International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 5 2997 - 3002

A Comparative Study of Different Log Analyzer Tools to Analyze User Behaviors

S. Bhuvaneswari
P.G Student, Department of CSE,
A.V.C College of Engineering,
Mayiladuthurai, TN, India.
bhuvanacse8@gmail.com

T. Anand
Professor &Head, Department of CSE,
A.V.C College of Engineering,
Mayiladuthurai, TN, India.
anandavcce@gmail.com

Abstract **-** With the explosive growth of information available on internet, WWW become the most powerful platform to broadcast, store and retrieve information. As many people move to internet to gather information, analyzing user behavior from web access logs can be helpful to create adaptive system, recommender system and intelligent e-commerce applications. Web access log files are the files that contain information about interaction between users and the websites with the use of internet. It contains the details like User name, IP Address, Time Stamp, Access Request, number of bytes transferred, result status, URL that referred. To analyze such user behavior, a variety of analyzer tools exist. This paper provides a comparative study between famous log analyzer tools based on their features and performance.

Keywords **-** *Web access logs, User behavior, Log analyzer, World Wide Web.*

*****

I. INTRODUCTION

The development of internet in recent decades made E-commerce websites to bring large records from users. Behavior information of the web users are concealed in the web access logs. It can be automatically created and maintained by the web server. This log file contains much information such as IP address, user name, time stamp, access request, result status, bytes transferred, etc. relative to the web user. Sample log file format is as follows:

```
31.184.238.152 - - [13/Oct/2014:05:32:33
+0530] "GET /logs/access.log
HTTP/1.0" 200 2314780
http://ordereriacta100mgonlinequickshi
pping.soup.io" "Opera/9.80 (Windows
NT 6.1; WOW64) Presto/2.12.388
Version/12.16" "homeplanguru.co.in"
```

Traditionally, four types of logs available in web server: transfer log, agent log, error log and referrer log. First two are standard whereas the remaining is optional. To analyze those access logs, one should follow the sequential steps such as preprocessing, user identification, session identification followed by clustering. A large variety of techniques have been proposed to do this task.

Another efficient way to extract the user behavior from log files is by making use of automated analysis tools. Web log analyzer software passes a server log file from a web server, and based on the values contained in the log file, derives indicators about when, how, and by whom a web server is visited. Usually reports are generated from the log files immediately, but the log files can alternatively be passed for a database and reports generated on demand. Features supported by log analysis packages may include "hit filters", which use pattern matching to examine selected log data.

II. RELATED WORK

Internet is used as information source and it is commonly known as web. Web is an open medium. Due to its Openness, it becomes tough for users to plough through the information [1].

It has been necessary to utilize automated tools to analyze and track the usage patterns. These make a need to create server-side and client-side intelligent systems that can effectively mine for knowledge [2].This can be done analyzing the web access logs which is stored in web servers. These logs enable the analyst to keep track the website and the user behavioral patterns [3].

Olfa nasraoui et al[4] proposed the Competitive Agglomeration for Relational Data (CARD) Algorithm , a clustering algorithm that is designed to organize user sessions into profiles, where each profile would highlight a particular type of user.

Michael shmuli-scheuer et al [5] proposed a scalable user profiling framework that is based on feature selection, where user profiles are represented by the textual content consumed or produced by different users and the aim is to weigh user profile terms according to their capability of representing the user's interests. For that purpose, a new feature selection method based on Kullback-Leibler (KL) divergence, tailored for the user behavior analysis task. But these methods became more complicated because the algorithms provided. Web analytic tools provide simple and effective solutions for the websites without involving any tough algorithms and methods [6].

III. TOOLS AVAILABLE

A variety of tools available in the internet to complete the task of web log analysis from access logs which produces effective reports as output. Some of the most widely used tools are:

A. Google analytics

It is a free utility provided by Google which mainly focuses on marketing. It helps to analyze visitor's traffic and provide a complete report about your audience and their requirements by tracing their path. It supports different file formats with unlimited size. It also supports mobile app analytics to assist the user effectively.

B. Deep Log Analyzer

Unlike the other tools, it has extensible capability to analyze different types of logs including FTP logs. It can create a list of keywords and the hits on web pages that holds keywords. It is very useful for search engine optimization.

