Predicting user behavior through Sessions using the Web log mining for an E commerce application

Sushmeendra N Rao[1], Rakesh B[2], Pallavi N Hegde[3], Anusha R kotur[4]
8th Semester, Department of Information Science and Engineering,
Gm institute of technology , Davangere

***Abstract*— It is the method to extract the user sessions from the given log files. Initially, each user is identified according to his/her IP address specified in the log file and corresponding user sessions are extracted. Two types of logs ie., server-side logs and client-side logs are commonly used for web usage and usability analysis. Server-side logs can be automatically generated by web servers, with each entry corresponding to a user request. Client-side logs can capture accurate, comprehensive usage data for usability analysis. Usability is defined as the satisfaction, efficiency and effectiveness with which specific users can complete specific tasks in a particular environment. This process includes 3 stages, namely Data cleaning, User identification, Session identification. In this paper, we are implementing these three phases. Depending upon the frequency of users visiting each page mining is performed. By finding the session of the user we can analyze the user behavior by the time spend on a particular page.**

Keywords*— Web log mining ,User identification , Session identification.

1. Introduction

The WWW as a largest information constructs has had much progress since its advent. As the WWW has become customary today, highlight content generated by users, interoperability and usability. A web may allow users to interact and collaborate with each other in a social media dialogue as creators of user-generated context in a virtual community. So, World Wide Web becomes more popular and user friendly for transferring information[7] [13]. Therefore, people are more interested in analyzing log files which can offer more useful insight into web site usage. Data mining is the extraction of knowledge from the huge amount of data sets, to find a relationship and patterns in data that have been not previously been discovered to summarize the data in original ways to make it understand and useful to the users. Web mining is one of the technique of data mining to extract useful information based on users' needs, under web mining, web usage mining is one of the application of data mining technology to extract information from weblog to analyze the user access to websites by [2] [14]. Web mining is the use of data mining technique to automatically discover and extract information from web documents and services. There are 3 general classes of information that can be discovered by web mining [4] :

- Web activity: From server logs and web browser activity tracking.
- Web graph: from links between pages, people and other data.
- Web content: for the data found a web page and inside of documents.

Major applications of web usage mining:

Web personalization: Web server logs are used to cluster web users having similar interests. It is also defined as adapting services and information which was available on a website to the needs and the expectations of a target user, the *active user*; the personalization task by [13] might benefit from the knowledge gained from an analysis of the user's navigational behavior combined with other features which are peculiar to a Web context, namely its structure and content.

System Improvement: load balancing, Web caching, network transmission or data distribution is the common application areas of web mining for improving the system performance [8].

Site reorganization: The link structure and content structure of any website are two significant factors for any web site. The recent development in web mining technologies goes towards shorter navigation sequences, for that purpose the ease to access target page in any web domain needs to be increased. The reorganization task can be performed with respect to the frequent patterns extracted. Web usage data also give information about the design of any web site with respect to users' behaviors [5] [8]. The website owner can redesign these pages and observe the behavior of users on these pages.

E-Commerce/Business intelligence: The Web usage mining usage allows different organizations to understand its customers and built customer profiles on the basis of customer's habits, based on interest and needs the companies can increase their profit by "cross selling" or selling items correlated to their demands. Hence, knowledge about the customers' preferences and needs make the CRM more effective. The main goals of companies [2] are retaining their old customer and attract new customers to beat their competitor's.

Usage Categorization: In this process the information stored in Web server logs is processed by applying various data mining techniques so as to (a) extracting statistical information and discovering interesting usage patterns, (b) according to the navigational behavior [1] the users are being clustered and (c) determine possible links between Web pages and user groups. Others data mining techniques are also used for finding useful patterns.

2. Web Log Files

Web log files are the files which contain complete information about the users browse activities on the web server by [2] 11] [4]. These web log files are created automatically by every user click to the corresponding web servers. These log files is in text format, most of the times and the size varies from 1KB to 100 MB.

2.1 Types of log files

There are three types of log files which are as follows:

- Web Server Logs
- Proxy Server Logs
- Browser Logs

2.1.1 Web Server Logs

History of web page requests is maintained as a log file. Web servers are the costly and the most common data source. They collect large volume of information in their log files. These logs contain name, IP, date, and time of the request, the request line exactly came from the client, etc. These data can be bound together as a single text file, or divided into different logs, like access log, referrer log, or error log. However, user specific information is not stored in the server logs [15].

