

# Session Based Web Usage Pattern Extraction through Web Server Log File

Dr. Ketan D. Patel<sup>1</sup>, Dr. Satyen M. Parikh<sup>2</sup>

<sup>1</sup>AMPICS, Ganpat University

<sup>2</sup>AMPICS, Ganpat University

(E-mail: ketan.patel@ganpatuniversity.ac.in)

**Abstract**— Web is considered as a powerful tool for information exchange between clients and business organizations. Hence it is important to design web interface so that end users can get the required details with minimum effort. For that it is crucial to identify web usage patterns. Such usage patterns helps web admins to identify the page visit patterns of their end users during their visit. Extracting the usage pattern allows web admins and designers to identify the bunch of pages visited by end users during their each visit. Web Server Log File is used in this research to extract session wise web usage patterns of visitors. Web usage pattern extraction approach used in this research, identifies and display series of web pages visited by end user during each of his/her visit.

**Keywords**—web usage pattern, web server log file, web usability, user session.

## I. INTRODUCTION

The increasing use of web needs to address the concern related to its usage. It is always in demand to analyze the usage patterns of different visitors while they interact with the website. It helps to identify the likes and dislikes of users, popularity of contents between users, popular path of user visits etc. Due to these benefits, visitors visit path and patterns should be identified and analyzed.

Web users requires easy to use and simple website. The web site should be usable in nature. Web usability simply refers to an ease with which user can use the website. It has five basic components such as Effectiveness – related to number of errors or mistakes done by visitors while they interact with the website, Efficiency – associated with the ability of user to perform the task quickly, Learnability – related to users ability to learn the design during their first visit, Memorability – associated with the easy recalling the functionality of the website while user returns to the website after long period of time, Satisfaction - related to the pleasure of user while they interact with the website to fulfill their purpose [1]. A website designed with above mentioned components helps user to complete their task in the website easily and effectively. Many factors that affect the usability of the website are Ease of use and Navigation, Site Contents, Identity of the site, Delay in downloading the contents, Made for the Medium, Responsiveness, Trust, and Site Promotion

& Emotions [2]. These components play significant role in overall usability of the site. To make website usable, its navigation structure should be easy to use and it should be as simple as possible. Navigation burden from the users should be minimized. It is good to have web pages accessible within three clicks [3]. Due to these user expectations, it is important to find out the way users are using the website, the pages they accessed frequently, the page visit pattern and sequences etc.

Web Server Log files are one of the authentic and good source to identify usage patterns among other log files [4]. Web user actions are stored in web server log files while they interact with the website. Server log file contains the user visit information like IP address of visitor, date and time of visit, pages visited, source page form where user arrived to the current page, number of bytes transferred during the visit, browser used for visit etc. These details help web admin to get valuable information about users visit. In this research work, web server log data of computer science department's website is used. The web server log data were collected, cleaned and then user visit patterns are extracted from it according to user session.

## II. METHODOLOGY

To extract session based web usage patterns, web server log file of computer science department is used. The log file was taken for one month time duration. Initially there were total 54852 log entries fetched. Server log files contain noisy and irrelevant data in large number of volume so it must be preprocessed before to apply any analysis technique on it. Preprocessing helps to remove irrelevant data from web server log file which normally exists in the form of web crawler or robot visit data, errors and failure entries, incomplete and missing entries and graphics as well as supporting file [5][6]. It is also important to identify user sessions after applying data cleaning. Preprocessing and session identification was done on the collected data using the preprocessing as well as session identification algorithms and approaches [7][8]. After applying preprocessing algorithms, 6065 entries were left for further processing. Table 1 shows the outcomes of preprocessing as well as session identification approaches applied on the collected log data.

visit of user. The output of this approach is displayed in following Fig. 2.

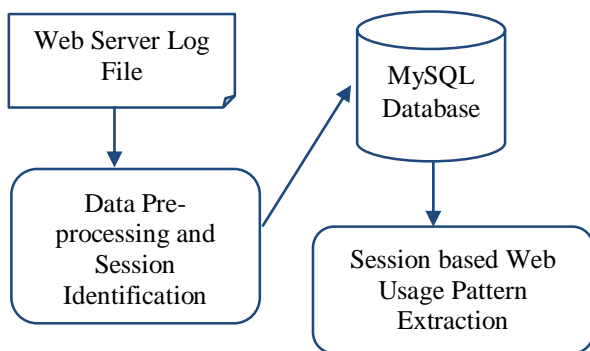
Original Log Entries	Entries Available after Preprocessing	No. of User Sessions Identified
54852	6065	3876

**Table 1: Preprocessing and Session Identification**

Table 1 show that before preprocessing, initially there were 54852 log entries were available. After applying preprocessing techniques the data reduces to 6065 entries. Along with that 3876 user sessions were identified using session identification approach.

