Journal of Theoretical and Applied Information Technology
20th September 2015. Vol.79. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PRIVATE SEARCHING KEYWORD FREQUENCY USING HOMOMORPHIC ENCRYPTION

[1]N. ANUSHA, [2]M. SUGANYA

[1]Assistant Professor,Department of computer science and Engineering,
Sathyabama University,OMR Road,Chennai,India

[2]PG Student Computer Science and engineering, Sathyabama University,
OMR Road, Chennai, India

E-mail: [1]anusha.nallapareddy@gmail.com, [2]msuganyabe2007@yahoo.co.in

ABSTRACT

With the rapid growth of internet technologies, the web has become the world's largest repository of knowledge. So, it is challenging task of the webmasters to organize the contents of the particular websites to gather the needs of the users. Optimised search engine is used for effectively searching keyword queries which are frequently visited by the user based on their interest. The most frequently used keyword queries are encrypted using homomorphic encryption which are semantically secured through searching mechanism. This paper presents a new framework for a semantic-enhanced Web Page Recommendation (WPR), and a suite of enabling techniques which include semantic network models of domain knowledge and Web usage knowledge, querying techniques, and Web-page recommendation strategies.Time based ranking technique is used to calculate the time taken by each user for visiting the recommended web-pages and also to keep record of the number of users. The privacy of the user is thus enhanced by using the clustered web-pages on the search engine. The proposed technique also provides a set of clustered web-pages which are effectively filtered through private searching keyword query and also provides better web-page features in order to satisfy the user's requirements.

KeyWords: *Semantic network, Web-Page Recommendation(WPR), Domain knowledge, knowledge representation, Web usage mining, Homomorphic encryption.*

1. INTRODUCTION

Web users are overwhelmed due to the rapid development in various websites which provides limitless information that is useful in day to day life. The web users spends long time to get the relevant web-pages by browsing irrelevant websites that provides huge amount of information on the WorldWideWeb(WWW). In order to predict the requirements of the web users and to evaluate the interest of the active users, recommended engine is employed so as to provide recommended lists of web-pages. In large-scale Web sites, it is typical that the content served to users comes from multiple Web or application servers. Many unauthorized users poses threat to the information provided by the server by accessing the web-pages without prior permission. To avoid such threats from occurring the users are requested to get registered with the server so as to access the recommended lists of web-pages which are suggested in an efficient manner.

In some cases, multiple servers with redundant content are used to reduce the load on any particular server. The overall Web usage mining process can be divided into three inter-dependent stages: data collection and pre-processing, pattern discovery, pattern analysis. Much of the research and practice in usage data preparation has been focused on pre-processing and integrating these data sources for different analysis. Usage data preparation presents a number of unique challenges which have led to a variety of algorithms and heuristic techniques for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pre-processing tasks such as data fusion and cleaning, user and session identification,page view identification.

2. RELATED WORK

[1] Recommends better web-pages by using semantic information from the knowledge gained by combining the domain and web data taken from such web-pages.[2] Provides web services in an more personalized and user-friendly manner. The domain knowledge of certain websites are not properly grouped together and thus clustering of similar websites become a difficult task to implement.[3] Implements more expensive models so as to provide effective recommender system that mines the data from a variety of channels in an personalised manner.[4] Captures variable length histories of the web pages to accurately find the user navigation on the web. The semantics of the webpage are not analysed properly in the probability models.[5] Develops a tool for the domain ontology which is based on the domain knowledge, concepts used in the ontology and taxonomic hierarchies.

[6] Describes the web personalization by integrating domain ontologies and the usage patterns from the websites. The semantic similarities among different ontologyies cannot be measured properly.[7] Performs mining process on the semantic web in order to enhance the quality of the mining results which are based on the interest and correlation measures of the visitor's browsing preferences.[8] Calculates the similarity between the web usage information and the domain knowledge in the context of the web page in order to improve the ranking of the recommendations.[9] Adopts a tree structure for storing frequent sequential patterns which are mined from the web log. However ,the average frequent sequence of the web access becomes longer and the database becomes larger.[10] Uses ontology of the website based on the concepts and terms extracted from the documents .The online recommendation are generated by semantically searching the frequent web pages from the mining process.[11]Performs recommendation process based on the acquisition of knowledge about the system ontology in multidimensional space.

