Research Article
Survey on Personalized Web Search With Privacy Protection

Arunpandi.V[1], Sundaramoorhty. S[2] and Varatharaj[3]

[1]PG Scholar/Department of CSE, Angel College of Engineering and Technology. Tiruppur,641665, India
[2]Asst. Proff/Department of CSE, Angel College of Engineering and Technology. Tiruppur,641665, India
[3]Asst. Proff/Department of ECE, Angel College of Engineering and Technology. Tiruppur,641665, India
Corresponding Author : *arunpandi077@address.com*

Keywords

Personalized Web Search,
Privacy protection,
Profile based search

Abstract

Personalized web search (PWS) is a technique used in the web search to make the query processing an effective thing. The PWS concept reduces the time that is required to search the data and produce search result to the user. Providing a protection for user profile in PWS is highly required to protect the user's confidential details on web. Because the PWS involves to create the user profile from the user's personal information's like the search histories (e.g. Click-based history).There are many concepts available for creating user profile implicitly and explicitly. The explicit method needs user interaction for the profile creation. But the implicit method automatically creates the profile from the available data like desktop document, email, browser history. The collected information is converted into hierarchical user profile with the preference. The generally available contents and specific contents are separated and arranged from top to bottom in hierarchical structure. The exposure of profile detail is designed by the user to make the privacy protection better. The results are retrieved from the server for the requested query by combining the user profile with the requested query before sending it to the server. Page ranking, re-ranking concepts are used for filtering the results that are received. Using the ranking concept the results are arranged and provided to the users. This paper provides the survey on the different personalization techniques and protecting the personal details from the user profile.

I. INTRODUCTION

Content searching in web has become increasingly difficult for users to find information on the web that satisfies their individual needs since information resources on the web continue to grow. Under these circumstances, Web search engines help users find useful information on the web. However, when the same query is submitted by different users, most search engines return the same results regardless of who submits the query. In general, each user has different information needs for his/her query. Personalized web search is introduced to improve the web search. Personalization is a playing an increasingly important role in creating better Internet experiences. Recent applications of personalization have focused on improving the search experience. An important aspect of personalization is creation of a user profile Personalized information retrieval and search promises to improve the Internet experience. An important requirement for building personalized web applications is to build user profiles that represent the users' interests.

II. LITERATURE SURVEY

Personalized search is a promising way to improve search quality by customizing search results for people with different information goals. Many recent research efforts have focused on this area. The profile creation involves, [6], (1) collecting information from users: All searches, for which at least one of the results was clicked, were logged per user. (2) Creation of user profiles. Two different sources of information were identified for this purpose: all queries submitted for which at least one of the results was visited and all snippets visited. Two profiles were created out of either queries or snippets. (3) Evaluation: The profiles created were used to calculate a new rank of results browsed by users. The average of this rank was compared with Google's rank. Many approaches create user profiles by capturing browsing histories through proxy servers or desktop activities through the installation of bots on a personal computer. These require participation of the user to install the proxy server or the both.

The user profile is constructed by observing the information from the browser web page cache. The user desktops contains large amount of personal data; richer profiles can be built using this data. Most of the prior efforts in creating user profiles use frequently occurring document words to represent the profile. The following problems may occur due to this kind of profile creation, [5]. (1) Irrelevant words ,(2) Polysemy and synonymy ,(3) Size of the profile, (4) The profile content may represent a mixture of recreational needs of the user, information and transactional.

In [2], ODP is one of the largest efforts to manually annotate web pages, exporting all this metadata information in RDF format. Over 65,000 editors are busy keeping the directory reasonably up-to-date, and the ODP now provides access to over 4 million web pages in the ODP catalog. Still, given the fact that Google now indexes more than 8 billion pages, the ODP effort still only covers about 0.05 percent of the Web pages indexed by Google. Investigates two ways to personalize search and makes the following contributions:

First, directly using ODP entries, we show how to generalize personalized search in catalogs such as ODP and Google Directory beyond the currently available search restricted to specific categories.

