of data records, unusual records dependencies.

Related Work

Context Sensitive Information Retrieval using Implicit feedback [6] [7]by Xuehua Shen & Bin Tan uses click log to improve retrieval accuracy in an interactive information retrieval setting. Personalizing Search via Automated Analysis of Interests and Activities[2] by Jaime Teevan, Susan T. Dumais uses click log method to re-rank the web results. Adaptive Web Search Based on User Profile Constructed without Any Effort from Users[5] by Kazunari Sugiyama, Kenji Hatano uses profile based method for collaborative filtering with detailed analysis of user's browsing history in one day.

Identification of User Interest For Personalized Search[8] by Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou uses click log method to learn a user's preference automatically based on her past click history and how it can use the user preference to personalize search results.

Privacy-Enhancing Personalized Web Search[10] by Yabo Xu & Ruihua Song uses a profile based method to summarize a user's interests into a hierarchical organization[12]according to specific interests. Two parameters for specifying privacy requirements are proposed to help the user to choose the content and degree of detail of the profile information that is exposed to the search engine

The generalization algorithm - GreedyIL algorithm

The GreedyIL algorithm improves the efficiency of the generalization using heuristics based on several findings. One important finding is that any prune-leaf operation reduces the discriminating power of the profile. In other words, the DP displays monotonicity by prune-leaf. we construct user profiles base on movement at the search site itself and study the decide user's interests and biases on different categories, which can then be used for personalization. Web search results should adapt to users with different data wants. In order to predict such information wants, there are numerous methods relate data mining techniques[4] to extract usage patterns from Web logs. In this paper propose a privacy-preserving personalized[9] web search structure UPS, which can simplify profiles for every query according to user-specified privacy requirements.

Existing methods

In our Existing System, Personalized web search (PWS)[3]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 query. But however,the disadvantages with existing system is that it only works on repeated queries from the same user and no customization is allowed.

The two techniques used in the existing system are:

a) Click log method

Click-log based simply impose bias to clicked pages in the user's query. It directly measures the effectiveness of hyperlinks in reallife user interactions. This involves first recording data from users and then interpreting that data. The raw data of a click log consists of a sequence of page-views (with timestamps) for each user. No detailed information is recorded about users' interaction with their browsers, such as the specific anchors that are clicked to generate each page-view

b) Profile based method

Profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. It tailors tailor the search results by referring to, often implicitly, a user profile that reveals an individual information goal.

Proposed system

A PWS framework called UPS (User customizable Privacy-preserving Search) is introduced so that it can adaptively generalize profiles by queries while respecting user specified privacy requirements. The key component for privacy protection is an online profile implemented as a search proxy running on the client machine itself. We present a greedy algorithm, namely Greedy IL, for runtime generalization.

PROPOSED SYSTEM ADVANTAGES

- Works on different types of queries from user.
- Customization of privacy requirements.

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. The generalization process is guided by considering two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles were administrator maintains all files and responsible for storing that files into cloud. User given query and the generalized user profile are sent together to the PWS server for personalized search. Query with related user preferences stored in a user profile with the aim of providing better search results. User given query based on privacy requirements and cost of profiling search results are checked whether to personalize or not.

