

# Web Logs – A Roadmap to Online Consumer

Sonia Sharma<sup>\*1</sup>, Dalip<sup>2</sup>

<sup>\*1</sup>Hindu Girls College, Jagadhri, Haryana, India

<sup>2</sup>Maharishi Markandeshwar Deemed to be University, Mullana, Ambala, Haryana, India

## ABSTRACT

In this web era internet users are increasing day by day and in this focused condition generally number of client get to the web to get data of their advantage and now days for a site proprietor, there is prerequisite to know the perusing conduct of online consumers. By knowing the interest of consumer, a site proprietor can change the structure of the website and can attain the objective of the organization by increasing their productivity. The need of particular client is found out through the examination of the client's surfing conduct on web. By utilizing web mining methods expectation of shopper conduct is effortlessly unsurprising. In web usage mining process weblogs plays a very important role, without weblog, prediction of behavior of online consumer is not possible. Weblogs act as an important key to reach up to the online consumer very efficiently. Therefore, this paper portray the substance of web log, log configuration and procedure of web use mining.

**Keywords :** Web logs, Web Usage Mining, Consumer, Prediction

## I. INTRODUCTION

Now a day's everybody is evaluating the web to get lots of data of their advantage and in the present focused condition, makers/specialist co-ops are quick to know whether they give the best administration/item to clients or whether clients anticipate get their administration or purchase item. Specialist co-ops should need to know how to make their sites or web application intriguing to clients and how to enhance promoting systems to pull in them. The advancement of Internet in late decades made E-trade sites to bring huge records from clients. Web usage mining(WUM) which is also known as Web Log Mining, is the best utilization of Mining of data strategy connected on substantial volume of information to separate helpful and intriguing examples from Web information, particularly from web logs, so as to enhance online applications[9][10][11][12]. The site

can be modified according to the intrigue and need of client. Log documents contain a rundown of activities that happened at whatever point client gets to the specialist co-op's site or web application. Each hit to the Website will be signed in a log document [9]. These log records are put away in web servers. The crude web log record is one line of content for each hit to the site and contains data about who went by the web page, where they originated from, and what they did on the site. These log records convey a helpful data for specialist organizations so breaking down these log documents can give them experiences about site movement designs, client action, client intrigue and so forth.

## II. RELATED WORK

Mobasher [1] recognized personalization and customization frameworks. The creator gave brief

depictions of regular methodologies of personalization of web which incorporates synergistic, content-based and lead based sifting frameworks. The Author outlined the process of web personalization with different periods of an information mining cycle, for example, data preprocessing, design. The different stages in web usage mining such as data preprocessing, pattern discovery and pattern analysis, in detail is described by Jaideep Srivastava et. al [2]. Authors additionally depicted the measurements and submissions territories where web usage can be utilized. L.K. Joshila Grace et.al. [4] in a very effective way highlight the usage of web logs in web usage mining process for web personalization Lavanya KS et.al [5] highlight the use of weblog in Hadoop environment and discussed in detail the how web logs can be used for analyzing behavior of user.

### III. WEB LOGS AND ITS CONTENTS

[3] When a web client communicates with the web and presents a demand, at that point his/her navigational data called as web get to log and it is put away in a web log document. Web servers, intermediary servers, and customer programs are the three distinct establishments of web log document. Log records are archives that rundown the exercises that have been occurred. These log records abide in the web server. PCs that pass on the site pages are called as web servers. The Web server stores most of the records essential to demonstrate the Webpages on the customers PC. All the individual site pages consolidate to shape the zenith of a Web website page. Pictures/practical records and any substance that impact dynamic parts of the site to work. , The program requests the data from the Web server, and using HTTP, the server passes on the data back to the program that had requested the website page. The program hence changes over, or gatherings, the records into a customer noticeable page. This gets appeared in the program.

Correspondingly the server can send the records to various client PCs meanwhile, empowering different clients to see a comparative page at the same time.

#### Contents of a Log File

The Log documents in various web servers keep up various sorts of data. The fundamental data show in the log record are

**User name:** This recognizes who had gone to the site. The recognizable proof of the client for the most part would be the IP address that is doled out by the Internet Service supplier (ISP). This might be a transitory address that has been relegated. Thusly here the extraordinary distinguishing proof of the client is slacking. In some sites the client distinguishing proof is made by getting the client profile and enables them to get to the site by utilizing a client name and watchword. In this sort of access the client is being distinguished interestingly so that the return to of the client can likewise be recognized.

