

The Role of Web Usage Mining in Web Based Learning Environments

J.Umarani¹, G.Thangaraju², S.Kavitha³

Research Scholar, Bharathiyar University, Coimbatore ^{1&3}

Research Scholar, Karpagam University, Coimbatore ²

umashenthaan@gmail.com

Abstract The Web offers several opportunities in the field of education. With the immense expansion of information available on the web, web mining has become appropriate for the web based learning systems. Learning through Online is one of the practical modes of education. Learning Websites, essential courses, Web supported instructional shells, and E-books are some of the modes of delivering the Online Learning. Web mining is the series of task used for mining or extracting useful information from the web pages or web sites. It provides intrinsic knowledge of teaching and learning process for effective education planning by applying various techniques/tools. This paper discusses the benefits and usefulness of web usage mining methods in online learning environment.

Key words: Web usage mining, E-Learning, Web Based Environment

I. INTRODUCTION

In Early stages of Learning takes place the class room learning, Other than class room those who are interested using the Library to gather more knowledge after that the usage of ICT are take place , for the improvement of web applications Learning process also enhanced and use the web technology for Learning process. From the Web based Learner to Web Based Learning content provider have the different components and use the different techniques to fulfill their needs as well for the success of technology and skill implementation.

The information provider or those who are willing to deliver the message to others either the organized sectors or individuals use the web technology to upload their content. After the uploaded process organization/individuals are awaiting the how the content should reach to the society/learner. It will should be analyzed, for that process the most powerful technology data mining in specific Web Mining WM are used that purpose. The Web mining are divided into further they are listed below

1. Web Content Mining 2. Web Structure Mining 3. Web Usage Mining

Web Usage Mining uses many tools for analyze the web logs, the web logs are available in different stages client-level logs, proxy-level logs, server-level logs and content-level logs the flow of logs are depicted as the following diagram.

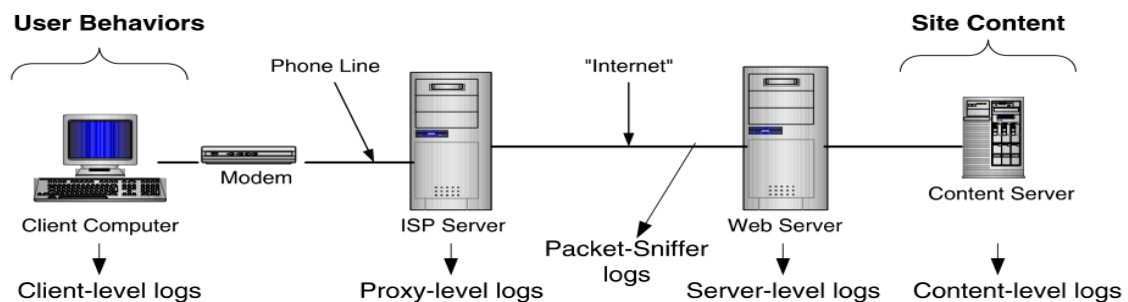


Figure. 1. Architecture of log flows in Web Environment

The Role of Web Usage Mining in web based learning is discussing in the methodology sections.

II. LITERATURE RIVEW

In current Scenario, all communications are made via online; here consider some of the applications done on online process E-commerce, E-Banking, E-Transport Booking and also E-Learning. The mentioned above applications are processed in different levels of operations, such as the sources of applications are have content or information's that is stored in any one of the DBMS, the data's are update after the transactions of each applications.

In [1]. The role of web usage mining in web applications evaluation are analyzed with the some steps of operations, they are data preprocessing, pattern discovery and pattern analysis. For the process of Web usage mining the Hybrid approaches is used in this article which combines the two methods they are the compact HPG (Hyper Probability Grammar) approach with the precise Online Analytical Process [OLAP] approach. In this representation data is stored into a database throughout the XML and Quilt Query. The conational constraints for the examination are built on the top of this database and data together with the constraint are used for modelling HPG, which are then mined using BFS based algorithm for mining the association rules.