EXPERIMENTAL SETUP

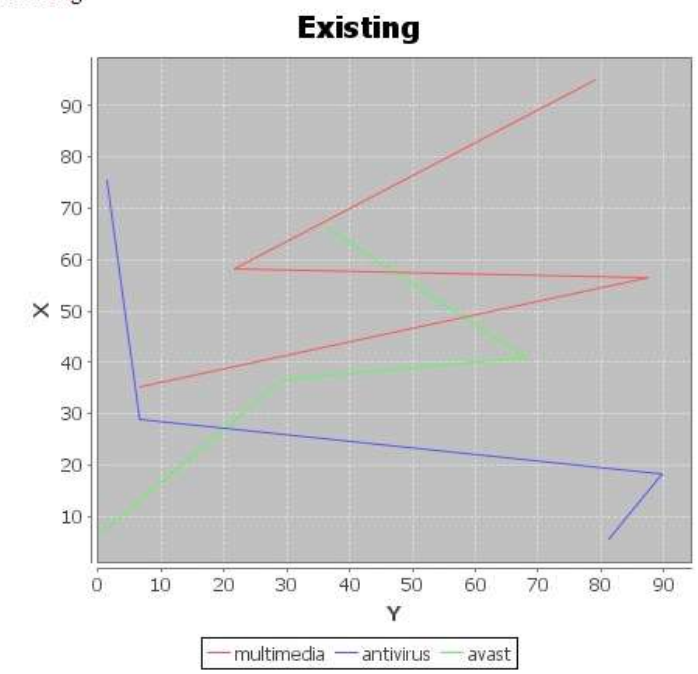
The implementation of our system is based on the Java compiler and analysis framework. In our system we use a translator from PQL to Data log and the database program analysis tool [WL04] to find security violations. We applied static analysis to look for all tainted object propagation problems described in this report, and we used a total of 28 source, 18 sink, and 29 derivation descriptors in our experiments. The derivation descriptors correspond to methods in classes such as String, String Buffer, String Tokenizer, etc. Source and sink descriptors correspond to methods declared in 19 different J2EE classes. We used four different variations of our static analysis, obtained by either enabling or disabling context sensitivity and improved object naming. Analysis times for the variations are listed in values are obtained on an windows 150 machine with 4 GB of memory running windows.

Contrary to intuition, we actually pay less for a more precise analysis. Imprecise answers are big and therefore take a long time to compute and represent. In fact, the context-insensitive analysis with default object naming runs significantly slower on the largest benchmarks than the most precise analysis. The most precise analysis version takes a total of less than 10 minutes on the largest application; we believe that this is acceptable given the quality of the results the analysis produces.

Existing:

Existing graph denotes that keywords searched by user with respective privacy level that are based on history and cookies,here we can't achieve user personalised web search.And if we are about to clear history we can't get relevent information that are previously searched by us.

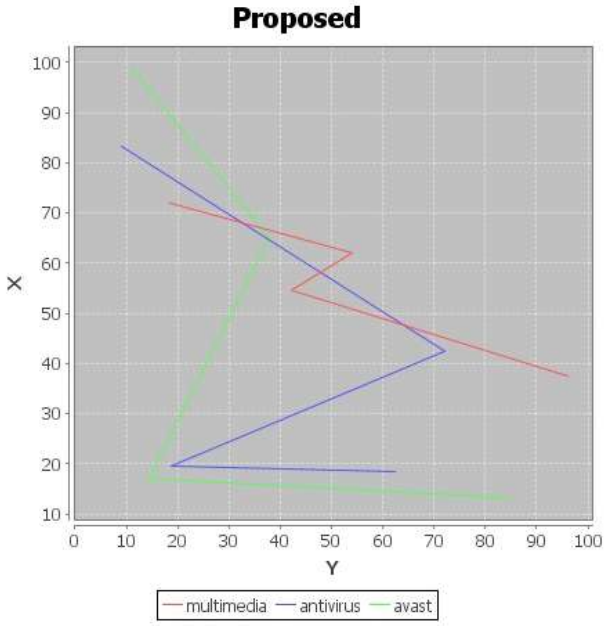
Existing:



Proposed:

Achieving more privacy level than existing method, by introducing user profile and content based recommandations,even at first time of searching also.

Proposed:



X-axis : denotes that privacy level of content.

y-axis : keywords.

Acknowledgment

I am highly indebted to my mentor **, Mr G.Senthil Kumar** , for his continuous support,motivation,supervision & guidance throughout the tenture of our project and helping us in classifying the concepts requiring knowledge and perception , handling critical situations & in understanding the objective of our work.

CONCLUSION

In this paper, presenting the experimental results of UPS, shows significant improvements in user search results. 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. UPS also performed online generalization on user profiles to protect the personal privacy without compromising the search quality.

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, 002.

[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, 007.

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

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