• **Visiting Path:** The way taken by the client while going to the site. This might be by utilizing the URL straightforwardly or by tapping on a connection or through a webindex.

• **Path Traversed:** This recognizes the way brought by the client within the site utilizing the different connections.

• **Time stamp:** The time spent by the client in each site page while surfing through the site. This is recognized as the session.

• **Page last visited:** The page that was gone to by the client before he or she leaves the site.

• **Success rate:** The achievement rate of the site can be controlled by the quantity of downloads made and the number replicating action under passed by

the client. In the event that any buy of things or programming made, this would like wise include the achievement rate.

•**User Agent:** This is only the program from where the client sends the demand to the web server. It's only a string depicting the sort and form of program programming beingutilized.

•**URL:** The asset got to by the client. It might be a HTML page, a CGI program, or acontent.

•**Request type:** The strategy utilized for data exchangeisnoted.ThestrategieslikeGET,POST.

These are the substance exhibit in the log document. This log record points of interest are utilized as a part of instance of web use mining process. As per web use mining it mines the exceptionally used site. The use would be the much of the time went to site or the site being used for longer time length. In this manner the quantitative use of the site can be investigated if the log record is brokedown.

#### IV. TYPE OF WEB LOGS AND LOG FORMAT

Traditionally, four types of logs are available in web server: transfer log, agent log, error log and referrer log. First two are mandatory whereasthe remaining two areoptional.

##### Error Log

```
[04-Jul-2017 10:53:27 Asia/Kolkata] PHP
Warning: Cannot modify header information -
headers already sent by (output started at
/home/swsplindia/public_html/phptextClass.
php:6) in
/home/swsplindia/public_html/phptextClass.ph
pon line 66
```

##### Access Log

The server get to log records all demands that are prepared by the server. The area and substance of the entrance log are controlled by the Custom Log mandate. The Custom Log order is utilized to log solicitations to the server. A log arrange is indicated, and the logging can alternatively be made restrictive on ask for attributes utilizing condition factors. The Log Format order can be utilized to disentangle the choice of the substance of the logs. This area portrays how to arrange the server to record data in the entrance log. Here are three log designs considered for get to logsections on account of Apache HTTP Server [6]

##### Common Log Format

The configuration of the common log format is given below

Log Format

```
"%h%l %u %t \"%r\"%>s %commonCustom Log
logs /access logcommon
```

FOR EXAMPLE

```
207.46.13.57 -- [31/Jul/2017:06:54:21
+0000] "GET /image/logo.png HTTP/1.1"
200 7563"
```

207.46.13.57 (%h)-This is the IP address of the customer which made the demand to the server.

- (%l) - The hyphen present in the log file entry next to the IP address indicates that the requested information is not available.

Frank (%u) - The user id of the person requesting the document as determined by HTTP [6] authentication

- [31/Jul/2017:06:54:21 +0000] (%t) -The Time format resembles like [day/month/year: hour: minute: second zone]

“GET /image/logo.png HTTP/1.1 (“\%r\”) is the request sent from the client is given in double quotes.

GET is the method used. Logo.png is the information requested by the client. The protocol used by the client is given as HTTP/1.0

200 (%>s) - This is the status code sent by the server. The codes beginning with 2 for successful response, 3 for redirection, 4 for error caused by the client, 5 for error in the server.

7563(%b) - The last entry indicates the size of the object returned to the client by the server, not including the response headers. If there is no content returned to the client, this value will be “-”

### Combined Log Format

The configuration of the combined log format is given below

#### Log Format

```
"%h %l %u %t |"%r|"%">s %b |"%{ Referrer} i|"
|"%{ User-agent} i|" " combined Custom Log
log/acces_log combined
```

The log file entries produced in combined log format will look something like this [8]

This gives the site that the client reports having been referred from.

## V. USAGE OF WEB LOGS IN WEB USAGE MINING PROCESS

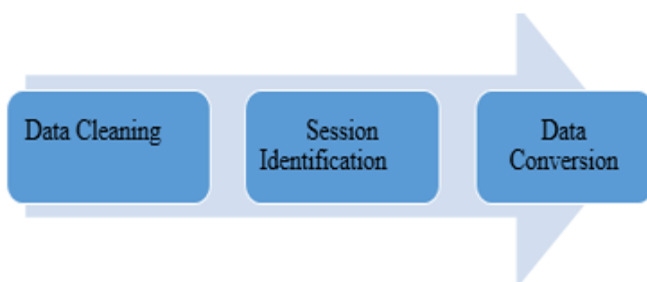


Figure-I: Data Preprocessing Steps

After getting weblogs of website of an e-business organization data preprocessing and pattern

discovery steps have to perform. Figure-I shows the steps of data preprocessing which is the main step under web usage mining process and the preprocessing of web logs provides hidden detail of consumer via which behaviour of consumer can be predicted very easily. after getting web logs, Next step is to preprocess the data and the first step in preprocessing is cleaning of data [7].after this requires session are identified and then data is converted into the required format for mining. The unwanted data are removed and a minimized log file is obtained.

