Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan Vol. 23 No. 1-2018
ISSN:1813-1786 (Print) 2313-7770 (Online)

Web Usage Analysis: A Case Study of Transforming ODL Programs into E-Learning

M. U. Ahmed[1], M. Aquil[2], S. Farid[3]

[1,2] Computer Science Department, Allama Iqbal Open University, Islamabad, Pakistan
[3]Computer Science Department,Bahauddin Zakariya University, Multan
[1]moiz.ahmed@aiou.edu.pk

***Abstract*-**Nowadays, the Open and Distance Learning is a widely used mode for dissemination of information to the distant learners. This emerging mode of distance learning is based on e-learning, a modern way of education through the use of Internet and world wide web. The design and structure of e-learning website plays a pivotal role in reducing the geographical distance among teachers and students. However, creating interactive and user-friendly website requires the information about the usage of website by the potential users. Web usage mining and analysis is a way to find the interesting pattern of web usage. This paper presents the web usage analysis of the distant learners while browsing the contents on e-learning website of Allama Iqbal Open University, Pakistan. The web usage log file is collected from the Ubuntu web server and is preprocessed for the pattern generation. The open source log file analyzer Web Log Expert 9.4 has been used for the web usage analysis. The analysis of AIOU e-learning website explores some intriguing patterns of e-learners in terms of number of visitors, number of hits, usage of bandwidth, and errors generation. The result shows the transformation of distance learning programs into e-learning has been successful, and students are actively using the e-learning website.

***Keywords*-**E-Learning, Web Usage Mining, Distance Learning, ODL, Log File, WebLog Tool, Online Learning Analysis.

I. Introduction

Open and Distance Learning (ODL) is a non-traditional field of education in which students and teachers do not interact in a class room environment. It is a flexible mode which can provide education at the doorstep of the learners. The impediment of geographical distance is overcome by mailing of educational contents using different modes of communication, making it accessible for the students [i]. However; the delivery and accessibility modes of ODL are rapidly transforming with the inclusion of Information and Communication (ICT) Technology. The development of information technology is gradually shifting ODL environment towards web-based learning.

Therefore, e-Learning is a self-directed mode of education in which students and teachers uses Information and Communication Technology (ICT) to acquire education and knowledge. With the initiation of e-learning, we are witnessing an increased activity of distance learners on various academic web sites. It requires continuous interaction and engagement [ii], resulting in accumulation of large amount of data [iii]. So, an earnest requirement is to enhance the available content on e-learning websites. In order to meet the changing needs of distant learners [iv]. In the given context, we can employ the web mining techniques to analyze the usage patterns of distant learners. Moreover, it can help to address the miscalculation and complains of web learners by following their desired links [v].

Web Usage Mining (WUM) in a web-based learning environment is “the application of data mining techniques to extract knowledge from web data”, including web documents, hyperlinks between documents and the usage of web sites log, etc. [vi]. Web usage analysis is the third category of web mining, which is used to discover the user's access pattern. It provides the users' interaction information while browsing the contents through the click series of hyperlinks [vii] which helps to investigate the preferences of users. Therefore, the apex educational institutions specially the universities have taken this initiative to evaluate the usage frequency of potential users, present & rearrange the digital contents to suit the styles of distant learners.

In Pakistan, Allama Iqbal Open University (AIOU) is the first distance learning university in Asia, and second in the world. University has more than half a million students enrolled in collection of courses and programs [viii]. Since the establishment of ICT, university is transforming its ODL programs into e-learning mode. University has established a directorate of overseas and e-learning for the promotion of technology integration and extension to out-reach its educational programs. In the first phase, university has selected graduate and post graduate programs for the e-learning services. These services are provided through an open source learning management system; i.e.; MOODLE (Modular Object-Oriented Dynamic Learning Environment) [ix].

To extend the e-learning services to other educational programs, it is necessary to investigate the web usage analysis of the e-learning activities. It is pertinent that any effective educational website can reduce the geographical distance aspect in e-learning environment [x]. In this regard, the primary objective of this paper is to study and analyze the AIOU e-learning website and discover some useful and significant usage patterns.

