

EXTRACTING THE PERFORMANCE ANALYSIS OF WEB LOG ANALYSIS USING WUM TECHNIQUE

¹Ms. J. Jothilakshmi, M.Phil Research Scholar, Department of Computer Science
Muthurangam Govt. Arts College

²Mrs.T.K.Ganga, Assistant Professor, Department of Computer Science, Muthurangam Govt. Arts
College, (Autonomous), Vellore District

ABSTRACT:

In this article, Web logs will become a comprehensive and easy reading analysis for customers and prospects. This identifies how visitors view the website, which pages view, ignore, how much time they spend on the site and shows where it comes from. It specifies the number of viewers, the number of individual IP addresses, the frequency of the bandwidth, and the success or hits of the site, the number of weeks, and the hours of the day. To learn more about visitor's access information, which web pages want to view, files downloaded, directories are accessed, and the images are watching. Reference information includes domains and URLs from viewers. Search Engine Performance Report shows search viewers site, words and phrases searched by visitors. Its purpose is to analyze the log files of the website from the Web server using the WUM technique. The data warehouse is built and populated, and any web application methods that identify various statistical and data mining techniques. The 123LogAnalyzer is currently an application that can be helpful in this research of innovation. These methods will then be analyzed, understood and used to determine how the web site is used. A graphical representation of these forms will be created. This is an easy way to minimize narrow precious things from the social network of internet mining concerns. There are many unwanted information in social media, using this review to allow valuable content, images, videos, audios, directories, etc. to access from the Internet easily.

Keywords: Performance, Visitors, Bandwidth, WUM technique, 123LogAnalyzer

1.INTRODUCTION

Web processing uses data mining techniques to automatically detect and retrieve information from websites. Web service is a popular technology for analyzing the audience's performance in power supply systems. Web mining technique plays an important role in the process of converting databases. This is used to filter out useful Web Registration information from the Web server. It provides an easy way to access the information we need from the Internet and sends content from the World Wide Web and displays the right path. Server provides error when surfing the majority but using this technique we can easily access internet.

Web mining research can be classified into 3 types:

- a) Web Content Mining (WCM)
- b) Web Structure Mining (WSM)
- c) Web Usage Mining (WUM).

a) Web Content Mining

Web Content Mining refers to finding useful information from web content including text, image, audio and video,

b) Web Structure Mining

Web Structure Mining outlines of mining web page and search engine results are used in the rankings.

c) Web Usage Mining

Web Usage Mining analyzes the web log files, search version and indicates user profiles.

1.1 SCOPE AND OBJECTIVES OF THE RESEARCH WORK

This research work aims to generalize the website's record / log files from the web server using the Attribute-Oriented Induction Technique.

CONTEXT:

It is used to identify the frequent access method for any website. Accordingly, the website can be upgraded.

WEB MINING:

Web Mining can be widely defined as the discovery and analysis of useful information from the World Wide Web. One hand, this broad definition, millions of sites and online databases from the available information and evidence, the Automatic Search and Recover describes the web content platform, and on the other hand, the user access methods for finding and analyzing or more web servers or online services, namely, Web Usage Mining.

II. LITERATURE SURVEY

Web Mining is a data mining application that has become a key feature of large-scale research into global web services in recent years. The web application mining web servers deals with the extraction of interesting knowledge from the sorting of information in the web mining area. The use of information available on the World Wide Web includes users' usage methods. This paper attempts to convey a brief overview of the web trends of the web mines and the web mining associated with its techniques, tools and applications [8].

The website tries to create the best and most efficient websites by understanding the behavioral patterns of website visitors. In this research, Web log files are analyzed to obtain the user access system of various web pages in websites. It shows the daily usage of the report, the day of the week, broke within hours of the day. To learn more about access to information about the audience, how many web pages have been downloaded, how many files were downloaded, all the directories were accessed, and which websites were viewed. The referrer information includes visitors and URLs from viewers [9].

Web Usage Mining (WUM) is a data mining system that can be used to detect user access methods from web record data. A lot of research has already been done in this area, and various applications are used in customization, computer development, and business intelligence. The WUM initiative involves three phases known as model innovation and method analysis. There are different techniques of WUM that have their own advantages and disadvantages. This study is conducted in some of the existing WUM technologies, showing how WUM can be used for web server logs [10].

III.SYSTEM IMPLEMENTATION

3.1 DETAILED PROCESS OF WUM

3.1.1 Data preprocessing

Data preprocessing has a fundamental role in web application mining applications . It has a variety of tasks:

(a) Data Cleaning

This move eliminates all data tracked on web logs that are useless for mining use.

(b) Session Identification and Reconstruction

This step (i) identifies multiple users 'sessions from the worst information available in log files, and (ii) reconfigure users' navigation path into identified sessions.

