

## Privacy Enhanced Web Personalization

BANDARU VENKATA RANADHEER<sup>1</sup>, CH. RAJU<sup>2</sup>

<sup>1</sup>PG Scholar, VRS & YRN College of Engineering & Technology, Chirala, Prakasam (Dt), Andhra Pradesh, India.

<sup>2</sup>Assoc Prof, VRS & YRN College of Engineering & Technology, Chirala, Prakasam (Dt), Andhra Pradesh, India.

**Abstract:** Personalized Web Search (PWS) briefly says about its accuracy in developing the eminence of different explores services throughout the world using net. Whatever, corroboration manifest that users reluctance to reveal their personal information during explore has become major impedance for the long proliferation of PWS. We analyze confidential protection in PWS applications that model user priorities as hierarchical user profiles. We propose a PWS framework called UPS that can flexible generalized profiles by queries while respecting user mentioned security requirements. Our runtime generalization point out at striking a handle between two predictive metrics that elaborate the utility of personalization and the privacy risk of revealing the generalized profile. We present two greedy algorithms, namely GreedyDP, GreedyIL, for runtime generalization. We also furnish an online prophecy mechanism for deciding whether personalized a query is useful. Extensive experiments establish the effectiveness of framework. The experimental results also reveal that Greedy IL significantly outperforms GreedyDP in terms of efficiency.

**Keywords:** Personalized Web Search (PWS), UPS, GreedyDP, GreedyIL.

### I. INTRODUCTION

Nowadays Internet is widely used by users to satisfy various information needs. However, ambiguous query/topic submitted to search engine doesn't satisfy user information needs, because different users may have different information needs on diverse aspects upon submission of same query/topic to search engine. So discovering different user search goals becomes complicated. The evaluation and depiction of user search goals can be very useful in improving search engine relevance and user knowledge [1]. Personalized Web search is a technique in order to provide better search results. It is a promising way to improve search quality by customizing search results for people with different information goals. 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. Apart from the personalized results, there is need of security in the personalized web search. Users are not keen to disclose their information during web search. This has become major issue in profiling the user in personalized web search. There should be a mechanism which considers profiles according to information provided by user. Actually more the search engine knows about user, more accurate search results will be obtained by search provider. But users cannot trust on search engine that information provided by user is not misused. Search engines can provide more accurate and specific data if users trust search engine and provide more information. Hence, search engines should provide security mechanism such that user will be ensured of its privacy and its information should be kept safe. In

personalized web search, user information is collected and analyzed in order to find intention behind issued query fired by user. Typically search is performed by providing queries to retrieval system in form of set of words. If different users enter same query, the system will produce same results without considering the user. But search results should be produced by taking the user in the equation, so that different users can get different search results for same query. By keeping track of user's personal information and interests.

### II. LITERATURE SURVEY

Personalized web search (PWS) differs from generic web search, which returns identical research results to all users for identical queries, regardless of varied user interests and information needs. PWS can be categorized into two types; one is click-log-based methods and other profile-based ones. The click log based methods are based on just selecting the clicked pages in the user's query history. The main drawback of this method is that it works on repeated set of queries by the users only. Profile based method has more effectiveness in improving the quality of web search with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history, browsing history, click-through data, bookmarks, user documents and so forth. [2]. The main drawback of this method is that it requires the user personal data to be send to the server, hence this privacy issue makes the user uncomfortable. To provide personalized search results to users, personalized web search maintains a user profile for each individual [3]. These profiles can be used in various ways to create an environment of personalized search. Some

of the methods to help in inferring user's information needs are:

- **Personalized Search Based on Content Analysis:** Personalized web search can be achieved by checking content similarity between web pages and user profiles [3]. When the user issues a query, each returned snippet/documents are filtered or re-ranked and classified. Chirita et al. [3] [4] use the ODP (Open Directory Project) hierarchy to implement personalized search. In [5], a user profile is built as a vector of distinct terms and is constructed by aggregating past user click history [3]. Shen et al. [6] first use language modeling to mine immediate search contextual and implicit feedback information [3]. Teevan et al. [7] and Chirita et al. [8] exploit rich models of user interests, built from both search-related information, and other information about the user. [3]
- **Personalized Web Search Based on Hyperlink Analysis:** Most generic web search approaches rank importance of documents based on the linkage structure of the web. A large group of these works focuses on personalized PageRank. PageRank, proposed by Page and Brin [9], is a popular link analysis algorithm used in web search. The fundamental motivation underlying PageRank is the recursive notion that important pages are those linked-to by many important pages [3]. Qiu and Cho [10] develop a method to automatically estimate a user's topic preferences based on Topic-Sensitive PageRank scores of the user's past clicked pages. The topic preferences are then used to bias future search results.
- **Community-based Personalized Web Search:** Some approaches that personalize search results for the preferences of a community of like-minded users [3] is known as community or collaborative based search. In community-based personalized web search, when a user issues a query, search histories of users who have similar interests to the user are used to filter or re-rank search results [3]. Sugiyama et al. [5] use a modified collaborative filtering algorithm to constructed user profiles to accomplish personalized search. Sun et al. [11] proposed a novel method named CubeSVD to apply personalized web search by analyzing correlations among users, queries, and web pages in clickthrough data. Smyth et al. [12] show that collaborative web search can be efficient in many search scenarios when natural communities of searchers can be identified.
- **Server-Side & Client-Side Implement Personalized:** web search can be implemented on either server side (in the search engine) or client side (in the user's computer or personalization agent). For server-side personalization, user profiles are built, updated, and stored on the search engine side [3]. For client-side personalization, user

information is collected and stored on the client side (in the user's computer or a personalization agent), usually by installing a client software or plug-in on a user's computer [3].

