A SURVEY: TOEKN CARRIED DATA MONITORING OVER BUSINESS FRAMEWORK

Anshu Mishra [#1], Shailendra Gupta [#2,] Pankaj Richhariya [#3]
 # Department of Computer science & Engineering, BITS, Bhopal, M.P., India
[1]anshmishra8817@gmail.com

ABSTRACT: ***In this era, large amount of the information in form of digital information is made available in digital form. For many years, people have held the hypothesis that using phrases for are a presentation of document and topic should perform better than terms. In this synopsis here investigate this fact with considering several states of art data mining methods that give satisfactory results to improve the effectiveness of the pattern. In the current base, the paper author has taken technique which makes use of signing data with token data technique which further use for their operation monitoring over the different level. They have worked with complete data lifecycle associated with a token from its production to consumption. The token carried data and their monitoring framework makes use of its state detection and operation. The approach token-log-based process discovery algorithm is utilized in the system to monitoring its activity. A further enhancement of technique stated as in variable related detail mining over the data.***

Keywords – Web services, Token based data, business framework, data monitoring, log based approach, web mining.

I. INTRODUCTION

Web mining is the collection of data gathered by traditional data mining methodologies and methods with information collected over the internet. This is used to know about customer behaviour, to evaluate the capability of a particular Web site, and help quantify the success of a business campaign. It also enables looking for patterns in data with content mining, usage mining, and structure mining. Content mining can be used to audit data, which is collected by search engines or web spiders. Structure mining is used to examine data related to the structure of a particular site on internet and Web Usage Mining is applied to many real world problems to discover interesting user navigation patterns for improvement of website design by making additional topic or recommendations observing user or customer behaviour.

Security during data transfer in cloud computing using token based data security along with its implementation is presented in this paper. The method auto generated token certificate basd activation method with SSL (Secure Socket Layer) provides an appropriate collaboration between the cloud client and the service provider. So that the user may become confident while he transfer data by utilizing various cloud applications and services. The chances of attacks may be reduced by implementing this TBDSA. This algorithm is designed to takes less time to run and increases the performance of the system.

Web Usage Mining is an application of data mining techniques to discover interesting usage patterns from Web data, in order to know and better serve the requirements of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered. They are web server data, application server data and application level data.

Web server data correspond to the user logs that are collected at a Web server. Some of the typical data collected at a Web server include IP addresses, page references, and access time of the users and is the main input to the present Research. This work concentrates on web usage mining and in particular focuses on discovering the web usage patterns of websites from the server log files.

The Web Usage Mining is a process, where user access patterns are discovered and analyzed by mining the log files and related data associated with a certain website. It is a kind of web mining which automatically discovers user usage patterns and is helpful in studying and analyzing user interests. Web usage mining mainly consist three stages, namely data pre-processing, pattern discovery and pattern analysis as shown in figure 1.

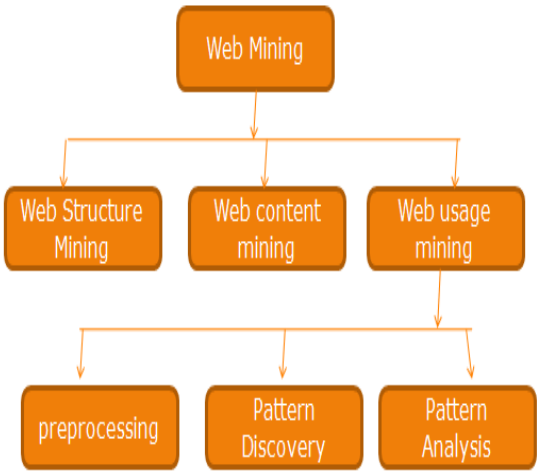

Figure 1: Stages of Web Mining

Information for data mining is gathered from different resources. Data is in various format following different conventions with multiple duplicates, inconsistencies, sometimes inadequate. Hence data preprocessing is very vital and is also most complex of all stages. It decreases the size of data radically thus improving the efficiency of mining. Pattern Discovery applies various methods on the preprocessed knowledge to discover frequent patterns like statistics analysis, clustering, association rule mining, sequential pattern, classification and so on. In the next stage, Pattern analysis, all the patterns made in the previous stages are analyzed to choose only the interesting patterns and rules sieving the useless patterns and rules.

While performing rule mining and various approaches over the dataset, there are terms which associate while working with the Data mining approach over dataset.

Briefly, the stages are noted as follows:

- Obtain data from various sources
- Data preprocessing
- Pattern Discovery
- Pattern Analysis

Obtaining data from different resources and then mine the knowledge from raw data is call process that is data mining. There are various stages in data mining process. Data pre-processing, Pattern discovery and pattern analysis are some of them.

A. Data Pre-processing

The data should be pre-processed to generate the efficiency and ease of the data mining process. The main stage of data pre-processing is to prune noisy and irrelevant data, and to reduce volume of the data for the discovery of pattern phase. Field Extraction process and data cleaning algorithms parse the web log records separating the fields and purging.

B. Pattern discovery

Few techniques to discover patterns from pre-processed data are listed like converting IP addresses to domain names, filtering, dynamic site analysis, cookies, path analysis, association rules, sequential patterns, clustering, decision trees etc.