2.1.2. Proxy Server Logs

It acts as an intervening level of catching lies between client browser and web servers. Proxy caching is used to decrease the loading time of a web page as well as the reduce network traffic at the server and client side. The actual HTTP request from multiple clients to multiple web servers are tracked by the proxy server [9]. The proxy server log is used as a data source for browsing behavior characterization of a group of unauthorized users sharing a common proxy server.

2.1.3. Browser Logs

On client side using JavaScript or Java applets the browsing history is collected. To implement client side data collection, user cooperation is needed. Here pre-processing discussed using Web Server Logs [11]. Web server logs are used in the web page recommendation to improve the E-Commerce usability.

2.2 Types of log files format

Web log file is a simple plain text file which record information about each user. Display of log file data in three different formats by [11] [14] [15].

- W3C Extended log file format
- NCSA common log file format
- IIS log file format

NCSA and IIS log file format the data logged for each request is fixed. W3C format allows user to choose properties, the user wants to log for each request.

W3C Extended log file format

The W3C log format is the default log file format on the IIS server. The field is separated by space, time is recorded as GMT (Greenwich Mean Time). It can be customized, that is administrators can add or remove fields depending on what information want to record. In the W3C format of year is YYYY-MM-DD [2]. Omitting unwanted attribute fields when log file size is limited.

An example is shown below.

#software - version of IIS that is running

#version - the log file format

#Date- recording date and time of first log entry.

#fields: date time c-ip cs-username s-ip cs-method cs-uri-stem cs-uri-query sc-status sc bytes cs-bytes time-taken cs-version cs(User-Agent) cs(Cookie) cs(Referrer)

NCSA common log file format

The NCSA Common log file format is a fixed ASCII text-based format, so you cannot customize it. The NCSA Common log file format is available for Web sites and for SMTP and NNTP services, but it is not available for FTP sites. Because HTTP.sys handles the NCSA Common log file format, this format records HTTP.sys kernel-mode cache hits.

An example is shown below.

IIS log file format

The IIS log file format is a fixed ASCII text-based format, so you cannot customize it. Because HTTP.sys handles the IIS log file format, this format records HTTP.sys kernel-mode cache hits.

An example is shown below.

There are some important technical issues that must be taken into consideration because it is necessary for Web log data to be prepared and preprocessed in order to use them in the consequent phases of the process by [9] [11].

3. METHODOLOGY

In the data preprocessing, it takes web log data as input and then process the web log data and gives the reliable data. The goal of preprocessing is to choose primary features, then remove unwanted information and finally transform raw data into sessions. To achieve its goal Data preprocessing is divided into Data Cleaning, user identification, and Session Identification [12] [2].

3.1. Data cleaning

The use of data cleaning procedure is to remove all the unwanted data used in data analysis and mining. To increase the mining efficiency data cleaning is very important. The cleaned data include removal of local and global noise, elimination of videos, graphic records and the format efficiency, elimination of HTTP status code records, robots cleaning.

3.1.1. The Records of- graphics, video and the format information:

In every record of URI field,JPEG, GIF, CSS filename extension is found, these extensions are going to be eliminated from the web log file. The files with these extensions are the documents embedded in the web page. So it is not necessary to include these files in identifying the user interested web pages. This process support to identify user interested sessions.

3.1.2. Failed HTTP- status code:

This cleaning process will reduce the evaluation time for finding the user's interested sessions. In this process the status field of every record in the web access log is checked and the status code over 299 or below 200 are removed.

3.1.3. Robots- Cleaning:

It is also called as spider or not, it is a software tool that scans a website periodically to extract the content. All the hyperlinks from a web page are automatically followed by WR. The uninterested session from the log file is removed automatically when WR is removed.

