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

A Framework to Protect Privacy during Personalized Web Search

Nandeshwara Reddy Katta
M.Tech Student,
Department of CSE,
Sri Mittapalli College of Engineering,
Guntur, A.P, India.

S.Suresh Babu
Assistant Professor,
Department of CSE,
Sri Mittapalli College of Engineering,
Guntur, A.P, India.

Abstract:
The web search engine has gained a lot of popularity and importance for users seeking information on the web. Since the contents available in web is very vast and ambiguous, users at times experience failure when an irrelevant result of user query is returned from the search engine. Web search engines (e.g. Google, Yahoo, Microsoft Live Search, etc.) are widely used to find certain data among a huge amount of information in a minimal amount of time. These useful tools also pose a privacy threat to the users. Web search engines profile their users on the basis of past searches submitted by them however, effective personalized search requires collecting and aggregating user information, which often raises serious concerns of privacy infringement for many users. Indeed, these concerns have become one of the main barriers for deploying personalized search applications, and how to do privacy-preserving personalization is a great challenge. This paper models preference of users as hierarchical user profiles. It proposes a framework called UPS which generalizes profile at the same time maintaining privacy requirement specified by user. Two greedy algorithms namely GreedyDP and GreedyIL are used for runtime generalization. Also, an online prediction mechanism to decide whether to personalize a query or not is provided in this paper.

Keywords: ***Privacy Protection, profile, personalized web search, risk, Search Engines***

Introduction:

Personalized search refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided. Pitkow et al. describe two general approaches to personalizing search results, one involving modifying the user's query and the other re-ranking search results.

Google introduced personalized search in 2004 and it was implemented in 2005 to Google search. Google has personalized search set up for not just those who have a Google account but everyone as well. There is not very much information on how exactly Google personalizes their searches, however, it is believed that they use user language, location, and web history.

Early search engines, like Yahoo! and AltaVista, found results based only on key words. Personalized search, as pioneered by Google, has become far more complex with the goal to "understand exactly what you mean and give you exactly what you want." Using mathematical algorithms, search engines are now able to return results based on the number of links to an from sites; the more links a site has, the higher it is placed on the page. Search engines have two degrees of expertise: the shallow expert and the deep expert.

An expert from the shallowest degree serves as a witness who knows some specific information on a given event. A deep expert, on the other hand, has comprehensible knowledge that gives it the capacity to deliver unique information that is relevant to each individual inquirer. If a person knows what he or she wants than the search engine will act as a shallow expert and simply locate that information. But search engines are also capable of deep expertise in that they rank results indicating that those near the top are more relevant to a user's wants than those below.

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

While many search engines take advantage of information about people in general, or about specific groups of people, personalized search depends on a user profile that is unique to the individual. Research systems that personalize search results model their users in different ways. Some rely on users explicitly specifying their interests or on demographic/cognitive characteristics. But user supplied information can be hard to collect and keep up to date. Others have built implicit user models based on content the user has read or their history of interaction with Web pages.

Benefits:

One of the most critical benefits personalized search has is to improve the quality of decisions consumers make. The internet has made the transaction cost of obtaining information significantly lower than ever. However, human's capability of processing information has not expanded much. When facing overwhelming amount of information, consumers need a sophisticated tool to help them make high quality decisions. Two studies examined the effects of personalized screening and ordering tools, and the results show positive correlation between personalized search and the quality of consumers' decisions.

The first study was conducted by Kristin Diehl from University of South Carolina. Her research discovered that reducing search cost led to lower quality choices.

The reason behind this discovery was that 'consumers make worse choices because lower search costs cause them to consider inferior options.' It also showed that if consumers have a specific goal in mind, they would further their search, resulting in an even worse decision. The study by Gerald Haubl from University of Alberta and Benedict G.C. Dellaert from Maastricht University mainly focused on recommendation systems. Both studies concluded that a personalized search and recommendation system significantly improved consumers' decision quality and reduced the number of products inspected.

