Personalized Search with User Privacy Protection

S.Poornima[1] Mr.R.Rangaraj[2]
[1]M.Phil Student [2]Assistant Professor & Head (M.Sc, M.Phil, M.Sc (Psy), Ph.D)
[1,2]Department of Computer Science
[1,2]Hindusthan College of Arts and Science, Coimbatore, India

Abstract— The quality of search services, personalized web search (PWS) provides better results in internet and has been attracting attention recently. Many researchers tailored for individual user needs in PSW. One of the most critical benefits personalized search is to improve the quality of decisions that consumers make. The major barrier of PWS is to disclosure the private information during the search. To protect user privacy in profile-based PWS, two contradicting effects are considered during the search process. On the one hand attempts are made to improve the search quality with the personalization utility of the user profile by hiding the privacy contents existing in the user profile to place the privacy risk under control. The proposed work is very effective and efficient and does not require iterative user interaction. The work enhances the stability of the search quality by avoiding the unnecessary exposure of the user profile. With this system admin can easily find out the ratings. Admin can cluster the user interest which allows search engines to sell more advertisements because companies understand that they will have a better opportunity to sell based on high percentage matched individuals than medium and low percentage matched individuals.

Key words: Personalized Web Search (PWS), Stability, Individual Attention, Clustering

I. INTRODUCTION

The term Personalized web search refers to search experiences that are tailored specifically to an individual's interests by gathering information about the individual beyond specific query provided [1]. Normally search engines are categorized by two degrees of expertise they are: 1) the shallow expert in which the expert from the shallowest degree, serves as a witness based on some specific information on a given event. 2) A deep expert contains comprehensible knowledge that gives the capacity to deliver unique information that is relevant to each individual inquirer. Based on the person needs, what he or she wants than the search engine will act as a shallow expert and simply locate that information. But deep expertise the search engine provides rank based results indicating that those near the top are more relevant to a user's wants than those below. One technique Google uses to personalize searches for users by tracking log in time and if the user has enabled web history in his browser.

II. NEED FOR PWS

Personalization of Web search is used to carry out retrieval for each user incorporating his/her interests. The Personalized Web Search provides a unique opportunity which consolidates and scrutinizes the work from industrial labs on personalizing web search by using user logged search behavior context by providing a fully anonymized dataset that contains anonymized user ids, query terms, queries, URLs, URL domains and clicks.

The shared dataset enables the researchers to study the problem of personalizing web search experience. It introduces potential privacy problems in which a user may not be aware that their search results are personalized for them. Since there is no threat of privacy invasion with the services, the search balance has are obtained to favor personalization over privacy, even when it comes to search.

Search engines put forth advertisements based on a large amount of collected data aggregated from an individual's web clicks, by segregating the interest of an individual. It helps to match up products and services to individuals by more specialized ones which in turn reduce the money spent on advertisements to reach the product to the consumers with PWS. This paper enhances the stability of the search quality by avoiding the unnecessary exposure of the user profile. With this system admin can easily find out the ratings. Admin can cluster the user interest which allows search engines to sell more advertisements because companies understand that they will have a better opportunity to sell based on high percentage matched individuals than medium and low percentage matched individuals.

III. LITERATURE SURVEY

Many research are been focused on information retrieval for personal search in internet. Personal detail has been used in the web search context in creating a personalized version of page rank [3, 4]. Many commercial information systems are available that retrieves information based on user interest [5, 6].The author [7] presents a large-scale evaluation framework deals with personalized search based on query logs and personalized search strategies using MSN query logs. And analyze significant improvement over common web search on some queries which are based on short and long term queries. To improve the accuracy of information search a decision theoretic framework is designed with intelligent client-side web search agent (UCAIR) [8].

XuehuaShen et al [9] explored at how to exploit implicit feedback information, including query history and also with click-through history in the same search session, to improve information retrieval performance. KL-divergence retrieval model has been proposed for context information retrieval, uses TREC AP Data that creates a test set for evaluating the implicit feedback models. Since a long time has been advocated in Information Retrieval an interactive activity, the implicit feedback techniques are applied for purpose of collecting user interests and preferences. The paper [10] analyze about a short synthesis of the implicit techniques used to define user models is presented.

Yabo Xu et al [11] proposed a notion of online anonymity that protects web users to maintain online anonymity through time. The approach makes use of a third party called the user pool and protects the user privacy, by

not allowing the user pool to be trusted. To systematically examine the issue of privacy preservation in personalized search [12] a client-side personalization has been designed that has advantage over server-side personalized search services in preserving privacy, and also envision possible future strategies to fully protect user privacy. Two greedy algorithms, GreedyDP and GreedyIL, are designed [13] for runtime generalization of user information and disclose the private information during search and reveals that GreedyIL outperforms than GreedyDP in terms of efficiency.