3. SYATEM ANALYSIS

In Web usage mining in the recommendation system, the most basic level of data abstraction is that of a page view. A page view is an aggregate representation of a collection of Web objects contributing to the display on a user's browser resulting from a single user action (such as a click-through). Conceptually, each page view can be viewed as a collection of Web objects or resources representing a specific "user event," e.g., reading an article, viewing a product page, or adding a product to the shopping cart.Session is the most basic level of behavioural abstraction at the user level. A session is a sequence of page-views by a single user during a single visit.The site content data also includes semantic or structural meta-data embedded within the site or individual pages, such as descriptive keywords, document attributes, semantic tags, or HTTP variables. The underlying domain ontology for the site is also considered part of the content data. Domain ontology's may include conceptual hierarchies over page contents, such as product categories, explicit representations of semantic content and relationships via an ontology language.

The most frequently recommended web-pages are marked by number of users currently visited and time taken to visit a particular website. The information about each user and their frequently visited webpage history are automatically updated in the database which is maintained by the server. The homomorphic encryption is used to encrypt the URL (Uniform Resource Locator) of user's frequently visited webpage on the search engine in order to provide privacy on search query so that the unauthorised users are not allowed to view or access the web-pages from the server.The homomorphic encryption technique plays an important role in protecting the information about private search keyword query of the web user such as Internet Protocol (IP) address of the system and user id. To view the encrypted web-page information, the web user should send request to the server and get permission to access the webpage in an decrypted form from the server. Time based ranking technique is used to calculate the time taken by each user for visiting the recommended web-pages and to also keep record of the number of users.

Journal of Theoretical and Applied Information Technology
20th September 2015. Vol.79. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. SYSTEM ARCHITECTURE

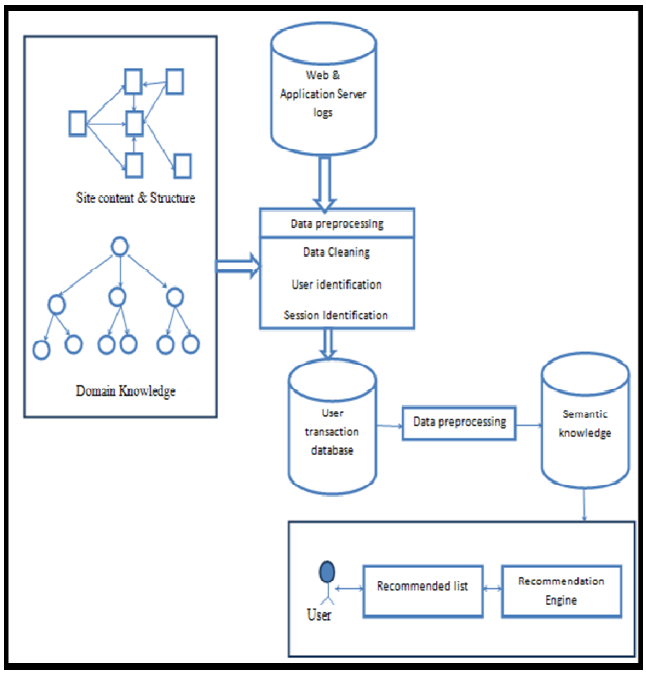

Figure 1: Architecture Diagram Of Recommendation Engine

Figure 1. describes web-pages extracted from the domain knowledge and semantic knowledge of website stored in the web server logs. The log data is automatically collected by the web and the application server represents the navigational behaviour of the web user. Each log entry in the server contains the time and the date of the hypertext transfer protocol(HTTP) request and internet protocol(IP) address of the client, the resource requested parameters invoking web application, request status and the user agent(browser, operating system type and version). The data pre-processing of a web-page includes three process which are data cleaning, session and user identification.The recommendation engine provides recommended lists of websites after data pre-processing steps to the user.

5. DATA FLOW DIAGRAM

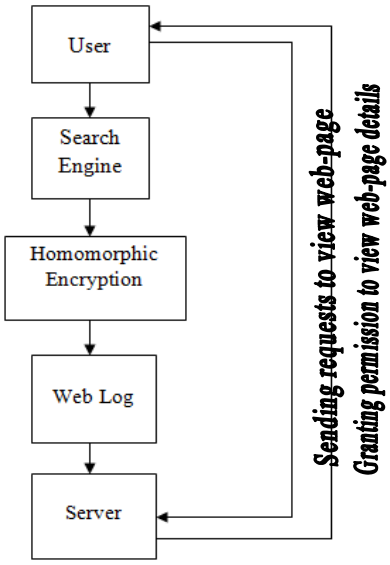

Figure2 : Homomorphic Encryption Used In Private Searching Keywords In Web Search Engine

Figure 2. describes data flow diagram of web search engine which allows searching frequent keywords visited by the user. Homomorphic encryption technique is used to encrypt details of frequently visited web-page automatically whenever the user searches frequent keywords in particular websites on the search engine. These encrypted details of the web-page are stored in the web log file in the server. The server sends requests to the user to view the encrypted details of the web-page. The user grants permission to the server who views the decrypted form of the details of the web-page.

