RESEARCHERID
THOMSON REUTERS

ISSN: 2277-9655
[Dwivedi*., 6(12): December, 2017] Impact Factor: 4.116
IC™ Value: 3.00 CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

USER'S ACTIVITY ANALYSIS FROM WEBLOG: WEB LOG EXPERT

Sunita Dwivedi
Computer Department, Makhanlal Chaturvedi National University of Journalism and Communication, Bhopal, India

DOI: 10.5281/zenodo.1130885

ABSTRACT

In this era of World Wide Web, where gigantic amount of data is being generated by visitors and automated application software's in the web. These generated data is not just the data but it contains the jewels of information. To discover the hidden information in data, a lot of efforts have been done by various researchers where numerous techniques on different web log data sets have been applied. However, It has been applied here the Web log expert software tool on web data log data sets to extract the fruitful information by performing the deep analysis.

KEYWORDS: Web Log, Web Usage Mining, Web Log Expert

I. INTRODUCTION

Users interact over and over again with diverse web sites and can access plenty of information on WWW. The World Wide Web is growing constantly and huge amount of data is produced due to users' various interactions with websites. Web Usage Mining (WUM) [1] [2] [3] is the use of data mining methods to discover the fruitful and interesting patterns from web log data. It tries to know frequently retrieved pages, guess user navigation, develop web site structure etc. The web usage (weblog) data consists of the data from browser logs, proxy server logs, web server logs, and user profiles. In WUM, data mining techniques are used to the pre-processed web usage data in order to uncover interesting and useful patterns. User's browsing activities is recorded into web server log. By analyzing web log files, various questions can be addressed like- Who is visiting which page? From what search engine are visitors visiting? What are the most recent visits for each page? What pages are being accessed regularly? Which operating systems and browser are most commonly used by visitors? And many more similar questions and their answers, various web log analysis tools have been developed so far to each the web usage mining process. Deep web log analyzer [5], webalizer, Google analytics, web log experts[8] etc.? Here in this research paper web log data has been analyzed using one of the most popular web log expert software tool.

The rest of this paper is organized as follows - section 2 presents brief literature review. Section 3 shows methodology for web log analysis. Section 4 presents the results and discussion. Section 5 presents a conclusion and future works.

II. LITERATURE REVIEW

This section focuses on some of the related works that has already been done in this area. A lot number of researchers has contributed their efforts in this very important research field. This is as follows-

V. Kumar *et* al. [5] have performed briefly an investigation on various popular web log analysis software tools available in the market. This critical investigation helps in choosing the appropriate analysis tool.

V. Kumar *et* al. [8] has performed the web log analysis over weblog data using the Deep weblog analyzer software. They performed this analysis over all HTTP requests made in one month to the NASA Kennedy Space Center World Wide Web server in Florida. This log file is 195 MB (205,242,368 bytes) in size and it has 1891715 lines of web requests which were hit by the visitors.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Kumar and Alagarsamy[11] in has discussed about the web usage mining, different file format of log data and carried out the analysis task with the help of web log expert tool.

S. Bhuvaneswari and T. Anand [12] have done a short exploration on various automated web log analyzing software such as - Web Log Expert, Google Analytics, Open Web Analytics, Deep Log Analyzer, Piwik Webalizer, and AWStats. These tools are calibrated for its virtues and demerits to choose the proper software for web log analysis task. They performed analysis under heads such as general statistics (GS), activity statistics (AS), Access statistics (AS), Visitor's information (VI), spiders/error reports and browser information.

A. Tyagi and S. Choudhary[15] have an analysis job using the Web log expert software tool. They have also discussed the logical order in the web usage mining task i.e. data collection, preprocessing, pattern discovery using the analyzer.

III. METHODOLOGY FOR WEB LOG ANALYSIS

This section discusses the methodology used in this process of web log analysis from the web log data. The web log analysis task starts from the collection of web log data [7][9] from the different sources. The sources may be the web servers, application servers and in some cases it may be in a proxy server that intermediately stores the log information at end of Internet service providers (ISP's). Thus log file is generally a text file which records the information about all the activities performed by the visitor during his visit on the website. The log file exists in various formats- such as- Apache; IIS log files Common Log Format (CLF), NCSA Common Log Format, and Extended Log file format, ASCII file format. Here, dataset- ClarkNet-HTTP [13] is used for analysis purpose. This contain of all HTTP requests to the ClarkNet WWW server. ClarkNet is an Internet access provider (IAP) for the Metro Baltimore-Washington DC reason. ASCII log file format is given for demonstration-

Log Format

Web log file taken here is in ASCII file format and each line of log is a request made by visitor and comprised of following fields [13]-

1. **Host making the request**. A hostname when possible, otherwise the Internet address if the name could not be resolved
2. **Timestamp in the format** "DAY MON DD HH:MM:SS YYYY"
3. **Time zone** is **-0400**
4. **Request** given in quotes
5. **HTTP reply code**
6. **Bytes in the reply**

Table 1: A sample snapshot from web log file

S No	User's request as stored in web log file
1	204.249.225.59 - - [28/Aug/1995:00:00:34 -0400] "GET /pub/rmharris/catalogs/dawsocat/intro.html HTTP/1.0" 200 3542
2	access9.accsyst.com - - [28/Aug/1995:00:00:35 -0400] "GET /pub/robert/past99.gif HTTP/1.0" 200 4993
3	access9.accsyst.com - - [28/Aug/1995:00:00:35 -0400] "GET /pub/robert/curr99.gif HTTP/1.0" 200 5836
4	world.std.com - - [28/Aug/1995:00:00:36 -0400] "GET /pub/atomicbk/catalog/sleazbk.html HTTP/1.0" 200 18338
5	cssu24.cs.ust.hk - - [28/Aug/1995:00:00:36 -0400] "GET /pub/job/vk/view17.jpg HTTP/1.0" 200 5944
6	er6.rutgers.edu - - [28/Aug/1995:00:00:37 -0400] "GET /pub/rjgula/network.htm HTTP/1.0" 200 2017
7	cyclom1-1-6.intersource.com - - [28/Aug/1995:00:00:37 -0400] "GET /pub/k2/jeep/jxj.htm HTTP/1.0" 200 3254