Second, extending the manual ODP classifications from its current 4 million entries to a 8 billion Web in an automated way is feasible, based on an analysis of how topic classifications for a small but important subset of a large page collection can be extended to this large collection via topic-sensitive biasing of PageRank values. This generalizes earlier approaches which already investigated topic-sensitive page ranks, but relied on very simple classifications using only 16 topics.

In [2],Page ranking, Topic sensitivity pageRanking, personalized page Pageranking are the methods to order the details in hierarchy to find the specific and general details. PageRank Compute Web page scores based on the graph inferred from the link structure of the Web. It is based on the idea that "a page has high rank if the sum of the ranks of its backlinks is high". Topic-sensitive PageRank Building a topic oriented PageRank, 16 main topics from the Open Directory Project is considered with a set of 16 Page-Rank vectors for each topics. Then, the similarity between a user query and similarity among these topics is computed, and the 16 vectors are now combined using appropriate weights. Personalized PageRank A more recent investigation uses a different approach: it focuses on user profiles. One Personalized PageRank Vector (PPV) is computed for the each user.

In [5], creating a user profile using Wikipedia requires the following three steps. (1) Web pages are mapped to Wikipedia concept, (2) Hierarchical profile created from this concept. Concept in profile is tagged in two ways. First, whether the

target is transaction or recreational. Second, how recent the users are in that topic.

In [1], users have to register personal information such as their interests, age, and so on, beforehand, or users have to provide feedback on relevant or irrelevant judgments, ratings on a scale from 1 (very bad) to 5 (very good), and so on. These types of registration, collecting feedback, or ratings are consumes some specific time and users prefer easier methods.

A. Hyperlink Based Personalized Web Search

The search web field involves in the information retrieval from the structure of hyperlinked web pages like google. This kind of engines having the following problems. (1) Allocated weight for Web page, and (2) Hyperlinked Web pages may have related contents that are not considered. The use of personalized PageRank to enable personalized Web searches was suggested as a modification of the global PageRank algorithm, which computes a universal notion of importance of a Web page. Experiments in this work concluded that the use of personalized PageRank scores can improve a Web search. However, browsing patterns, bookmarks, and so on were not considered for the experiment.

B. Personalized web site

Personalized Web site is constructed using the contents that present in the web pages, the structure of the contents, the link topologies that are used in web pages. The link personalization and content personalization are the types in which the web site personalization takes place. The link personalization deals with the site URLs and the links given in those web pages. The content personalization involves in the content analyzing.

C. Recommender system

It has become increasingly difficult to search for useful information on the Web because the amount of information on the Web continues to grow. Therefore, we get the feeling of being overwhelmed by the number of choices. This is known as "information overload." An approaches to reduce this overload, recommender systems have emerged in domains such as E-commerce, digital libraries, and knowledge management. These systems provide personalized suggestions based on user preferences. This system collects feedback from the users in the form of finding the similarities between the contents, similarity between user profiles and rating the items in a particular domain.

III. PRIVACY PROTECTION

In [3] & [4], a search process involving many such interaction cycles, a user thus potentially reveals the following three kinds of personal information: 1. *User identity:* This could be a personal user ID in the case when the user has to register an account, or the IP address of the machine that the user is using. 2. *Queries:* This includes all the queries the user has

submitted to the search engine. 3. *Viewed results:* This includes all the viewed web pages by the user.

A. Level I: Pseudo Identity

Level I privacy protection (Pseudo Identity) in a personalized web search system ensures: a) The pseudo identity contains less information than the user identity. So the ID(U) is replaced by IDp(U). b) The TEXT (N; i) can be collected according to IDp(U) at the search engine side.

B. Level II: Group Identity

Level II privacy protection (Group Identity) in a personalized web search system ensures: a) A single user identity ID(U) is shared by group of users. b) The TEXT(N; i) is collected at the group level according to ID(U).