In [2], this article represents the four types of problems they are 1. Incomplete or Limited Information Problem, 2. Incorrect Information problem 3. Persistence Problem 4. Incorrect recommendation. To solve the mentioned above issues the implements the following strategy, User Request- User request is processed for search engine to obtain the results. (2) top n Results Extraction – Top n results are extracted from search engine based on the user query. (3) Content mining - Statistical parameters such as a term frequency (TF) are calculated. For this every result is individually analyzed based on keywords and content. The calculations depend on the user query. Every result of the keywords and content words are compared by full word matching. If a match is found then particular weight is awarded to each word. Likewise each link is given the final matching score. (4) Page Reranking- At last, the normalized value of each result is sorted in descending order to get the most relevant content for the user query. Re-ordered results are sent back to the user so that the top most pages are more relevant for the user query.

In [3], this article represents the overall concepts of web mining and in particular describes about the functions of web usage mining preprocessing, pattern analysis and pattern discovery. And describes the applications that depends on the web usage mining are E-Commerce, E-governance and E-Learning Web usage mining is becoming an active interesting field of research because of its prospective commercial benefits. Finally the author analyzed the following factors visitor's behaviour, web logs, web services and e-services providers tool that satisfy the customer needs.

In [4], this article represents the web usage mining basic operations and also represents source of data for web usage mining, it includes the Server Level Collection, Client Level Collection, and Proxy Level Collection. Finally it was described the operations of web usage mining.

In [5], the chapter 12 of the book Data mining and Its Application represents the following essential concepts basics of web usage mining, process of web usage mining, Key Elements of Web Usage Data Pre-Processing and its sub contents Data Fusion and Cleaning, Page view Identification, User Identification, Sessionization, Path Completion, Data Integration. The further concepts of web usage mining is Data Modeling for Web Usage Mining and Discovery and Analysis of Web Usage Patterns and its sub content is Session and Visitor Analysis, Cluster Analysis and Visitor Segmentation, Association and Correlation Analysis, Analysis of Sequential and Navigational Patterns, Classification and Prediction based on Web User Transactions.

In [6], the author proposed new frame work for the proposed methodology they are depicted as follows;