C. Web Log Expert:

It is one of the traditional tools used for web log analysis from the log files which can be either IIS or Apache format. Reverse DNS Lookup is the extra-ordinary feature resided in this tool which helps to find domain name of the source IP addresses found in logs. It has built-in database.

D. . Webalizer

It is a command line operable tool famous for web analytics in Linux/Unix environment. It has own configuration language used for reading and parsing the log files. As it contains extensive features, it can be scheduled daily to perform analysis and automatic report formation.

E. PIWIK:

It is the fastest log analysis tool released in the year of 2015. Apart from the web analysis, Piwik has a set of plug-in to enhance the reporting formats. It has own interface using python to get the reports.

F. Open Web Analytics

It is capable of processing really large logs and can optionally fetch those directly from a database format too. Unlike many other professional tools, this open source version can provide a click-stream report. This helps website code troubleshooter, to know exactly what the web user did, and can try repeating those steps to replicate the problem. It can also create heatmap type of report whereby the website statistics is segregated into most-hit and least-hit pages, shown in the form of color gradients for easy understanding.

G. AWStas:

AWStats is a free powerful and featureful tool that generates advanced web, streaming, ftp or mail server statistics, graphically. This log analyzer works as a CGI or from command line and shows you all possible information your log contains, in few graphical web pages. It uses a partial information file to be able to process large log files, often and quickly.

The following table describes the comparison of different web log analyzer tools such as Google analytics, Deep log analyzer, web log expert, webalizer, piwik, open web analytics and AWStats. Every tool has heterogeneous feature to perform log analysis. Some of the common features are analyzed below:

Features	Google analytics	Deep log analyzer	Web log expert	Webalizer	Piwik	Open web analytics	AW Stats
Vendor	Google	Deep software Inc.	Alentum software lmtd.	Webalizer	Piwik Inc	Open web analytics	AW Stats Inc.
Current version	Single	6.0	8.6	2.23-08	2.13.1	1.5.7	7.4
Installation	No need to install. Google account is enough	Easy to install	Easy to install	Easy to install	Easy to install	Easy to install	Easy to install
Log file formats	CLF,XLF,ELF	Apache, IIS	Apache, IIS	CLF,XLF,ELF,FTP	Apache,IIS,Nginx	CLF,XLF,ELF	CLF,XLF,ELF ,W3C,etc.
Website linkage	possible	Log files should be imported	Log files should be imported	Log files should be imported	Log files should be imported	Log files should be imported	Log files should be imported
Price	Free	Starts from $199.95	Starts from $99.00	Free	Free	Free	Free
Language	Built-in	Built-in	Built-in	C	Python/Ruby	PHP	Perl
Database	Built-in	MsAccess	Built-in	GeoDB	MySQL bundled with WAMP	MYSQL	XML/XSLT
User interface	Simple	simple	Simple	Simple	Simple	simple	simple
OS	Windows/Mac / Linux/	windows	Windows	Linux/Mac/Solaris	Windows/Mac / Linux/	windows	windows

	Solaris				Solaris		
Report format	HTML/PDF/CSV	HTML/Ms-Excel	HTML/PDF/CSV	HTML	HTML/PDF	HTML	HTML/PDF
Dynamic reports	Available	Available	Available	Not Available	Available	Not Available	Available
E-Mail report facility	Available	Not Available	Available	Not Available	Available	Not Available	Available
Report scheduler	External	Buit-in	Built-in	External	External	External	External
Real time analysis	Available	Available	Not Available	Not Available	Available	Not Available	Available
Mobile tracking	Available	Not Available	Not Available	Not Available	Available	Not Available	Not Available
Website	http://www.google.com/analytics/	http://www.deep-software.com/	http://www.weblogexpert.com/	http://www.webalizer.org/	http://www.piwik.org/	http://www.openwebanalytics.com/	http://www.awstats.org/

Table 1 Comparison of the features of different web log analysis tools

IV. RESULTS AND INTERPRETATION

Log analyzer tools will produce results such as general statistics, activity statistics, access statistics, visitor's information, browser information and spiders/ error reports. Sample reports generated by different tools are summarized below:

A.General statistics

It produces the general information such as total number of hits per day, the average number of hits per day, page views, etc. it lists all the necessary information one should know about the website**.**

