

## A METHODOLOGY ON PREDICTING STUDENT'S BEHAVIOUR ON WEB USAGE MINING

**M R RAMYA**, Registration No: 19123152282022, Research Scholar, Department of Computer Science, ST Hindu College, Nagercoil, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-2, Tamil Nadu, India.

[mrramya\\_1984@yahoo.co.in](mailto:mrramya_1984@yahoo.co.in)

**DR. R RAJALAKSHMI**, Assistant Professor, Department of Computer Science, Noorul Islam College of Arts and Science, Kumaracoil, Kanyakumari District, Tamil Nadu, India.

[rajalakshminatesh@gmail.com](mailto:rajalakshminatesh@gmail.com)

**DR. N. CHENTHALIRINDRA**, Assistant Professor, Department of Computer Science, S.T. Hindu College, Nagercoil, Kanyakumari District, Tamil Nadu, India.

[chenthalirindra@gmail.com](mailto:chenthalirindra@gmail.com)

### ABSTRACT

The paper focuses on Web usage mining (WUM) among students categories of accessing web contents for their education purposes. The study focuses on finding the potential knowledge from browsing patterns of the students. It leads us to find the correlation between web pages in the students visit. The primary data source used in web usage mining is the university server log-files (web-logs). Browsing web pages by the students leaves a lot of information in the log-file. Analysing log-files information drives us to understand the behaviour of the student in academic basis. Web-logs include web server access logs, proxy server logs, client logs and application server logs. Our paper focuses on the use of web mining techniques to cluster web pages according to students visit. This cluster process helps us to understand the student's behaviour in accessing the network.

**KEYWORDS:** Web Mining, Information, Knowledge, Clustering, Data Mining, Web Usage, Web Content, Web Structure

### 1. INTRODUCTION

Web mining is the application of data mining techniques to extract knowledge from web data; it includes web content mining, web structure mining and web usage mining. The web mining is depicted in the Fig.1. Web Usage Mining (WUM) is a special type of web mining tool. That can discover the knowledge in the hidden browsing patterns and analyses the visiting characteristics of the user. It is a complete process that includes various stages of data mining cycle, including Data Pre-processing, Data Cleaning, and Pattern Discovery & Pattern Analysis [6]. Initially, the web log is pre-processed to clean, integrate and transform into a common log file. Data mining techniques were applied to discover the interesting characteristics in the hidden patterns. Pattern Analysis is the final stage of web usage mining which can validate interested patterns from the output of pattern discovery [1,2]. In the pattern discovery stage, statistical, database, and machine learning operations are performed to obtain hidden patterns reflecting the typical behaviour of users, as well as summary statistics on web resources, sessions, and users.

### 2. WEB USAGE MINING

Web usage mining is used to investigate useful data, information, knowledge from the weblog records, and assists in recognizing the user access patterns for web pages. In Mining, the usage of web resources, the individual visit pages, records of requests of visitors of a website that are collected, as web server logs. While the content and structure of the collection of web pages follow the intentions of the authors of the pages, the individual requests demonstrate how the consumers see these pages. Web usage mining may disclose relationships that were not proposed by the creator of the pages. It clearly identifies the habit of users accessing scenario. Some of the methods to identify and analyze the web usage patterns are given below.

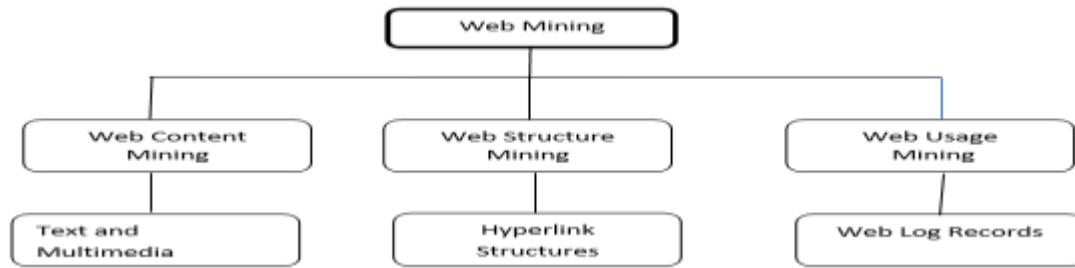


Fig. 1: Web usage mining process

**2.1. Session analysis:**

The analysis of preprocessed data can be accomplished in session analysis, which incorporates the visitor records, days, time, sessions, etc. This data can be utilized to analyze the visitor's behavior. The document is created after this analysis, which contains the details of repeatedly visited web pages, common entry, and exit. A standard log-file has the information like remote host, log name, user name, date, request, status, bytes