Algorithm 1: Data cleaning

```
Input: log_table
Output: refine_log_table
Begin
   Read records in log_table
       For each record in log_table
           Read fields (Status code)
             If Status code=200, Then Get all fields.
               Ifsuffix.URL_Link={*.gif,*.jpg,*.css,*.ico} then,
                   Remove suffix.URL_link
                   Save fields in new table.
               End if
               Else
                   Next record
             End if
End
```

3.2. User identification

Each different user accessing the website is identified in the user identification process. The aim of this process is to retrieve every user’s access characteristics, then make user clustering and provide recommendation service for the users. Different users are identified by different ip addresses.

Algorithm 2:User identification

```
Input: refine_log_table
Output: identification of user
Begin
    Read records in log_table
    for each record in dataset do
        If current IP is not in ListOfIP then add the current IP in
        ListOfIP mark whole record as a new user and assign
        userID
        else assign the old userID.
        End else
        End if
End
```

3.3. Session Identification

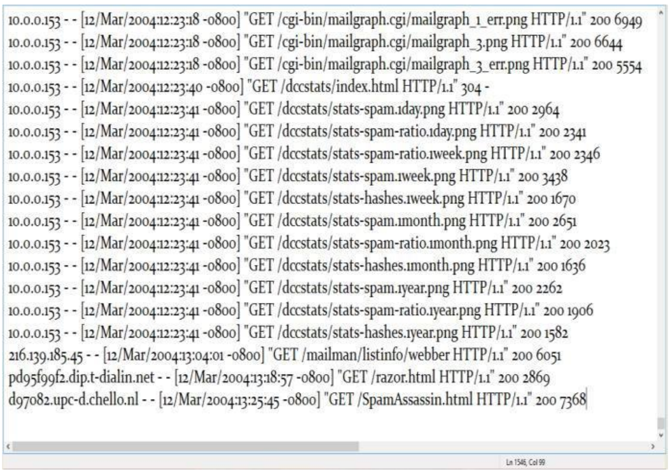
A sequence of pages viewed by a user during one visit is known as the Session. The session is recorded in the log file. In pre-processing it is necessary to find session of each user. It defines the number of times the user has accessed a web page. It takes all the page reference of a given user in a log and divides them into user sessions. These sessions can be used as an input data vector in classification, clustering, prediction and other tasks. Based on a uniform fixed timeout a traditional session identification algorithm is used. A new session is identified when the interval between two sequential requests exceeds the one hour.

Algorithm 3: Session identification

```
Input: user identified table
Output: identified sessions
Begin
    Read records in log_table
        for each record in dataset do
            if time_required > one hour assign new sessionID for that log entry
                increment sesssionID
            else
                assign the old sessionID.
            End else
            End if
End
```

4. *EXPERIMENTAL RESULTS*

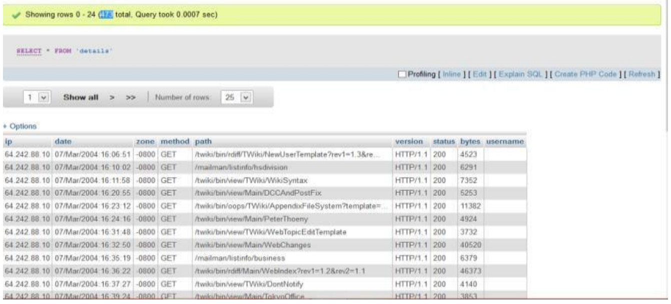
We have conducted several experiments on log files collected from web server. During Data cleansing step all irrelevant entries are removed. Sample raw web log file is as below:



```
10.0.0.153 - - [12/Mar/2004:12:23:18 -0800] "GET /cgi-bin/mailgraph.cgi/mailgraph_1_err.png HTTP/1.1" 200 6949
10.0.0.153 - - [12/Mar/2004:12:23:18 -0800] "GET /cgi-bin/mailgraph.cgi/mailgraph_3.png HTTP/1.1" 200 6644
10.0.0.153 - - [12/Mar/2004:12:23:18 -0800] "GET /cgi-bin/mailgraph.cgi/mailgraph_3_err.png HTTP/1.1" 200 5554
10.0.0.153 - - [12/Mar/2004:12:23:40 -0800] "GET /dccstats/index.html HTTP/1.1" 304 -
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam.1day.png HTTP/1.1" 200 2964
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam-ratio.1day.png HTTP/1.1" 200 2341
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam-ratio.1week.png HTTP/1.1" 200 2346
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam.1week.png HTTP/1.1" 200 3438
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-hashes.1week.png HTTP/1.1" 200 1670
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam.1month.png HTTP/1.1" 200 2651
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam-ratio.1month.png HTTP/1.1" 200 2023
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-hashes.1month.png HTTP/1.1" 200 1636
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam.1year.png HTTP/1.1" 200 2262
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-spam-ratio.1year.png HTTP/1.1" 200 1906
10.0.0.153 - - [12/Mar/2004:12:23:41 -0800] "GET /dccstats/stats-hashes.1year.png HTTP/1.1" 200 1582
216.139.185.45 - - [12/Mar/2004:13:04:01 -0800] "GET /mailman/listinfo/webber HTTP/1.1" 200 6051
pd95f99f2.dip.t-dialin.net - - [12/Mar/2004:13:18:57 -0800] "GET /razor.html HTTP/1.1" 200 2869
d97082.upc-d.chello.nl - - [12/Mar/2004:13:25:45 -0800] "GET /SpamAssassin.html HTTP/1.1" 200 7368
```

