VOL. 11, NO. 9, MAY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

PROFILE BASED PERSONALIZED WEB SEARCH USING GREEDY ALGORITHMS

B. SekharBabu, P. Lakshmi Prasanna, D. Rajeswara Rao, J. LakshmiAnusha, A. Pratyusha and A. Ravi Chand
Department of Computer Science Engineering, K L University, Andhra Pradesh, India
E-Mail: pprasanna@kluniversity.in

ABSTRACT

Internet usage is being increased as it provides information to all the users. The required information is retrieved by the search engines. The currently working search engines using sophisticated algorithms will not always provide relevant information to user's requirements. To resolve the issue, Personalized web search is used that will improve the quality of the search result by reordering the search results. This web search is done to provide relevant results using the user profile. The proposed UPS framework will dynamically generate a user profile for a user's query prioritizing the user's privacy. To acquire this we are implementing Greedy DP and Greedy IL Algorithms that are used for runtime generalization.

Keywords: personalized web search, greedy DP, greedy IL, UPS.

INTRODUCTION

A web search Engine is a software system that enables to search for the information on web (www) [2]. The information mined can be images, documents or any other type of file. The usefulness or performance of the search engine depends on the relevance of results it mines. Some search engines use ranking [5] method to the web pages for better results but the ordering varies from one engine to another. The search Engines mainly can be two types where the search engine creates a list of results by using software called CRAWL, a crawler based search engine like GOOGLE and the other depends on the human editors like YAHOO [3], Google directory generally said to be human powered directories. The directory search is good to reveal relevant results in a different topic search. If the user gives a query named "virus" to the search engine expecting the result of a malware program but the engine might present a result of a parasite or an infectious agent. Based on the perspective of the user the result is not acquired in such situations. The same continues with the query "Kingfisher" that can be a bird or an airlines resulting in occurrence of ambiguity [4]. The web information systems are thus facing many challenges. The data is being increased day by day resulting in information overloading is one of the major challenges. In such situations behavior and the content measures and considered. However all the search engines work with a specific goal of listing the relevant results implementing different algorithms?

Web Personalization will resolve some of the issues that were being faced by information systems. Generic search follows "one size fits all" model that cannot retrieve results dynamically like a personalized web search. Main goal is to retrieve the relevant results for any kind of user based on one's interest. The solution to personalized web search (PWS) can be profile based and click-log based. Click log based method is implemented by referring to a user's query history. But this can only work on repeated queries provided by same user which is a limitation. Profile based method is implemented by using user interest model that is generated by user profiling. But this method is unstable in some situations though it performs effectively for any sort of query. PWS can improve its effectiveness by generating an accurate user profile [10], gathering the information from history, bookmarks, used documents. By extracting such information may lead to loss of privacy. Privacy issue mainly rises from lack of protection to personal data. A PWS is said to be efficient only if it provides privacy to the user's profile. The loss of privacy can be observed in shopping behavior where a customer provides frequent shopper card disclosing his detailed profile. Thus people compromises privacy for their economic benefit. Another observation is that detailed information is not necessary to hold the user's interests at general level. Some unnecessary detailed information like location, time becomes noise in some search task. Personal data like emails, documents are mostly unstructured that is difficult task to measure privacy. It is necessary to summarize and organize the user's information to structured form [6]. Some things can be shared by one user where as the other may hide it. So the user must be given control over which information must be disclosed to the server. Personalization will provide such benefits to the user [7]: Offers an extendible way to build a hierarchical user profile on client side. Every user might not be interested to specify one's interests. Implementing an algorithm that will automatically collect the information that satisfied the goal. Hierarchical structure of user profile determines the more general at a higher level and the infrequent are at low-level. The profile is built by exploring histories, personal documents and emails.

Reduces difficulty to measure privacy. By the Hierarchical user profile the private information is hidden using the parameters. *MinDetail* determines the protected user profile. *ExpRatio,* measures the considerable private information that is closed or disclosed for a specific *minDetail.*

Related work

Previous works mainly focused on improving searches result on profile based PWS compared to click-

log method. Representation of profile is of many types like listings or key words to represent profile and a well known and well implemented hierarchical profile. Hierarchical profile is generated by analyzing the frequency of term on user data[19]. To build the profile, personalized information is required where the user provides it from the feedback, from the system where users specify their interest and also from the browsing history.

The solution for PWS in particular is in representation of profiles and adequacy measure of personalization. Some methods use a pack of words or vectors to interact to one's profile. Most of the profile representation is done with existing weighed point chain. [8]. Privacy protection problems are identified to be two classes [9] where one class treats personal or closed information to identify an individual and the other deals with data sensitivity considering it as privacy.