**Data cleaning:** In this process the entries made in the log file for the unwanted view of images, graphics, Multimedia etc., made by the users are removed. Once these data are removed, the size of the file is minimized to a greater extent.

**Session Identification:** This done by using the time stamp details of the web pages. The total time used by each user of each web page. This can also be done by noting down the user id those who have visited the web page and had traversed through the link of the webpage. Session is the time duration spent in the web page.

**Data conversion:** In this log file after cleaning and identification process are converted into the form which is required for mining algorithm.

After the conversion of the data in the log file into a formatted data the pattern discovery process is under gone. With the existing data of the log files many useful patterns are discovered either with user id's, session details, time outs etc., For pattern discovery , different techniques such as Path Analysis, Association Rules, Sequential Pattern, clustering, classification can be used.

## VI. FORMATION OF AN PROTRACTED LOGFILE

The log document substance would be significantly progressively effective on the off chance that it gives the subtleties of the snaps made by the client while visiting the site. On the off chance that the client

opens a specific site and does some other work outside the framework then it might likewise be considered as the utilization of the site. The insights about the snaps made by the client and the time the person in question looked over or did some other task can likewise be noted for successful mining of the web use information. We will think of one as more circumstance were the client snaps to open a site and furthermore works with some other site in an alternate program window. In this circumstance we can examine that the client can just peruse subtleties of just a single site at any given moment. At that point it is comprehended that the other site is said to be perfect. Be that as it may, even in this circumstance the subtleties in the log document would take note of the contribution as the website page is being utilized. By taking these little contrasts in the time or the session of the site page being utilized, still a proficient web mining should be possible.

**VII. EXPERIMENTS AND RESULTS**

To analyze the weblogs of web site data is collected for the period of 31 July 2017 to 25 December 2017 from the website [www.viralsach.xyz](http://www.viralsach.xyz) .Figure-II shows the sample file of web logs. After collecting weblogs, for analyzing the weblogs of site weblog explorer 9.2 is used and only human entries are considered for the work. Figure –III shows the operating system used by the consumer for visiting the website. Figure-IV presents unique IP. It shows that 50974 hits are obtained by 2643 visitor. In this way an owner of the website can analyze the weblogs of consumers visiting their website and can easily predict the behaviour of consumer.