**3. WEB USAGE MINING PHASE:**

Web Usage Mining has been defined as the application of data mining techniques to large Web data repositories in order to extract usage patterns, namely the visitor behaviour. As further step, pattern discovery and pattern analysis allows profiling users and their preferences. For that statistical method plays a fundamental role. Main task of user profiling is to organize data to collect consistent information about the users here students. A distinction can be made between explicit information and implicit information: the former is derived from the data provided by the user filling any type of application form, whereas the latter can be induced by the log-files and the cookies as well as any other method of tracking the web navigation. The data can be collected at two levels. First level is at the server level where data are memorized into log-files, which will be recording all the activities of the server, for instance; the web page selections. There are four main mining techniques that can be applied to web access logs to extract knowledge, but we will focus on algorithms based on clustering.

**4. WEB LOGS**

A web server log is an important source for performing WUM because it explicitly records the browsing behaviour of site visitors. The data recorded in server logs reflects the (possibly concurrent) access of a web site by multiple users. However, the site usage data recorded by server logs may not be completely reliable due to the presence of various levels of caching within the web environment. Cached page views are not stored in a server log. Cookies are tokens generated by the web server for individual client browsers in order to automatically track the site visitors. Tracking of individual users is not an easy task due to the stateless connection model of the HTTP protocol.

**5. IMPLEMENTATION AND RESULTS:**

In this paper we transformed the dataset which represented as a server log-file into a proper form to be used in our implementation code, the following steps explains the achievements of our implementation. The data used for this research were collected and extracted from the University servers.

Table.2: Summary report

Hits Summary based on Educational Web site	Total
Number of Hits	350
Number of Successful Hits	320
Outgoing traffic	8 Mb
Incoming Traffic	2Mb

<b>Visits Summary</b>	<b>Total</b>
Number of Visits	60
Average number of Visits	15
Average Visit Duration	8 hr
<b>Visitors Summary</b>	<b>Total</b>
Number of Unique Visitors using educational site	50
One time Visitors	35
Repeat Visitors	5
<b>Web Page View Summary</b>	<b>Total</b>
Total Page Views	250
Most Popular Page	50
Most Popular Download	15
<b>Time and Place Summary</b>	<b>Total</b>
Time	>10:00 am
Place	Location

The main objective of the web usage mining technique is to generate statistical reports as output results that can use to detect some valuable information. After analyzing the work we focused on the data extraction from our web server log file as an input concerning the visitors and the user's behaviors to generate an investigation report concerning the web server status. In this work we extracted a potential knowledge from the log-files or server log-files while browsing patterns via users accessing web pages types. We used clustering techniques to group web pages that have been visited by users. For instance, if the user visited many pages randomly these pages can be classified in many profiling classes such as sports, economic, political ...etc. Suppose that this user visited sports pages 3 times, economic, political pages one time, and then we will be able to say that this user has a sporty behaviour. Association rule mining is used as well to find the correlation between these web pages.

## 7. CONCLUSION

This paper has tried to provide the important of the rapidly growing area of web usage mining. With the growth of web-based applications, particularly, there is a significant interest in analysing web usage data to better understand web usage data, and apply the knowledge to better serve users. The web data source used in web usage mining is the server log-files (weblogs). Our proposed system is used to analyze the user sessions from which information regarding the problems occurred to the users and usage of the website can be obtained within particular intervals of time. This is used to configure the server and adjust the Web site which is highly useful for administrators.

## REFERENCES

- [1] Abraham and V. Ramos. Web usage mining using artificial ant colony clustering and genetic programming. In Procs. Int. Conf. CEC03 on Evolutionary Computation, pages 1384–1391. IEEE Press, 2003.
- [2] Cyrus Shahabi, Amir M Zarkesh, Jafar Adibi, and Vishal Shah. Knowledge discovery from users webpage navigation. In Workshop on Research Issues in Data Engineering, Birmingham, England, 2005.
- [3] J. D. Velásquez, H. Yasuda, T. Aoki and R. Weber, A new similarity measure to understand visitor behavior in a web site, IEICE Transactions on Information and Systems, Special Issues on Information Processing Technology for web utilization, E87-D(2): 389–396, 2004.