Existing system

Personalized web search

The previous works of privacy preserving PWS is not favorable. The existing PWS have problems: [1]

Run-time profiling is not supported in the existing profile-based PWS. "One size fits all" strategy is used. This approach will put the user's privacy at risk. A better approach for this is to make online decision that can choose the extent of exposure of user profile and whether to implement personalization of query.

Customization of privacy requirements is not considered in the existing methods. This leads to over protection of some user's privacy while other's lack protection. *Surprisal* based on information theory,is a metric used to detect sensitive topics presuming the interest with less document support are high sensitive. A counter example will misgive the assumption: If larger numbers of documents are browsed with the query "sex", the metric might conclude that "sex" is very general.

Iterative user interactions are required to create personalized search results in many PWS techniques. Refining search results is performed with some metrics that requires several user interactions like ranking [11], average rank, rank scoring. This interface is not practical for run time profiling.

Disadvantages in existing system

a) Usage of surprisal based on the information theory to identify sensitivity.

b) The existing profile-based PWS do not assist runtime profiling.

c) Customization of privacy requirements is not reflected.

d) Personalization techniques require iterative user interactions to produce personalized search results.

Proposed system

This Proposed UPS (User Customizable Privacy preserving Search) framework [1] [12] will generalize a profile for any query as per user privacy detail. Two simple and effective algorithms that support runtime profiling are GreedyDP (Discriminating Power) and GreedyIL (Information loss) mainly to make advantage of DP and to diminish the IL.

For better user interest results we can use re-ranking strategies in which results will be based on relevancy. Based on user interests, personalization system ranks the search results. Some of the following methods for re-ranking are:

Scoring methods: This method is responsible for assigning weights. Whenever a user visit the HTML page then a weight will be assigned to it. More the accessing keywords match between two users then more the rating will be scored.

Rank and visit scoring: More weight value will get top rank in the search results. Another method for re-ranking is it checks the logs of similar users. Based on the entries found snippets will be created which is considered as second most high ranking result.

Some of the generalized algorithms used for personalized web search:

Greedy algorithm: Brute force optimal algorithm is NP- Hard. Greedy algorithms namely GreedyDP, GreedyIL algorithms [1] better performs.

GreedyDP algorithm: In this algorithm we have an operator called prune leaf which indicates the removal of a leaf topic from a profile. Gredy DP algorithm works in bottom -up approach. But in this main problem is it requires recomputation of candidate profiles.

GreedyIL algorithm: In prune leaf operation, it reduces the discriminating power of the profile. It finds the information loss instead of discriminating power; it saves a lot of computational cost. Considering all this, GreedyIL is better than GreedyDP.

The Proposed Algorithms works efficiently for any search of query but the Experimental Results comparatively evaluates that which algorithm better performs in different situations.

PERSONALIZATION PROCESS

The Web personalization process can be divided into four distinct phases:

Collection of Web data - Implicit data includes past activities as recorded in Web server logs via cookies or session tracking modules. Explicit data usually comes from registration forms and rating questionnaires. Additional data such as demographic and application data (for example, e-commerce transactions) can also be used. In some cases, Web content, structure, and application data can be added as additional sources of data, to shed more light on the next stages.

Preprocessing of Web data - Data is frequently pre-processed to put it into a format that is compatible with the analysis technique to be used in the next step. Preprocessing may include cleaning data of inconsistencies, filtering out irrelevant information

according to the goal of analysis (example: automatically generated requests to embedded graphics will be recorded in web server logs, even though they add little information about user interests), and completing the missing links (due to caching) in incomplete click through paths. Most importantly, unique sessions need to be identified from the different requests, based on a heuristic, such as requests originating from an identical IP address within a given time period. Analysis of Web data - Also known as Web Usage Mining, this step applies machine learning or Data Mining techniques to discover interesting usage patterns and statistical correlations between web pages and user groups. This step frequently results in automatic user profiling, and is typically applied offline, so that it does not add a burden on the web server.

Decision making/Final Recommendation Phase - The last phase in personalization makes use of the results of the previous analysis step to deliver recommendations to the user. The recommendation process typically involves generating dynamic Web content on the fly, such as adding hyperlinks to the last web page requested by the user.

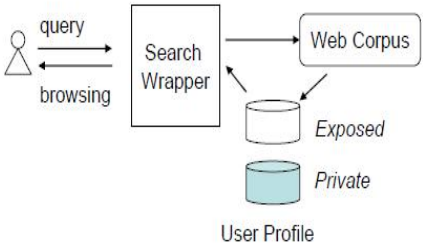
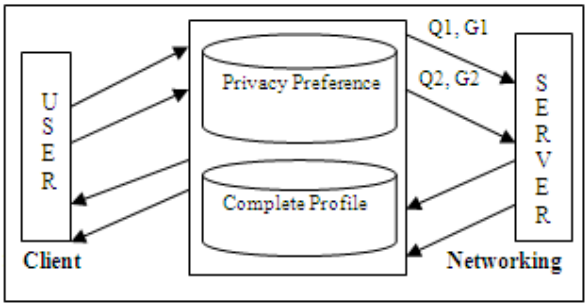


Figure. System overview.