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

The next step in web log analysis is the pre-processing [4] of the web log data where various techniques such as – cleaning, transformation, sessionisation, user identification is performed to go further for analysis task [6]. Here, the preprocessing task is carried out automatically by the web log expert software tool internally. The next step core task, the major action is performed i.e. Analysis- it involves some intermediate functions- such as pattern discovery, classification, filtering, and clustering [5]. The web log expert does it all in the as a background process when web analyzer is run over web datasets. In the final step, the visualization of the results obtained in the previous step is done. This helps in better interpretation and presentation of the analysis which generally exists in forms of some statistical data.web log expert does it very well, and facilitates to demonstrate the result in interactive graphical and tabular form. The result can be visualized and interpreted by drilling down and up by navigating among levels of data ranging from most summarized to most detailed form.

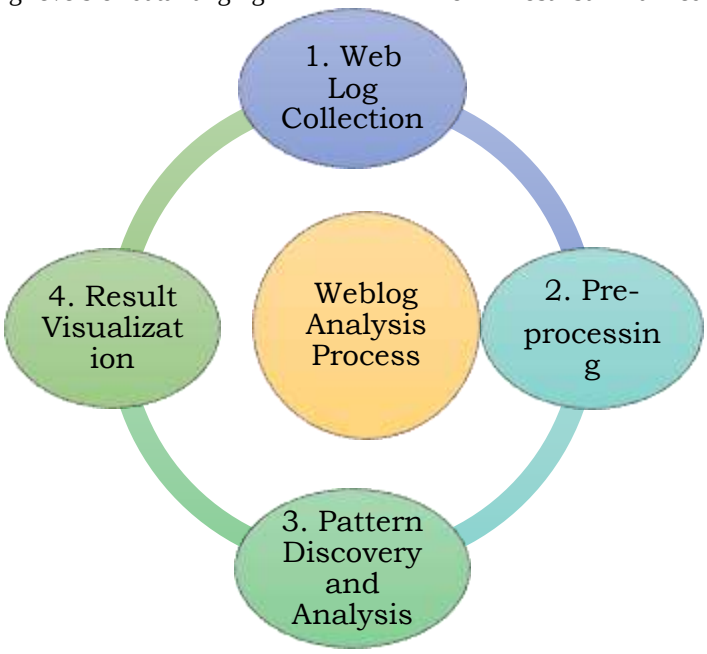

Figure 1: Weblog analysis process

Web Log expert –Web Log Analysis Software

WebLog Expert [14] software tool is a commercial, fast and powerful web log analyzer. It will furnish the information about any website's visitors: activity statistics, accessed files, paths through the site, information about referring pages, search engines, browsers, operating systems, and more. The tool generates easily communicable reports that incorporate both charts and text information (tables). The log analyzer can generate information in PDF HTML, and CSV formats. It comes under three edition- standard /Enterprise/ and Professional editions it latest release version is 9.4 (released on 8/11/2017) with memory size 17.7 MB.

IV. RESULTS AND DISCUSSION

When the analysis over the access log file was performed using the weblog Expert tool, the following report were generated. Although, Web Log Expert tool populates rigorous report under different heads but here only some of the important reports are included due to the space limitation.

General Statistics Report

This report demonstrates the website's significant statistical result giving an instant snapshot of activities performed over the website. This summary report[5] (shown in table 2) is the starting point of the website analysis.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Time range: 8/28/1995 09:30:34 - 9/4/1995 09:29:00	
Summary	
Hits	
Total Hits	16,54,579
Visitor Hits	16,54,579
Spider Hits	0
Average Hits per Day	2,06,822
Average Hits per Visitor	11.84
Cached Requests	1,33,950
Failed Requests	36,936
Page Views	
Total Page Views	3,61,125
Average Page Views per Day	45,140
Average Page Views per Visitor	2.58
Visitors	
Total Visitors	1,39,721
Average Visitors per Day	17,465
Total Unique IPs	90,501
Bandwidth	
Total Bandwidth	13.46 GB
Visitor Bandwidth	13.46 GB
Spider Bandwidth	0 B
Average Bandwidth per Day	1.68 GB
Average Bandwidth per Hit	8.53 KB
Average Bandwidth per Visitor	101.03 KB

Activity analysis at various granular levels

This subsection demonstrates the graphical representation for the analysis of the results obtained from the web log data. Figure 2 displays the activities performed by the day of weak by visitor. On Thursday, more visitors hit the website and the lowest hits on Sunday.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

