



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

# Review on Secure Advanced Web Search Personalization Using Domain Knowledge

Amrish A. Patil, Prof. Amrit Priyadarshi

Student, Dept of I.T. , Dattakala Group of Institution's Faculty of Engineering, Swami-Chincholi, Daund, Pune-,  
Maharashtra, India

Asst. Professor, Dept of I.T., Dattakala Group of Institution's Faculty of Engineering, Swami-Chincholi, Daund, Pune  
Maharashtra, India

**ABSTRACT:** General web search engine are generally used for gathering huge information from web. However these internet users generally does not require generic model. Every web users wants specific information that he wants. In this paper we are attempt to improve customized web search and additionally privacy of search query. Customer's Profile gives a basic data to performing customized web search. We propose a PWS system. Where we are going to make propelled client profile utilizing client browsing history, and advancing it with domain knowledge, learner programs are used to learn the user interest. User profile are created and stored at Server side To accomplish security and privacy we are going to encrypt the client profile, and stored at server side. We likewise give the security to the query which is asked for from customer side, by encrypting it at customer side and decode at server side. OWASP security guidelines rules are taken after while outlining the framework. We likewise give online framework to choose which query session is run of the mill and which one is atypical. That is framework will naturally choose which question to be customized.

**KEYWORDS:** Personalized Web Search, User Modelling, Domain Knowledge, Enhanced User Profile, Confidentiality, Privacy

### I. INTRODUCTION

The web search engine is good source for ordinary people to looking for useful information on the web. However, users experiences are sometimes bad when search engines return results that do not match with its needs . Such irrelevance is largely due to the enormous variety of users' contexts and backgrounds, as well as the ambiguity of texts. Personalized web search (PWS) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the cost, client data must be gathered and dissected to make sense of the client goal behind the requested query. The answers for PWS can for the most part be categorized into two sorts, to be specifically click log-based methods and profile-based ones. The click log based methods are clear—they just force inclination to clicked pages in the client's query history. In spite of the fact that this procedure has been exhibited to perform reliably and impressively well, it can just deal with repeated queries from the same client, which is a strong limitation confining its applicability. Interestingly, profile-based methods enhance the search experience with confused client interest models created from client profiling strategies. Profile-based methods can be possibly compelling for a wide range of questions, however are accounted for to be insecure under a few circumstances. Although there are pros and cons for both types of PWS techniques, In fact, privacy concerns have become the major barrier for wide proliferation of PWS services.

A typical search engine do not properly handles the ambiguity of various search query for example if one user wishes to search computer virus and another user searches for virus from medical background, both user get the information about computer as well as medical virus Here both user may feel unhappy because of unwanted ads and search results because service providers revenue is depends on the ads. One way to solve this ambiguity is provide more personal information to server so that user will get the appropriate result as per the requirement. To perform Personalized Web



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

search it is essential to model User's need/interest. Development of user profile is an essential part for customized web search.

User profiles are built to model user's need focused around his/her web use information. We are going to user server side profiling. One of the natural issues with personalized search is that clients are often frail about giving private or individual data in regards to themselves to search providers. Naturally, the more that a search provider thinks around a particular client, the more precise their search results can be custom-made for them, yet how are the clients to trust that the data that the search provider keeps up about them won't be misused, lost, or vindictively utilized? We also aim to make user profile more precise and advance by using users browsing history and Domain knowledge and in order to enhance the privacy, this paper will look at philosophies and methods to optimize the privacy that users are given when using a typical personalized search service, User Profile Encryption schema is used to achieve confidentiality and privacy. Here we assumed possible security threats for example A2-Broken Authentication and Session Management, A3-Cross-Site Scripting (XsSS), A6-sensitive data exposurer, A1-Injection and to cope with it we are using security guideline from OWASP top 10. it also gives the protection from possible pollution attacks. And to differentiate typical session and atypical query session we are going to use query and user profile based on this system will decide which query should be personalized.

## II. RELATED WORK

In this section, we overview the related works. We focus on the literature of various web search personalization and privacy techniques. In 2014 Rakesh Sharma and Aditi sharan[15] proposed a framework for PWS. which uses browsing history to create the user profile. Domain knowledge is used to store different categories which is prepared by DMOZ. information prepared in user profile are categorized in to this DMOZ categories and to do this alchemy API is used. To create Advanced user profile first we take url(document) from user profile, add this url into advanced user profile find the cosine similarity of this URL with urls present into specific domain, retrieve first 20 url after arranging in descending order. then calculate average cosine similarity of this 20 url. find out the url which have value greater than average value and add it to the enhanced user profile. This framework give good result when we compare normal user profile with enhanced user profile.