Fig 1: Raw log file

Thus, after completion of Data Cleansing Web Server Log file is cleaned and is prepared for data to be loaded into a relational database. Here the data are loaded & stored in MYSQL Server. The result obtained after performing data cleaning phase is shown below.



Showing rows 0 - 24 (473 total, Query took 0.0007 sec)

SELECT * FROM 'details'

□ Profiling [Inline] [Edit] [Explain SQL] [Create PHP Code] [Refresh]

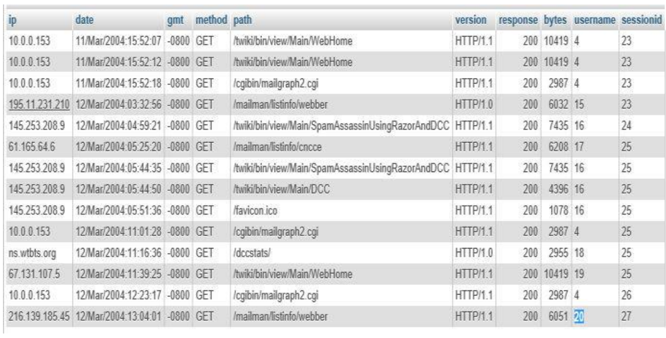
1 Show all > >> | Number of rows: 25

+ Options

ip	date	zone	method	path	version	status	bytes	username
64.242.88.10	07/Mar/2004:16:06:51	-0800	GET	/twiki/bin/rdiff/TWiki/NewUserTemplate?rev1=1.3&re...	HTTP/1.1	200	4523	
64.242.88.10	07/Mar/2004:16:10:02	-0800	GET	/mailman/listinfo/hsdivision	HTTP/1.1	200	6291	
64.242.88.10	07/Mar/2004:16:11:58	-0800	GET	/twiki/bin/view/TWiki/WikiSyntax	HTTP/1.1	200	7352	
64.242.88.10	07/Mar/2004:16:20:55	-0800	GET	/twiki/bin/view/Main/DCCAndPostFix	HTTP/1.1	200	5253	
64.242.88.10	07/Mar/2004:16:23:12	-0800	GET	/twiki/bin/oops/TWiki/AppendixFileSystem?template=...	HTTP/1.1	200	11382	
64.242.88.10	07/Mar/2004:16:24:16	-0800	GET	/twiki/bin/view/Main/PeterThoeny	HTTP/1.1	200	4924	
64.242.88.10	07/Mar/2004:16:31:48	-0800	GET	/twiki/bin/view/TWiki/WebTopicEditTemplate	HTTP/1.1	200	3732	
64.242.88.10	07/Mar/2004:16:32:50	-0800	GET	/twiki/bin/view/Main/WebChanges	HTTP/1.1	200	40520	
64.242.88.10	07/Mar/2004:16:35:19	-0800	GET	/mailman/listinfo/business	HTTP/1.1	200	6379	
64.242.88.10	07/Mar/2004:16:36:22	-0800	GET	/twiki/bin/rdiff/Main/WebIndex?rev1=1.2&rev2=1.1	HTTP/1.1	200	46373	
64.242.88.10	07/Mar/2004:16:37:27	-0800	GET	/twiki/bin/view/TWiki/DontNotify	HTTP/1.1	200	4140	
64.242.88.10	07/Mar/2004:16:39:24	-0800	GET	/twiki/bin/view/Main/TokyoOffice	HTTP/1.1	200	3853	