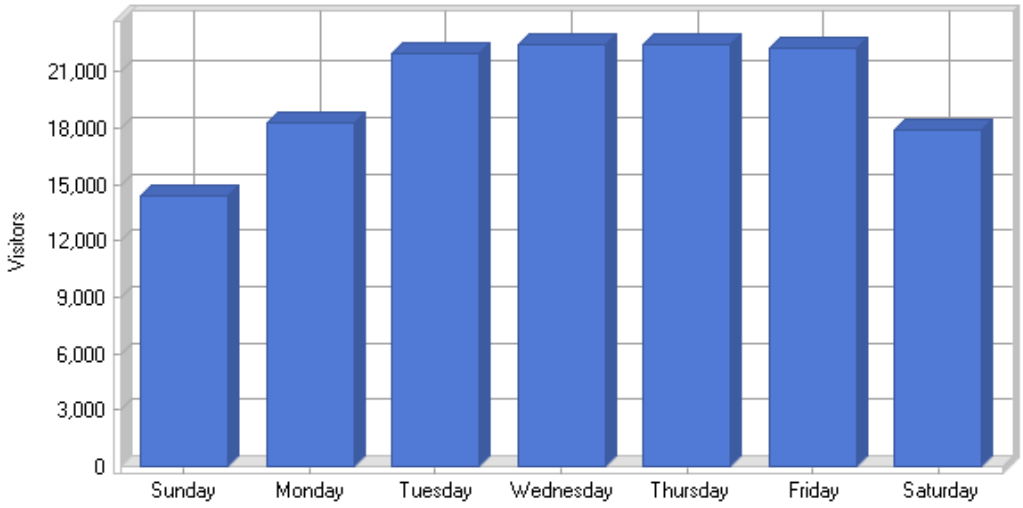

Figure 2 : Activity by Day of Week

With the help of figure 3, it can be stated that in 24 hours, in beginning hours more traffic is experienced on the website. Website experiences the minimum traffic at 15th hour in the whole day. Figure 4 and figure 5 show the visitor's activity in a little higher level i.e. monthly and weekly basis.