Another author[1] have proposed a personalized web search show that unites community based and substance constructed confirmations arranged in light of novel ranking procedure. The author attempted to explain this issue through this model which deliver results on the premise of inclination and enthusiasm of the client. In this paper, author proposed an uncommon way to deal with find the investment and slant of the customer. It's a two way approach, first it will find out the activities of customer through his/her profile in social networking sites. Moreover, it will check information from what the long range social communication site provide for the customer through companions and community. In light of the results, customer's interest will be organized by the web search or it is personalized.[13] has proposed a method that the Client profiles, depictions of client interest, can be utilized via search engines to give personalized search results. Numerous ways to deal with making client profiles gather client data through proxy servers (to catch perusing histories) or desktop bots (to catch exercises on a PC). Both these methods oblige investment of the client to install the intermediary server or the bot. In this study, we investigate the study of a less-obtrusive method for social occasion client data for personalized search. Specifically, we assemble client profiles in light of movement at the search site itself and study the utilization of these profiles to give personalized search results. By actualizing a wrapper around the Google search engine, we had the capacity gather data about individual client search activities. Specifically, we gathered the questions for which no less than one search result was inspected, and the pieces (titles and rundowns) for each analyzed result. Client profiles were made by characterizing the gathered data (inquiries or pieces) into ideas in a reference idea order. These profiles were then used to re-rank the search results and the rank-order of the client inspected results prior and then afterward re-positioning were looked at.

G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao proposed method in [14], they shows that Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' is not feeling comfortable for reveling their private information to server, this is major problem of personalized web. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We have propose a PWS framework called UPS that can generalized profile at runtimes per



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

user security requirement. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a query is beneficial.

## III. PROPOSED SYSTEM

### A) *Description of the Proposed system:*

In this paper we try to enhance personalized web search as well as privacy of search. Client's Profile gives a critical information to performing personalized web search. We propose a PWS framework. Where we are going to create advanced user profile using user browsing history, gathered by learning agents and enriching it with domain knowledge, we map the user interest into various DMOZ categories, to find out most interest. Weight of each categories shows user interest in that category, and this information is used to find out most relative urls for current user query. User profiles are later encrypted to achieve security. We also provide the security to the query which is requested from client side, by encrypting it at client side and decrypt at server side. Result will also encrypted at server side and decrypted at client side. OWASP security guidelines are followed while designing the system. We also provide online system to decide which query session is typical and which one is atypical. That is system will automatically decide which query to be personalized

## IV. SYSTEM ARCHITECTURE

### A) *Domain Knowledge Modeling:*

Domain Knowledge Modeling information is the foundation learning that we used to upgrade the client profile. The source which we have utilized for preparing Domain Knowledge is DMOZ directory. To prepare Domain Knowledge, first we have crawled the website pages from DMOZ directory for some pointed out classifications, where every class is spoken to by gathering of URL's available in that classification. We have extricated the keywords from the web sites and build the term category matrix. Presently we structure a term category matrix, which determines weight of each one term in each category. The weight nothing but occurrence of the term in that category. Apache Nutch server and Apache Solr server has been used for crawling and indexing purpose.

### B) *User Profile Modeling:*

User profile shows the user interest in particular domain. To create user profile we analyze web search history and assign it in various domain categories. AlchemyAPI has been used for classifying web pages. AlchemyAPI classifies a web page by giving it a particular category along with confidence (numerical value) which shows its probability of belonging to that particular category. If the web page is classified with confidence above the specified threshold level then we confirm that page is belongs to that category. We are using DMOZ directory as domain knowledge later we map the AlchemyAPI categories into DMOZ categories. For creating user profile vector plus model is used.

### C) *Enhanced User Profile*

Enhanced User Profile is an important part in our framework. An Enhanced User Profile make user profile more powerful by adding more relative url's from domain knowledge. For preparing the Enhanced User Profile we have considered each URL of the User Profile, match it with Domain Knowledge URLs and add most relevant URLs to the Enhanced User Profile. Following steps explain the process of preparing the Enhanced User Profile. Perform the following steps for each document (URL) in user profile: Select the URL from the User Profile. Add the URL to the Enhanced User Profile. Find the cosine similarity of this URL with the URLs present in user specific categories from the Domain Knowledge base. Rank the URLs on descending order of cosine similarity. Retrieve top 20 URLs. Calculate the average of the cosine similarity of these top 20 URLs. From the top 20 URLs add only those URLs to the enhanced user profile whose similarity value is above the average value. To summarize the process, for each URL (form user profile) most relevant URLs from the user specific Domain Knowledge category are added to prepare enhanced user profile. The cosine formula used for the similarity of the URL  $u$  in User Profile to each web pages  $d_j$  in Domain Knowledge is as follows: A cosine similarity measure is the angle between the web page in User Profile  $u$  and the document vector  $d_j$ .