Despite of having various advantages of personalized search, there is no large-scale use of personalized search services currently. Personalized web search faces several challenges that hinder its real-world large-scale applications:

- Privacy is an issue.
- Users are not static.
- Queries should not be handled in the same manner with regard to personalization [3].

In order to improve performance of the web search we need to take care of privacy issues in personalized web search. As personalizing search requires gathering and processing of user information, which leads to privacy issue. This is becoming the main obstacle in deploying personalized web search applications. Adequate work has been proposed in order to maintain privacy in personalized web search:

- Chaum proposed in [18] the use of an anonymity network which consists of several routers that act as anonymizers. It is a technique which is based on public key cryptography that allows an electronic mail system to hide who a participant communicates with as well as the content of the communication - in spite of an unsecured underlying telecommunication system.  
Drawback: The main drawback of this approach was that the process of submitting a query to the WSE and receiving the answer through an anonymous channel is very time-consuming.
- Krause and Horvitz in [19] employ statistical techniques to learn a probabilistic model, and then use this model to generate the near-optimal partial profile. They had introduced and explore an economics of privacy in personalization, where people can opt to share personal information, in a standing or on-demand manner, in return for expected enhancements in the quality of an online service  
Drawback: Limitation in this work was 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. in [15] proposed a privacy protection solution for PWS based on hierarchical profiles. Using a user-specified threshold, a generalized profile is obtained in effect as a rooted subtree of the complete profile. These profiles summarize a user's interests into a hierarchical organization 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.

Drawback: The main drawback of this approach was that this work does not address the query utility, which is crucial for the service quality of PWS.

- Xiao and Tao proposed Privacy-Preserving Data Publishing (PPDP). A person can specify the degree of privacy protection for her/his sensitive values by specifying “guarding nodes” in the taxonomy of the sensitive attribute. [13]

Drawback: The greedy algorithm presented in this paper was not optimal and also did not support runtime profiling.

- Teevan et al. collect a set of features of the query to classify queries by their click entropy. He first examined the variability in user intent for a large number of queries using both implicit and explicit measures. Then study was carried out to show variation in the implicit measures predicts variation in the explicit measures, and look at what other factors can account for variation in the implicit measures. Queries are characterized using a variety of features of the query, the results returned for the query, and the query’s interaction history. Using these features predictive models were built to identify the queries that will benefit most from personalization, and explore which features are the most valuable for prediction [14].

Drawback: This works motivates in questioning whether to personalize or not to, they assume the availability of massive user query logs and user feedback.

All the existing profile-based Personalized Web Search does not support runtime profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminately. 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. [17] The existing methods do not take into account the customization of privacy requirements. This probably makes some user privacy to be overprotected while others insufficiently protected. [2] For example, in, 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. Many personalization techniques require iterative user interactions when creating personalized search results. 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, there is a need of predictive metrics to measure the search quality and breach risk after personalization, without incurring iterative user interaction.

### III. CONCLUSION

This paper presents a survey report of different methods to help in inferring user’s information needs of Personalized Web Search. It also covers issues like need of personalized web search, how personalized web search can be implemented, what are challenges in it, privacy and security issue of it and existing system of personalized web search. This paper also gives a survey report of different ways to maintain privacy in personalized web environment. It also tells about the drawbacks of the existing privacy issues. The future scope of our paper will be to overcome the existing system drawbacks and design a framework to maintain a complete privacy of the users so that they can work without any fear of working in personalized web environment.

### IV. 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.

[14] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschütter,  
“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. Järvelin and J. Kekäläinen, “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’s Profile:**



**Bandaru Venkata Ranadheer** received B.TECH. Degree from St. Ann's college of Engineering & Technology Chirala, Prakasam, Andhra Pradesh. and currently pursuing M.Tech in Computer Science Engineering at VRS & YRN College of Engineering & Technology, Chirala; Prakasam (Dt), Andhra Pradesh. His areas

of interest include in DATA MINING.



**Mr.Ch. Raju** is presently working as Associate Professor in VRS & YRN College of Engineering & Technology, Chirala, AP, India. He did his B.Tech. degree in Computer Science & Engineering from JNTUH, Hyderabad. And then completed his M.Tech.in Software Engineering as his specialization from

JNTUK, Kakinada. He has a teaching experience of 8 years. He published papers in 4 International Journals and 1 National Conference.