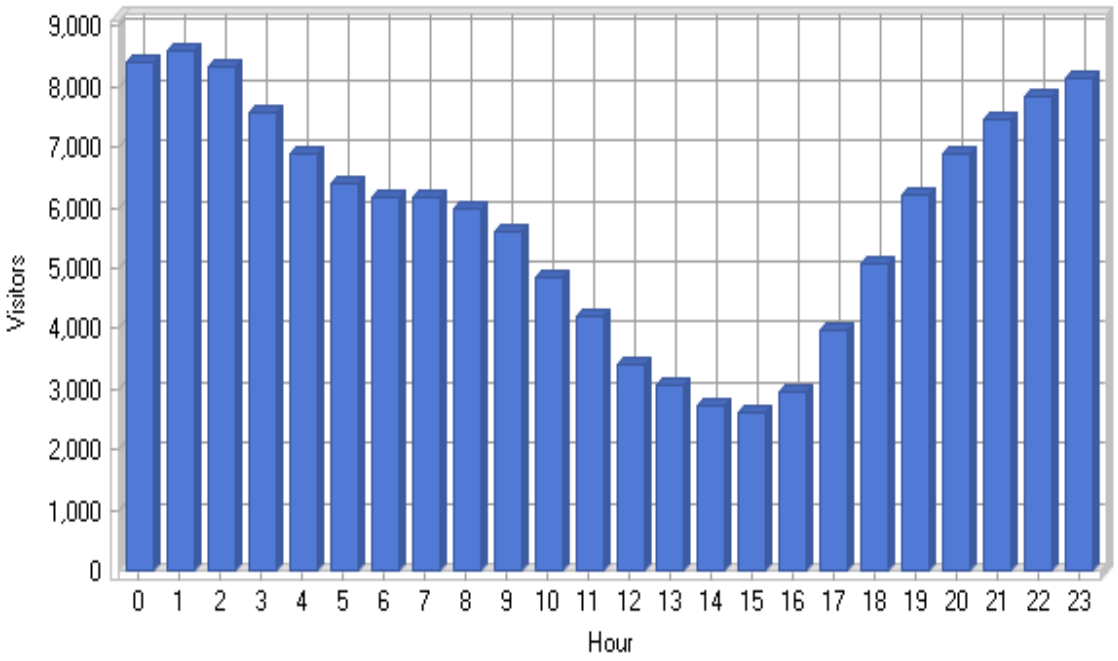

Figure 3: Activity by hour of day

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

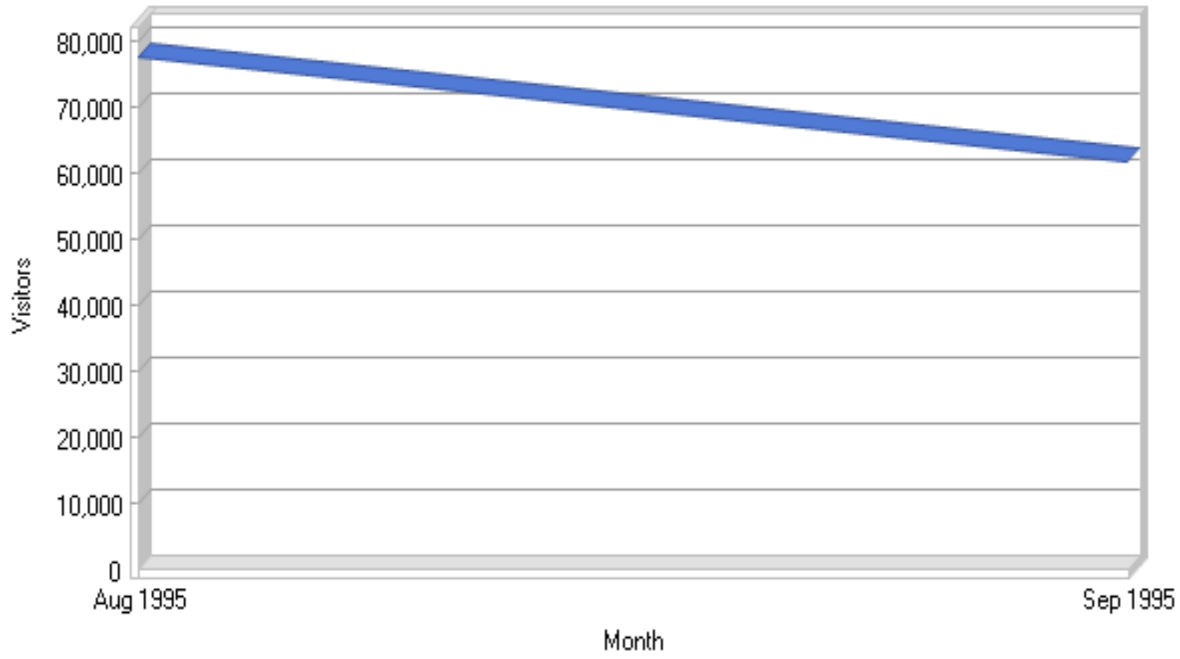

Figure 4: Activity by Month

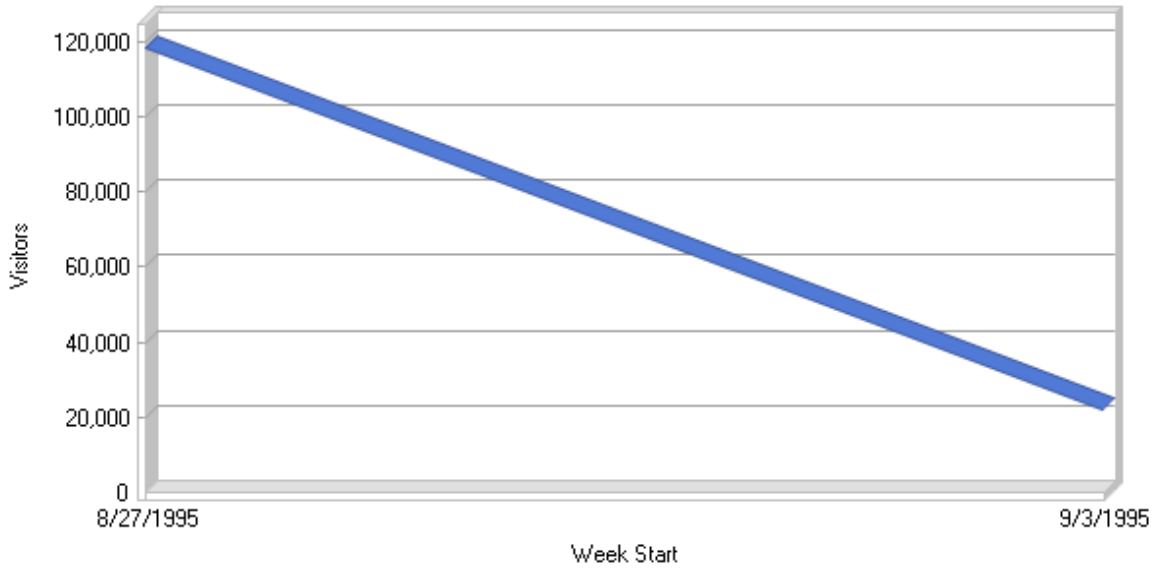

Figure 5 : Activity by Week

Figure 6, presents a very crucial result i.e. bandwidth consumption on each and every day that helps to monitor and plan the traffic management over the website for better service to the visitors.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

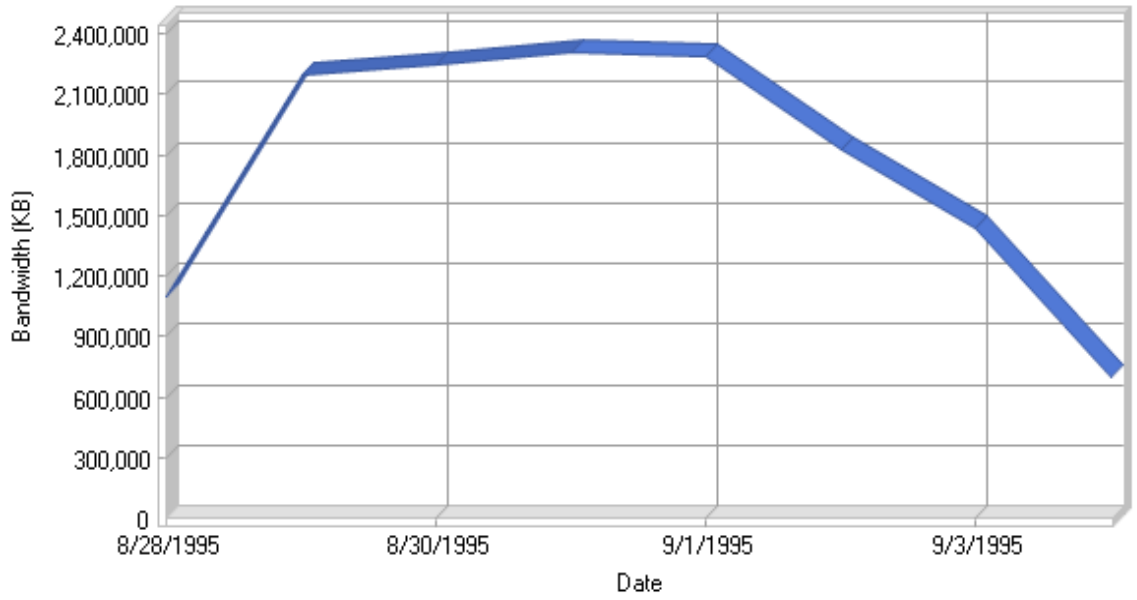

Figure 6 : Daily Bandwidth Consumed

Granular Level daily activity analysis

Figure 7 and Figure 13 displays the error and types of error that occurs during the users visit over the website. These figures depicts that 404 errors has occurred in most of the cases when users have tried to access the resource in the website. This error can be reduced by affixing the unavailability of resources in the website.