6. RESULTS & DISCUSSIONS

The administrator will upload file details like (file key, file sub key, secondary key, web URL and field) in to the server as shown in Figure 3. Then administrator can view all user history details and most number of clicked sequence on each and every web pages.

Journal of Theoretical and Applied Information Technology
20th September 2015. Vol.79. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

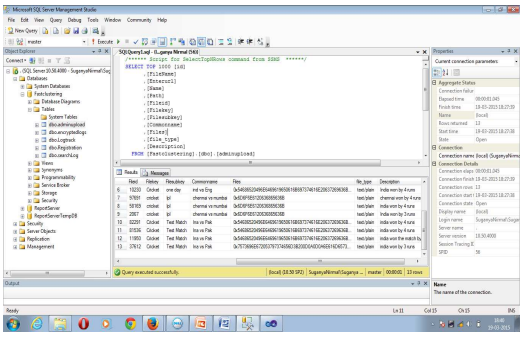

Figure 3: Administrator Configuration

Based on the user's application logic, user gives the different inputs of query to the query Processor as shown in Figure 4(a),(b)&(C).It may be a keyword or content then searching results are retrieved by clusters and that results are filtered by usage. Sequential pattern mining is an important data mining problem with broad applications. It is challenging since one may need to examine a combinatorial Explosive number of possible subsequence patterns.

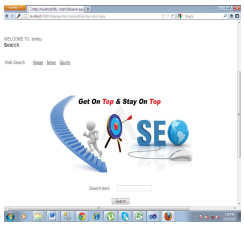

Figure 4(a)

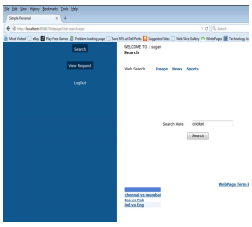

Figure 4(b)

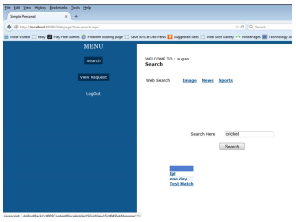

Figure 4(c)

Figure 4(A),(B)&(C): Searching Procedure

This type initiates the data search at server side. Query processing is checking the user query these results are retrieved from the database. Query processing results are combination of Web pages and relationships. And all these queries are checked by the processor for log creation and comparison. This gives the related data's as shown in Figure 5(a).

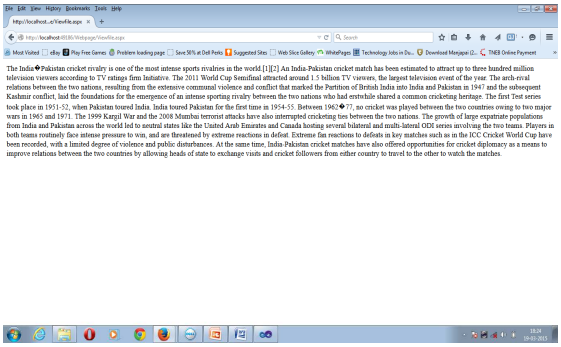

Figure 5(A): Recommended Web Pages

Using time based ranking the search result will be displayed according to how much time they have spent to view that particular webpage and then using term net WP algorithm will be applied to high recommended list webpage this web page will extract all related content from these webpage and display user recommended list urls as shown in Figure 5(b).

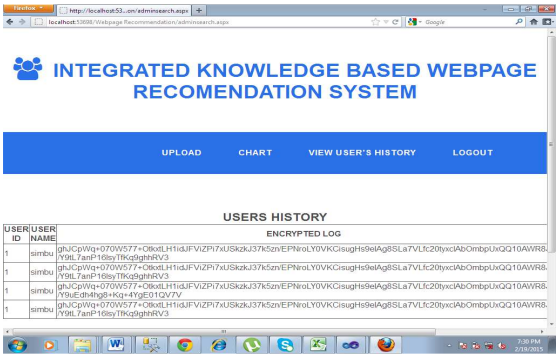

Figure 5(B): Description Of Web-Pages

7. PERFORMANCE ANALYSIS

The performance of effective keyword searching is based on measurement of recommended web-page accessed by the user and time taken for visiting frequent web-page based on keywords in the optimised search engine is implemented using the time based ranking technique. Figure 7. describes how many users have frequently visited web-pages which is measured in milliseconds of time.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

