International Journal of scientific research and management (IJSRM)
\||Volume||3||Issue||7||Pages|| 3247-3251||2015||
Website: www.ijsrm.in ISSN (e): 2321-3418

Privacy Intensification And Profile Based Search In Web

S.Deivanai[1] , Mr.Pradeepkumar sahoo[2]
[1](cse,sri sairam college of engineering,india)
[2](cse,sri sairam college of engineering,india)

ABSTRACT: Personalized search engine helps to extract useful information from world wide web based on the users area of interest specified in the profile which helps to rank user preferred pages at the first.In this paper when an search item is entered by the user search engine will retrieve the results based on the previous search history of clicked pages which can be obtained by analyzing web logs present in the server and also area of interest specified in the profile. Page ranking is done by content based ranking algorithm which is done based on visitor count, page rank , access time length , feedback (likes/dislikes) ,Term weighing technique , https . Crawlers internet protocol address is monitored regularly by the administrator for providing security and an alert message is sent to crawlers mail if an unauthorized access happens .Location based search is done by automatically locating the latitude and longitude of the user by google apis and all theSearch items containing the addresses are located within the radius and reverse geocoding technique is used to convert Geographic coordinates into addresses.

KEY WORD—*Personalized web search(pws) ,Ranking, privacy , ontology mining*

I.INTRODUCTION

World wide web contains numerous amount of hyperlinks. Getting user preferred link at the top most is a tedious task . But doing so can help to save time consumption of search.[1][2][5][12].To outcome this issue many research has been done in this field such as community level using previous browsing history.[1][2][3][4][5][6] and finally personalized search based on area of interest present in user profile to find an solution .[11] but still there is a need for better enhancement of personalized search.

Example: This can be better explained with the following: a scientist and a chemical engineer may need information about "Mercury" even though their fields are entirely different our normal search engine will retrieve same kind of results to both of them . Here Scientist is searching “Mercury" as a planet and chemical engineer is searching mercury as an chemical but search engine will return only same kind of results for both of them.[12]

In our proposed system framework constructing the user profile is the primary step. In the next Step search is done by previously clicked through pages and details specified in the profile by giving consideration to the user feedback. Here maintenance of browsing history , user profile and feedback stored in server is done by admin. Even though querying based on user profile return personalized search results to all sorts of query there is an need for high level security to the sensitive information disclosed to avoid issues.

In our framework construction of user profile is the primary step. After constructing the user profile, search can be done based on user previously clicked through pages and details specified in the profile and giving consideration to users feedback.. Admin maintains the Browsing History , Profile of the users and feedbacks which are stored in the server. Querying based on user profile return personalized search results to almost all sorts of query. Need security for the sensitive information to avoid issues . [1]

In this paper we are going to enhance the security by sending alert messages to crawlers mail if an unauthorized person tries to access their account. web pages ranking is improved by using content based ranking algorithm which includes feedbacks, Term weighing technique,https,access time length ,page rank along with visitor count.

II. RELATED WORK

Quality of service in PWS is improved by construction of hierarchial user profile.[1] GreedyDp algorithm is used for better profile construction and greedy IL is used for providing security. Size of the query used to obtain the search results is very large. “one profile fits for all queries “barrier to return search results in an effective manner.

An bottom-up approach is performed to study the web dynamics based on users feedback.[2] Also, search can be done based on user preferences which helps to reduce go through multiple pages to retrieve the desired result. Extended page rank algorithm is used which consists of user preferences and link analysis. Better link analysis is achieved through this experimental research. The shortcoming in this system is it can be applied only to groups not based on individual users.

Automatic rule acquisition method is proposed which helps to mine relevant web sites rather than mining from scratch .[3] This involuntary rule acquisition process uses a rule ontology RuleToOnto, which signifies data about the rule components and their structures. The rule achievement process consists of the rule component identification step and the rule composition step. A* algorithm is used for rule composition, but there is no other approach which helps to state that experiment sounds better.

Determining the semantic resemblance between words is important for performing numerous tasks on the web such as comparative abstraction, community mining, document clustering, and automatic metadata extraction.[4] A new pattern extraction algorithm and a pattern clustering algorithm is used for finding resemblances between words. The best combination of page counts-based co-occurrence measures and lexical pattern clusters is studied using support vector machines. Besides, this method considerably improves the precision in a community mining task. But still we need the support at a level of individual search .