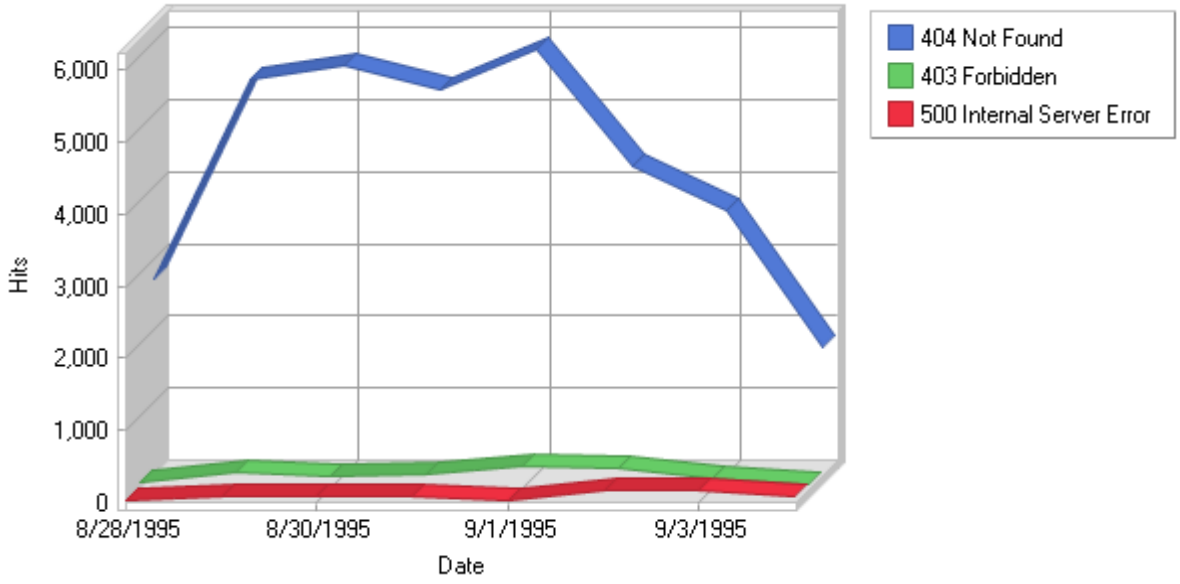

Figure 7 : Daily Error Types

Figure 8 brings out the pictorial statistics of number of visitors visiting the website daily form the different countries of the world. United States gives the more visitors on the website. Figure 9 shows the visitors coming from the various domains to the website. Figure 12 demonstrate the number of files that are accessed from the website.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

