

Survey on Security in Personalized Web Search

Kiran Jagdish Khandelwal¹, L. J. Sankpal²

¹Pune University, Stesos Sinhgad Academy of Engineering, Kondhwa, Pune, 411048, Maharashtra, India

²Professor, Pune University, Stesos Sinhgad Academy of Engineering, Kondhwa, Pune, 411048, Maharashtra, India

Abstract: *Nowadays, web has become vast source of Information. In process of acquiring that information, search engines play an important role. The use of World Wide Web is increasing rapidly over the years. Number of search results are obtained or showed by search engines, but less accuracy and poor quality of search results make hard for user to gain information that is needed. In recent years search engines has come up with advanced techniques. Users are not interested to spend time for crucial queries for search. It has found that if search results are late and inaccurate then users are not interested. Require of current situation is to provide immediate and accurate search results. This need can be solved implementing personalized web environments. As it is becoming important aspect to provide personalized web environments many techniques and approaches have developed. But at the same time security of personalized web searches has also gained significance in which personal/private information can not be disclosed by web searches. Personalized web search is proving its effectiveness but also raising matter of privacy and securing personal information. Many personalization methods have been exposed and put into practice. But it is not sure that those methods will make sure their efficiency in dissimilar queries for different users. In this paper we will discuss on require of personalized web search and securing personalized information.*

Keywords: Personalized web search, Personalization Techniques, Privacy, Information Retrieval, Immediate and Accurate Results.

1. Introduction

In uncontrolled growth of Internet use, obtain necessary or future information is also a significant topic. Results of search should be dependent on user behavior. For that implementation of personalized web environment is required. Web search results should adapt to user requirements. Main issue in search engines is same results are generated for different users. Sometimes these results are short and confusing [1].

Importance should be given to areas like question answer methodology, better browsing, virtual results, localization and modified web search. Search engines should display results depending upon the user. Many technologies have been implemented for different needs of users.

Currently search engines are serving all users same search results regardless of special needs of user. Though indexing algorithms are used in traditional search engines has difficulty in achieving efficiency expected by user. To obtain data related to user's interest personalization is used. Generic web search differs from personalized web search because it provides same search results for different users for similar queries. Information intended from different users may be different while using same query. Mixed type of search results will be retrieved. Because of this time will be consumed to choose information that user wants. By knowing some things about user, a search engine might refine user results to make them more relevant is the concept behind personalized search. The web search engine has long become the most important portal for ordinary people looking for useful information on the web.

Personalizing web search is the process of obtaining web search results depending upon history or past behavior of user. Depending upon conditions, Personalization can be done at server or client side [12]. User may be aiming to achieve many goals in single query. Hence efficiency of personalization of web search is depended on user behavior

and query. In web personalization, reading minds of users is challenging task to perform. Also there is limitation of words used to search. Two three words are used to search over web. That is increasing the challenge in personalizing web search.

There is need of security in the personalized web search. Users are not keen to disclose their information during web search [13]. This has become major issue in profiling the user in personalized web search. There should be a mechanism which generalizes profiles according to information provided by user [7]. 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 [8].

There are mainly two areas where research can be focused. 1) By using personalization of the user profile, improve search result quality. 2) For keeping privacy risk under control, they need to hide private information of user present in user profile.

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.

2. Literature Review

This paper focuses on a personalized web search and techniques to achieve personalization of user's web search.

1) **Personalized Web Search:** There are mainly two types of personalized web search they are Click-log-based and Profile-based personalized web search.

A. Click-Log-Based

In this type of personalized web search, personalization is carried out on the basis clicks. The data get recorded through clicks in query logs to simulate user experience. In this approach, the web pages frequently clicked by user in past history for a particular query is recorded and personalized score is computed for particular web page and web search results are according to that score[9].

Click log based personalized web search approaches will perform consistent and considerably well when it is working on frequent queries. drawback of click log based approach is when a never asked query is entered by user; it will not create precise and related search results. It has found that one third of queries are frequent and it will work well on that [1].

B. Profile-Based:

In this approach, search results are sorted on the basis of personal interest of user profile. Mainly there are two strategies are used for creating user profile. One is that using words which are frequently used in documents. huge profiles are created by this approach. It gives less accuracy and less efficiency in search results. Second approach is that using pre defined ontology such as DMOZ [1] [2]. This approach eases formation and preservation of problem. This has found that profile based personalized web search has become unstable when users history increases.

2) Pursuit of Personalization

At search engine, results must be downloaded where user profile is stored or a user profile must be communicated to server where web corpus is stored, to include information about user. There are few reasons for re-ranking of results locally, 1) as we are working on relatively small data; inclusion of computationally-intensive procedures is feasible, 2) for privacy reasons as users are uncomfortable with sending their data on the internet or to an unknown destination, re-ranking of search results is done locally, 3) simple evaluation can be provided by re-ranking methods. Instead of collecting rating for all documents on the internet which is infeasible in nature, we need to consider only top results [3].

Three important components of our model are [3]:

1. Corpus Representation
2. User Representation
3. Document/Query Representation

1. Corpus Representation

As we are dealing with web search, our corpus will be web. There are certain parameters like number documents present on web, number documents that contains searched keyword. As users do not have direct access to the details of web it is turning to be a disadvantage of personalization. From all documents on web or documents related to search keyword statistics of web can be generated. The corpus representation can be focused by the query entered from user.