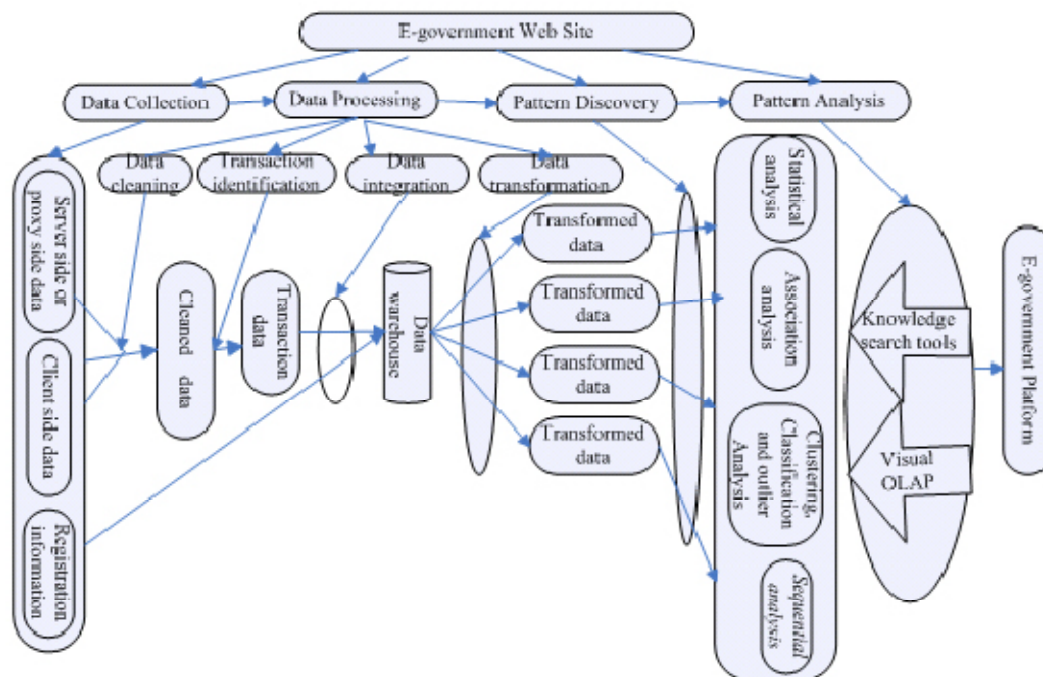


Figure.2. Frame work for web usage mining in E-government

In [7], the author proposed a new architecture for providing E-Learning in effective the architecture consists the following components they are 0 to 6, component 0 represents about the Active Session Extractor, components 1 represents the Sliding Window(w pages) implicit query extractor, components 2 represents Term vector builder, components 3 represents learner modeling, components 4 represents Content modeling, components 5 represents Recommendation Engine and component 6 represents the configuration module.

In [8], In this paper, our main aim is to carry out experimental work on web log data collected from NASA web server to find out useful browsing patterns. Performances of web server improved due to the results are extracted from this work. There are a number of web usage mining tools available in the market but here Web Log Explorer (WLE) tool is used for the implementation of our work. It is used to determine the number of accesses to the server and to individual files, the times of visits and the domain names, and URLs of users. The input of this tool is web log files collected from the web servers. In this work, we have also carried out comparative analysis of JPG and GIF image file types using results generated through MATLAB. In comparative work, we analyzed the effect of image file types for bandwidth usage per hit as parameter

In [9], the e-learning system consist three parts. Teaching resource library, learning platform and user. Education resource library is a storage server to store different types of resource which is related to education. The learner of that web based system is the user. Web server is the Learning platform that gives web based learning platform to user. The E-learning system which is based on web mining will progress the learning because it will supply learning substance according to the user's delicate information. The e-learning is also used to analyze the web logs and site files, personal information of learners Learning results, learning behaviour, and use data mining to meet the needs of different user. E-learning websites contain user information, learning results, behaviour of learning by the use of web mining.

In [10], this article represents the e-learning platform usage analysis with the log file formats Common Log Format (CLF), Extended Log Format (ELF), Cookie Log Format (CKLF) and Forensic Log Format (FLF). And also uses the Indexes and Metrics calculation to perform the analysis of web usage analysis.

In [11], Web usage mining has lot of contributions in e-learning domain such as, (a) Dynamic personalization like providing real time recommendations for e-learners (b) Generally referenced web pages are cached in proxy servers.(c) Structuring the site structure based on the interest of learner's. (d) Creating access shortcuts for interested pages to enhance user friendliness. (e) Updating course content of web site based on the previous usage information. (f) Identifying groups of learners of similar interest and sending custom-made course materials to fascinated groups.

In [12], In this research, we explained the use of Web mining approaches in CMS and identified some illustrative learning patterns that can be found by using Web mining approaches. Although some retesting patterns were found, the exploratory state of Web mining tools in education suggests replication and confirmation from other forms of research to build a context for understanding and drawing implications from the data. The primary findings of this research are to suggest that Web mining can be an approach that educational researchers can use, and when combined with other forms of data collection has potential for adding to the way we build knowledge about E-learning. A second contribution of the current study is to draw implications for how to improve the process of Web mining e-learning data sets.

In [13], the awareness of the potential advantages of integrated web usage mining and the insufficient data recorded by web servers, there is a need for more specific logs from the relevance side to improve the information previously logged by the web server. This additional value by precise event recorded.

In [14],proposed a new methodology Moodle Log Analysis, Mdl_log is an unstructured table, which records any user action on Moodle is achieved by a user (login) on a given course, and in a specific activity or resource. It provides the terminology as it follows user, session, visit, activity, Episode. The variables are mentioned in Session level and Chapter level. Finally it provides the results by presenting the values of variables Learner_ID, Gender, Dg_Type, UML_score, Global_score, DtFirstAccess, and DtEndAccess.