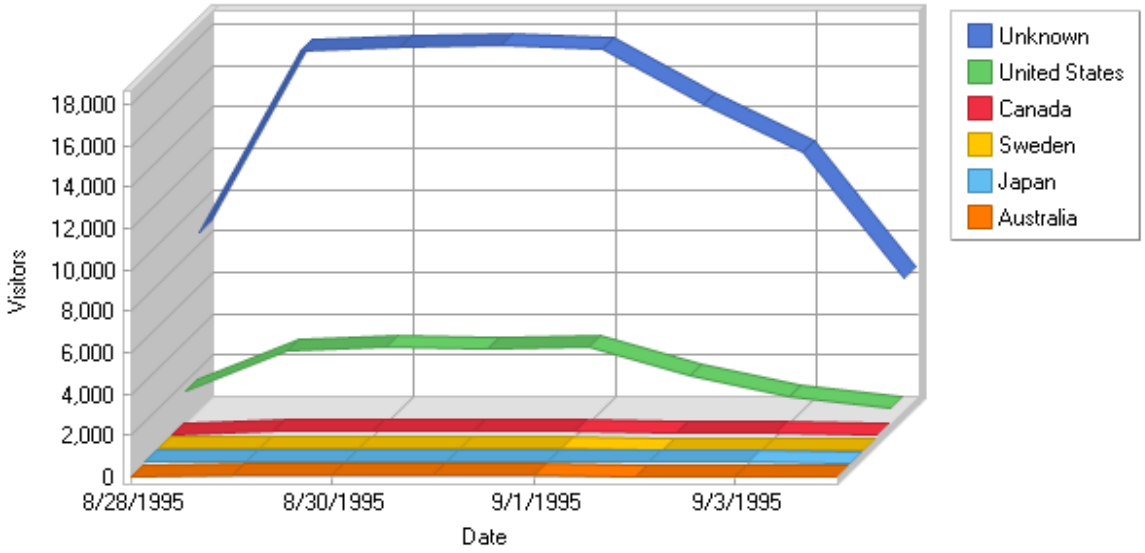

Figure 8: Daily Countries Activities

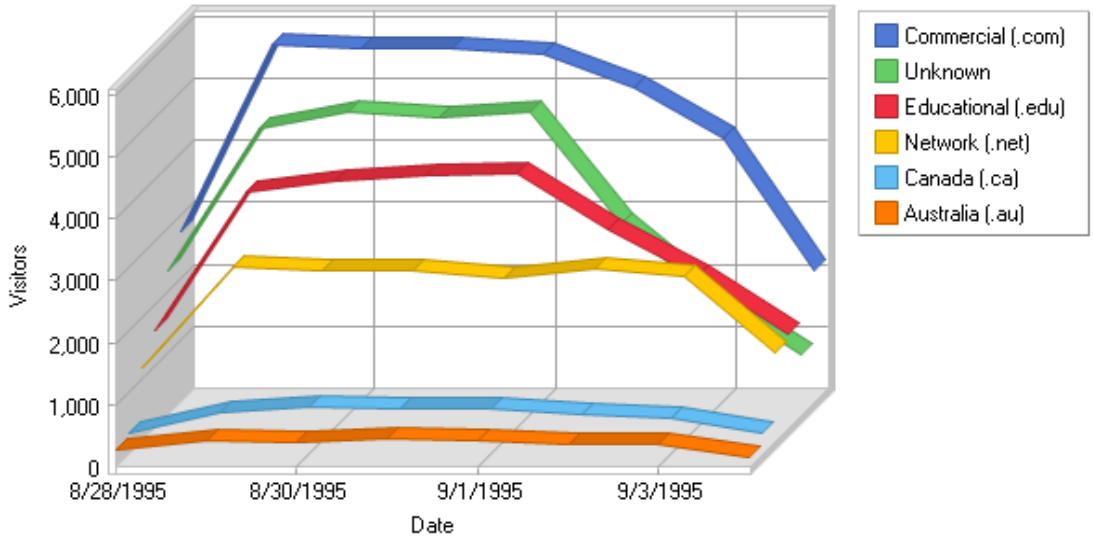

Figure 9 Daily Top Level Domains

Figure 10 and Figure 12 shows the daily, how many times a particular page and file is accessed from the website by the visitors. This provides the information about the most and moderate and least important pages in the website. It is helpful in restructuring the website for better design of user interface for visitors. Figure 11 shows the number of hits experienced by the website on daily basis.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