The rest of the paper is organized as follows; the literature review is described in section II, followed by our proposed model as elaborated in section III. Section IV presents and discusses the results. Finally, section V deals with the conclusion and implies future extension of the proposed research.

II. Literature Review

WUM is the most elaborated and researched category of Data Mining. It analyzes the web log documents to discover the interesting patterns of website usage [xi]. Currently, With the evolution of internet technology, web usage analysis has gained a lot of attention from the academic researchers and professionals as well. Scholars have explored the various patterns of web mining particularly web usage analysis [xii]. In the given context, the existing literature on web mining deals with the usage analysis, techniques, and domain of business knowledge. Although, the collection of data is more complex in WUM because data is stored in a web server configured for hosting web application [xiii].

Initially, the WUM was focused on the analysis of business websites, in order to increase the sale and to gain more profit. But, later it shifted its focus towards education and e-learning websites to improve the learning and teaching system [xiv]. As a result, a number of research studies have been conducted in the field of web usage analysis and mining. The study conducted by Aldekhail, highlighted that the quality and quantity of web pages can only be improved by following the 'click sequencer' and the 'time spend' on the pages. The study further revealed that the designer can rely on the WUM technique to gather the useful information about visitor's behavior in order to redesign and reform the existing website [xv].

Khan, Singh and Sharma in 2018, conducted an empirical study on a company's website. The objective was to analyze the server log files of the website. It took one-month data from the website log files and capturing interesting statistical analysis by using the web log expert tool. During these researches it was found that users were more interested in clicking the footer links rather than header links. This would help them in redesigning the website for better results [xvi]. Dharmarajan Proposed two algorithms for discovering the interested pages. It was done by calculating the log file parameters of timestamp and memory size of the web pages. He further applied those algorithms on the web server log file and get the statistical analysis which are beneficial for personalizing the services to e-learners[xvii].

In another experimental study, Singh et al. argued that the browsing sequence is helpful for the analysis of web usage. The result shows that the browsing pattern can be used to determine the number of accesses to the server and individual files, the time of visits, the domain names, and URLs of users. In addition, it also explored that how much bandwidth consumes while downloading the images and animations [xviii]. Furthermore, Chinnaiyan and Ilango conducted research on the educational website and used the Web Log Expert Tool to produce a website statistical information. The main purpose of this research was to analyze the visitor's behavior by mining the web access log files [xix]. Similarly, Kaur and Aggarwal conducted a research in 2015, on an educational website to analyze the log file of the website. Researchers find out the popular hour of the usage in a particular day and also investigated the favorite content browsed by the users. Web log Expert Lite 8.6 tool was used for analyzing the log file [xx].

In 2016, Meghawal and sharma used WebLog Expert Lite 7.8 tool on an educational website server log file to find out statistical data of web site. Researchers focused on website errors, broken links, and corrupt data, and they discovered very interesting facts which would help web masters in maintaining their websites [xxi]. Sunil and Doja analyzed the web log file of "History of the Computers" blog. The access patterns of learners are collected with the help of recommender agents called web widgets named flag counter and feedjit. It is a client-side browser-based application. It analyzed the different types of information provided by the website and gives recommendations to improve the website [xxii].

Nonetheless, web analysis did not confine itself in entrepreneur web world. Researchers widened their area of interests and started utilizing web mining techniques in educational websites. But universities and educational institutes are still struggling to design user friendly and information focused websites for the visitors. It is a herculean task to manage huge amount of information in a compact way that can be easily graspable and accessible for a large community of users. Especially, universities contemplated for distance and e-learning education confronted with the major challenges to create comprehensive and engaging e-learning websites [xxiii].