OBJECTIVE:

The main objective of this project is to improve the quality search by maintaining the user's privacy control. The quality search is obtained by using the user profile built by using user's history and searches. The loss of sensitive data must be controlled during the process of query processing. The implementation of Greedy Algorithms namely Greedy DP and Greedy IL would evaluate the performance. A comparative graph between the algorithms will show the better performance of algorithms.

System architecture



The component to protect privacy is generating an online profile that is put into effect on a search proxy running on a client machine itself. This proxy will have the hierarchical user profile and customized privacy requirements.

Phases in this Architecture consists both online and offline phase. Hierarchical generation of user profile on client side and customized privacy requirements specified by the user are handled.

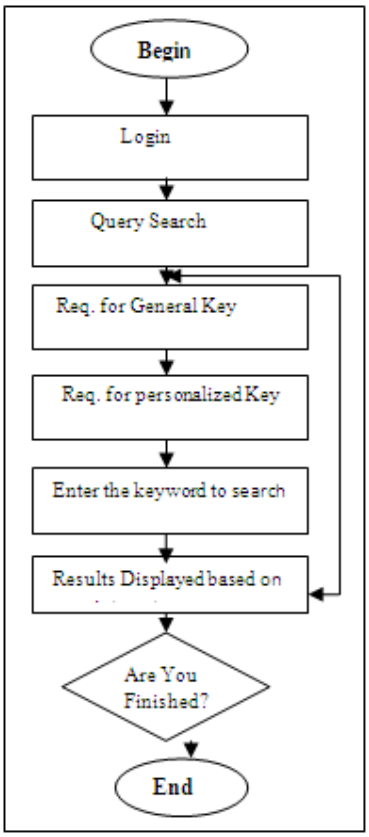
The above mentioned working and query handling is observed in online phase as:

1.User issues a query Q1 on the client, search proxy will generate a user profile in runtime resulting the generalized user profile G^1 fulfilling the privacy requirements.

2. Both the query and generalized user profile are sent to the server for the personalized search to retrieve the relevant results.

3. The result is personified with the profile and is sent to the query proxy where the proxy will present the results or re-ranks them according to user profile.

4. The Proposed Algorithms works efficiently for any search of query but the Experimental Results comparatively evaluates that which algorithm



RESULTS AND DISCUSSIONS

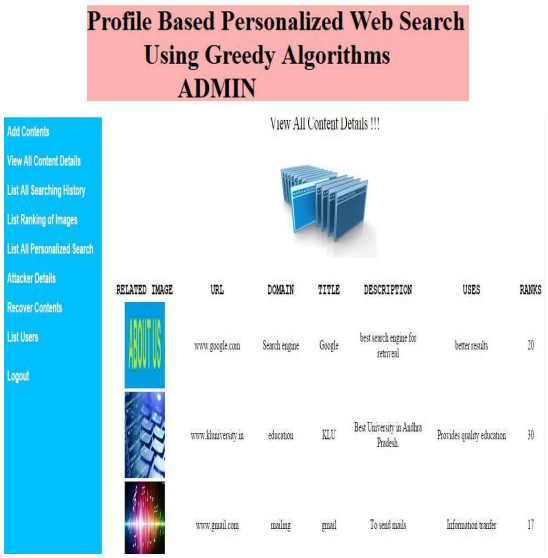
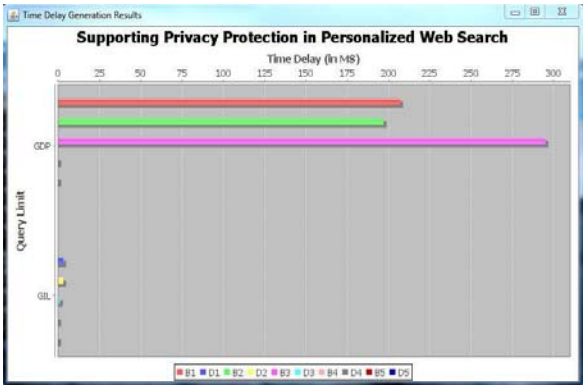


Figure. Add contents.

Figure. List all personalized search.