Personalized search gains popularity because of the demand for more relevant information. Research has indicated low success rates among major search engines in providing relevant results; in 52% of 20,000 queries, searchers did not find any relevant results within the documents that Google returned. Personalized search can improve search quality significantly and there are mainly two ways to achieve this goal.

The first model available is based on the users' historical searches and search locations. People are probably familiar with this model since they often find the results reflecting their current location and previous searches.

There is another way to personalize search results. In Bracha Shapira and Boaz Zabar's "Personalized Search: Integrating Collaboration and Social Networks", Shapira and Zabar focused on a model that utilizes a recommendation system.[36] This model shows results of other users who have searched for similar keywords. The authors examined keyword search, the recommendation system, and the recommendation system with social network working separately and compares the results in terms of search quality. The results show that a personalized search engine with the recommendation system produces better quality results than the standard search engine, and that the recommendation system with social network even improves more.

EXISTING SYSTEM:

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

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

with complicated user-interest models generated from user profiling techniques. Profile-based methods can be poten-tially effective for almost all sorts of queries, but arereported to be unstable under some circumstances.

DISADVANTAGES OF EXISTING SYSTEM:

- 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.
- Generally there are two classes of privacy protection problems for PWS. One class includes those treat privacy as the identification of an individual, as described. The other includes those consider the sensitivity of the data, particularly the user profiles, exposed to the PWS server.

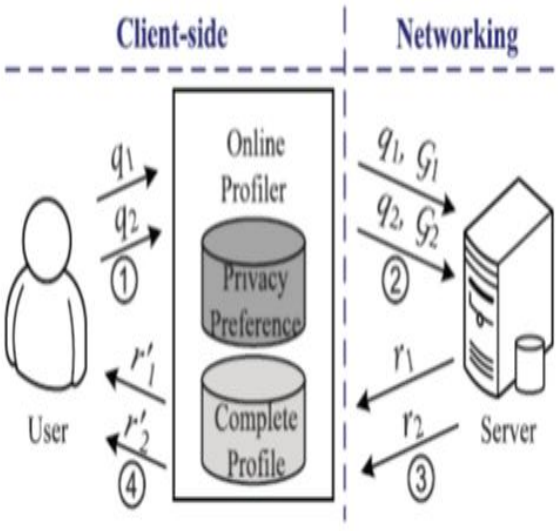
PROPOSED SYSTEM:

- We propose 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 conflicting metrics, namely personalization utility and privacy risk, for hierarchical user profile, we formulate the problem of privacy-preserving personalized search as #-Risk Profile Generalization, 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 discriminating power (DP), the latter attempts to minimize the information loss (IL). By exploiting a number of heuristics, GreedyIL out performs GreedyDP significantly.
- We provide 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.
- Our extensive experiments demonstrate the efficiency and effectiveness of our UPS framework.

ADVANTAGES OF PROPOSED SYSTEM:

- ❖ Increasing usage of personal and behaviour information to profile its users, which is usually gathered implicitly from query history, browsing history, click-through data bookmarks, user documents, and so forth.
- ❖ 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.

SYSTEM ARCHITECTURE:



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

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

MODULES:

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

IMPLEMENTATION

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as add contents, view all

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

contents, list all searching history, list ranking of images, list of all personalized search, attacker details, recover contents, list of all user and logout.

Add contents

In this module, the admin can add n-number of contents. If the admin want to add a new content, then admin will enter a URL, domain, title, description, uses, related images of the particular content ,then submit and that data will stored in data base. If admin want view to the newly added content, then click on view contents button, it will display the all contents & with their tags, the initially rank will be zero.

List of users

In this module, the Admin can view list of all users. Here all register users are stored with the details such as user ID, user name, E mail ID, mobile no, Location, date of birth, address, pin code, general key and personalized key.