Fig 2: cleaned log file

Then individual user is identified based on the ip address. After identifying user the required result is shown in the below figure 3.



ip	date	gmt	method	path	version	response	bytes	username	sessionid
10.0.0.153	11/Mar/2004:15:52:07	-0800	GET	/twiki/bin/view/Main/WebHome	HTTP/1.1	200	10419	4	23
10.0.0.153	11/Mar/2004:15:52:12	-0800	GET	/twiki/bin/view/Main/WebHome	HTTP/1.1	200	10419	4	23
10.0.0.153	11/Mar/2004:15:52:18	-0800	GET	/cgibin/mailgraph2.cgi	HTTP/1.1	200	2987	4	23
195.11.231.210	12/Mar/2004:03:32:56	-0800	GET	/mailman/listinfo/webber	HTTP/1.0	200	6032	15	23
145.253.208.9	12/Mar/2004:04:59:21	-0800	GET	/twiki/bin/view/Main/SpamAssassinUsingRazorAndDCC	HTTP/1.1	200	7435	16	24
61.165.64.6	12/Mar/2004:05:25:20	-0800	GET	/mailman/listinfo/cncce	HTTP/1.1	200	6208	17	25
145.253.208.9	12/Mar/2004:05:44:35	-0800	GET	/twiki/bin/view/Main/SpamAssassinUsingRazorAndDCC	HTTP/1.1	200	7435	16	25
145.253.208.9	12/Mar/2004:05:44:50	-0800	GET	/twiki/bin/view/Main/DCC	HTTP/1.1	200	4396	16	25
145.253.208.9	12/Mar/2004:05:51:36	-0800	GET	/favicon.ico	HTTP/1.1	200	1078	16	25
10.0.0.153	12/Mar/2004:11:01:28	-0800	GET	/cgibin/mailgraph2.cgi	HTTP/1.1	200	2987	4	25
ns.wtbts.org	12/Mar/2004:11:16:36	-0800	GET	/dccstats/	HTTP/1.0	200	2955	18	25
67.131.107.5	12/Mar/2004:11:39:25	-0800	GET	/twiki/bin/view/Main/WebHome	HTTP/1.1	200	10419	19	25
10.0.0.153	12/Mar/2004:12:23:17	-0800	GET	/cgibin/mailgraph2.cgi	HTTP/1.1	200	2987	4	26
216.139.185.45	12/Mar/2004:13:04:01	-0800	GET	/mailman/listinfo/webber	HTTP/1.1	200	6051	20	27

Fig 3: user identification

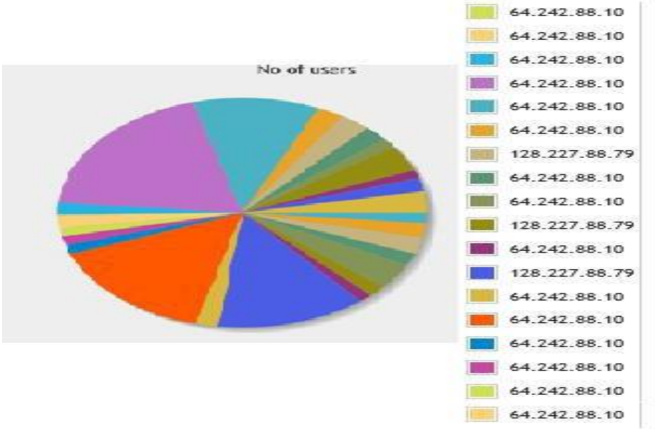
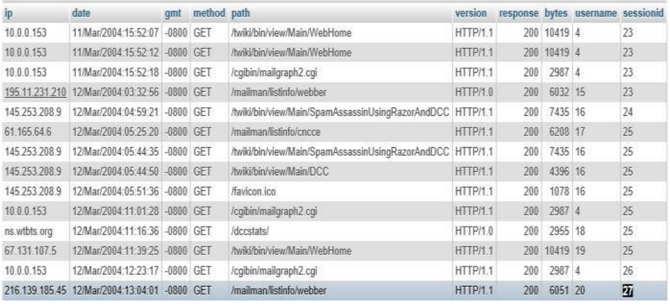