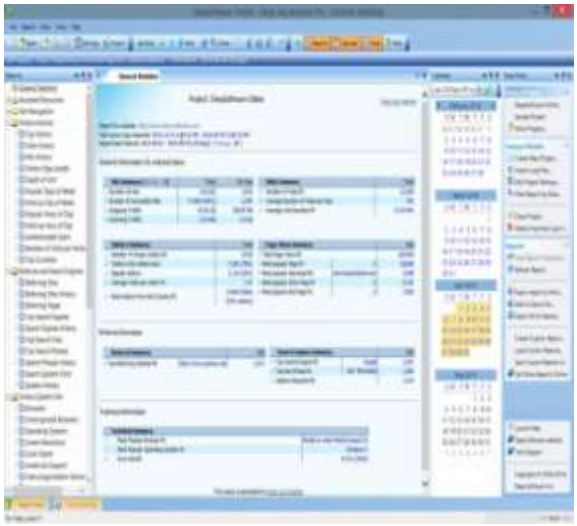


Fig 1: General statistics of Deep log analyzer

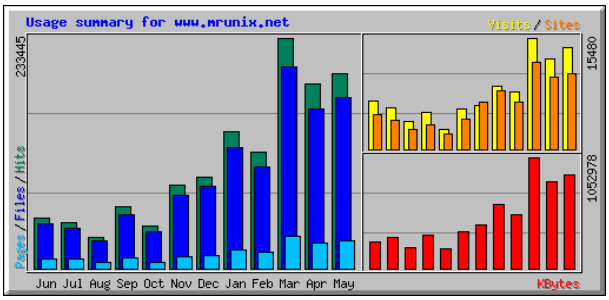


Fig 2: G.S of Webalizer

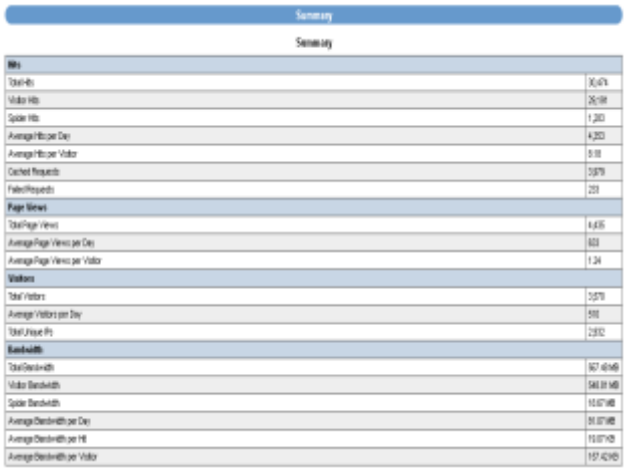


Fig 3: General statistics of web log export

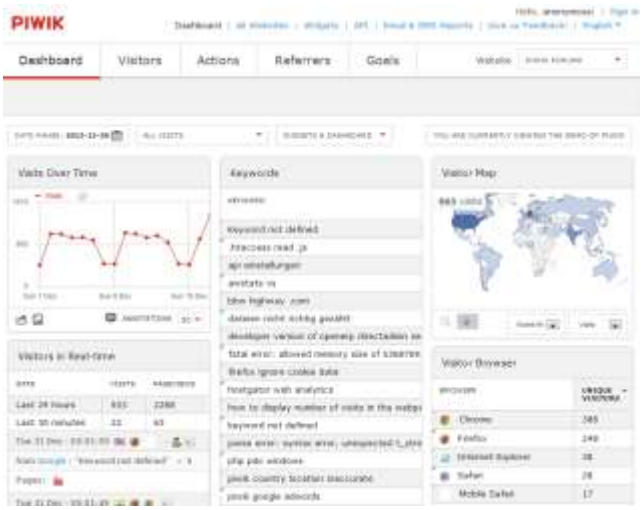


Fig 4: G.S of PIWIK

B. Activity statistics

Monthly, Daily and Hourly basis activity statistics can be provided to the user. Some of the tools provide the activity report in graphical format whereas some of which provide it in tabular report format. From this information, one can increase the visitors count by adapting some features.