In [15], E-Learning process uses the web usage mining resources for making an e-learning as adaptive. The following methodology is used to represent the WIM in Adaptive E-learning.

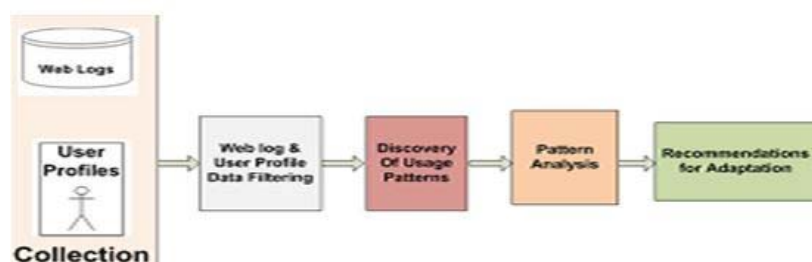


Figure.3. Methodology for WUM for the e-learning site.

All the mentioned above methods are have some of the drawbacks, we plan to rectify the all drawbacks, proposed a new methodology for the Roles of web usage mining in E-Learning.

III. METHODOLOGY

The Proposed methodology of the Role of Web Usage Mining in Web Based Learning Environments is depicted in the following figure. Roles are categorized in different levels here five levels are used in the entire process.

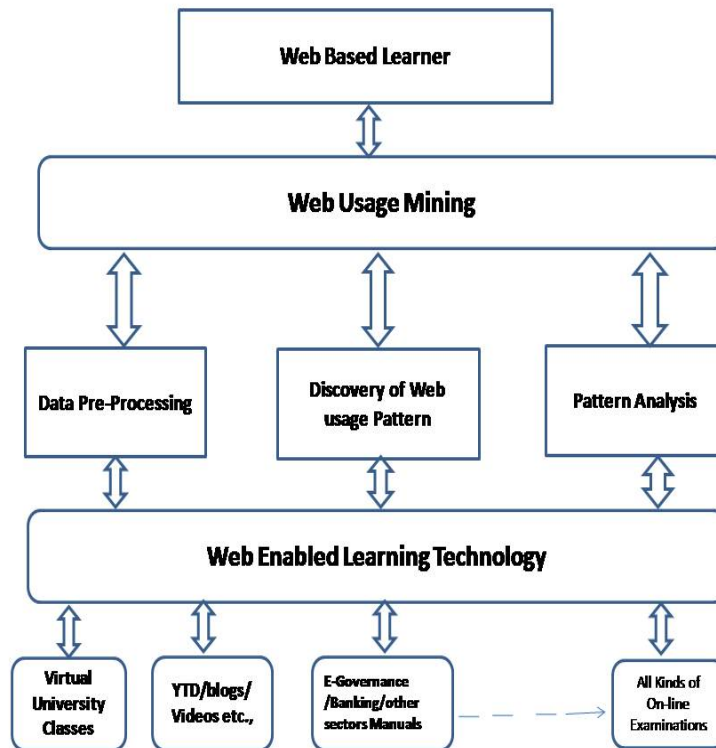


Figure. 4. Over all architecture for the Role of WUM in WBLE

A. WEB BASED LEARNERS

Web Based Learners are a individual/groups of people/a unit of organizations/ a classroom student strength/smart phone user/anybody who access the web based learning content. It is denoted as *Lns*.

B. WEB USAGE MINING

WUM consist of the three major steps for mining the webpage or web content access they are;

1. Data Pre-processin
2. Discovery of Web Usage Pattern
3. Pattern Analysis.

1. Data Pre-Precorsseing

This process takes the inputs from the log files kept in their system in all level of application the data should be stored in any one the location. The stored data can be retrieved and it is used for the next level operations. The data's categorized as Usage Data, Content Data, Structure Data, and User Data. In this stage the following process are usedie. Data Cleaning, Learner and Session Identification, Path Completion.The following Table 1. Represents the user details which include Browser type, URL, References, Agent, IP and Time.The table 2.Represents the Session details which include the session Id, time, IP, URL, Reference and agent. The table 3 represents the Attribute Definitions for Each Learner .