Fig 4: Representation of no. user identification

Then each session is identified based on the time spent on each web page. After identifying session the required result is shown in the below figure 5.



ip	date	gmt	method	path	version	response	bytes	username	sessionid
10.0.0.153	11/Mar/2004:15:52:07	-0800	GET	/twiki/bin/view/Main/WebHome	HTTP/1.1	200	10419	4	23
10.0.0.153	11/Mar/2004:15:52:12	-0800	GET	/twiki/bin/view/Main/WebHome	HTTP/1.1	200	10419	4	23
10.0.0.153	11/Mar/2004:15:52:18	-0800	GET	/cgibin/mailgraph2.cgi	HTTP/1.1	200	2987	4	23
195.11.231.210	12/Mar/2004:03:32:56	-0800	GET	/mailman/listinfo/webber	HTTP/1.0	200	6032	15	23
145.253.208.9	12/Mar/2004:04:59:21	-0800	GET	/twiki/bin/view/Main/SpamAssassinUsingRazorAndDCC	HTTP/1.1	200	7435	16	24
61.165.64.6	12/Mar/2004:05:25:20	-0800	GET	/mailman/listinfo/cncce	HTTP/1.1	200	6208	17	25
145.253.208.9	12/Mar/2004:05:44:35	-0800	GET	/twiki/bin/view/Main/SpamAssassinUsingRazorAndDCC	HTTP/1.1	200	7435	16	25
145.253.208.9	12/Mar/2004:05:44:50	-0800	GET	/twiki/bin/view/Main/DCC	HTTP/1.1	200	4396	16	25
145.253.208.9	12/Mar/2004:05:51:36	-0800	GET	/favicon.ico	HTTP/1.1	200	1078	16	25
10.0.0.153	12/Mar/2004:11:01:28	-0800	GET	/cgibin/mailgraph2.cgi	HTTP/1.1	200	2987	4	25
ns.wtbts.org	12/Mar/2004:11:16:36	-0800	GET	/dccstats/	HTTP/1.0	200	2955	18	25
67.131.107.5	12/Mar/2004:11:39:25	-0800	GET	/twiki/bin/view/Main/WebHome	HTTP/1.1	200	10419	19	25
10.0.0.153	12/Mar/2004:12:23:17	-0800	GET	/cgibin/mailgraph2.cgi	HTTP/1.1	200	2987	4	26
216.139.185.45	12/Mar/2004:13:04:01	-0800	GET	/mailman/listinfo/webber	HTTP/1.1	200	6051	20	27

Fig 5: Session identification

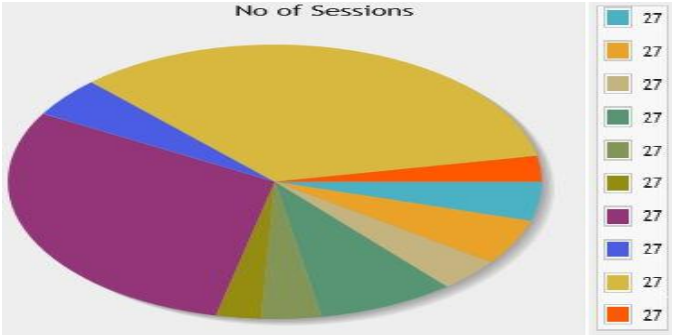


Fig 6: Representation of no. session identification

The figure shows No.of sessions done by different users, which was represented with different colors, each color indicates individual user.

Finally, following table shows clear cut idea about the work, here we have taken 1546 rows of sample weblog dataset and after applying data cleaning algorithm 1 the number of logs are going to reduce to 439 rows which consists of cleaned data shown at figure 2. After getting the cleaned data the individual users are identified by applying algorithm 2 and the result is shown in figure 3 & 4. After getting the user identification the No.of sessions are identified for the individual user by applying algorithm 3 and the result was shown in figure 5 & 6. The overall result was shown in the following table 1.

