Detection of User Failure Entries from Web Server Log Files to Enhance User Experience

Dr. Ketan D. Patel[1*], Dr. Satyen M. Parikh[2]

[1]AMPICS, Ganpat University

[2]AMPICS, Ganpat University

[1]ketan.patel@ganpatuniversity.ac.in, [2]satyen.parikh@ganpatuniversity.ac.in

Abstract:

Web sites are important source of information exchange among web users and business organisations now days. Web users expect that web resources should be accessible with minimum efforts and errors. Unavailability of web resources like broken links creates negative impression to the web users and they quit which leads to overall downfall of web traffic. Hence it is important to identify such user's errors and failures during their visit to the web site. Web Server Log file helps to identify such errors and failure attempts. In this research, an error detection approach is implemented to detect such errors and failure attempts from server log file. With this approach, error entries are detected from log files which help web admins, organisations and web designers to enhance overall structure of web navigation which leads to better user experience.

Keywords: Web Server Log File, Error Logs, Broken Links, Web Usability.

1. Introduction

In today's digital era people are heavily dependent on websites to do their routine tasks online, ranging from learning to exam, product search to purchase, fund inquiry to transfer, facilities searching to booking, utility bill generation to bill payment etc. Due to these heavy dependence and trust, web users require web sites to be usable in nature. Web usability plays important role to enhance web traffic. According to ISO 9241 usability means "The extent to which a product or a service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." [1]. In terms of website, usability is defined as the ease with which novice user can perform operations on the website to fulfil their day to day needs. Web usability in general refers to an ease with which visitor use the website, Complete his/her task, gets the needed information accurately and efficiently and pleased with his/her visit [2].

To enhance web usability, web navigation should be as simple as possible as well as error free. To do so it is important to identify the failures attempts of users while they interact with the website. Such failures normally include errors like error code 404 – page not found. Such errors discourage end users and it becomes the common reason for web users to leave that site. Due to this, web sites should avoid such broken links as much as possible. To identify such broken links or paths, web usage data of users is required. Web Server Log Files are one of the essential sources to identify web usage.

In this research web server log file data is collected and then error detection approach is implemented with the use of PHP scripting language. Error entries belong to HTTP code of 400 series is detected by

this approach. It identifies entries belongs to such codes as well as it identify and display the URL of the pages associated with such errors codes.

2. Related Work

Some of the methods and approaches are implemented by researchers to identify errors from log files. Meghwal and Dr. Sharma in their study identify system errors from server log files. Authors used web log analyser tool to detect system errors from web server log files. Study also finds the general visitors statistics such as total and visitor hits, average hits per day and per visitors, failed request, total visitors etc [3]. Suneetha and Dr. Krishnamoorthi use NASA server log files to identify top errors as well as potential users of the site. Authors identify important information from server log files which includes: top errors, number of visitors, unique visitors and hit ratio [4]. Pamutha et al., in their study implemented algorithms for data cleaning as well as session identification. Study generates statistical information regarding usage data such as uniques IPs and pages, total sessions with durations, failed and robot requests etc [5].

Researchers have significantly contributed to detect error entries from server log files using different approaches and tools. The proposed research work identifies error entries as per http status code as well as it also displays the path where these errors are encountered. It helps web admin and web designer to identify potential pages of the web site where user might get such errors.

3. Methodology

Web server log files are one of the important sources to identify user behaviours of particular web site [6]. It is expected that web users get required information from the web without any hurdle and failure. Because of that reason web admins and web designers are always interested to find out the usage pattern of their users. Errors during the visit are not accepted and hence it has become crucial to detect and remove such error entries causes. This research is an attempt to identify and list system errors from server log files.

To implement the same, web server log data is collected for the educational web site for the one month time duration form 31-12-2016 to 31-01-2017. Total 54852 log data were retrieved for this duration. The log file due to its nature is not directly suitable for any processing hence it was converted into .csv file. A PHP script is implemented to retrieve data from csv file. Figure 1 describes the outline of the research work.

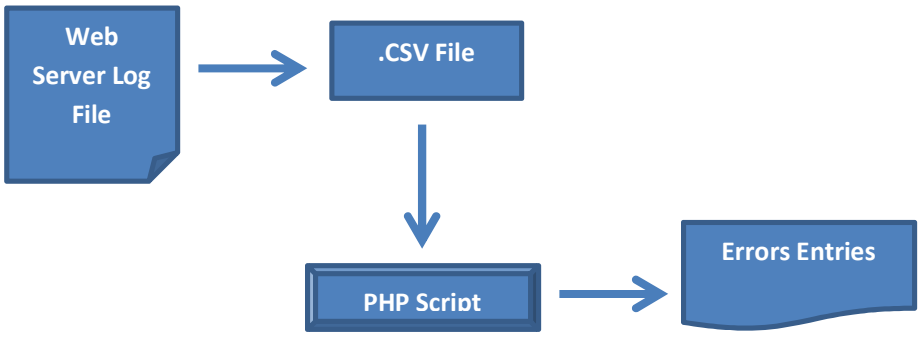

Figure 1. User failure and error entries detection process

Once the data is available in .csv format, PHP script is executed to read one by one entry from log file, and to identify error entries such as broken links as well as page not found entries from the collected

