International Journal of Technology and Engineering System (IJTES)
Vol 8. No.1 – Jan-March 2016 Pp. 44-49
©gopalax Journals, Singapore
available at : www.ijcns.com
ISSN: 0976-1345

ENHANCING PERSONALIZED WEB SEARCH USING GREEDY ALGORITHM

A.Smilien Rophie,
PG Scholar,
Department of Information Technology,
Francis Xavier Engineering College,
Tirunelveli, Tamilnadu, India
smilienrophie@gmail.com

Dr. A. Anitha,
Professor,
Department of Information Technology,
Francis Xavier Engineering College,
Tirunelveli, Tamilnadu, India
dr.aanitha@yahoo.com

ABSTRACT

The web search engines are used to get the information from a large variety of data within a very less amount of time. For searching or getting any information all the users prefer web search engines. Some of the result returned by search engines are irrelevant to user. To provide useful and relevant results to user, it is proposed to design Personalized Web Search (PWS) engines. By using this PWS, the quality of the web search can be increased. The user profile consists of history, bookmarks, clicked links and follows the hierarchical structure. This user profile improves the search result. If personalization is done in server, it affects user privacy. To overcome this, the server give back their results and stored in PWS client. At that time, when the client update the query, then the results are get stored in the click through database on the client side. The server dose not know any user information because all the data can be stored only in the PWS client. By this framework, it attain the user privacy in a certain level. Greedy algorithms are runtime generalization. In this paper it is proposed to strike a balance between the personalization quality and privacy.

Keywords- *Privacy; Personalized Web Search; Profile; Greedy algorithm; click-through data;*

I. INTRODUCTION

World Wide Web contains large amount of data which makes information retrieval of a complicated task. The major objective of web mining is to introduce the new and also relevant information by analysing the structure, content and usage of web. It is also used to learn about web users and their websites interaction. The identity of various Web users can be captured by web usage data along with browsing behaviour.

Web mining can be divided into three types:

- Web content mining: It is used to scan and mine the text , image of a Web page. It determines the contents based on relevance to the query.
- Web structure mining: It is used to generate the structured summaries of the information on web pages by direct link connection of web pages.
- Web usage mining: All the web usage information can be kept stored and collected for analysing the user profile.

A. Personalization

The process of representing the correct information to the correct user at the correct time based on the user needs is called personalization. For learning the individual user, the system need to collect all personal information, store it, and analyze the results of the user profile . Nowadays the term personalization is used by everyone.

Because it retrieves only the needed and relevant information based on user's interest.. The PWS can be classified as click-log-based methods and profile-based methods[13]. In the click log based methods, the user click details are only stored. It also retrieve the same result whoever submits the query. But the profile based methods are improving the search quality along with user profile techniques. It can also increases the personal usage and behavioural data by gathering the information from query history, user documents, click-through data, etc.,

II. EXISTING SYSTEM

The personalization method needs iterative user interactions at the time of producing the personalized web search results. They actually refine all the personalized search results with any metrics such as average rank, rank scoring and so on. The existing method does not helps runtime profiling and also lacking in privacy concepts. It does not take the account of user privacy requirements. The user interests having very less document support then it is more sensitive and vice-versa. If all the links stored in the server side, then it affects the user privacy.

III. PROPOSED SYSTEM

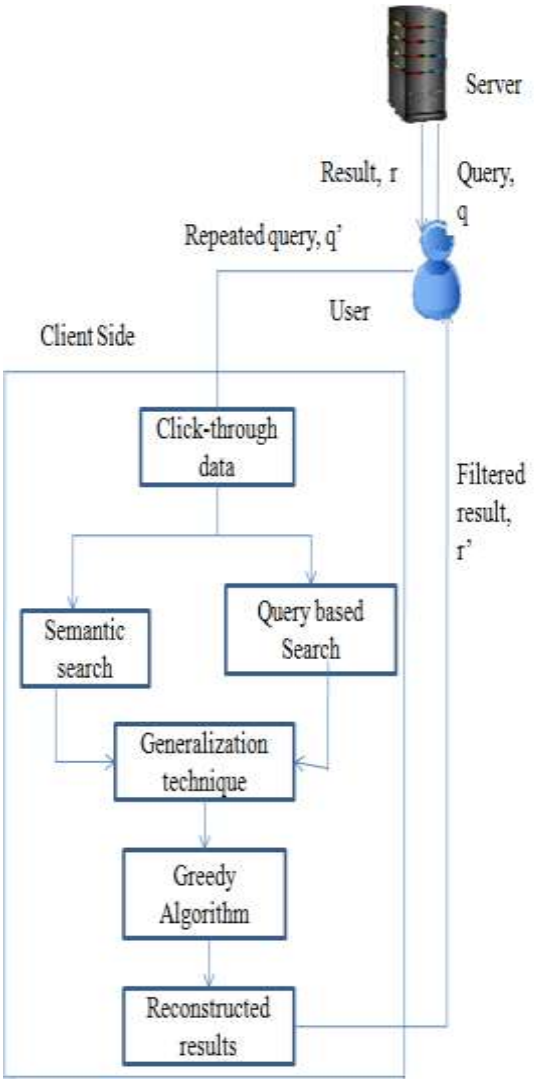

Fig. 1. System architecture of PWS

This web search system is user for searching the data in a personalized way. It also increases the quality of the web search.