A personalized mobile search engine (PMSE) is proposed which captures the preferences of search users in the method of concepts through mining their data by click through methods[5] Using GPS location are tracked and ranking of user profile is done by means of ontology .For diversity between concepts and user preferences four entropies are introduced which helps to equilibrium the weights between the content and location concepts.in the client-server model, the client gathers and stores locally the click through data to guard the secrecy, while substantial tasks such as concept abstraction, training, and re-ranking are performed at the PMSE server. PMSE server uses two confidentiality parameters which helps to improve the accuracy.

III. SYSTEM ARCHITECTURE

The Proposed System architecture consists of two phases offline and online phases. Creation of Ontology Member is done in the offline phases.[14] After the profile creation. Ontology based user profiling technique of capturing user interest and using the server to extract content and location preferences[7][8][9] based on profile is done. Figure 1 shows the general process of our approach, which consists of two major activities such as Ranking and Updation of profile

When a user enters a query, the search results are obtained from the back end search engines such us Google, Yahoo. The results obtained from the search are combined and re ranked according to the user's area of interest specified in the profile. Content based ranking algorithm is used for ranking the web results which consists of https ,term weighing technique ,page rank, visitor Count, feedbacks(likes/dislikes),access time length.

In the online phase after the search results are acquired from the backend search engine , Feedback of the user is obtained frequently and query processing is done based on the profile content . Based on the number of likes and dislike count for a single link Feedback count is calculated by the administrator of the database and search preferences are updated .Search results are stored in the database which helps for future references.

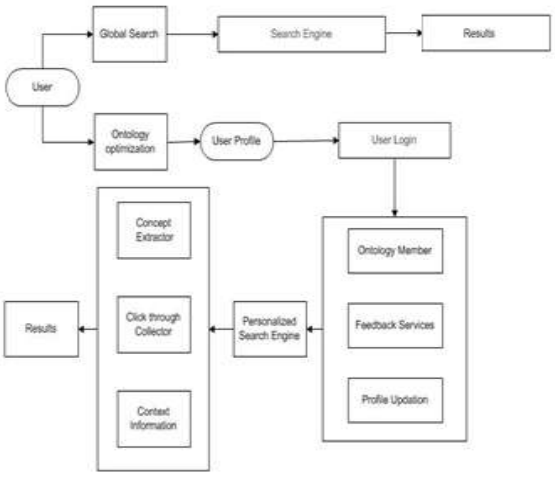

Fig1:Personalized web search architect

IV. PROPOSED SYSTEM

In our proposed system Framework, it consists of two phases online and offline phases. In the offline phase user profile is constructed hierarchicaly and in online phase query processing and privacy preservation is done. This Framework consists of *A) Ontology Mining B) Ranking* C) *Privacy for Sensitive Information D) Location based search*

A. Ontology Mining:

Ontology mining is a way of extracting concepts based on domain knowledge hiearachiay[10].[11][12][13] It consists of parent child

relationship, similarities and differences,(ie)Part of and kind of relationship between topics. We capture the following type of relationship for mining the content:

• Similarity: Two concepts may return similar kind of results which belongs to same topical interest.if (ci, cj) > δ1 (δ1 is a threshold), then ci and cj are considered as similar.

• Parent-Child Relationship: Hierarchical representation of relationship between concepts. Thus, if pr(cj|ci) > δ2 (δ2 is a threshold), we mark ci as cj 's child

B.CONTENT BASED RANKING ALGORITHM

Web pages retrieved using Various techniques should be Ranked. Here ranking helps to make the user preferred link at the top most which is done based on the following attributes .Ranking preference of the web page is based on the following attributes such us PageRank , Term Weighting Technique [TWT] User's Feedback , Visitor Count ,Access time length , Https.

The PageRank algorithm checks the entire link structure of the network and calculates the PR value of web page. The PR value of pages only depends on the number of in-links and out-links of a page.