(c) Content and Structure Retrieving

Web content refers to finding useful information from web content including text, image, audio, and video. Restoration of the structure helps analyze the outbound links of the web page, which is used in the search engine result rankings.

(d) Data Formatting

Once the successful completion of the previous phases, data is properly designed before using mining technologies. Therefore stored data stored in a corresponding database from Web logs.

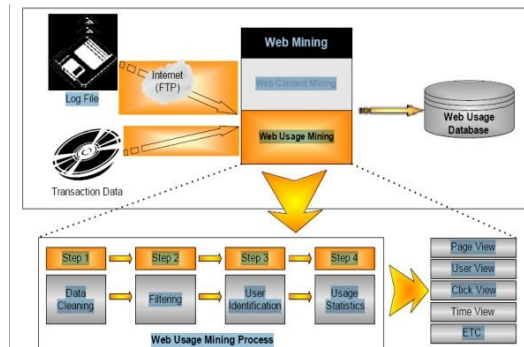


Figure 3.1.1: Phases of WUM

3.1.2 Mining Algorithms

(a) Statistical Analysis

A website has the most common method of statistical techniques for acquiring knowledge about the audience. By analyzing the Session File, one can perform descriptive statistical analyzes (frequency, mean, average, etc.) in the variants such as the page views, the viewing time and the length of the extension path. Many web traffic analysis tools often issue a specific report that contains statistical information such as pages [2], an average page view or average length of a page. Knowledge of this type is useful for improving computer performance, improving the security of the system, facilitating the task of modifying the site and supporting marketing results.

(b) Clustering

Clustering is a characteristic. In the web application field, two types of interesting clues were found. (i.e) the use of clusters and page clusters. Clustering users creates groups of users who display similar browser types. This knowledge is especially useful for marketing partitions in e-commerce applications or to put user statistics to provide user customized web content. On the other hand, a set of pages will find groups of pages with related content. This information is useful for Internet search engines and web help providers. In both applications, permanent or dynamic HTML pages are created to suggest relevant hyperlinks according to previous history of user query or information needs.

(c) Classification

Classification is the task of modifying a data item in many predefined classes. Web Domain is interested in developing a profile of users of a specific class or category. It should be included in the extraction and selection of features that describe the characteristics of the given class or type.

(d) Association Rules

Association rule generation is often used to relate pages referred to in one of the server sessions. See the packages of pages that are accessible with more support than some specific entry in the context of web application macon, association rules. These pages may not be directly connected via hyperlinks. For

example, the association's use of Apriori Algorithm [2] (or its variants) may reveal a connection between users who visit a page that includes electronic products to access a page about sports equipment. Web designers restructure their website. Terms of use for pre-set of document rules to reduce the delayed user being loaded from a remote site.

(e) Sequential Patterns

The technique sequential pattern discovery attempts to find inter-session patterns such that the presence of a set of items is followed by another item in a time-ordered set of sessions or episodes. By using this approach, web advertisers can predict future visit guidelines, which can be useful in placing ads in certain user groups.

(f) Dependency Modeling

Dependency modeling is another useful modeling task in the web mining. The goal here is to create a model that has the ability to represent significant dependencies among the different variables on the Internet.

3.1.3 Pattern Analysis

Pattern analysis is the last step in the overall web application mining process described in Figure 3.4. The motivation behind the algorithm is to filter out unwanted rules or shapes from the system found at the point of detection. [2] An accurate analysis system is normally administered by the application for web mining processing. The most common form of sample analysis is the knowledge technique like SQL.

3.2 METHODOLOGY

3.2.1 Modules Descriptions

It is a step in the study of the theoretical design changes in the working system. This is the process of changing a new system operation. The functionality of the software development is concerned with translating design specifications in the source code. In the most important phases, the new system is optimistic for users who are effective and efficient. Some Modules are defined as follows:

a) Extracting web log files

Extract log files with various formats from various web servers.

b) Converting web log files

Converting information from text files (which is a file generated by log analyzer) and storing the webs that are available in the database in the directory.

c) Generalization web log data

Send all data to correct tuples

d) Bar chart generation

Based on the information available on the database, it is necessary to build the required chart.
(For example: daily, weekly, web pages viewed etc.)

e) Table generation

Based on the information available on the database, the database is going to produce the necessary information in the table. Example. daily, weekly, web pages viewed etc.

3.3 SYSTEM ARCHITECTURE

Systems Architecture is the process of art of defining the data of an organization to meet architecture, components, modules, interfaces, and specific requirements. One could see this as the use of system theory for product development.

