International Journal of Computer Science and Software Engineering
Volume 1, Number 1 (2015), pp. 33-38

http://www.irphouse.com

Personalized Web Search With Supporting Privacy

Ashwini Suresh Lohar, Abhijit Kulkarni ,Rahul Ghotekar, Surekha Karanjkar
Guided by prof Shailesh Patil

Shri Chattarpati Shivajiraje college of engineering & Computer Department&
ShavitriBhai PhuleUniversity Pune,India
ashwinilohar41@gmail.com
surekhakaranjkar93@gmail.com
abhijitkulkarni205@gmail.com
rghotekar01@gmail.com

Abstract

Personalization is effective way of improving the quality of search result .Personalization means "Obtain the result according to each user's interest and needs.
Personalization deals with information overloaded. Web search is useful only context based and interest oriented. After the development of world wide web search engine become useful for retrieving the information from the web. They help in retrieving information as fast as possible but improvement is still necessary. Current search engine don't consider about the individual need of user and give the same result to all users. Search engines give information based on web popularity rather than user interest.

Keywords— Personalised Web Search, Domain Knowledge ,User Profile ,Enhanced User

I. INTRODUCTION

Amount of information on the web is increasing day by day .It has become major problem for all the users because even a good query can return not just tens, but thousands of "relevant" documents .Personalization is a way to find most relevant documents using information about user's goals, knowledge, preferences, navigation history, etc. As current search engine do not consider about the specific need of user and treat every user equally.' One sizes fits all' is main concept of generic search engine .When no of users give the same query same result will be given by search engine without considering what users exactly want. We can prove this fact by some examples: Suppose there are two users .One user is student who want information

about mouse which is computer hardware another user want information about mouse which is a rodent. For both the user generic search engine will provide the same result as same query was entered by both the users. But users want information according to their need in less time. If it is not so then he will get frustrate.

II. RELATED WORK

Micro speretta[1] proposed a system in which there is wrapper around the search site which helps in collecting information about user's search history and by the collected information user profile will be generated .All these profile are used for re-ranking the results of search engine .

Rakesh Kumar [3] identified that generic search engine are not fit for all and serve each user equally .Personalized search engine is promising way to solve the problems of getting irrelevant data. The terms like user profile modeling , enhancing the user profile they have used in [3] to get the relevant data from the search engines.

Chanchala Joshi [4] analysis the various drawbacks of Personalized web search .They have deals with various kind of search technique like Collaborative search engine ,User's location based search engine, Content based search engine, Hyperlink based search engine and they have found issues related to that technique .

Lidan Shou[6] Proposed a system client side privacy protection known as UPS (user customizable privacy preserving search) which can be adopted by any PWS(personalized web search)that captures user profiles in a hierarchical taxonomy.

III. FRAMEWORK

In our framework we have focused on individual interest of the person and we are providing the relevant web pages to the user. Because of this relevant pages user can save his time .

Our system is divided into several modules. This system considers user profile and domain knowledge.

User profile is based on user weblog and navigation and we are using domain knowledge in order to perform personalized web search. Domain knowledge will help us to store information on the basis on different domain/categories. We have proposed a suitable model which relevant information retrieval. Once the users inputs query ,the system provide a relevant data because of enhanced user profile and users search history will also be secure because of encryption on search data. Our model is also helpful as re-ranking is being done on the search data based on the enhanced user profile.

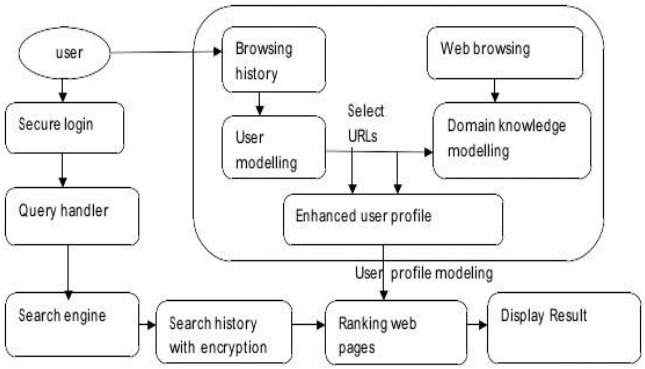

Figure 3.1 General Architecture of proposed Framework

A. ***Domain knowledge Modeling***

Domain knowledge is background knowledge which we are using for enhancing the user profile. By using DMOZ directory we are preparing domain knowledge. In this we will first crawled the web pages for some specified categories in which each categories are collection of URLs present in that categories.

After crawling, keywords will get extracted from crawled web pages, this collection will helps us in forming the vocabulary for the crawled web pages. Second term is category matrix which is nothing but weight of each term in each category .The weight is represented by frequency of each term in that category.