III. Web log Analysis Proposed Model

During the first phase, the university has started the integration of e-learning services with the M.Phil. and Ph.D programs. These programs are selected as a first step because most of the M.Phil. and Ph.D research scholars have their own laptops and available internet connections. E-learning website of AIOU (http://www.olive.aiou.edu.pk) is hosted on a web server at computer center of the university. Computer center is running university's main web portal [xxiv] and other websites including the e-learning portal by managing the web servers. E-Learning website is hosted on an apache-based web server, which stores the web transactions in a log file. This log file is important to get technical information about visitor's access pattern. It is a default text file which consists the record of transactions of the website hosted on Ubuntu (a Linux based operating system).The proposed model of web usage analysis is shown in Fig. 1. The model has four important components: data collection, data preprocessing, pattern discovery, and pattern analysis as shown in Fig. 1.

1. Data Collection: The First step of web usage mining is to collect data for analysis from the available sources. In this case data is collected from the web log file which is located at the university's web server. Web log file is generated from the 'click sequence' of users by visiting the website. It generates the access pattern and web usage of website by the visitor. It is containing the visitors' data and track their activity till the time visitor leave the website.
2. Data Preprocessing: The collected data is unstructured and in a raw form. To bring data in some meaningful shape, it needs to be clean. During data preprocessing the unwanted and irrelevant data is removed from the log file. At this stage the information about the user, geographical location, time-stamp is gathered and data is converted into an organized form for further analysis.
3. Pattern Discovery: This method is performed on the preprocessed data. At this stage the various techniques are applied to detect the organized patterns. Different techniques such as, pattern recognition, statistical analysis, association rules algorithms, clustering, classification and sequential pattern analysis are some techniques used for pattern discovery.
4. Pattern Analysis: This final phase compares and creates links between the achieved results. The discovered patterns are perceived to dig out the useful information about the access log parameters and remove the uninterested patterns. [xxv].

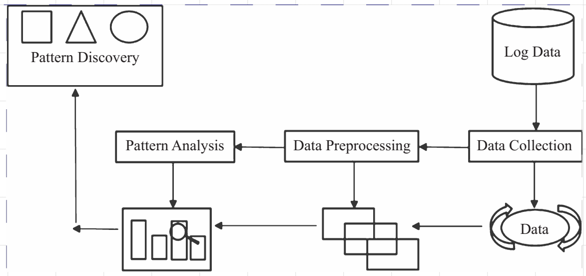

Fig. 1. Proposed Model of Web Log Analysis

IV. RESULTS

The data of 38 educational programs is selected, in which 1500 students are enrolled with an average of 40 students per program. The log file is collected for the activities executed in these programs for Spring 2017 semester. Furthermore, an open source web usage mining tool, Web Log Expert 9.4 is downloaded and configured for the analysis of log data. The general activity statistics is shown in Table I. There are three parts of the usage statistics about number of visitors, number of hits, and bandwidth consumed. The general statistics shows that there are 2,993,998 hits, and 32,077 visitors who have visited the website during Spring 2017. The utilized bandwidth during semester activities is 82.54 GB with an average of 626.12MB per day. The number of hits have been further separated into smaller units such as; visitor hits, spider (robot) hits, average hits per day and average hits per visitor. Furthermore, cached request and failed requests are also presented in general statistics. Similarly, the bandwidth has also been defined in terms of visitor bandwidth, spider bandwidth, and average bandwidth per day, per visitor and per hit respectively.

TABLE I
GENERAL ACTIVITY STATISTICS

Visitors	
Total Visitors	32,077
Average Visitors per Day	237
Total Unique Ips	15,295
Hits	
Total Hits	2,993,998
Visitor Hits	1,315,919
Spider Hits	1,678,079
Average Hits per Day	22,177
Average Hits per Visitor	41.02
Cached Requests	64,387
Failed Requests	20,192
Bandwidth	
Total Bandwidth	82.54 GB
Visitor Bandwidth	72.53 GB
Spider Bandwidth	10.01 GB
Average Bandwidth per Day	626.12 MB
Average Bandwidth per Hit	28.91 KB
Average Bandwidth per Visitor	2.32 MB

Graphical representation of important parameter as given in Table I, is shown from Fig. 2 to Fig. 8. The Fig. 2 shows the number of visitors per month. A visitor is normally traced by a URL or by a cookie stored on his/her computer. The results show that, September is the most popular month of a semester. During this month, 9,088 visitors visited the e-learning website. It is due to the fact that in this time period, university announces the new admission schedule and it is also the peak time of the running semester for the already enrolled students. The final examinations are held during the month of December, that is why, website has less visitors during this time period.