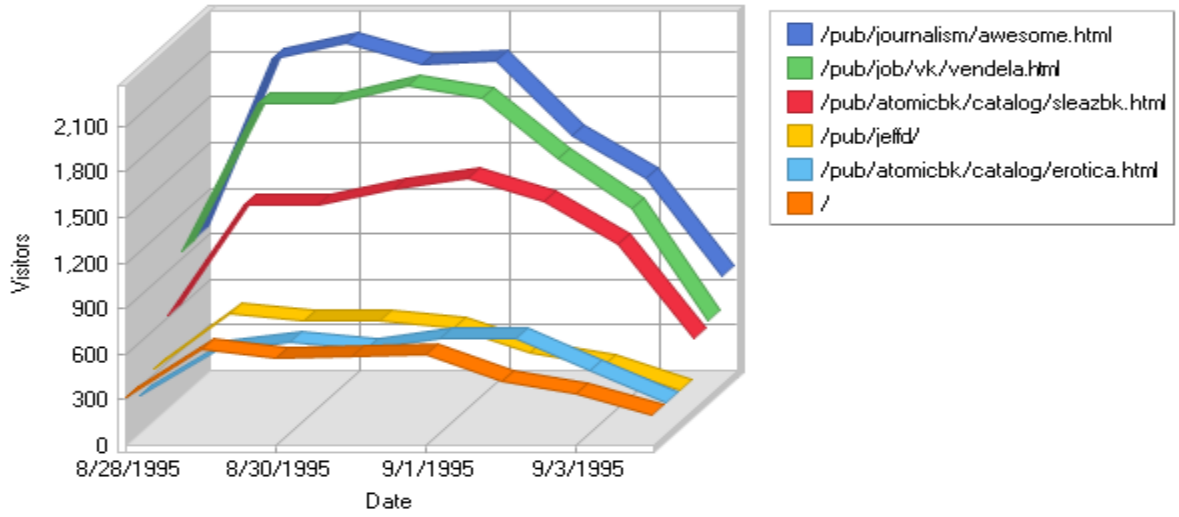

Figure 10: Daily Page Access

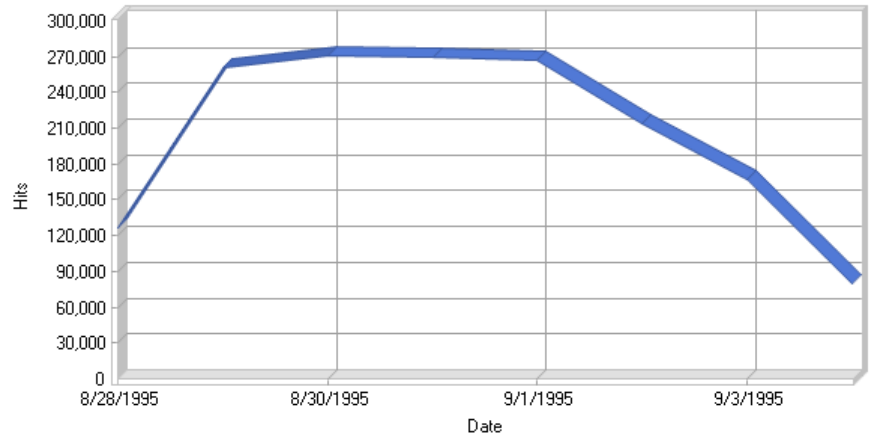

Figure 11: Daily Visitors

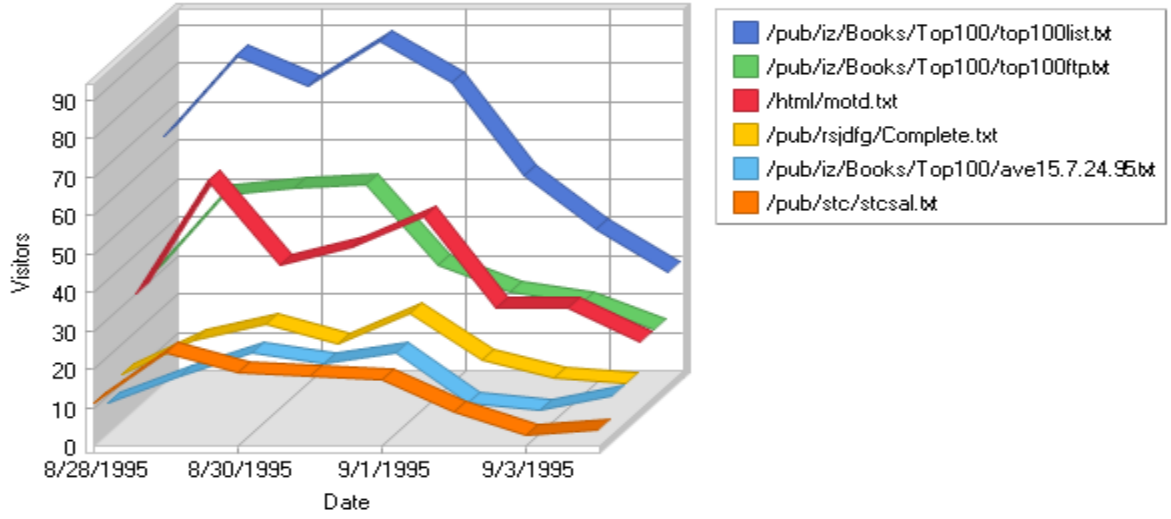

Figure 12: Daily File Access

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

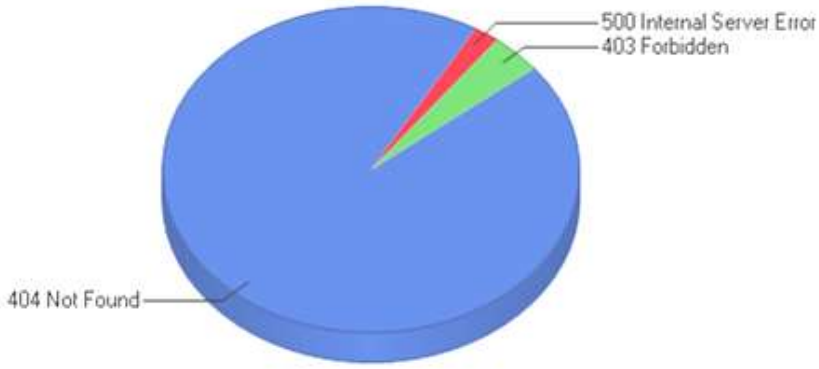

Figure 13: Error Types

Figure 14 shows the name of the most participating states against the number of visitors. California State gives the highest number of visitors to website visit while Georgia sate of united state gives the lowest number of visitors.

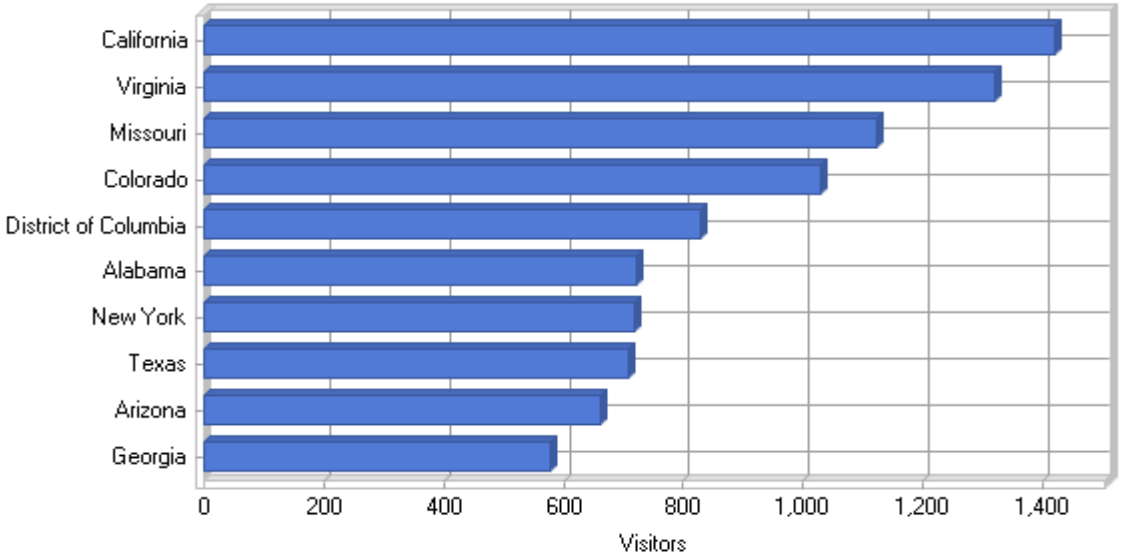

Figure 14: Most Active States

Figure 15 shows the most downloaded files by the visitors from the website in a given period of time. The Country location, resources of website, traffic/bandwidth all collectively can be properly managed to better service delivery to the users by the website.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