Xiaokui Xiao et al [14] develops's a new generalization framework that takes into account customized privacy requirements and successfully prevents privacy intrusion even in scenarios which results in generalized of tables to permit accurate aggregate analysis.

IV. PROBLEM DEFINITION

The existing personalized web searches do not support run time profiling. A user profile is typically generalized for only once offline, and used to personalize all queries from a same user indiscriminatingly. Such "one profile fits all" strategy certainly has drawbacks given the variety of queries. Existing 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. This probably makes some user privacy to be overprotected while others insufficiently protected.

– Inability to link different technologies,
– Poor data quality and
– Lack of relevant technology.

V. PROPOSED SYSTEM

The framework works in two phases, namely the offline and online phase, for each user. During the offline phase, a hierarchical user profile is constructed and customized with the user specified privacy requirements. In online phase, if the user submits the query, search engine shows the corresponding links. If the user clicks on the particular link, the session count will be increased. The query and the generalized user profile are sent together to the PWS server for personalized search. The search results are personalized with the profile and delivered back to the query proxy. Finally, the proxy either presents the raw results to the user, or re-ranks them with the complete user profile.
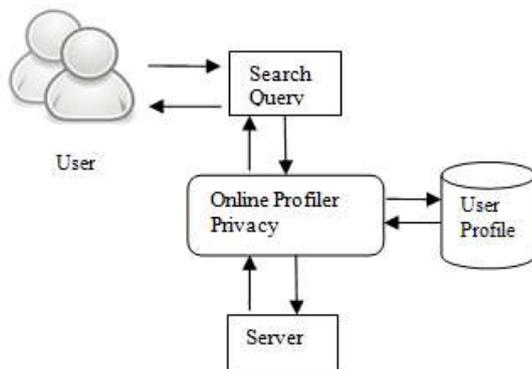
VI. SYSTEM ARCHITECTURE

Fig 1: System Architecture

VII. USER PROFILE CREATION

New users profile creation is carried out when a new user enter into the system, he/she should register with their details before login which is stored in the admin database. Then the private profile is converted into generalized profile after which the user gets the username and password from admin for login. This process gets the query from the user which is noted with the particular search time and it is stored in the database. The current user is identified by session attributes.

VIII. ADMIN LOGIN

Admin can login to the system with the respective username and password. Admin can add the domains and enter the details about the domain link, domain description and the keyword. Admin can also view the report of user interest that is chose by the users and cluster the individual user behavior. Admin views the ratings of every query given by the user which is helpful to identify the individual attention and collaborative attention.

IX. GENERALIZING USER PROFILE

The generalization process has to meet specific prerequisites to handle the user profile. This is achieved by preprocessing the user profile. At first, the process initializes the user profile by taking the user profile into account. Thereafter the process loads the data for the foreground and the background of the map according to the described selection in the user profile. While user send request to the server, the user private profile is hidden and the server send back the processed request as service or response to the clients with the generalized profile. For every request send by the user, the session count of the user is stored in the database.

X. PREDICTING USER INTEREST

User entered queries are typically found within the query logs of a search engine. Each query group contains closely related and relevant queries and clicks. By using this approach makes use of search logs in order to determine the relevance between query groups more effectively. Based on the session counts, admin can predict the user interest of each individual query.

XI. CLUSTERING QUERIES

The process groups or clusters the particular result based on single query and thus admin knows the session count for the particular query which helps in finding out the user behavior for a particular query and it can be used for particular advertisement that is user's ontology can be achieved in this process.

XII. MINING RESULT

In this process, the admin mine the results for frequently used links and infrequently used links separately. So the admin can get benefit to find out the user behavior and predict the user interest easily. Figure 2 shows the coding window to update the user record.

Fig 2: code window to update user.

XIII. RESULT ANALYSIS

Our major hypothesis is that using search context can help improve search accuracy. In particular, the search context can provide extra information to help us estimate a better query model than using just the current query. The metrics proposed in this work are rating, mining and clustering of information based on which the user's interest can be predicted easily.

Fig 3: Rating Page

Fig 3 shows the Rating page of the users for the particular pages with the hit count of each user when the pages are viewed. With which we can identify the user's interest.
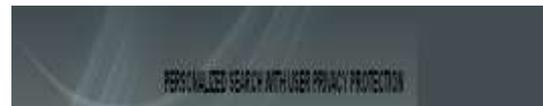

Fig 4: Analyzing user behavior

Fig 4 shows the analyzing behavior of the user with the links by searching with the most related keywords and rarely used infrequent links with which the user can be clustered with a group of same category.