Once the data is cleaned and sessions are identified, it is stored in database called MySQL for further processing. On MySQL database, session based web usage pattern extraction approach is applied.

The entire session based usage pattern extraction approach is explained using Fig. 1.

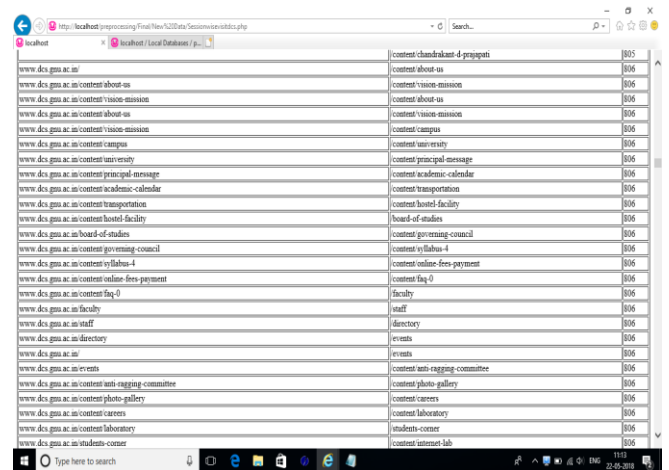


**Figure 1: Session based usage pattern extraction approach**

First it collects the web server log data then data preprocessing techniques are applied to clean irrelevant data. After that to identify sessions, session identification approach is applied. The cleaned data is stored in MySQL data base on which finally session based web usage pattern extraction approach is applied.

**III. RESULT AND DISCUSSION**

The above mentioned approach is implemented using PHP scripting language. Script reads each entry from the MySQL database and extracts and generate the page visit path as an output as per user sessions. It combines and displays the page visit entries which have assigned same session id. The pages visited by same user in 30 minutes time interval are assigned same session id as they are considered as a single



**Figure 2: Session wise page visit patterns.**

Fig. 2 shows the session wise pages visited by end users. In Fig. 2, page visit path of single session is displayed. It shows that user with session number 806 have visited several pages in his or her single visit. The page visit pattern for that particular session is like about-us – vision-mission – campus – university –principal message – academic calendar and so on. Means user associated with session 806 has visited above mentioned pages in that particular fashion. Proposed approach displays page visit path of each session in similar fashion. By analyzing path for each session, web admin have an idea regarding the usage of the website. Findings help web admins to discover the actual usage of the website. It further helps web admin or designer to modify the navigation structure if required. Modification is required if majority of users follows the path which requires more effort to reach to the desire content of the website.

**IV. CONCLUSION AND FUTURE WORK**

To retain user traffic on the website is always a challenging task for any organization. Web users demand website to be easy to use and usable hence it is important to identify web usage patterns. In this research session wise usage patterns of end users are discovered using web server log files. In this research, Session based usage pattern extraction approach is implemented that identify the different pages visited by end user during their individual visit. Such findings help web admins and designers to discover the navigation path of their end users when they interact with the website. They can redesign the web navigation so that users can have more ease to use the website.

In future frequency of such page visits patterns should be identified and based on that, popular paths of users can be extracted and compared with actual ideal page visit path to find any deviation in page visit. Web navigation should be designed accordingly.

## REFERENCES

- [1] J. Nielsen, "Usability 101: Introduction to Usability", 2012.
- [2] C. E. Downing and Chang Liu, "Assessing Web Site Usability in Retail Electronic Commerce", 35th Annual Computer Software and Applications Conference, IEEE 2011.
- [3] R. Ahmad, Li Zhang and F. Azam, "Measuring 'Navigational Burden'", Proceedings of the fourth International Conference on Software Engineering Research, Management and Applications, IEEE, 2006.
- [4] T. Hussain, Dr. S. Asghar and Dr. N. Masood, "Web Usage Mining: A Survey on Preprocessing of Web Log File", International Conference on Information and Emerging Technologies, IEEE, 2010.
- [5] G. Neelima, and Dr. S. Rodda, "Predicting user behavior through sessions using the web log mining", International Conference on Advances in Human Machine Interaction, IEEE 2016.
- [6] K. R. Suneetha, and Dr. R. Krishnamoorthi, "Identifying User Behavior by Analyzing Web Server Access Log File", IJCSNS International Journal of Computer Science and Network Security, volume 9, No. 4, April 2009.
- [7] K. D. Patel and S. M. Parikh, "Preprocessing on Web Server Log Data for Web Usage Pattern Discovery", International Journal of Computer Applications, volume 165(10), 29-32, May 2017.
- [8] K. D. Patel and S. M. Parikh, "Detection of User Failure Entries from Web Server Log Files to Enhance User Experience", International Journal of Management, Technology and Engineering, volume 8, issue V, May 2018.