There are three main parameters used in calculating TWT. The parameters are document length, document frequency and term frequency. The Content based ranking algorithm method takes user's feedback into account in the form of like and dislikes count. Like and dislike count are taken as the positive e or negative response respectively to the web page and helps to rank the web page. Number of Hits on the web page is considered as the visitor count. More popularity of a page depends on more number of hits on the page. Access time length Depends on how much time user is interested in spending time in visiting a single link. Encrypted sites are ranked first which provides secured connections to the user

B. PRIVACY FOR SENSITIVE INFORMATION

Privacy protection is done based on two metrics namely personalization utility and privacy risk.An generalized algorithm called GreedyIL (IL-Information Loss) is developed to minimize information loss and to protect privacy. For any kind of complex problem, greedy algorithm helps to find an optimal solution [1]for even an travelling salesman problems,it helps to take us quick decision. Quality of the Personalized search engine is determined by the discriminative power ie utility power and information loss ie security preservation. Sensitive topics needs security , so that topics has to be removed from exposing it to the server by forbidding the topics to unauthorized user this can be done by getting sensitive topics from user and setting guarding nodes .

For example: Given a user profile U , the sensitive nodes are a set of crawler specified sensitive topics S belongs to U . Here Man in the middle attack is possible by eaves dropping the sensitive information and they can utilize the information in their organization. Security enhancement is done by sending alert messages to client's mail when anonymized user access on account from unusual ip address and setting guarding nodes to the sensitive topics.

As initial stage, we introduce an operative _t called prune-leaf operative, which postulates the exclusion of a leaf topic t from a profile which seems to be a sensitive topic and setting forbidding option by using guarding nodes to the unauthorized user, if they try to access that sensitive topics. This operation can be denoted by $G_i \xrightarrow{-t} G_{I+1}$ this is called the method of pruning leaf t from Gi to obtain G_{I+1}.

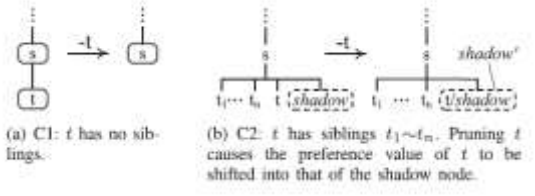

Fig 3: Prune Leaf Operation

This process of providing security to the sensitive topics doesn't provide security to its sibling topics (ie) in parent – child relationship ,security is given more to the parent topic rather than its child topic. But performing prune leaf operation reduces the utility power so in order to maximize the utility power and to reduce information loss we are going to maintain a priority queue of the prune leaf operation in descending order. The GreedyIL algorithm improves the effectiveness of the utility power based on the following findings.

Finding 1.

Whenever risk is satisfied iterative process of performing removal of sensitive topics and setting guarding nodes can be terminated.

Finding 2

Once a leaf topic t is pruned, we only need to recompute the IL(information loss) values for operators attempting to prune t's sibling topics.

Algorithim:

Greedy IL(H,q,R)
Input: Seed profile g;Query q;Privacy threshold μ

Output: Generalized Profile g* satisfying risk R
I the iteration index is initialized to 0 and Q be the priority queue of IP

Step-1:
online decision is made for whether to personalize the query q

Step -2:
if dp(q,R)<μ then
Obtain the seed profile g and insert (t,IL(t)) into Q for all t belongs to $T_H(q)$ and No need for performing prune leaf operation

Step-3:
While dp(q,R)>μ then
Perform a pop up operation called prune-leaf from Q
Process prune leaf $G_i \xrightarrow{-t} G_{i+1}$
If the process has no siblings then return generalized profile

Step-4:
Else if t has siblings then again perform prune leaf operation and merge t into shadow sibling.

Step-5
:Update the IL values for all operations on t's siblings in Q;
I←I+1;

Step-6:
Return generalized profile g*

D) LOCATION BASED SEARCH

Based on the user location and the search term nearby addresses containing the search term is filtered.Latitude and longitude of the search location is obtained by googleAPIs .Find Node(ls,k) is used for search for the k closest peers within the radius.Java Script object notation (JSON) is used for storing the data in the server and exchanging the data with the web page. Reverse geocoding is used to convert the geographic coordinates into addresses.