The web search model is proposed in four major stages.

- User profile.
- Customized privacy requirements.
- Generalizing user profile.
- Online decision.

A. User Profile

The profile generator makes the client profiles that contains the user's interests. Every client profile receives a various levelled structure[13]. The client profiles store the whole subject area of human information.

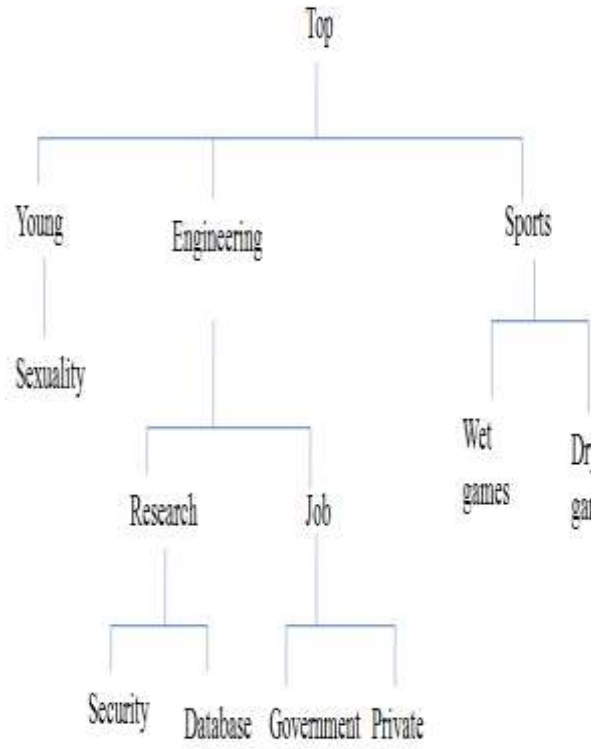

Fig. 2. User Profile

B. Click-through Data

The click–through data are strongly dependent between three sections of (q, r, c). The exhibited ranking r relies on upon the query q as controlled by the retrieval function executed in the search engine. Besides, the set c of clicked-on links fully depends on query q and exhibited ranking r. First, a user click on a link, if it is relevant to q . Based on the clicks and ranking it removes the unnecessary information. The main aim of click-through data is to store the relevant, user clicked links and it is also easy to collect the necessary information which is an optimal one.

C. Customized Privacy Requirements

In PWS, profile generalization is done by verifying previous user requirements on same query. Whenever a new query is submitted at the client side it is checked for personalization. If it is customized / personalized earlier, client will send reply. Otherwise server response is needed. In that case, the server give back their results and stored in PWS client. At that time, when the client update the query, then the results are get stored in the click through database on the client side. The server dose not know any user information because all the data can be stored only in the PWS client. By this framework, it attain the user privacy in a certain level.

D. Generalizing User Profile

Generalizing process wants to meet particular precondition to examine the user profile [10]. This generalizing user profile is unrealistic to perform offline generalization. The execution of offline generalization returns unnecessary information to the user query. A more adaptable solution needs online generalization which mainly depends only on the queries. Online generalization used to avoid irrelevant data and noisy data. The noisy data are removed based on stemming algorithm. The personalized search result can be reranked by RSVM technique. All the results are updated in the user profile. Because the results can be used by multiple times.

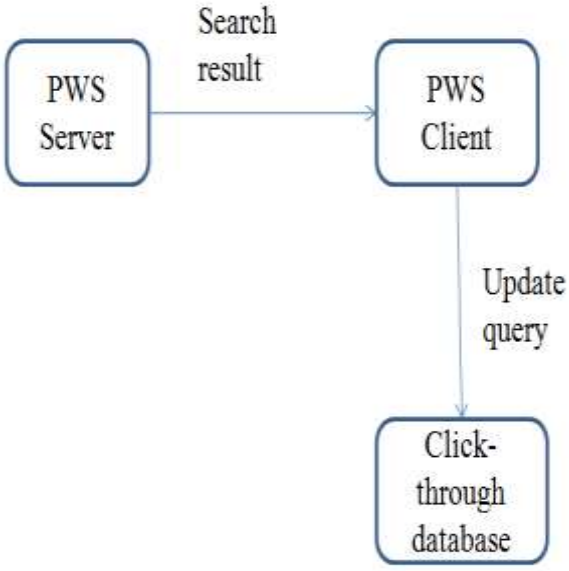

Fig. 3. Privacy Protection

E. Online Decision

- The user sends a query to the client, then the proxy maintains the user profile. At that time, the generalized user profile generated.