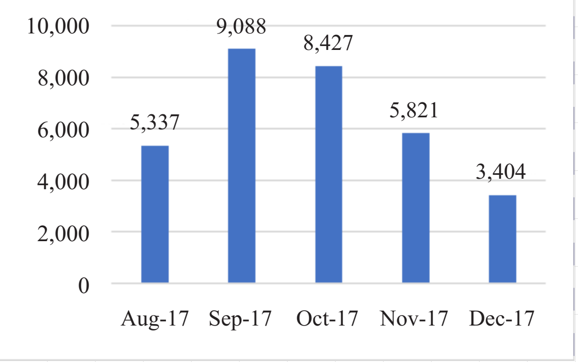

Fig. 2. Monthly E-Learning Website Visitors

The Fig. 3 shows the number of hits per month, which is obviously larger than the number of visitors per month. There are more than 30 links on the e-learning website, in which visitors are very much keen to click and get required information. Web server records each and every click in the log file, which is assumed as a separate hit on the webpage. The result shows that in the month of November the number of hits is maximum as compared to numbers of visitors shown in Fig. 2. It is due to the reason that students have to complete and submit assignments before the final examination. They upload and download the questions and answers frequently.

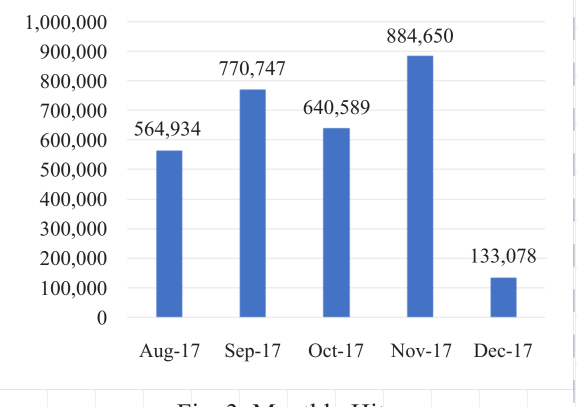

Fig. 3. Monthly Hits

The Fig. 4 shows the bandwidth consumed per month of the semester. The graph reveals that, during the month of October 29,087 MB bandwidth was used, it is showing the highest bandwidth consumption month. During this month maximum contents have been browsed and downloaded by the e-learning students.

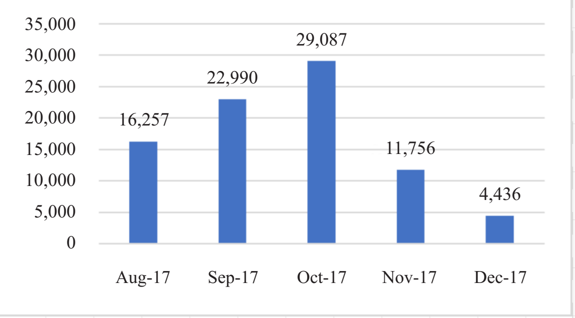

Fig. 4. Monthly Bandwidth Used During Spring Semester

The Fig. 5 shows the most active week of the semester. The graph shows that week 2 to week 10 are more active as compared to other weeks of the semester. The most active weeks are also synchronized with the number of visitors shown in Fig. 2. The Week 5 has been observed as the most active week of the semester. During this week 2,801 visitors visited the website. This is the maximum number of visitors throughout the spring semester, as shown in Fig. 5.