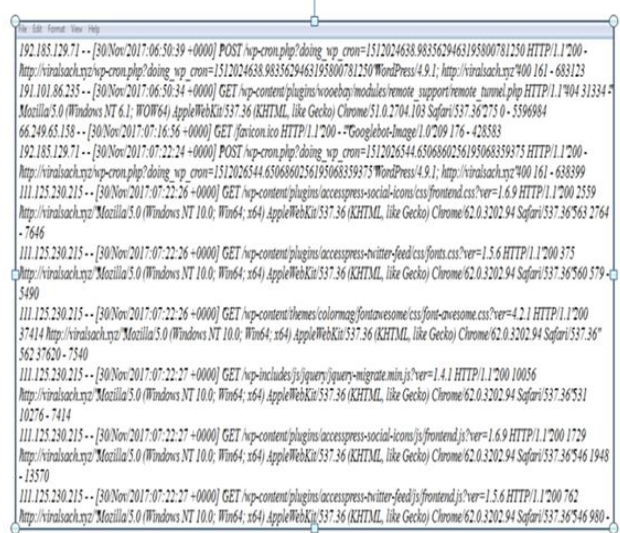


Figure-II: Sample Web Log

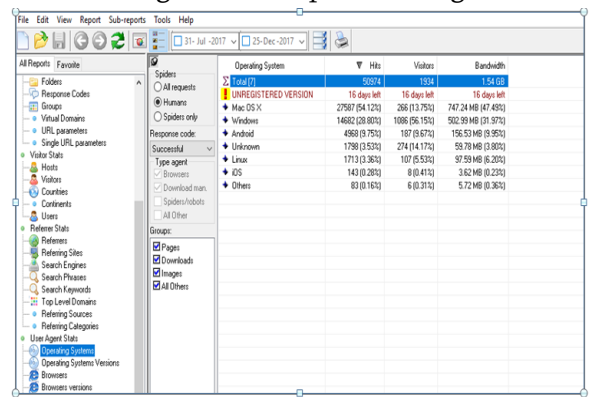


Figure-III: Operating System used by visitors

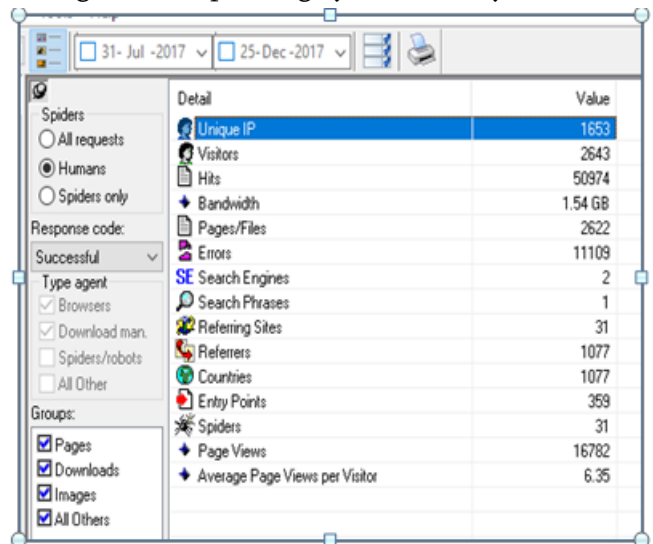


Figure-IV: Summary of Web logs

## VIII. CONCLUSION

The Paper gives a detailed look about the web log file, its contents, its types, its location etc. By understanding the weblog detail and web usage mining process any web site owner can predict the browsing behaviors of customer very easily and can make a **website** in a very effective way and can maintain a good relationship with their customer. Not only an owner can get the detailed report of user agent, the number of hits on the specific page, page view, time spent on a particular page but also the location of customer, month wise report, operating system ,browser used by a customer, bandwidth etc.

## IX. REFERENCES

- [1]. Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, SIGKDD Explorations, ACM SIGKDD, Vol. 1, Issue 2, Jan 2000pp.12-23.
- [2]. Bamshad Mobasher, "Data Mining for Web Personalization" The Adaptive Web, LNCS 4321, Springer-Verlag Berlin Heidelberg, 2007, pp.90–135.
- [3]. Krishana kant P.Adhiya,Satish R.Kolhe, "An Efficient And Novel Approach for Web Search personalization using Web Usage Mining", Journal of Theoretical and Applied Information Technology , Vol.73 ,2015 ISSN: 1992-8645.
- [4]. L.K. Joshila Grace1, V.Maheswari, Dhinaharan Nagamalai, "Analysis of Web Logs and Web User in Web Mining ",International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January2011.
- [5]. Lavanya KS, Srinivasa R, "Customer behavior analysis of web server logs using Hive in Hadoop Framework", International Journal of Advanced Networking & Applications (IJANA) ISSN: 0975-0282.
- [6]. Apache HTTP Server Version 1.3, <http://httpd.apache.org/docs/1.3/logs.html>
- [7]. Sonia Sharma, Munishwar Rai, "Web Mining: Roadmap to Customer", Proceedings of the International Conference INDIACom-2017; IEEE Conference ID: 40353, 2017 ISSN 0973-7529; ISBN 978-93-80544-24-3
- [8]. [www.viralsach.xyz](http://www.viralsach.xyz).
- [9]. Sonia Sharma, Munishwar sRai, "Customer Behaviour Analysis using Web Usage Mining," International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.6, pp.47-50, 2017.
- [10]. Jose Roberto de Freitas Boullosa. Architecture for Web Usage Mining"
- [11]. Arvind Kumar Sharma, P.C. Gupta, - "Exploration of efficient methodologies for the improvement in web mining techniques-A survey", International Journal of Research in IT & Management (ISSN 2231-4334) Vol.1, Issue 3, July 2011.
- [12]. Ashirr K Kashyap, IflahNaseem, Dheeraj Mandloi, "Web Mining an Approach to Evaluate the Web," International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.79-85, 2017.

### Cite this article as :

Sonia Sharma, Dalip, "Web Logs - A Roadmap to Online Consumer", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 1, pp. 576-581, January-February 2019.  
Journal URL : <https://ijsrset.com/IJSRSET192112>