2. User Representation

A rich index of personal content is used to represent a user. An index is consisting of user's interests and computational activities. Email messages which are viewed or sent, calendar entries, web pages visited by user and documents present in client machine this can be included in index content. By using this information, rich but unstructured profile of user can be created. By treating every document as a source of evidence related the user's interests, independent of the query, this index can be used.

3. Document and Query Representation

The representation of document is essential for determining what terms are included and how often they are used. To access full text of documents it will take time so by accessing only title and snippets of documents is used to create representation of documents. Snippets of documents obtained by search engines are based on query.

3) Information Retrieval

There are many Information Retrieval systems. In most of these algorithms, information retrieval problem treated one single query and documents sets. Most of existing retrieval models makes decision based only on query and document collection. Information of actual user and search context is largely ignored in existing information retrieval systems. Web search engines provide search results depending upon the query submitted by user. To improve retrieval accuracy, additional context information should be exploited by an optimal retrieval system whenever it is available. Major challenge in information retrieval is context-sensitive retrieval of information [4] [6].

4) Privacy in Personalized web search

Personalized web search is gaining more and more popularity. But maintaining privacy is serious issue 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. Personalized anonymity is a security technique which is implemented to provide privacy in personalized web search in which person can specify degree of privacy [5]. Anonymizing user profile is also technique by which privacy of user can be maintained [11].

In Personalizing Search Based on User Search Histories (Mirco Speretta and Susan Gauch, 2000) [9] investigate the utilization of a less-obtrusive method for social event client data for customized inquiry. In particular, they manufacture client profiles focused around movement at the pursuit site itself and study the utilization of these profiles to give customized inquiry results. By actualizing a wrapper around the Google internet searcher, we had the capacity gather data about individual client seek exercises.

3. Proposed System

We suggest a privacy-preserving personalized web search framework UPS. According to user-specified privacy requirements it can simplify profiles for each query. We also formulate the problem of privacy-preserving personalized

search as δ -Risk Profile Generalization, with its NP-hardness proved relying on the meaning of two contradictory metrics, namely personalization utility and privacy risk, for hierarchical user profile. To support runtime profiling a GreedyDP and GreedyIL generalization algorithms develop which is simple but effective. While the former tries to maximize the discriminating power (DP), the latter attempts to minimize the information loss (IL). GreedyIL outperforms GreedyDP considerably, by developing a number of heuristics. We give a reasonably priced mechanism for the client to make a decision whether to personalize a query in UPS. This choice can be made previous to each runtime profiling to improve the constancy of the search results while keep away from the needless contact of the profile. Our extensive experiments show the competence and efficiency of our UPS framework.

The UPS framework works on two phases that is offline and online.

1. Offline Phase

In offline phase a hierarchical user profile is build and modified with the user-specified privacy supplies.

2. Online Phase

In online phase the proxy generates a user profile in runtime in the light of query terms when a user issues a query q_i on the client. Generalized user profile G_i pleasing the privacy supplies is the output of this step. By bearing in mind two conflicting metrics, namely the personalization utility and the privacy risk, both defined for user profiles are directed by generalized process. Subsequently, for personalized search, the query and the generalized user profile are sent together to the PWS server. After that the search results are personalized with the profile and bring back to the query proxy. Lastly, the proxy either reranks them with the complete user profile or presents the raw results to the user and provides the security.

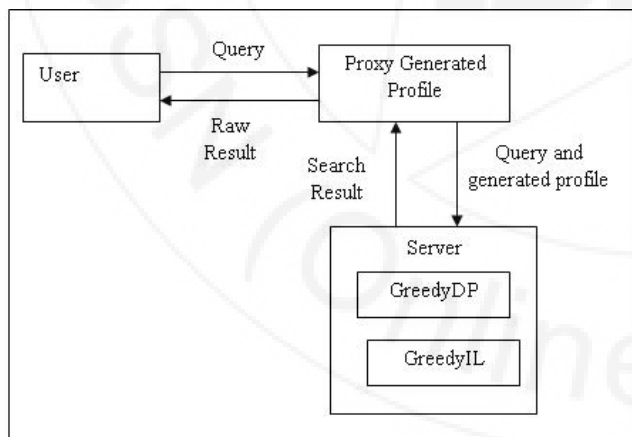


Figure: System Architecture

4. Conclusion

We have prepared a survey report for different topics of Personalized Web Search. This report 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. Based on such report, we surveyed Click-log-based PWS, Profile-based PWS, Pursuit of personalization, Information Retrieval, Privacy in Personalized web search.

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] K. Ramanathan, J. Giraudi, and A. Gupta, "Creating Hierarchical User Profiles Using Wikipedia," HP Labs, 2008.
- [3] 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.
- [4] 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.
- [5] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.
- [6] K. Ja'rvelin and J. Keka'lainen, "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.
- [7] 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.
- [8] X. Shen, B. Tan, and C. Zhai, "Privacy Protection in Personalized Search," SIGIR Forum, vol. 41, no. 1, pp. 4-17, 2007.
- [9] M. Spertta and S. Gach, "Personalizing Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI), 2005.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] X. Xiao and Y. Tao, "Personalized Privacy Preservation," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2006.

Author Profile

Kiran Jagdish Khandelwal received the B.E. degree in Computer Science and Engineering from Anuradha College of Engineering in 2013. Now doing PG in Computer Engineering, from Sinhgad Academy Of Engineering, Kondhwa, Pune, Maharashtra, India