C. Level III: No Identity

Level III privacy protection (No Identity) in a personalized web search system ensures: a) The search engine is not able to find the user identity ID(U). b) The TEXT(N; i) cannot be collected on the search engine side, even at the group level.

D. Level IV: No Personal Information

Level IV privacy protection (No Personal Information) in a personalized web search system ensures: The user identity ID(U) is available or both user identity ID(U) and TEXT(N) is not available to the search engine.

E. Meta search Engine

There are quite a few Meta search engines on the Web such as Dogpile, Looksmart and ixquick .A meta-search engine sends user requests to several autonomous search engines and reranks search results returned from each one. When meta search engines is used, then autonomous search engines only receive all user queries from the single meta search engine. Thus there is the Level III privacy protection to that underlying autonomous search engine.

IV. PERSONALIZED WEB SEARCH

The personalized web search is takes place in three ways: (1) client side personalization, (2) Server side personalization and (3) client-server cooperative personalization.

A. Server side personalization

Level II privacy protection can be achieved. But when the search engine uses the user login ID to collect user information, this method will not achieve Level II privacy protection; when the search engine only uses the IP address to aggregate the user information, this method works.

Sometimes, search engines group users randomly or according to some criteria before they release the search engine logs. Then we will also have Level II privacy protection to those third parties which receive the search engine logs. It is impossible to implement Level III or Level IV privacy protection if personalization is done on the server side.

B. Client side personalization

A client-side personalized search agent can do query expansion to generate a new query before sending the query to the search engine. The sensitive contextual information is generally not a major concern since it is strictly stored and used on the client side and the overhead in computation and storage for personalization can be distributed among the clients. A main drawback of personalization on the client side is that the personalization algorithm cannot use some knowledge that is only available on the server side (e.g., PageRank score of a result document).

C. Client-server cooperative personalization

The user profile is still stored on the client side, but the server also participates in personalization. When a query is given to the search engine then the client extracts contextual information from the user profile. The combination of extracted information from the profile and query is sends it to the search engine. The search engine then does personalization with the received context. The contextual information sent to the server specifies the user's search preferences (e.g., query expansion terms, topic weight vector). This architecture provides the same level of privacy protection as server-side personalization. However, the personally identifiable information collectable on the server side is less than in the case of pure server-side personalization.

V. CONCLUSION

The above researches shows that many techniques available for making the web search as personalized one. The personalization involve in collecting the user interests implicitly or explicitly called as user profile. Collecting the user interests in explicit manner is not so easy. Because the users are not interested to provide their interest to the server. So we go for other methods to create user profile in implicit manner. The implicit profile creation considers the user documents in the user desktops, browser history and click through links. The profile creation using the document is not effective because the profile may very large, irrelevant words; a word with different meaning and etc .., explicit profile creation is the best to provide the good personalized search result. But the personalization through the profile creation leads to the loss of privacy to the user information. Providing the protection for user data from the profile is not defined in any of the above research. The personalized web search requires the privacy protection for the user profile.

REFERENCES

1. Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa," Adaptive Web Search Based on User Profile Constructed without Any Effort from Users" in WWW2007.
2. P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschutter, "Using ODP Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIRConf. Research and Development Information Retrieval (SIGIR), 2005.
3. A. Kobsa, "Privacy enhanced personalization", CACM vol. 50, no. 8, August 2007.
4. Y. Xu, B. Zhang, Z. Chen and K. Wang, "Privacy enhancing-personalized web search", *16th Int'l conf. world wide web (WWW) pp. 591-600, 2007.*
5. Krishnan Ramanathan, Julien Giraudi, Ajay Gupta "Creating hierarchical user profiles using Wikipedia," inHPL-2008-127.
6. M. Spertta and S. Gach, "Personalizing Search Based on UserSearch Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence(WI), 2005.
7. Xuehua Shen, Bin Tan, ChengXiang Zhai, "Privacy Protection in Personalized Search", in ACM SIGIR Forum Vol.41 No.1 June 2007.