Results obtained after search

COMPARISON AMONG GREEDY ALGORITHMS



The above graph displays the comparative results of Greedy DP and Greedy IL where the time delay is high in Greedy DP compared to Greedy IL. Greedy IL performs best with respect to time.

CONCLUSIONS

Personalized search is a promising way to improve search quality. However, this approach requires users to grant the server full access to personal information on Internet, which violates users' privacy. In order to provide each user with more relevant information, several approaches were proposed to adapting search results according to each user's information need**.** Although there are several search engines currently present, it has been observed that they fails to capture user's preference and behavior and hence the search results may or may not be related with the profile of the user. A client side privacy protection framework called UPS i.e. User customizable Privacy preserving Search is proposed. 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. This achieves better search results protecting privacy. It is our firm belief that User Personalization will

add great value to the relevancy and the way in which web search is being performed.

REFERENCES

[1] Lidan Shou, He Bai, Ke Chen and Gang Chen. 2014. Supporting Privacy Protection in Personalized Web Search. IEEE. 26(2).

[2] Sandesh Pare, Bharati Vasgi. 2014. Personalization of the Web Search. 4(10).

[3] Charanjeet Dadiyala, Prof.Pragati Patil, Prof, Girish Agrawal. 2013. Personalized Web Search. International Journal of Advanced Research in Computer Science and Software Engineering. 3(6).

[4] Indu Chawla. 2011. An overview of personalization in web search. 978-1-4244-8679-3/11, IEEE.

[5] Mangesh Bedekar, Dr. Bharat Deshpande, Ramprasad Joshi. Web Search Personalization by User Profiling. First International Conference on Emerging Trends in Engineering and Technology.

[6] A.Jebaraj RatnaKumar. 2005-2010. An Implementation of Web Personalization using Web Mining Techniques. Journal of Theoretical and Applied Information Technology, JATIT.

[7] Yabo Xu, Benyu Zhang, Zheng Chen, Ke Wang. Privacy-Enhancing Personalized Web Search.

[8] Radhika. M, V. VIjaya Chamundeeswari, R.Ramya Devi. 2014. A Survey on Personalization Of web Search Using Generalized Profile. IRF International Conference.

[9] Neha Dewangan, Rugraj. 2014. Supporting Privacy Protection in Personalized web Search- A review. International Journal of Computer Engineering and Applications. VIII (II).

[10] Sandesh S. Pare, Bharati P. Vasgi. 2014. A Short Review on Different Personalization Schemes.International Journal of Engineering Research and Development. 10(3).

[11] J.Jayanthi, Dr. K S Jayakumar. 2011. An Integrated Page Ranking Algorithm for Personalized Web Search. International Journal of Computer Applications. 12(11).

[12] Ms. A. S.Patili, Prof. M.M.Ghonge, Dr. M.V Sarode. 2014. User Customizable Privacy-Preserving Search Framework-UPS for Personalized Web Search. International Journal of Research in Advent Technology. 2(4).

[13] Alexander Pretschner, Susan Gauch. Ontology Based Personalized Search. Supported by National Science Foundation Career Award number 97-03307.

[14] Jayanthi. J, Dr. K. S Jayakumar, Sruthi Surendran. 2011. Generation of Ontology Based User Profiles for Personalized Web Search. 978-1-4244-8679-3/11 IEEE.

[15] Fang Liu, Clement Yu, Senior Member, IEEEand Weiyi Meng, Member, IEEE. 2004. Personalized Web Search for Improving Retrieval Effectiveness. IEEE Transactions on Knowledge and Data Engineering. 16(1).

[16] Snigdha Gupta, Saral Jain, Mohammad Kazi, Bharat Deshpande, Mangesh Bedekar, Komal Kapoor. 2008. Personalization of Web Search Results Based on User Profiling. First International Conference on Emerging Trends in Engineering and Technology, 978-0-7695-3267-7/08 IEEE.

[17] Sarika N.Zaware, Deepali Shah, Taahir Hamgi, Payal Khinvasara, Rasika Tribhuwan, Supporting Privacy Protection in Personalized Web Search. 2014. International Journal of Sientific Progress and Research. 05(02).

[18] Yaowu Xu, Eli Saber, A Murat Tekalp. Hierarchical Content Description and Object Formation by Learning.

[19] P. Lakshmi Prasanna, P. Vidyullatha, B. Sekhar Babu, M. Narmadha, samhitharoy Bhuri, varsha kamireddy. 2015. Big data for Mobile applications in retail Market. ARPN Journal of Engineering and Applied Sciences.

[20] Tarannum Bibi, Pratiksha Dixit, RutujaGhula, RohiniJadhav. 2014. Web Search Personalization Using Machine Learning Techniques. 978-1-4799-2572-8/14 IEEE.