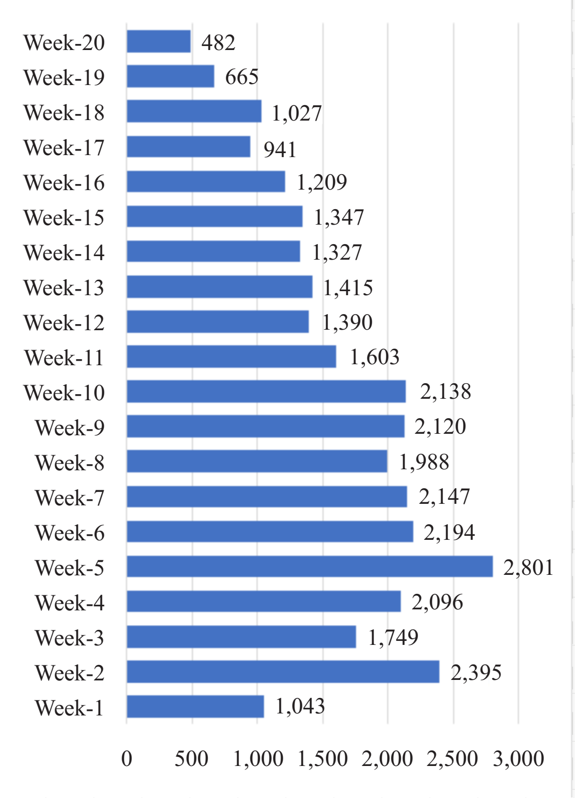

Fig. 5. Most Active Week of Visitors

The Fig. 6 shows the most active day of the week for the whole semester. The graph reveals that Monday is the most active day as compared to other days of the week. There are 4,877 visitors who accessed the website on first day of the week. It is due to the reason that after relaxing during week-end holidays, the students participate in e-learning activities with more enthusiasm and zeal.

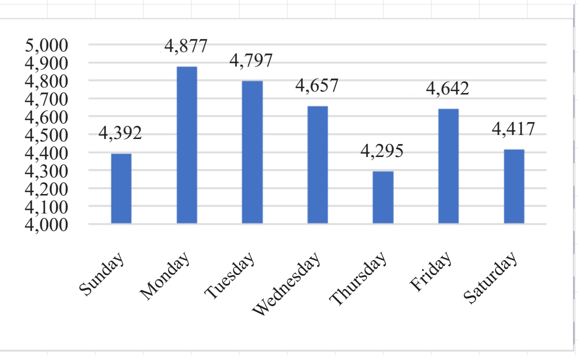

Fig. 6. Most Active Day of a Week for a Whole Semester

Sometimes, website is unresponsive while accessing it. There can be many factors but over here we will only discuss the errors generated by the website. If the request of a user is not fulfilled, an error occurs, which returns a HTTP status code in response to the request. If status code is less than 200 or greater than 299, it means that an error leads towards failure of server response. The values of error code show the type of error generated as a result of failed response. Fig. 7 shows five types of common errors that occurred during the access of website for the spring semester. The most commonly occurred error is, 404 error (35%) which shows that the link is not found. It is due to the reason that the server maintenance is performed regularly without considering the peak days and time. The second most error is "403" (18%) which is forbidden (link is not public and is password protected) [xxvi]. Besides, these errors, some other categories of errors also occurred, such as; 407, 500, and 503 and are shown in Fig. 7.

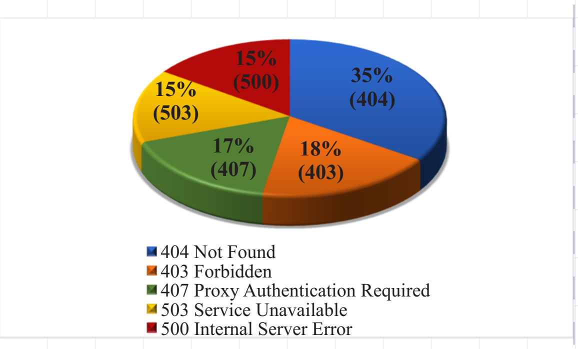

Fig. 7. Pie Chart Showing the Frequently Occurred Errors

Fig. 8 shows the frequency of users from the most active countries that are accessing the e-learning website of AIOU. Although there are more than 100 countries' from where the website is being accessed, but only the top 10 countries' data has been shown in the graph. Pakistan is the most active country as usual, however; there are users from other countries that are continuously browsing the website. This result shows that university can extend its educational programs for overseas students across the continents as well.