V .EXPERIMENTAL EVALUATION

Ranking of web pages is improved in the proposed system based on user feedbacks,access time spent for a single link by the user ,document strength ,number of inlink and outlinks of the webpage and visitor count. But in previous findings they have used page rank and visitor count which may make many fraud websites to increase their inlinks and outlinks and make their site rank at the top. Ranking of web pages for population explosion in india is done according to the algorithm is presented in a tabulated form is as follows:

S.NO	URL	COUNT
1	www.1bt.bridgeport.com	225
2	www.wikipedia.org/wiki/Demographics	178
3	www.wikipedia.org/wiki/Human_overpopulation	132
4	www.preservearticles.com	115
5	www.latimes.com/world/population	101
6	www.allprojectreports.com	56
7	http://goodpalhubpages.com	15
8	www.importantindia.com	5

Table 1: Rank preference calculation of each url

CONCLUSION AND FUTURE ENHANCEMENT

Personalized web search (pws) has been improved by using content based ranking which helps to make user preferred pages at the top most. Privacy protection is done by sending alert mail to the unauthorized user account and also by forbidding the access of the unauthorized user. Location based search is done based on latitude and longitude of the search location by google apis. In the feature enhancement personalized web search can be improved by sending their search details to others immedietly by instant messaging.

REFERENCES

[1]. Lidan Shou,He Bai,Ke Chen,And Gang Chen,"Supporting privacy preservation in personalized web search" IEEE Transactions On Knowledge And Data Engineering pp :453-467, Vol 26 , No.2 Feb 2014

[2]. Athanasios Papagelis and Christos Zaroliagis ,"A Collobrative decentralized approach to web search " IEEE Transactions On Systems, Man, And Cybernetics—Part A: Systems And Humans, pp : 1271-1290, Vol. 42, No. 5, September 2012

[3]. Sangun Park and Juyoung Kang "Using rule ontology in repeated rule acquisition from

similar websites" IEEE Transactions On Knowledge And Data Engineering, pp :1106-1119, Vol. 24, No. 6, June 2012

[4] Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka, "A Web search engine based approach to measure semantic similarity between words" IEEE Transactions On Knowledge And Data Engineering, p p : 977 -990 ,Vol. 23, No. 7, July 2011.

[5] Kenneth Wai-Ting Leung, Dik Lun Lee, and Wang-Chien Lee,"A Personalized mobile search engine" IEEE Transactions On Knowledge And Data Engineering, pp: 820-834, Vol. 25, No. 4, April 2013

[6] Alexandre Viejo, Jordi Castell`a-Roca, Oriol Bernad´o, Josep M. Mateo- Sanz ,"single party private web search" 2012 pp :1-8 ,Tenth Annual International Conference on Privacy, Security and threats

[7] Ms.R.Priyadarshini , S.Aishwarya, A.Ajaaz Ahmed "Search engine vulnerabilities and threats a survey and proposed solution for a secured censored search platform "Proceedings of the International Conference on Communication and Computational Intelligence – 2010,Kongu Engineering College, Perundurai, Erode, T.N.,India.27 – 29 December,2010. pp.535-53

[8] Kamlesh Makvana, Pinal Shah*Parth Shah," A Novel Approach to Personalize Web Search through User Profiling and Query Reformulation",data mining and intelligent computing 2014 international conference.

[9] Yufei Tao and Cheng Sheng ,"Fast Nearest Neighbour Search with Keywords" IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 4, April 2014

[10] Neha Batra , Ashok Kumar , Dr. Dheerendra Singh , Dr. R.N. Rajotia,"content based hidden web ranking algorithm" 978-1-4799-2572-8/14/$31.00_c 2014 IEEE.

[11]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)

[12] R Divya , Dr. C. R .Rene Robin ," Onto-Search: An Ontology Based Personalized Mobile Search Engine" IEEE

[13] Pretschner, A., Gauch, S "Ontology-based personalized search"Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence pp. 391—398 (1999).

[14] Kenneth Wai-Ting Leung , Dik Lun Lee , Wang-Chien Lee" Personalized Web Search with Location Preferences" ICDE Conference 2010.pp :701-712.

[15] TarannumBibi , Pratiksha Dixit "Web Search Personalization Using Machine Learning Techniques" 2014 IEEE International Advance Computing Conference (IACC) 1299

[16] Ashish Nanda , Rohit Omanwar, Bharat Deshpande "Implicitly Learning a User Interest Profile for Personalization of Web Search using Collaborative filtering" 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)

[17] Hannarin Kruajirayu, Ake Tangsomboon and Teerapong Leelanupab "CoZpace: A Proposal for Collaborative Web Search for Sharing Search Records and Interactions" 2014 Third ICT International Student Project Conference (ICT-ISPC2014)