Fig 5: Frequently used Queries

Fig 5 shows the result of a single user and their frequently used queries with the related keyword are displayed and the login in time of the page is noted with the session time field. With which the user's behavior can be identified.

Fig. 6: Clustering Users

Fig 6 shows the clustering users based on the frequent used queries with the keywords ,date and time used with the ID and number of times used with the session count. So with this result we can categorize the cluster of user with the same interest. So that the advertising can be done based on the user's category by minimizing the money sent for advertising and reaching the consumers easily

without scattering the adverting to the infrequent users of particular type. With the results we have better opportunity to sell based on high percentage matched individuals than medium and low percentage matched individuals.

XIV. CONCLUSION

Personalized search is a promising way which improves search quality in internet. This approach requires users to provide personal information on Internet and produce the general ID and Password for the access. First, the personal information is collected from the user and is organized. With the user hit the user's interest can be predicted by the administrator and clusters are form based on the category of the user interest in any particular field or part. The frequently used links and infrequently used links are identified that determines the user interest.

This makes easy for advertising purpose with the identification of the user interest easily. With this system admin can easily find out the ratings. Admin can cluster the user interest which allows search engines to sell more advertisements because companies understand that they will have a better opportunity to sell based on high percentage matched individuals than medium and low percentage matched individuals. The experimental results verified our hypothesis that there is an opportunity for users to expose small portion of private information to get a relatively high quality search. Offering general information is a greater impact for improving search quality which is achieved in the system.

REFERENCES

[1] J. Pitkow, H. Schuetze, T. Cass, R. Cooley, D. Turnbull, A.Edmonds, E. Adar, and T. Breuel. Personalized search.Communications of the ACM, 45(9):50-55, 2002.

[2] Google personalized search: http://www.google.com/psearch.

[3] Glen Jeh, and Jennifer Widom. Scaling personalized web search. In Proc. of the 12th International World Wide WebConference (WWW), Budapest, Hungary, May 2003.

[4] T.H. Haveliwala. Topic-sensitive PageRank. In Proc. of the11th International World Wide Web Conference (WWW) Honolulu, Hawaii, May 2002.

[5] Paolo Ferragina, and Antonio Gulli. A personalized search engine based on Web-Snippet hierarchical clustering. In Proc. of the 14th International World Wide Web Conference (WWW), Chiba, Japan, May 2005.

[6] P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschutter. Using ODP metadata to personalize search. In the Prof. of 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Salvador, Brazil, August, 2005.

[7] Zhicheng Dou, Ruihua Song and Ji-Rong Wen, "A Large-scale Evaluation and Analysis of Personalized Search Strategies", (IW3C2), WWW 2007 , May 8–12, 2007, Banff, Alberta, Canada,ACM 978-1-59593-654-7/07/0005.

[8] Xuehua Shen, Bin Tan, ChengXiang Zhai, "Implicit User Modeling for Personalized Search", CIKM'05,October 31–November 5, 2005, Bremen, Germany,Copyright 2005 ACM 1-59593-140-6/05/0010.

[9] XuehuaShen, BinT An, ChengXiang Zhai," Context-sensitive Information Retrieval Using Implicit FeedbacK", SIGIR'05,August15–19,2005,Salvador, Brazil. Copyright2005ACM1-59593-034-5/05/0008

[10] Gabriella Pasi, "Implicit Feedback through User-system Interactions for Defining User Models in Personalized Search", The 6th international conference on Intelligent Human Computer Interaction, IHCI 2014, Volume 39, 2014, Pages 8–1.

[11] Yabo Xu, Ke Wang, Guoliang Yang, Ada W. C. Fu, "Online Anonymity for Personalized Web Service" CIKM'09, November2 –6, 2009,Hong Kong, China. Copyright 2009 ACM 978 -1-60558-512-3/09/11.

[12] Xuehua Shen, Bin Tan, ChengXiang Zha "Privacy Protection in Personalized Search" ACM SIGIR, Vol 41. No.1 June 2007.

[13] Lidan Shou, He Bai , Ke Chen , Gang Chen, "Supporting Privacy Protection in Personalized Web Search "Knowledge and Data Engineering, IEEE Transactions on (Volume:26 , Issue: 2) Feb. 2014, DOI:10.1109/TKDE.2012.201

[14] Xiaokui Xiao, Yufei Tao,"Personalized Privacy Preservation" SIGMOD 2006, June 27–29, 2006, Chicago, Illinois, USA. Copyright 2006 ACM 1-59593-256-9/06/0006.