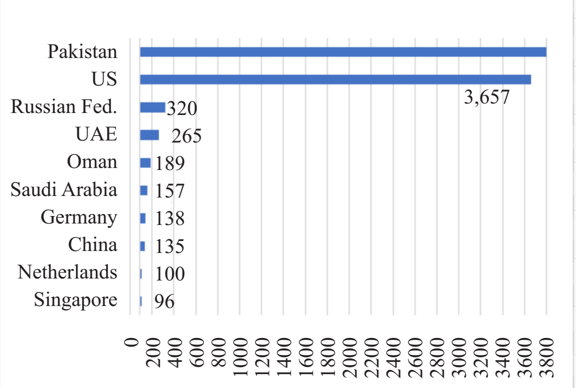

Fig. 8. Most Active Countries in respect of visitors

V. Conclusion

The conversion of distance learning into online mode has given rise to the use of Internet and world wide web. It has resulted in huge flood of information on e-learning website by structuring digital contents for teaching and support services. This information explosion on e-learning website requires the usage analysis for the potential users. Web usage mining is the area that deals with the discovery of hidden patterns in order to improve the web-based services.

This paper has presented the analysis of web usage by the distant learners of AIOU. It has described the steps of web usage analysis for the discovery of interesting patterns. The log file was collected from the university web server. An open source tool was downloaded and configured for the log file of ubuntu apache server. The pattern discovery was focused on number of visitors, number of hits, bandwidth consumed, error generation and the most active countries in respect of visitors. The results show that the e-learning initiatives at AIOU has proven successful as number of visitors and number of hits are significant. The result has highlighted some interesting information about the active months and weeks during the spring semester.

The research is beneficial for the university to improve the web services keeping in view the number of visitors and number of hits. The bandwidth consumption can be used to upgrade the system while planning the expansion of e-learning services. The errors codes can be focused to address the notable issues like non-availability of website during the peak time. The maintenance schedules can be managed during low usage time period within months, weeks and days. The usage frequency from other countries is encouraging to expand the programs for overseas students. The research may be beneficial for other universities while reviewing their websites for improving the quality and usability. The future area will focus on micro level analysis of individual program and courses from each of the four faculties of the university.

REFERENCES

[i] S. Kocdar, A. Karadeniz, A. Bozkurt, K. Buyuk, “Measuring Self-Regulation in Self-Paced Open and Distance Learning Environments”, The International Review of Research in Open and Distributed Learning, Vol.19, No.1, pp. 25-43, 2018.

[ii] P. A. Danaher, A. Umar, “Teacher education through open and distance learning”, Commonwealth of Learning (COL), pp. 1-6, 2010.

[iii] N. Tyilo, “E-Learning as Instructional Innovation in Higher Education Institutions (HEI's): Lessons Learnt from the Literature”, Journal of Communication, Vol.8, No.1, pp. 87-93, 2017.

[iv] M. U. Ahmed, N. A. Sangi, A. Mahmood, “A Learner Model for Adaptable e-Learning Interaction”, International Journal of Advanced Computer Science and Applications, Vol.8, No.6, pp. 139-147, 2017.

[v] W. Premchaiswadi, P. Porouhan, “Process modelling and decision mining in a collaborative distance learning environment”, Decision Analytics, Vol.2, No.1, pp. 1-34, 2015.

[vi] T. Srivastava, P. Desikan, V. Kumar, “Web mining–concepts, applications and research directions”, In Foundations and advances in data mining, Springer, pp. 275-307, 2005.

[vii] R. Cooley, J. Srivastava, M. Deshpande, P. N. Tan, “Web usage mining: Discovery and applications of usage patterns from web data”, AcmSigkdd Explorations Newsletter, Vol.1, No.2, pp. 12-23, 2000.

[viii] Allama Iqbal Open University, Available: http://www.aiou.edu.pk/overview.asp, Retrieved: 12-2-2018.

[ix] G. C. Oproiu, “A study about using e-learning platform (Moodle) in university teaching process”, Procedia-Social and Behavioral Sciences, Vol. 180, pp. 426-432, 2015.