C. Pattern Analysis

Following statistics are a few listed ones which are the end products of analysis such as the frequency of visits per document, most recent visit per documents, he is visiting which documents, frequency of use of each hyperlink, and most recent use of each hyperlink. The common techniques used for analysis of patterns are visualization methods, OLAP techniques, Data & Knowledge Querying, Usability Analysis.

II. LITERATURE REVIEW

1. Huiping Peng, 2010

In this author stated the use of FP-growth algorithm for processing the web log records, obtaining a set of frequent access patterns, then using the combination of search engine interestingness and site topology allure of association rules for web mining [1]. Analysis of web usage mining by using Web Log analyzer tool, "Web Log Expert" was carried out by Sanjay Kumar Malik in 2010. He focused on the creation of Ontology for an intelligent or efficient web and it's relation with web usage mining. Finally, they also summarize some other research challenges towards an intelligent machine and web environment [2].

2. Hao Yan, Bo Zhang, Yibo Zhang, Fang Liu, Zhenming Lei, 2010,

In this proposed a two-step clustering algorithm K-means, to find user groups in realistic information collected from WAN. They have given some useful practical conclusions to facilitate design of targeting and recommending applications , K-mean is efficient algorithm which is already proven for the best clustering approach , further more K-mean is extended to K-Medoit and other scheme to make further more efficient clustering from the available data and document from the user [4].

3. Han J., Pei J., Yin Y. and Mao R.,2004

In this paper author proposed a novel frequent-pattern tree structure, which is an extended prefix-tree structure for storing compressed, crucial information about frequent patterns, and develop an efficient FP-tree based mining method, FP-growth, for mining the

complete set of frequent patterns by pattern fragment growth [5].

4. Chuchra et al., 2012,

In this author discussed about the problems occur when cloud client send data from one end to another end. There are large numbers of attacks have encountered to give security from the attackers and that becomes mandatory task for any cloud service provider and also gave a methodology to provide security on the cloud at the client site. Here, cloud client and cloud service provider took a joint action for providing data security. The digital signature with autogenerated token-number by cloud service provider during membership on the cloud is used [6].

5. MohdHelmyAbdWahab, MohdNorzali Haji Mohd, Mohamad Mohsin 2008

In this paper author describes the pre-processing techniques on IIS Web Server Logs ranging from the raw log file until before mining process can be performed [7].

6. C.P Sumathiin, R. Padmajavalli, 2011

In this paper author resented an overview of the various steps involved in the preprocessing stage [8].

7. RenataIvancsy, IstvanVajk,2006

In this paper they investigated three pattern mining approaches from the web usage mining point of view. Also author has done the in-depth analysis of Web Log Data of NASA website to find information about a web site, top errors, potential visitors of the site etc. which help the system administrator to improve their system and web designer to improve system by determining occurred systems errors, corrupted and broken links by using web using mining [9].

8. Vaibhav Kant singh, Vijay Shah, Yogendra Kumar Jain,2008

In this paper author shows that how the various approaches achieve the objective of frequent mining. They also look for hardware approach of cache coherence to improve efficiency of the above process [10].

III. CONCLUSION

Web usage mining is the application of data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis. One of the algorithms which are very simple to use and easy to implement is the Apriori algorithm. In this survey paper, existing methods are discussed and discover the web usage patterns of websites from the server log files with the foundation of clustering and improved the token based algorithm. Applying the token based data monitoring with pattern and request analysis with the hidden rules over social media can use for betterment. A detail variable analysis over the available data is going to observe and further improvement on mechanism can perform over the data.

REFERENCES

[1]. Chuanyi Li, Jidong Ge, Zhongjin Li, Liguo Huang, Hongji Yang, and Bin Luo, “Monitoring Interactions across Multi Business Processes with Token Carried Data”, IEEE 2016.

[2]. Huiping Peng "Discovery of Interesting Association Rules Based on Web Usage Mining" 2010 International Conference.

[3]. Sanjay Kumar Malik, Nupur Prakash, S.A.M. Rizvi" Ontology and Web Usage Mining towards an Intelligent Web focusing web logs" 2010 International Conference.

[4]. Hao Yan, Bo Zhang, Yibo Zhang, Fang Liu, Zhenming Lei, "Web usage mining based on WAN users' behaviours" 2010 International Conference.

[5]. Han J., Pei J., Yin Y. and Mao R., "Mining frequent patterns without candidate generation: A frequent-pattern tree approach" Data Mining and Knowledge Discovery, 2004.

[6]. Rimmy Chuchra, 'Data security in cloud computing, International Journal of societal applications of computer science, oct-2012.

[7]. MohdHelmyAbdWahab, MohdNorzali Haji Mohd, Mohamad Mohsin, "Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm" 2008.

[8]. C.P. Sumathi, r. padmajavalli,"An overview of preprocessing of web log files for web usage mining" 2011.

[9]. RenataIvancsy, IstvanVajk "Frequent Pattern Mining in Web Log Data"2006.

[10]. Vaibhav Kant Singh, Vijay Shah, Yogendra Kumar Jain, "Proposing an Efficient Method for Frequent Pattern Mining" 2008