Final Result:

Rows in the Web Log file	Rows after Preprocessing	Total No. of users	Total No. Of sessions
1546	439	20	27

Table 1: Final result

6. CONCLUSION

Web usage mining is indeed one of the emerging areas of research and important sub-domain of data mining and its techniques. In order to take full advantage of web usage mining and its all techniques, it is important to carry out preprocessing stage efficiently and effectively. This paper tries to deliver areas of preprocessing, including data cleansing, session identification, user identification. Once the preprocessing stage is well-performed, we can apply data mining techniques like clustering, association, classification etc for applications of web usage mining such as business intelligence, e-commerce, e-learning, personalization, etc. Web log mining is one of the recent areas of research in Data mining. Web Usage Mining becomes an important aspect in today's era because the quantity of data is continuously increasing. We deal with the web server logs which maintain the history of page requests

Web log file analysis began with the purpose to offer to Web site administrators a way to ensure adequate bandwidth and server capacity to their organization. By analyzing these logs, it is possible to discover various kinds of knowledge, which can be applied behavior analysis of users. Our proposed system is used to analyze the user sessions from which information regarding the problems occurred to the users and usage of the website can be obtained within particular intervals of time. This is used to configure the server and adjust the Web site which is highly useful for administrators.

References

[1] Ruili Geng, and Jeff Tian "Improving Web Navigation Usability by Comparing Actual and Anticipated Usage" IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 45, NO. 1, FEBRUARY 2015.
[2] G. Neelima and Sireesha Rodda, "An Overview on Web Usage Mining", Springer International Publishing Switzerland December 2015.
[3] Gan Teck Wei, Shirly Kho, Wahidah Husain, Zurinahni Zainol " A Study of Customer Behaviour Through Web Mining" Volume 2, Issue 1 available at www.scitecresearch.com/journals/index.php/jisct/index, February, 2015.
[4] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 10, OCTOBER 2014.
[5] Zhen Liao, Yang Song, Yalou Huang, Li-wei He, and Qi He "Task Trail: An Effective Segmentation of User Search Behavior", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 12, DECEMBER 2014.
[6] George Gkotsis · Karen Stepanyan · Alexandra I. Christie · Mike Joy," Entropy-based automated wrapper generation for weblog data extraction", Received: 31 October 2012 / Revised: 24 October 2013 Accepted: 4 November 2013 / Published online: 21 November 2013 © Springer Science+Business Media New York 2013.
[7] V. S. Dixit • Shveta Kundra Bhatia ," Refinement and evaluation of web session cluster quality", Springer transaction Received: 20 February 2014 / Revised: 2 May 2014.
[8] Renuka Mahajan & J. S. Sodhi & Vishal Mahajan ," Usage patterns discovery from a web log in an Indian e-learning site: A case study", Springer Science+Business Media New York 2014.
[9] Muhammad Muzammal · Rajeev Raman," Mining sequential patterns from probabilistic databases", Received: 11 April 2013 / Revised: 11 May 2014 / Accepted: 3 July 2014 © Springer-Verlag London 2014.
[10] Tomas Arce , Pablo E. Roman , Juan Velasquez , Victor Parada ," Identifying web sessions with simulated annealing", Expert Systems with Applications 41 (2014) 1593–1600.
[11] Hongzhou Sha, Tingwen Liu, Peng Qin, Yong Sun, Qingyun Liu,"EPLogCleaner: Improving Data Quality of Enterprise Proxy Logfor Efficient Web Usage Mining" , Available online at www.sciencedirect.com , Information Technology and Quantitative Management ITQM 2013.
[12] Pani, S.K., Panigrahy, L.: Web Usage Mining: A Survey on Pattern Extraction from Web Logs. International Journal of Instrumentation, Control & Automation (IJICA) 1(1) (2011)
[13] Romero, C., Ventura, S., Zafra, A., de Bra, P.: Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems (received January 8, 2009) (received in revised form May 4, 2009) (accepted May 4, 2009)
[14] Siau, K.: Health Care Informatics. IEEE Transactions on Information Technology in Biomedicine 7(1) (March 2003).
[15] R.Shanthi, Dr.S.P.Rajagopalan, "An Efficient Web Mining Algorithm To Mine Web Log Information", IJIRCCE Vol. 1, Issue 7, September 2013.