[x] P. Gaur, G. B. Nagar, “Importance of Websites in Open and Distance Learning: A Study of Priorities and Expectations of Mass Communication Students from their Institutions' Websites” IRA-International Journal of Management & Social Sciences, Vol.4, No.2, pp. 439-446, 2016.

[xi] M. Perkowitz, O. Etzioni, “Adaptive web sites: Automatically synthesizing web pages”, AAAI/IAAI, pp. 727-732, July, 1998.

[xii] L. Bing, “Web data mining: exploring hyperlinks, contents, and usage data”, Springer Science & Business Media, 2007.

[xiii] A. K. Sharma, P. C. Gupta, “Enhancing the Performance of the Website through Web Log Analysis and Improvement”, International Journal of Computer Science and Technology (IJCST) Vol.3, No.4, pp. 478-485, Oct-Dec, 2012.

[xiv] O. R. Zaiane, J. Luo, “Towards evaluating learners' behavior in a web-based distance learning environment. In Advanced Learning Technologies”, Proceedings IEEE International Conference, pp. 357-360, 2001.

[xv] M. Aldekhail, “Application and Significance of Web Usage Mining in the 21st Century: A Literature Review”, International Journal of Computer Theory and Engineering, Vol.8, No.1, pp. 41-47, 2016.

[xvi] S. Khan, Y. Singh, K. Sharma, “Role of Web Usage Mining Technique for Website Structure Redesign”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 3, No.1, 2018.

[xvii] K. Dharmarajan, M. A. Dorairangaswamy, “Discovering Student E-Learning Preferred Navigation Paths Using Selection Page and Time Preference Algorithm”, International Journal of Emerging Technologies (iJET), Vol.12, No.10, pp. 202-211, 2017.

[xviii] N. Singh, A. Jain, R. S. Raw, “Comparison Analysis of Web Usage Mining Using Pattern Recognition Techniques”, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.4, pp. 137-147, July, 2013.

[xix] R. Chinnaiyan, and V. Ilango. "Analyzing the user behaviors by mining web access log files", International Journal of Advanced Studies in Computers, Science and Engineering, Vol.4, No.11, pp. 7-14, 2015

[xx] N. Kaur, H. Aggarwal, “Web Log Analysis for Identifying the Number of Visitors and their Behavior to Enhance the Accessibility and Usability of Website”, International Journal of Computer Applications, Vol. 110, No.4, pp. 25-30, January, 2015.

[xxi] A. R. Meghwal, A. K. Sharma, “Identifying System Errors through Web Server Log Files in Web Log Mining”, International Journal of Advanced Research in Computer Science and Software Engineering (IJCST), Vol.7, No.1, pp. 57-61, Jan – March, 2016.

[xxii] Sunil, M. N. Doja, “Recommender System Based on Web Usage Mining for System Based On Web Usage Mining For Restructuring of E-Learning Websites And Blogs”, International Journal of Research in Engineering & Advanced Technology (IJREAT), Vol.5, No.1, pp. 1-7, Feb - Mar, 2017.

[xxiii] C. L., Mugali, A. Maniyar, A. P. P Dandannavar, “Pre-Processing and Analysis of Web Server Logs”, International Journal of Innovative Research in Advanced Engineering (IJIRAE), Vol.2, No.8, pp. 46-55, 2014.

[xxiv] M. U. Ahmed, A. Mahmood, "Web usage mining: discovery and use of AIOU web usage patterns." International Journal of Technology Diffusion (IJTD), Vol.3, No.3, pp. 1-12, 2012.

[xxv] J. Mehra, R. S. Thakur, “An Effective method for Web Log Preprocessing and Page Access Frequency using Web Usage Mining”, International Journal of Applied Engineering Research, Vol.13, No.2, pp.1227-1232, 2018.

[xxvi] D. Pierrakos, G. Paliouras, C. Papatheodorou, C. D. Spyropoulos, “Web usage mining as a tool for personalization: A survey. User modeling and user-adapted interaction”, Vol. 13, No.4, pp. 311-372, 2003.

[xxvii] HTTP response status codes, Available: https://developer.mozilla.org/en-US/docs/Web/HTTP/Status, Retrieved: 05-2-2018