2. Discovery of web Usage Pattern

In this stage the accessed web content based on the type either in videos/text/audio which user can access the data, which type of data can be used. Numericaltechnique is used to abstract information about the website visitors. Then from this abstracted knowledge *Association rule* generates the association between frequently referenced pages and *Sequential pattern* tools helps in predicting future visit patterns. *DataClustering tools* group's comparable characteristics items together, most concerned groups in web usage mining tasks are image group, cluster, and page group,cluster, and *Classification* tool do the generalization process and combine together into one predefined class.

Table. I. User identification using browser type and IP address

Time	IP	URL	Ref	Agent		Time	IP	URL	Ref
0.05	2.3.4.5	A	B	IE8;Win2k		0.05	2.3.4.5	A	B
0.08	1.2.3.4	B	--	IE8;WinXP;SP1		0.08	1.2.3.4	B	--
0.13	1.2.3.4	C	A	IE8;Win2k		0.13	1.2.3.4	C	A
0.14	1.2.3.4	B	C	IE8;Win2k		0.14	1.2.3.4	B	C
0.17	1.2.3.4	E	--	IE7;Win2k		0.17	1.2.3.4	E	--
0.16	2.3.4.5	C	B	IE8;Win2k		0.16	2.3.4.5	C	B
0.21	2.3.4.5	D	A	IE7;WinXP;SP1		0.21	2.3.4.5	D	A
0.21	2.3.4.5	A	--	IE8;WinXP;SP1		0.21	2.3.4.5	A	--
0.24	1.2.3.4	E	C	IE8;WinXP;SP1					
0.24	1.2.3.4	C	A	IE8;Win2k	User 1				
0.35	2.3.4.5	B	B	IE8;Win2k					
0.59	1.2.3.4	D	C	IE7;WinXP;SP1					
1.21	1.2.3.4	E	B	IE8;Win2k					
1.25	1.2.3.4	A	D	IE7;WinXP;SP1					
1.26	2.3.4.5	C	--	IE8;WinXP;SP1					
1.27	2.3.4.5	F	C	IE7;WinXP;SP1					
1.26	1.2.3.4	F	C	IE8;Win2k					
1.31	1.2.3.4	B	A	IE8;Win2k					
1.37	1.2.3.4	D	B	IE7;WinXP;SP1					

Table.2. Example of identification of sessions

Session 1	0.24	1.2.3.4	E	C	IE8;WinXP;SP1
	0.24	1.2.3.4	C	A	IE8;Win2k
	0.35	2.3.4.5	B	B	IE8;Win2k
	0.59	1.2.3.4	D	C	IE7;WinXP;SP1
Session 2	1.21	1.2.3.4	E	B	IE8;Win2k
	1.25	1.2.3.4	A	D	IE7;WinXP;SP1
	1.26	2.3.4.5	C	--	IE8;WinXP;SP1
	1.27	2.3.4.5	F	C	IE7;WinXP;SP1

Table.3. Attribute Definitions for Each Learner

Name	Description
Average time	It is the average of the time spent by students on the specific topic, within a subject, in the selected data sample registered for the course.
Average Score	It is the average of the marks scored in assessment of each topic, within a subject, by the students in the selected data sample registered for the course
Total Attempts	These are the numbers of times a specific topic within a subject is referred by students in the selected data sample registered for the course
Marks Scored	These are the marks the student obtained in the assessment of a specific topic within a subject
n_assessment	Number of assessments done
n_assess_p	Number of assessments passed
n_assess_f	Number of assessments failed

3. Pattern Analysis

Pattern analysis is the last part of Web Usage Mining. This phase will filter out all unimportant patterns from the set found in the pattern discovery. Information query mechanism, such as Structured Query Language, is the most common form of pattern investigation method. Content and pattern information are also for filtering out patterns containing pages of usage types, content types or pages that match a certain hyperlink structure.