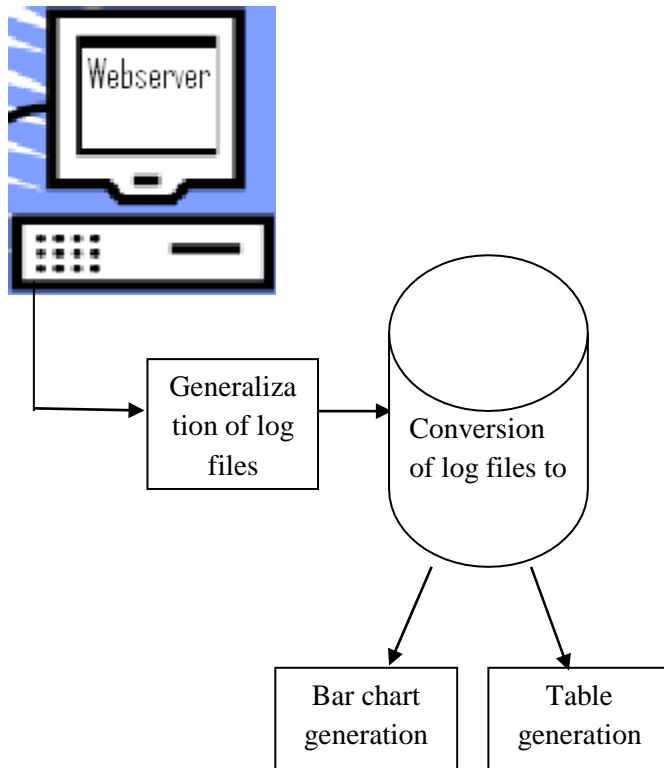


Figure 3.3 : Overall System Architecture

3.4 ACTIVITIES OF WUM

The main function of the WUM technique is to destroy useful Web Registration information from the Web server. This will take place in several processes such as preprocessing, mining algorithm and pattern analysis. This process first takes the source record file, receives useful information, and finally provides interesting forms, such as rules, patterns and statistics, by obtaining a user session file. This intelligence can be statistically specified, and then users can easily identify the monthly facts of everyday life from the daily, web server. Based on these guidelines, everyone can specify the right web page to get valuable data.

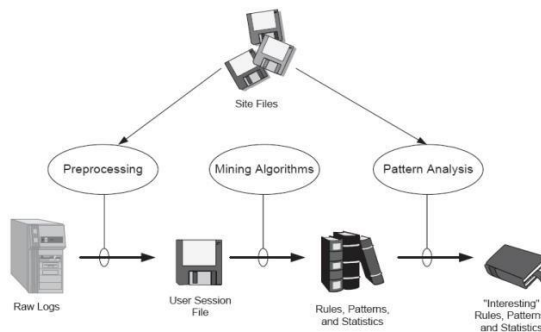
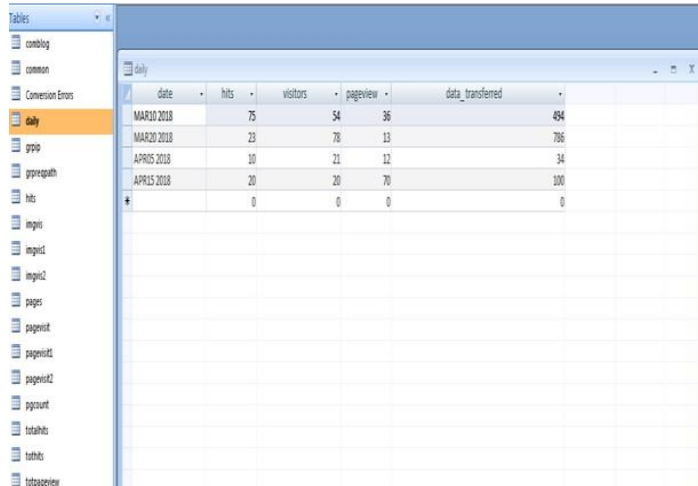


Figure 3.4 Activities of WUM

IV.EVALUATION RESULTS

Web Usage Mining mines web logs to learn web access to web pages. Identifying and exploring customers for e-commerce improves the quality of Internet information services to end user and to improve web server performance. Text file contents are derived from the text file and the tokens are split using the string Tokenizer. The contents are then stored in a database. Unwanted Tuples are then removed and stored in another table. Aggregate functions are used to extract the necessary tuples. SQL queries are sent to the database.