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

## D) Privacy And Confidentiality Of User Profile

Security and confidentiality are the main issues for PWS we are focusing these issues by encrypting user profile at server side. RSA encryption algorithm with key size of 2048 is used for this purpose, bigger key size gives good protection against external and internal threats. Security is also provided to the data transportation between client and server by using SSL. While developing a system we followed all security guidelines from OWASP top 10 security cheat sheet. This framework provides the security from various vulnerabilities for example Broken Authentication and Session Management, Cross-Site Scripting (XsSS), sensitive data exposer, and Injection.

## V. RESULTS

In our Experiment, we have used the browsing history of different users from our collage, we have crawled the datasets from DMOZ for the selected topics using Apache Nutch, while Apache Solr has been used for indexing crawled pages. By setting crawling parameters of Nutch we have restricted the crawling to specific DMOZ topic. We tested same query on normal search engine and compared result with our system. Our system suggested more related url than normal search engine. We also analyzed the result which is generated by normal user profile and advanced user profile. Advanced user profile gives better result than normal user profile.

## VI. CONCLUSION AND FUTURE WORK

In this paper we have proposed a framework for secure personalized web search. Here we build the user profile by using domain knowledge. We also proposed a method to maintain the privacy and confidentiality by encrypting the user profile at the server side. Security is also provided to transportation of the data. Our system is also able to detect the which session is typical and which is not. We performed some experiments that show better search result when we use advanced user profile as compared with simple user profile on same queries. In the future we would try to enhance the search quality based on user search preference and also aim to provide more security from the adversaries.

## REFERENCES

1. O. Shafiq, R. Alhajj and I. G. Rokne, "Community Aware Personalized Web search", International Conference on Advances in Social Networks Analysis and Mining, pp. 3351 - 3355, 2010
2. G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao, "Ups: Efficient Privacy Protection in Personalized Web Search," Proc. 34th Int'l ACM SIGIR Conf. Research and Development in Information, pp. 615-624, 2011.
3. Yabo Xu, Benyu Zhang, Zheng Chen, KeWang, "Privacy-Enhancing Personalized Web Search", WWW 2007, May 8-12, 2007
4. Namita Mittal, Richi Nayak, MC Govil, KC Jain "A Hybrid Approach of Personalized Web Information Retrieval" 010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 978-0-7695-4191-4/10
5. Carsten Eickhoff, Kevyn Collins-Thompson "Personalizing Atypical Web Search Sessions" *WSDM'13*, February 4-8, 2012.
6. K.W.T. Leung, D.L. Lee and Wang-Chien Lee, "Personalized Web search with location preferences", IEEE 26th International Conference on Data Engineering, pp. 701 - 712, 2010 .
7. Zhicheng Dou, Ruihua Song, Ji-Rong Wen, and Xiaojie Yuan "Evaluating the Effectiveness of Personalized Web Search" *Ieee Transactions On Knowledge And Data Engineering*, Vol. 21, No. 8, August 2009
8. Namita Mittal, Richi Nayak, MC Govil, KC Jain "A Hybrid Approach of Personalized Web Information Retrieval" 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.
9. J. Teevan, S.T. Dumais, and D.J. Liebling, "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, 2008.
10. F. Qiu and J. Cho, "Automatic Identification of User Interest for Personalized Search," Proc. 15th Int'l Conf. World Wide Web (WWW), pp. 727-736, 2006.
11. Wei Meng, Xinyu Xing, Anmol Sheth, Udi Weinsberg, Wenke Lee "Your Online Interests - Pwned! A Pollution Attack Against Targeted Advertising" CCS'14, November 3-7, 2014, Scottsdale, Arizona, USA.
12. Milad Shokouhi, Ryan W. White, Paul Bennett, Filip Radlinski, "Fighting Search Engine Amnesia: Reranking Repeated Results", SIGIR'13, July 28-August 1, 2013
13. M Speretta and S Gauch, "Personalized Search Based on User Search Histories", *Proceeding Of International Conference on Web Intelligence*, pp. 622-628, 2005
14. G. Chen, H. Bai, L. Shou, K. Chen, and Y. Gao "Supporting Privacy Protection in Personalized Web Search" *iee transactions on knowledge and data engineering* vol:26 no:2 year 2014.
15. 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)



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

Vol. 3, Issue 7, July 2015

## BIOGRAPHY

**Amrish A.Patil** is a Master of Engineering student in the Information Technology Department, Dattakala Group Of Institution's Faculty of Engineering Swami-Chincholi, Daund. He received B.E.in Information Technology degree in 2010 from DKTE , Ichalkaranji , MS, India. His research interests are Computer Networks (wireless Networks), Data mining, Security etc

**Prof. Amrit Priyadarshi** is an Assistant.Professor in Information technology department, Dattakala Group Of Institution's Faculty of Engineering Swami-Chincholi, Daund, Pune-413 133.He received Masters in Engineering (CSE) degree, and having 7 years experience in teaching. His research interest are Database System, Data Mining and cloud computing.