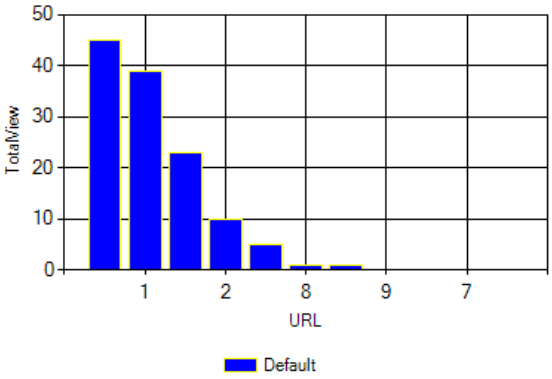

Figure7: Performance Chart

8. CONCLUSION

An efficient private searching keyword on frequently visited web-page which ensures user authentication and access control in a privacy preserving way using homomorphic encryption technique has been implemented. The features of websites are enhanced according to the user's interest on web based applications in an optimised web search engine. The Web-page recommendation is developed to offer Web users the top-*N* most commonly visited Web-pages from the currently visited Webpage. The knowledge bases used in the system, includes the website domain and Web usage knowledge bases, are represented by ontological-style semantic networks which can be implemented consistently in a formal Web Ontology Language. The current system works with static Web-pages. With the advancement in Web technology, pages have been evolving into pages with dynamic structures. To offer more effective Web-page recommendations, it will be highly desirable to develop advanced tools to identify and collect more appropriate Web usage data than Web logs, such as click stream data. Websites have been evolving over time therefore the knowledge bases, i.e. domain and Web usage knowledge bases, need to be updated accordingly. Considering the traditional Web usage data source, which is the Web log file, the system can only take a limited segment of the log file to build the Web usage knowledge base due to the fact that the size of the log file can be huge. The future work can be focused on the discovered Web usage knowledge is up-to date, new methods need to be developed to dynamically update the knowledge bases.

REFERENCES:

[1] A. Bose, K. Beemanapalli, J. Srivastava, and S. Sahar, "Incorporating Concept Hierarchies into Usage Mining Based Recommendations," in Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis Philadelphia, PA, USA: Springer- Verlag, 2007, pp. 110-126.

[2] A. Loizou and S. Dasmahapatra, "Recommender Systems for the Semantic Web," in ECAI 2006 Recommender Systems Workshop Trento, Italy, 2006.

[3] B. Mobasher, "Data Mining for Web Personalization," in The Adaptive Web. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl,Eds.: Springer-Verlag Berlin, Heidelberg, 2007, pp. 90-135.

[4] B. Liu, B. Mobasher, and O. Nasraoui, "Web Usage Mining," in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed.: Springer-Verlag Berlin Heidelberg, 2011, pp. 527-603.

[5] D. Dzemydiene and L.Tankeleviciene, "On the Development of Domain Ontology for Distance Learning Course," in 20th EURO Mini Conference "Continuous Optimization and Knowledge-Based Technologies" Neringa, LITHUANIA, 2008, pp. 474–479.

[6] H. Dai and B. Mobasher, "Integrating Semantic Knowledge with Web Usage Mining for Personalization," in Web Mining: Applications and Techniques, A. Scime, Ed. Hershey, PA, USA: IGI Global, 2005, pp. 276 - 306.

[7] J. Borges and M. Levene, "Generating Dynamic Higher-Order Markov Models in Web Usage Mining," in Knowledge Discovery in Databases: PKDD 2005. vol. 3721: Springer Berlin /Heidelberg, 2005, pp. 34-45.

[8] L. Wei and S. Lei, "Integrated Recommender Systems Based on Ontology and Usage Mining," in Active Media Technology. vol.5820, J. Liu, J. Wu, Y. Yao, and T. Nishida, Eds.: Springer-Verlag Berlin Heidelberg, 2009, pp. 114–125.

[9] S. A. Rios and J. D. Velasquez, "Semantic Web Usage Mining by a Concept-Based Approach for Off-line Web Site Enhancements," in Web Intelligence and Intelligent Agent Technology, 2008. WI-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IAT '08. IEEE/WIC/ACM International Conference on, 2008, pp. 234-241.

[10] Thi Thanh Sang Nguyen, Hai Yan Lu, Jie Lu "Web-page recommendation based on Web Usage and Domain Knowledge" 2014.

[11] Y. Lu, "Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP-Tree," Data Mining and Knowledge Discovery, vol. 10, pp. 5-38, 2005.