Table 3.1 Terms-Category Matrix(TCM) ref[3]

Term/Category	Cate1	Cate2	Cate3	……	Cate $_n$
t1	W11	W12	W13		W_{1n}
t2	W21	W22	W23		W_{2n}
t3	W31	W32	W33		W_{3n}
.					
T_m	W_{m1}	W_{m2}	W_{m3}		W_{mn}

B. ***User profile modeling***

User profile is created to get the knowledge of specific needs of the user .User profile helps in dealing with ambiguous queries. Alchemy API is used classify web pages. In this ,Alchemy API classifies a webpage by providing it particular category with some numerical value which shows it is belonging to particular category.

Because of using DMOZ for background knowledge, we have to map these Alchemy categories to DMOZ categories. Thus in our system , a User Profile is nothing but a category preference vector, in which weight of each category is used to represent user's interest in that category. As shown in the Figure 3.1, users

browsing history is used to generate user profile .When web pages browsed by the user grows above the given threshold, the learning agent updates user profile. User interest is represented by fix number of categories weights. It can be denoted by

U={cw1,cw2,cw3.....cwm}

Where, cwj will be the number of web pages of category visited by that user, normalized by maximum number of page visits among all categories.

By using Vector Space Model (VSM) we model user profile .All the web pages present in user's browsing history is taken into consideration. In this each web page is corresponds to a specific document .TDM (term document matrix) is output of vector space model and it represent each webpage/document as a feature vector of terms.

Here we consider each document as a URL

Table 3.2 Alchemy API to DMOZ Category mapping [ref 3]

Alchemy Categories	DMOZ Categories
Arts & Entertainment	Arts
Business	Business
Computer & Internet	Computers
Culture & Politics	Regional
Gaming	Game
Health	Health
Law & Crime	Society
Religion	Society
Recreation	Recreation
Science & Technology	Science
Sports	Sports
Weather	News

Table 3.3 Terms-Document Matrix(TDM) [ref 3]

	d1	d2	d3	…….	d_n
t1	W11	W12	W13		W_{1n}
t2	W21	W22	W23		W_{2n}
t3	W31	W32	W33		W_{3n}
.					
t_m	W_{m1}	W_{m2}	W_{m3}		W_{mn}

C. *Enhance user profile modeling*

In enhanced user profile each URL of user profile is matched with the Domain knowledge URLs and find the most relevant URLs which is then added to the Enhanced User Profile. From the domain knowledgebase find the cosine similarity of this URL with URLs present user specific categories from the domain knowledge .These URLs are ranked in decending order of cosine similarity .Average of cosine similarity of top 20 URLs are calculated .

From the top 20 URLs only those URLs are added to the enhanced user profile where cosine similarity value is greater then the average value.

$$Cosine(dj,u)=<dj*u>/||dj||*||u||$$

Cosine similarity is angle between the webpage in User pro file u and the document vector dj.

D. *Supporting privacy modeling*

In supporting privacy modeling ,encryption is done on the user's data to protect it from the others .Some cryptographic algorithms like DES (Data Encryption Standard) and RSA are used for providing encryption on user profile and search history to secure user data and made it confidential so that user will feel comfortable while searching .In Our system we are also applying login timeout facility so that user data can get protected .

IV. CONCLUSIONS

Information is crucial to the users and thus personalization is promising approach for searching the relevant document. By this Overall search efficiency get increased and also it reduces time wasted sorting through irrelevant users The main advantage of our system is that it resolves ambiguity. The search data is also safe as we are providing the encryption to the user profile and search history so that no one can misuse it .In future this framework can be applied for non textual data and collaborative filtering will also be applied for personalized web search.

REFERENCES

[1]. M Speretta and S Gauch, "Personalized Search Based on User Search Histories", Proceeding Of International Conference on Web Intelligence, pp. 622-628, 2005.

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

[3]. Rakesh Kumar ,Aditi Sharan .Personalized Web Search Using Browsing History And Domain Knowledge, 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)

[4]. Chanchala Joshi, Teena Jaiswal, Himanshu Gaur.An Overview study of personalized web search .International Journal of Scientific and Research

Publications, Volume 3, Issue 1, January 2013 1 ISSN 2250-3153

[5]. Zhomiyang Ma, Gautam Pant, and Olivia R. Liu Sheng The University of Utah. Interest-Based Personalized Search

[6]. Lidan Shou, He Bai, Ke Chen, and Gang Chen, "Supporting Privacy Protection in Personalized Web Search", IEEE Transactions on Knowledge and Data Engineering, vol.26, no.2, February 2014.