View list all searching history

This is controlled by admin; the admin can view the all searching history. If admin clicks on search history button, then the server will display the all searching history with their tags such as user name, key word used, field searched, time & date.

Attacker details

In this module, the admin can view the attacker details. If admin clicks on attacker details button, the admin will get attacker information with their tags such as attacker name, attacked content URL and attacked content ID. After attacking content, the admin will recover the content.

User

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations such as view my details, query search, personalized search, personalized search comparisons, attack content details, request for general key, request for personalized key and logout. If user clicks on my details button, then the server will give response to the user with their tags such as user ID, name, mobile no, address, pin code and email ID.

Query Search

In this module, the user can search query. Before searching any query, the user should request general key, then admin will provide a general key. Then enter general key, select field to search, enter key word and search, it will display all related contents with their tags. After searching a content rank will be increased.

Personalized Search

In this module, the user can search contents. Before searching contents, the user should request personalized key, then admin will provide personalized key, then enter key and enter keyword, then user will get a related contents with their tags. After searching content the rank will be increased.

Personalized Search Comparison

In this module, the user can view the comparison between greedy DP & greedy IL. After personalized searching, the greedy IL will be generated. If the user clicks on personalized search button, it will display all personalized search details with their tags such as user name, keyword used, date, time, using greedy DP and using greedy IL.

Time delay Generation chart

In this module, we can view the time delay Generation chart results. This chart shows the time delay by using greedy DP and time delay using greedy IP. After viewing or search the content, rank will be increased and also the time delay will be display, the time variation can be shown in this chart.

Attack content

In this module, user can attack contents, and then user should enter content URL to attack, then user will get all information about content, then user can add

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

malicious data and click on attack button. After attacking successful, the attacker details will send to admin.

Conclusion:

Disadvantage of search personalization is that internet companies such as Google are gathering and potentially selling your internet interests and histories to other companies. This raises a privacy issue. The issue is if people are content with companies gather and selling their internet information without their consent or knowledge. Many web users are unaware of the use of search personalization and even fewer have knowledge that user data is a valuable commodity for internet companies. A client side privacy protection framework called UPS i.e User customizable Privacy preserving Search is presented in the paper. Any PWS can adapt UPS for creating user profile in hierarchical taxonomy. UPS allows user to specify the privacy requirement and thus the personal information of user profile is kept private without compromising the search quality. UPS framework implements two greedy algorithms for this purpose, namely GreedyDP and GreedyIL.

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] 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.

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

[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.

[21] Y.Xu, K. Wang, G. Yang, and A.W.-C. Fu, "Online Anonymity for Personalized Web Services," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500, 2009.

[22] Y. Zhu, L. Xiong, and C. Verdery, "Anonymizing User Profiles for Personalized Web Search," Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226, 2010.

[23] J.Castellı´-Roca, A. Viejo, and J. Herrera-Joancomartı´, "Preserving User's Privacy in Web Search Engines," Computer Comm., vol. 32, no. 13/14, pp. 1541-1551, 2009.

[24] A.Viejo and J. Castell_a-Roca, "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines," Computer Networks, vol. 54, no. 9, pp. 1343-1357, 2010.

[25] X.Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.

[26] J. Teevan, S.T. Dumais, and D.J. Liebling, "To Personalize or Not to Personalize: Modeling Queries with Variation in User Intent," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), pp. 163-170, 2008.

[27] G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "Ups: Efficient Privacy Protection in Personalized Web Search," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information, pp. 615-624, 2011.

[28] J. Conrath, "Semantic Similarity based on Corpus Statistics and Lexical Taxonomy," Proc. Int'l Conf. Research Computational Linguistics (ROCLING X), 1997.

[29] D. Xing, G.-R. Xue, Q. Yang, and Y. Yu, "Deep Classifier: Automatically Categorizing Search Results into Large-Scale Hierarchies," Proc. Int'l Conf. Web Search and Data Mining (WSDM), pp. 139-148, 2008.