C. WEB ENABLED LEARNING TECHNOLOGY

The learning source content are videos and audio/power point presentation are stored in web server which permits to access the user or clients in 24x7 basis, the sources are ready to available to play in multimedia software or power point. Some source provider can give their content or information in free some of them collect some amount. The information accesses are maintained in log file here we follow the operations of log file.

1. LOG FILES

- ✓ **Web Server Log Files:** These log files resides in web server and notes activity of the user browsing website. There are four types of web server logs i.e., transfer logs, agent logs, error logs and referrer logs.
- ✓ **Web Proxy Server Log Files:** These log files contains information about the proxy server from which user request came to the web server.
- ✓ **Client browser Log Files:** These log files resides in client's browser and to store them special software are used

2. LOG FILES PARAMETERS

- ✓ Log files contain various parameters which are very useful in recognizing user browsing patterns .Below is the list of some of the parameters.
- ✓ **User Name:** Identifies the user who has visited the website and this identification normally is IP address.
- ✓ **Visiting Path:** It is the path taken by the user while visiting the website.
- ✓ **Path Traversed:** It is the path taken by the user within the website.
- ✓ **Time Stamp:** It is the time spent by user on each page and is normally known as session.
- ✓ **Page Last Visited:** It is the page last visited by the user while leaving the website.
- ✓ **Success Rate:** It is measured by downloads and copying activity carried out on the website.
- ✓ **User Agent:** It is the browser that user uses to send the request to the server.
- ✓ **URL:** It is the resource that is accessed by the user and it may be of any format like HTML, CGI etc.
- ✓ **Request Type:** It is the method that is used by the user to send the request to the server and it can be either GET or POST method.

3. TYPES OF LOG FILE FORMAT

There are mainly three types of log file formats that are used by majority of the servers.

1. **Common Log File Format:** It is the standardized text file format that is used by most of the web servers to generate the log files. The configuration of common log file format is given below in the box.

```
"%h %l %u %t \"%r\" %>s %b" common CustomLog logs/access_log  
common  
eg: 127.0.0.1 RFC 1413 frank [10/Oct/2000:13:55:36 -0700] "GET  
/apache_pb.gif HTTP/1.0" 200 2326
```

2. **Combined Log Format:** It is same as the common log file format but with three additional fields i.e., referral field, the user_agent field, and the cookie field. The configuration of combined log format is given below in the box.