- If the query is a new query, then the server responds to the individual user.
- If the query is a previous search query, then that query and generalized user profile are sent back to the client for search the content along with personalization.

F. Greedy Algorithm

A greedy method is a problem solving technique that follows the local optimum for every stage to find out the global optimum. By using this greedy algorithm, the discriminating power can be maximised and information loss can be minimised.

1) GreedyDP Algorithm

In GreedyDP, from a user profile the leaf topic is eliminated by introducing the prune-leaf operation. After every iteration, the generalized profile can be maintained. By this method, it is proving to increase the utility of present iteration. The main drawback is that it needs to be recomputed for all candidate profiles.This causes enormous memory necessities and also computational cost.It helps online profiling primarily based on metrics of utility of personalization and privacy [1]. The best profile can be reached at every iteration.

2) GreedyIL Algorithm

The efficiency of the generalization can be improved by the Greedy IL algorithm. The discriminating power demonstrates monotonicity by prune-leaf operation [1]. If the threshold value is high, then the iteration can be reduced. The different steps of GreedyIL algorithm are as following:

Step 1: The leaf topic are removed by applying pruning operation like DP(q; G) ≥ DP(q, G').

Step 2: The priority queue maintains op = (t, IL (t, Gi)), leaf topic can be pruned from op and IL (t, Gi).

Step 3: This procedure goes until the delta-risk is reached.

Step 4: The expression (TS(q, G)) stays unchanged of all the operations of pruning.

Step 5: In C1, t comprise no siblings, whereas C2 contains siblings. It is easy to develop the case C1. **Step 6:** The case c2 merge the shadow with the leaf t.

Step 7: The candidate operator is updated in Q.

IV. RESULT

The benefit of GreedyIL over GreedyDP is more clear regarding response time. Because the GreedyDP needs more recomputation, which suffer much logarithmic operations. The issue exacerbates as the query turns out to be more ambiguous. By this way the GreedyIL leads to be more efficient than GreedyIL. The GreedyIL achieves the scalability but the GreedyDP grows in an exponential manner. So the GreedyIL increases the search quality also.

Fig. 4. Search Result

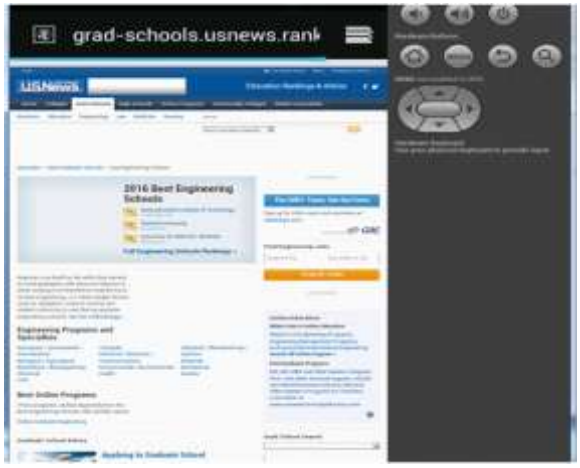

Fig. 5. Link of previous search result

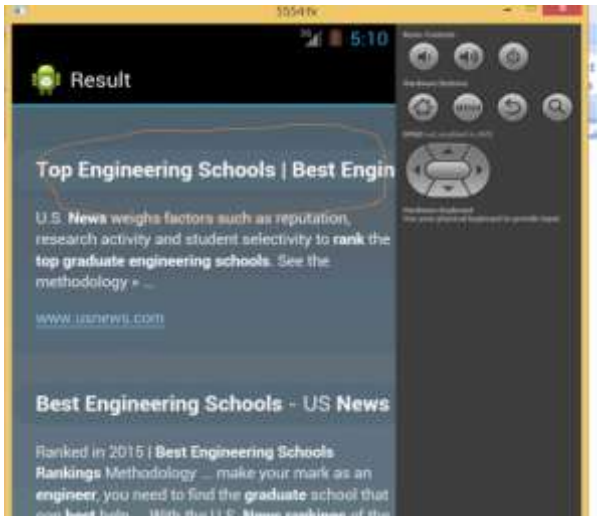

Fig. 6. Result for next relevant query

V. CONCLUSION

The use of Internet in the recent years is growing quickly which makes the need of a technique and they can give accurate and relevant results to the user. It gives an experience of personalized web search and eventually users can get what they want in a crisp manner in fewer time and fewer clicks as well. This project is mainly used for technically lagged people and also solve the web search problem by personalization. This project mainly focus on client side privacy method. This framework called UPS and it is used to increase the quality of search engines with the personalization application of the user. It can be utilized by any PWS that sum-ups the user profiles as a hierarchical levelled structure. Thus the search accuracy can be attained at a certain degree.