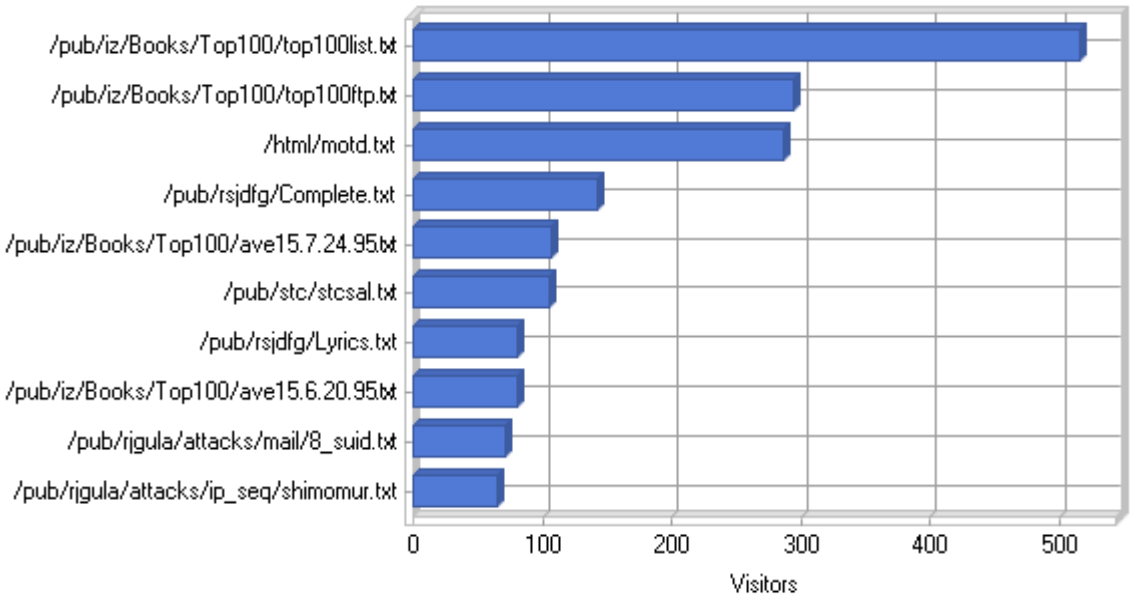

Figure 15: Most Downloaded Files

V. CONCLUSION

Web log analysis is the process to find useful and interesting information from web usage data. It is helpful in e-commerce to enhance structure of website, identifying customer behavior, business intelligence, and personalization. The web usage data contains data from web server logs; browser logs, proxy server logs etc. In this paper, the process of web log analysis has been described which consists of steps: Data Collection, pre-processing, pattern discovery, pattern analysis and result analysis visualization with web log expert software tool. The process is helpful in better designing of website, identifying the most profitable place for marketing, reducing the bouncing rate of visitors, and traffic management. In future, the comparative study of web log analysis tool will be done.

VI. ACKNOWLEDGEMENTS

This work was supported by Makhanlal Chaturvedi National University of Journalism and Communication, Bhopal, India for carrying out my research work.

VII. REFERENCES

[1] C.J. Carmona, et al. (2012). Web usage mining to improve the design of an e-commerce website: OrOliveSur.com, Expert Systems with Applications 39, 11243–11249.
[2] Neha Goel and C.K. Jha (2013) “Analyzing Users Behavior from Web Access Logs using Automated Log Analyzer Tool”, International Journal of Computer Applications , Volume 62– No.2, January.
[3] Etzioni, O. (1996). The World Wide Web: Quagmine or gold mine. Communications of the ACM, 39, 65–68.
[4] L. K. Joshila Grace, V. Maheswari, Dhinaharan Nagamlai. (2011).Web Log Data Analysis and Mining, Advanced Computing Communications in Computer and Information Science, Vol. 133 pp 459-469.
[5] Vinod Kumar, R S Thakur, "Exploring Behavior of Visitors Activity at Granular Level from Web Log Data using Deep Log Analyzer”, International Journal of System and Software Engineering, IJSASE, ISSN: 2321-6017, Vol. 4(1), pp.16-26, 2016.
[6] Mobasher, B. (2005). Web usage mining and personalization. CRC Press, LLC. Moral-Pajares, E., & Lanzas-Molina, J. R. (2009). La exportacion de aceite de oliva virgen en Andalucia: Dinamica y factores determinantes. Revista de Estudios Regionales, 86.
[7] Liu, B. (2006). Web data mining: Exploring hyperlinks, contents, and usage data (datacentric systems and applications). Springer-Verlag.
[8] Vinod Kumar, RS Thakur, “A Brief Investigation on Web Usage Mining Tools (WUM)”, Saudi Journal of Engineering and Technology, ISSN: 2415-6264, Vol. No. 2 (1), pp. 1-11, 2017

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

[9] Facca, F. M., & Lanzi, P. L. (2005). Mining Interesting Knowledge from Weblogs: A Survey, 53, 225–241.
[10] F. Bounch, F. Giannotti, C. Gozzi, G. Manco, M. Nanni, D. Pedreschi, C. Renso, S. Ruggier (2001). Web log data warehousing and mining for intelligent web caching, J. Data Knowledge Eng. 36- 165–189, PH: S0169-023x (01)00038-6.
[11] V. Jaya kumar, Dr. K. Alagarsamy (2012).Analyzing server log file using web log expert in web,data mining ,International Journal of Science, Environment and Technology, ISSN 2278-3687(O)Vol.2, Issue No.5,2013, pp.1008–1016.
[12] S. Bhuvaneswari, T. Anand. (2015).A Comparative Study of Different Log Analyzer Tools to Analyze User Behaviors, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Vol. 3 Issue 5, 2015, pp. 2997-3002.
[13] Web Log Data (1995)," http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html". (Accessed December 10, 2017)
[14] Web Log Expert https://www.weblogexpert.com/ (Accessed December 10, 2017)
[15] Arti Tyagi,Sunita Choudhary. (March 2015). Web Usage Mining Using Web Log Expert Tool, International Journal of Advanced Research in Computer Science and Software Engineering,ISSN: 2277 128X,Vol. 5, Issue 3.

CITE AN ARTICLE

Dwivedi, S. (n.d.). USER'S ACTIVITY ANALYSIS FROM WEBLOG: WEB LOG EXPERT. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(12), 547-558.