4.1 Sample Screen Shots:



The screenshot shows a web analytics interface with a table titled 'daily'. The table has columns for 'date', 'hits', 'visitors', 'pageview', and 'data_transferred'. The data is as follows:

date	hits	visitors	pageview	data_transferred
MAR10 2018	75	54	36	484
MAR20 2018	23	78	13	786
APR03 2018	30	21	12	34
APR15 2018	20	20	70	100
*	0	0	0	0

Fig.4.1.1 Generalization of Daily visit report

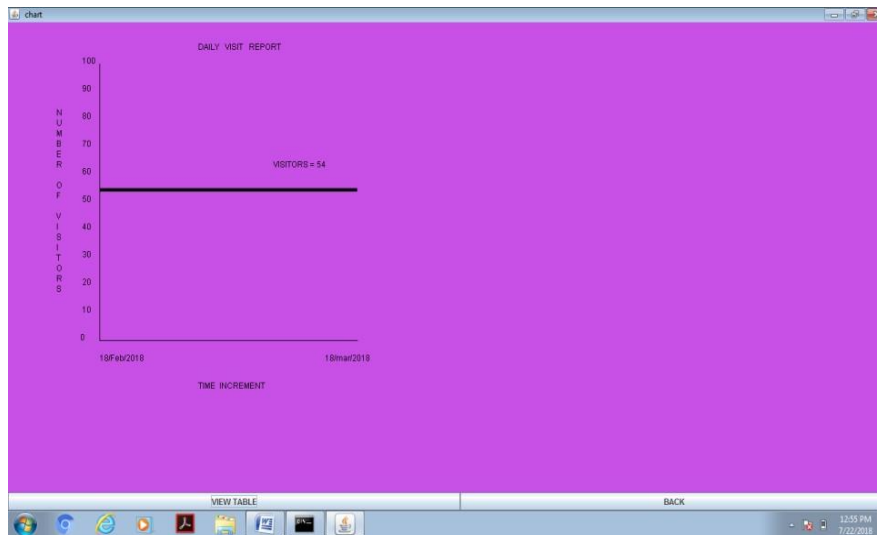


Fig.4.1.2. Bar/ Line chart of Daily Visit Report

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Web Application Mining research and web Usage Mining applications is an active field used in some popular websites. This research provides Web Usage Mining processing. Web Server log files are mined in order to examine Web Usage pattern . The methodology employs data uses Data Preprocessing,

Mining Algorithms and Pattern Analysis. Data processing phase for Web Usage Mining is a challenging task. The relationship between the pages that are accessed can be cut by using the mining instructions for the web log file. Web usage patterns and user behavior mining algorithms are analyzed. Web administrators and web masters can use the results of this research to improve web servers and performance by improving web sites and improving their contents, structure, presentation and delivery.

5.2 Future work

As the future enhancement of this research, web pages can be pre-fetched depending on the usage patterns. Pre-fetching can improve the web performance at the best level. Furthermore, it can use a comprehensive data analytics system to be the most appropriate alternative to web log access study, using Association rules and extracting knowledge from web log data. Finally the research can be extended to access and process the external web servers with appropriate access rights.

REFERENCES

- [1] Bianka M. M. T. Gonçalves¹, Isabel Cristina Italiano², and João Eduardo Ferreira¹, Data updating between the operational and analytical databases through dw-log algorithm, Proceedings of the 9th International Database Engineering & Application Symposium (IDEAS'05), 1098-8068/05, IEEE, 2005.
- [2] Cooley: Jaideep, Srivastava t, Robert, Mukund Deshpande, Pang-Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web to 23.Data", SIGKDD Explorations. Jan 2000. Volume 1, Issue 2 - page 12
- [3] Hisayoshi Kato, Hironori Hiraishi and Fumio Mizoguchi, "Log summarizing for web access data using data mining techniques", IEEE transaction 2001
- [4] Jan Kerkhofs Prof. Dr. Koen Vanhoof Danny Pannemans, A Case Study of Web Usage Mining on Proxy Servers, Published Limburg University Centre, July 30, 2001.
- [5] ANITHA TALAKOKKULA Computer Engineering and Intelligent Systems ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol.6, No.2, 2015
- [6] M. Aldekhail, International Journal of Computer Theory and Engineering, Vol. 8, No. 1, February 2016

[7]Mr. Akshay Upadhyay, Mr. Balram Purswani International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013 ISSN 2250-3153 International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056

Volume: 04 Issue: 04 | April -2017 www.irjet.net p-ISSN: 2395-0072

[8] Mr.Sahaj Chavda, Saurabh Jain, Nikunj Panchal, Manisha Valera, International Research Journal of Engineering and Technology (IRJET) “Recent Trends and Novel Approaches in Web Usage Mining “

Volume: 04 Issue: 04 , April -2017 e-ISSN: 2395 -0056 p-ISSN: 2395-0072 www.irjet.net

[9] M. SARAVANAN, B. VALARAMATHI, “Generalization of Web Log Datas Using WUM Technique”

[10] Maryam Jafari, Farzad SoleymaniSabzchi, Shahram Jamali “Extracting Users’ Navigational Behavior from Web Log Data: a Survey”, Journal of Computer Sciences and Applications, 2013, Vol. 1, No. 3, 39-45