ISSN: 2321-8169
2997 - 3002

Fig 5 : Activity statistics of Web log expert

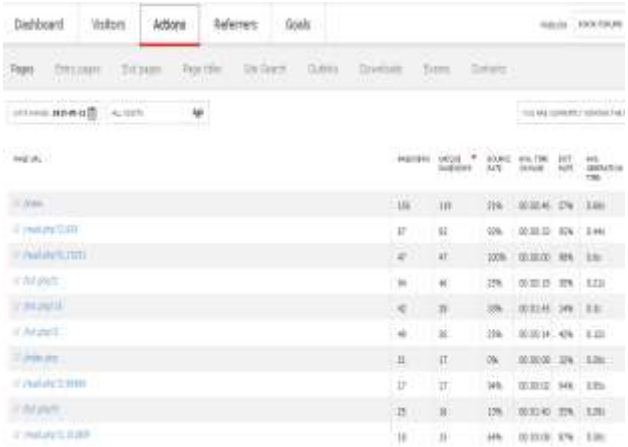


Fig 6 : Activity statistics of PIWIK

C. Access statistics:

It is the most important part of analysis by which one can get which page hits the maximum number of times, navigational behavior of the users, etc. By analyzing it, one can easily get overall idea for their website.

Fig 7: Access statistics of Web log expert

Fig 8: Access statistics of PIWIK

D. Visitor information

The visitor section will help to determine who are all accessed the website. The report contains the information such as IP Address, country of the visitor, number of times visited, etc**.**

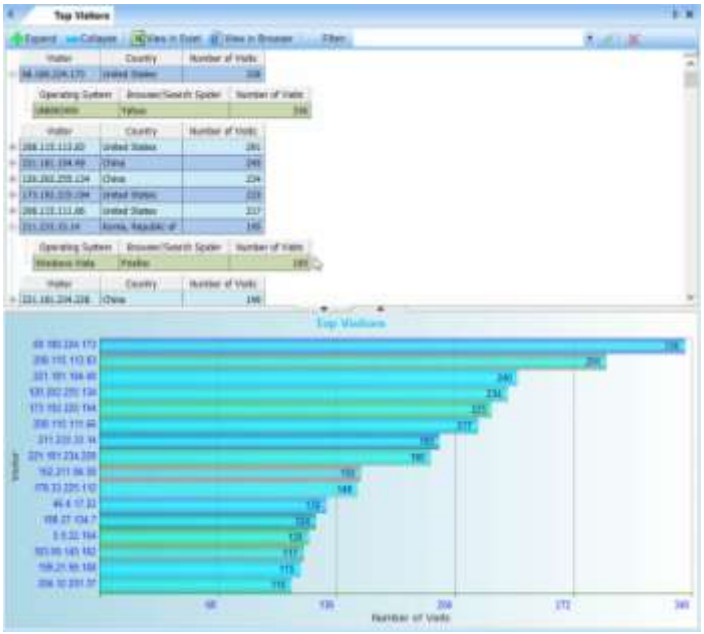


Fig 9: visitor section of Deep Log Analyzer

Fig 10 : Geo-location of visitors in OWA

 ISSN: 2321-8169
2997 - 3002

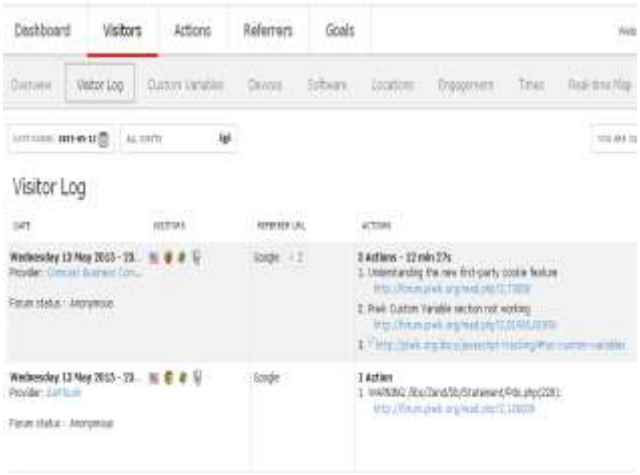


Fig 11: visitors section of PIWIK

E. Browsers

It helps to determine which browser is mostly preferred by users so that one can make the website better compatible to that browser. Format of the browser report may differ based on tool**.**