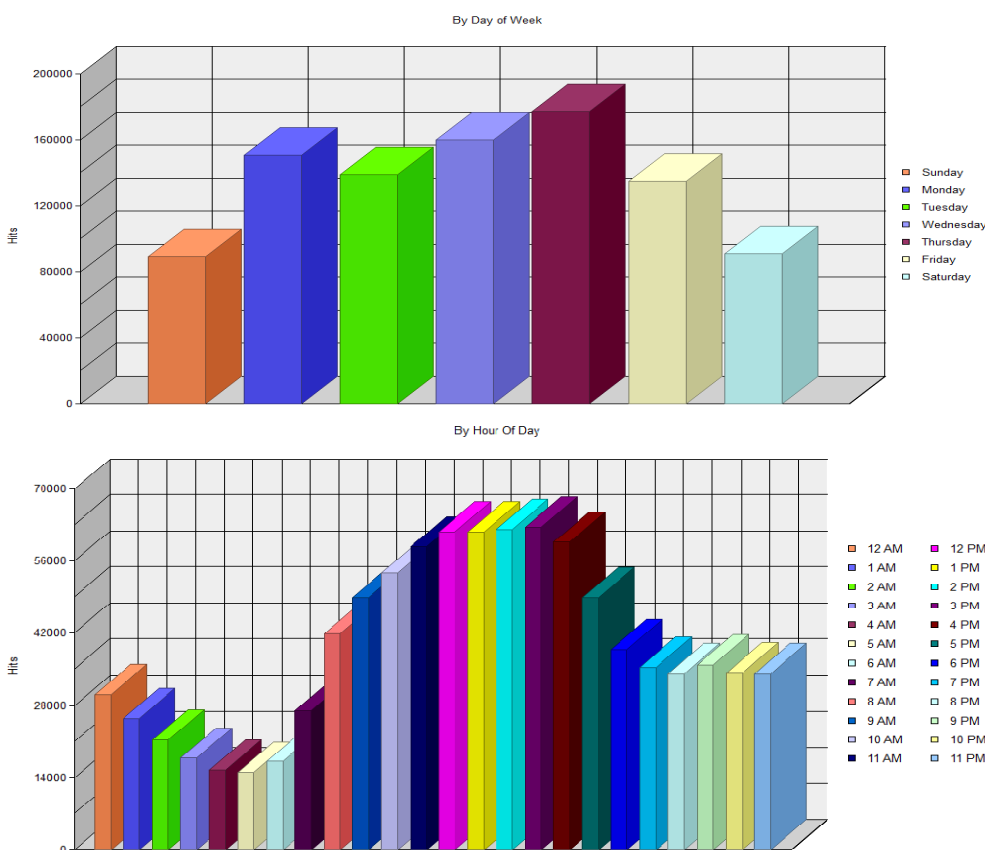
```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}N\" \"%{User-agent}N\"" combined CustomLog log/access_log combined  
eg: 127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET  
/apache_pb.gif HTTP/1.0" 200 2326  
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I  
;Nav)"
```

3. **Multiple Access Logs:** It is the combination of common log format and combined logfile format but in this format multiple directories can be created for access logs. Configuration of multiple access logs is given below in the box.

```
LogFormat "%h %l %u %t \"%r\" %>s %b" common
CustomLog logs/access_log common
CustomLog logs/referer_log "%{Referer}i -> %U"
CustomLog logs/agent_log "%{User-agent}i"
```

IV. RESULTS AND DISCUSSION

Analysis of web server log file from NEFT server resulted in recognition of various patterns. Technique “Converting IP address to domain name” helps in identification of visitor from the country they are sending request to the web server. Pattern recognized from grouping of visitors based on response code is helpful in identifying the visitors causing unnecessary traffic by requesting the web pages that are not available



V. CONCLUSION

Web usage mining is a non-trivial process of extracting useful implicit and previously unknown patterns from the usage of the web. The Significant research is invested to discover these useful patterns to increase effectiveness of e-learning sites. However, the goals of these applications and methods, turning visitors into purchasers, are different from the goals in E-learning: turning learners into efficient better learners.” We have seen some examples where data mining techniques can enhance on-line education for the educators as well as the learners. While some tools using data mining techniques to help educators and learners are being developed, the research is still in its infancy.

In addition, with the consciousness of the potential advantages of incorporated web usage mining and the inadequate data recorded by web servers, there is a need for more concentrate logs from the relevance side to enrich the information already logged by the web server. This added value by precise event recording on the E-learning side will give clickstreams and the patterns revealed a better significance and elucidation.