log file. The script identifies the http status code for each entry and identifies the rows which have http status code falls in 400 series. Error codes to these series belongs to the errors such as unauthorised access, forbidden, Page not found, server errors, unavailability of the service etc.

Error detection approach not only identifies such entries from the server log data but also displays the relevant path where such errors occurred. Figure 2 shows the output of the script.

Figure 2. Error entries with relevant path

Figure 2 shows individual error code like 403 and 404 along with the path of the pages requested by the user. By looking at the path, web admins would have the idea to look into the structure of the website where users are getting difficulties and where the changes are required. It also gives an idea regarding the web resources which are tried to access most frequently without proper authentication.

Script also identifies the total entries in the web log file pertaining to http error codes. Table 1 describes the error code along with their total existence in the particular time frame.

Table 1. Error Codes with Total Entries

Sr. No	Error Code	Number of entries
1	403	216
2	404	394

Table 1 shows that out of total entries, 216 entries found for error code 403 which means forbidden error. This error belongs to the access permission of the resources requested by the user. Similarly table shows that out of total entries, 394 entries found for error code 404. This code is related to error page not found. This error occurs when the requested page is not found.

Figure 3 shows the number of error entries identified in graphical forms.

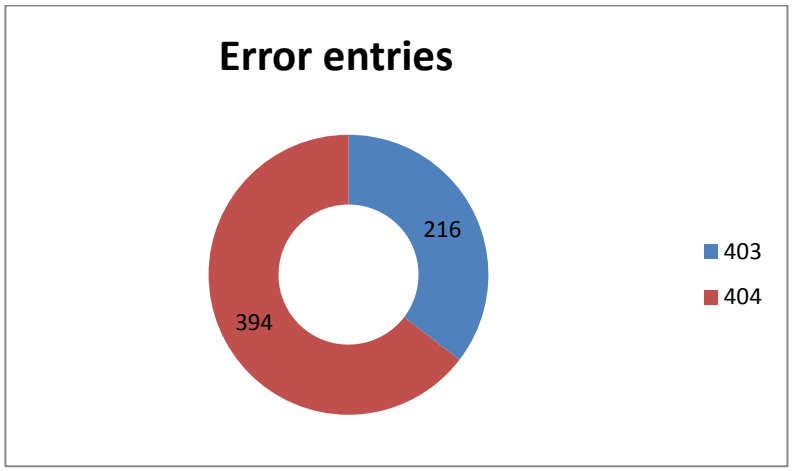

Figure 3. Total Error entries

Detection of such error entries helps web admin to identify what's going wrong when user visited the website and what improvements are required to make the visit pleasant one. Such improvements makes website to be more usable and it overall enhance the usability of the site.

4. Conclusion:

Website helps peoples to fulfil their daily needs in few clicks as well as provide a platform for information exchange. These demands that website should be usable, error free and easy to use. Errors like page not found should be avoided to achieve this. In this research such errors are detected from server log files. Study identifies errors as well their frequency of occur along with path of requested pages. This helps web admins to restructure the web navigation in such a way that such errors should be avoided. In future accurate prediction model can be developed from such findings which offer other options of similar pages to web user when user gets page not found or other such errors.

5. References:

1. ISO 9241-11:1998(en), [Online] Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en.
2. J. Nielsen, "Return on Investment (ROI) for usability", (2003).
3. A. Meghwal and Dr. A. Sharma, "Identifying System Errors through Web Server Log Files in Web Log Mining", IJCST, Vol. 7, Issue 1, (2016).
4. K. R. Suneetha and Dr. R. Krishnamoorthi, "Identifying User Behaviour by Analyzing Web Server Access Log File", International Journal of Computer Science and Network Security (IJCSNS), Vol. 9, No. 4, (2009).
5. T. Pamutha, S. Chimphlee, C. Kimpon and P. Sanguansat, "Data Preprocessing on Web Server Log Files for Mining Users Access Patterns", International Journal of Research and Reviews in Wireless Communications (IJRRWC), Vol. 2, No. 2, (2012).
6. T. Hussain, Dr. S. Asghar, Dr. N. Masood, "Web Usage Mining: A Survey on Preprocessing of Web Log File", International Conference on Information and Emerging Technologies, IEEE, (2010).