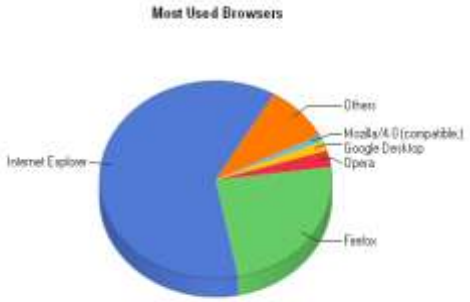


Fig 12: Browser section of Web log expert

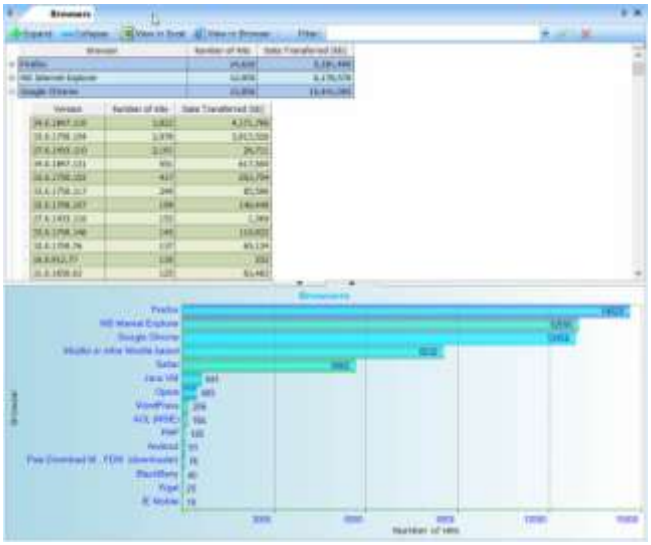


Fig 13: Visitor section of deep log analyzer

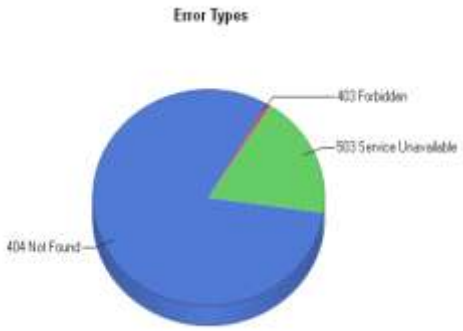
F. Referrers

It contains the information such as type of the referrer and the evolution over some period in the form of graph.

Fig 14: referrer section

G. Errors

It finds out what kinds of error people face when they look into the website. For error feature, both the tabular and graphical form of representation is available**.**



V. CONCLUSION

Automated web log analyzer tools have great impact on web analytics. They take web access log file as input; analyze it and the different domains of reports. As varieties of tools are available, some of the most popular tools are taken to analyze. Every tool offered some or the other feature which was better than the rest. The results were examined by incorporating the website with those tools. Such log analyzer tools should be widely used and they help a lot gain and understand the behavior of the customer or user.

REFERENCES

[1] G.K Gupta, "Introduction to data mining with case studies: web data mining", PHI Learning private limited,pp.231-233,2011.

[2] R.Cooley et al, "Web Mining: Information and Pattern discovery on the world wide web", Proceedings of ICTAI, 1997.

[3] Hillol kargupta, Anupam joshi, Krishnamoorthy sivakumar, Yelena yesha, "Data mining: next generation challenges and future directions: Web mining – concepts, applications and research directions", PHI Learning private limited,pp.405-409.

[4] Olfa nasraoui et al., "Extracting web user profiles using relational competitive fuzzy clustering", International journal on artificial intelligence.

[5] Michal Shmueli-Scheuer et al, "Extracting User Profiles from large scale data", ACM, 2010.

[6] Neha Goel, C.K.Jha, "Analyzing User behavior from web access logs using automated log analyzer tool", International journal of computer applications,2013.

[7] Kanchan Sharadchandra Rahate et al., "A Novel Technique for Parallelization of Genetic Algorithm using Hadoop," International Journal of Engineering Trends and Technology (IJETT), vol.4, issue 8, August 2013.

[8] Silvia Schiaffino et al., "Intelligent User profiling" ,Springer, 2009.