VI. REFERENCES

[1] Aruna V., Palanivel K. (2014), "A Privacy-Preserving Personalized Web Search(PWS) Framework UPS", Global Journal of Advanced Engineering Technologies, Vol. 3

[2] Castella Roca J., Viejo A. and Herrera Joancomarti J. (2009), "Preserving User's Privacy in Web Search Engines", Computer Comm., Vol. 32, No. 13-14, pp. 1541-1551.

[3] Chirita P.A., Nejdl W., Paiu R. and Kohlschutter C. (2005), "Using ODP Metadata to Personalize Search", Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development Information Retrieval (SIGIR).

[4] Dou Z., Song R. and Wen J.R. (2007), "A Large-Scale Evaluation and Analysis of Personalized Search Strategies", Proc. Int'l Conf. World Wide Web (WWW), pp. 581-590.

[5] Madhu G. and Govardhan A. and Rajinikanth T.V.(2011), "Intelligent Semantic Web Search Engines: A Brief Survey", International journal of Web & Semantic Technology (IJWesT) Vol.2, No.1

[6] Priyanka Z. and Todmal S.R. (2014), "Privacy Protection Using CPS Framework in Personalized Web Search", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 2, pp. 181-184.

[7] Qiu F. and Cho J. (2006), "Automatic Identification of User Interest for Personalized Search", Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736.

[8] Radhika M. and Vijaya Chamundeeswari V. (2015), "Privacy Protection In Personalized Web Search Using Generalized Profile", ARPN Journal of Engineering and Applied Sciences, Vol. 10, No. 7, pp. 4-17.

[9] Ramya V. and Gowthami S. (2014), "Enhance Privacy Search In Web Search Engine Using Greedy Algorithm", International Journal of Scientific Research Engineering & Technology (IJSRET).

[10] Sai Kumar V., Pavan Kumar P.N.V.S. (2015), "A UPS Framework for Providing Privacy Protection in Personalized Web Search", International Journal of Innovative Research in Science, Engineering and Technology Vol. 4

[11] Shen X., Tan B. and Zhai C. (2005), "Implicit User Modeling for Personalized Search", Proc. 14th ACM Int'l Conf. Information and Knowledge Management (CIKM).

[12] Shen X., Tan B. and Zhai C. (2007), "Privacy Protection in Personalized Search", SIGIR Forum, Vol. 41, No. 1, pp. 4-17.

[13] Shou L., Bai H., Chen K., and Chen G.(2014), "Supporting Privacy Protection in Personalized Web Search", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2

[14] Srivastava J., Cooley R., Deshpande M. and Tan P.N. (2006), "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", SIGKDD Explorations, Vol. 1, pp. 12-17.

[15] Teevan J., Dumais S.T. and Horvitz E. (2005), "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.

[16] Teevan V., Dumais S.T. and Liebling D.J. (2008), "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.

[17] Viejo A. and Castella-Roca J. (2010), "Using Social Networks to Distort Users' Profiles Generated by Web Search Engines", Computer Networks, Vol. 54, No. 9, pp. 1343-1357.

[18] Xiao X. and Tao Y. (2006), "Personalized Privacy Preservation", Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD).

[19] Xu Y., Wang K., Zhang B. and Chen Z. (2007), "Privacy-Enhancing Personalized Web Search", Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 591-600.

[20] Xu Y., Wang K., Yang G. and Fu A.W.C. (2009), "Online Anonymity for Personalized Web Services", Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 1497-1500.

[21] Yuan X., Song R., Wen J.R. and Dou Z. (2009), "Evaluating the Effectiveness of Personalized Web Search", IEEE Transactions On Knowledge And Data Engineering, Vol. 21, No. 8, pp. 1178-1190.

[22] Zhu Y., Xiong L. and Verdery C. (2010), "Anonymizing User Profiles for Personalized Web Search", Proc. 19th Int'l Conf. World Wide Web (WWW), pp. 1225-1226.

[23] Jaideep S., Cooley R., Deshpande M. and Tan P.N. (2006), "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", SIGKDD Explorations, Vol. 1, pp. 12-17.