REFERENCES

- [1] SašaBošnjak,MirjanaMarić and ZitaBošnjak, "The Role of Web Usage Mining in Web Applications Evaluation", Management Information Systems, Vol. 5 [2010], No. 1, Page. 031-036, UDC 005.21:004.738.5.
- [2] Ms.Shital C. Patil and Prof. R. R. Keole, "The Role of Web Content Mining and Web Usage Mining in Improving Search Result Delivery".International Journal of Computer Science and Mobile Computing. IJCSMC, Vol. 3, Issue. 3, March 2014, pg.7 – 14.ISSN 2320–088X.
- [3] Anupama and Prasanth, "Web Usage Mining – Its Application in E-Services", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 3, Issue 2, February 2013, pp:572-576.
- [4] SunilI and Prof. M. N. Doja, "A Review Paper On Identifying Students Interest In E-Learning Using WebUsage Mining".International Journal of Latest Trends in Engineering and Technology,Vol.(8)Issue(1), pp.520-525, DOI: <http://dx.doi.org/10.21172/1.81.067>, e-ISSN:2278-621X.
- [5] Bamshadand Mobasher, Web Usage Mining, Chapter 12.Datamininig and Its Application.
- [6] Ping Zhou and Zhongjian Le, "A Framework for Web Usage Mining in Electronic Government, School of Information Management, JiangXi University of Finance and Economic, NanChang ,China 330013.PP:1169-1176.
- [7] Mohamed Koutheair Khribi1, Mohamed Jemni and OlfaNasraoui," Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval", Educational Technology & Society, 12 (4), 30–42.
- [8] Nanhay Singh, Achin Jain and Ram Shringar Raw, "Comparison Analysis Of Web Usage Mining Using Pattern Recognition Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.4, July 2013. DOI : 10.5121/ijdkp.2013.3410
- [9] K.Umadevi, B.UmaMaheswari and P.Nithya, "Design of E-Learning Application through Web Mining, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 8, August 2014. ISSN(Online): 2320-9801, ISSN (Print): 2320-9798.
- [10] Stavros Valsamidis and SotiriosKontogiannis," E-Learning Platform Usage Analysis", Interdisciplinary Journal of E-Learning and Learning Objects Volume 7, 2011,pp:185-204.
- [11] ShimaaAbd Elkader AbdElalal, "E-learning using data mining", Chinese-Egyptian Research Journal, Helwan University,pp:10-25.
- [12] Jiye Ai and James Laffey," Web Mining as a Tool for Understanding Online Learning",MERLOT Journal of Online Learning and Teaching,, Vol. 3, No. 2, June 2007'.PP:160-169.
- [13] Osmar R. Za'iane, "Web Usage Mining for a Better Web-Based Learning Environment", Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada.
- [14] NawalSael,AbdelazizMarzak, and HichamBehja, "Web Usage Mining Data Pre-processing and Multi Level analysis on Moodle", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013, ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784, www.IJCSI.org.
- [15] RenukaMahajan, J. S. Sodhi and Vishal Mahajan," Web Usage Mining for Building an Adaptive e-Learning Site: A Case Study", International Journal of e-Education, e-Business, e-Management and e-Learning, Manuscript submitted July 10, 2014; accepted August 29, 2014.doi: 10.7763/ijeeee.2014.V4.343.

AUTHOR PROFILE



Ms.J.Umarani is working as Assistant Professor in the Department of Computer Applications, Thanthai Hans Roever College, Perambalur, and Tamilnadu, India. She has 11 years of experience in teaching. He has published many research articles in the National / International Conferences and Journals. She is currently pursuing doctor of programme in Computer Science at Bharathiyar University, Coimbatore, and Tamilnadu, India. Her current area of research interests Data warehousing, Distributed Data Base and software metrics.



Mr.G.Thangaraju is working as Assistant Professor in the Department of Computer Sciences, Government Arts and Science College, Veppanthattai, Perambalur, and Tamilnadu, India. He has 20 years of experience in teaching. He has published many research articles in the National / International Conferences and Journals. He is currently pursuing doctor of programme in Computer Science at Karpagam Academy of Higher Education, Coimbatore, and Tamilnadu, India. His current area of research interests Data warehousing, Distributed Data Base and software metrics



Ms.S.Kavitha is working as Assistant Professor in the Department of Computer Applications, Dwaraka Doss Goverdhan Doss Vaishnav College (Autonomous), Arumbakkam, Chennai-600 106, and Tamilnadu, India. She has 18 years of experience in teaching. She has published many research articles in the National / International Conferences and Journals. She is currently pursuing doctor of programme in Computer Science at Bharathiyar University, Coimbatore, and Tamilnadu, India. Her current area of research interests Data Mining and warehousing, Distributed Data Base and software metrics.