

Web Usage Mining: Analysis of WebLog Software's

Sudip Kumar Suklabaidya¹, Pradeep Sharma²

¹ CDAC NOIDA, ² Assam down town University, Guwahati

Abstract — World Wide Web is becoming one of the prime modes of information in present time. Websites are the nuts and bolts of World Wide Web where all the information lies. It is very important for the website builders and marketers to focus on learning about web users and their interactions with websites. Information on navigation paths is available in log files. Web log files are automatically created and maintained by a web server. Every hit to the website, including each view of html document, image or other object is logged. There are many software's available to analyze these log files. In this paper, we propose an evaluation of some of these software available on scores measured on various factors of web usage mining. Our analysis can be helpful for all the website builders and marketers for selecting software available for web usage mining.

Keywords—Web mining; Web usage mining; Weblog; Web servers; web mining software; E-commerce.

1. INTRODUCTION

Web mining today has become a crucial area of research because of the mammoth amount of data available on the World Wide Web [6]. Web usage mining is a section of web mining where data mining techniques are used to automatically discover usage patterns from web server log files, user queries, and registration data [2][3][5]. These usage patterns are very helpful for extracting hidden useful information like visitors profiles measuring online marketing efforts, launching marketing campaigns [4], etc.

```
<Date> <time> <s-ip> <cs-method> <cs-uri-stem>
<cs-uri-query> <s-port> <cs-username> <c-ip>
<cs(User-Agent)> <sc-status> <sc-substatus>
<sc-win32-status> <time-taken>
```

These information are very effective for E-commerce which has now become one of the most emerging trend in recent times [10]. There are many software's available for web usage mining. The marketer's has to select software according to their need, which can be very tough with so many software's around. This paper analyzes few of the software's available in the market on various factors and ranked them accordingly, which can be very useful for the marketers to get the appropriate software according to their need, which can be, the platform they want to execute the software, the input files, output files, output format, cost

and other information they can get from the software.

2. INTRODUCTION TO WEB LOG FILE

A web log is a file that lists all the actions that have occurred during a client server transaction. Every time you visit a site, the web server will generate a record of the HTTP transaction into a log file. Weblog files are created date wise, Each time a client requests a resource the server of that resource may record the following in its log files:

- The name & IP address of the client computer.
- The time of the request.
- The URL that was requested.
- The time it took to send the resource.
- If HTTP authentication used; the username of the user of the client will be recorded.
- Any errors that occurred.
- The referrer link.
- The kind of web browser that was used.

The most popular log file formats are the Common Log Format (CLF). Example of a log file format of a research institution as shown below.

There are basically four types of log files:

- Access log: The server access log records all requests processed by the server. The name of the log file is specified either by `-l logfile` command line option, or with `AccessLog` directive; log file can be either an absolute path:

`AccessLog /absolute/path/logfile`

or relative to `ServerRoot`:

`AccessLog logs/logfile`

The format of the access log is highly configurable.

A typical configuration for the access log might look as follows:

```
LogFormat "%h %l %u %t \"%r\" %>s %b"
common CustomLog logs/access_log common
```

This defines the *nickname* common and associates it with a particular log format string. The format string consists of percent directives, each of which tell the server to log a particular piece of information.

- Referrer log: The referrer log indicates the page where the visitor was located when making the next request. From referrer log one can tell how the site is categorized. By this way one can also see the key words by the visitors to find one's website. And path of the visitor's navigation can also be known.
- Agent log: The agent log has information about the browser version, and operating system of the visitor.
- Error log: Most error happened with images such as .gif, .bmp, .jpg etc. If the same file appears repeatedly as non-existent, there is a serious problem. administrators should check whether the link is still valid or the server is down. Only by monitoring the content of error log they can make the site run properly. A typical example of an error log:

```
[Wed Oct 11 14:32:52 2000] [error] [client 127.0.0.1] client denied by server configuration: /export/home/live/ap/htdocs/test
```

Access log, agent log and referrer log are always together that is called extended log file. However some server turn off the agent log and referrer log, only leave the access log which is called common log file.

3. WEB USAGE MINING

Discovery of meaningful patterns from data generated by client-server transactions on one or more Web servers [1].

Automatically generated data stored in server logs, referrer logs, agent logs, and client-side cookies. These log files are processed to gain all the useful information about the user. Different processes of web usage mining are [3][7] shown in Figure 1.

- Preprocessing: conversion of the raw data into the data abstraction (users, sessions, episodes, clickstreams, and page views) necessary for further applying the data mining algorithm [8]. Some of the major tasks of data preprocessing are: data cleaning (fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies), data integration (integration of multiple databases, data cubes, or files), data transformation (normalization and aggregation), data reduction (Obtains reduced representation in volume but produces the same or similar analytical results), data discretization (Part of data reduction but with particular importance, especially for numerical data).
- Pattern Discovery: is the key component of Web Usage Mining (WUM), which converges the

algorithms and techniques from data mining, machine learning, statistics and pattern recognition etc. research categories [8].

- Pattern Analysis: Validation and interpretation of the mined patterns [8]. This process answers questions like "How are people using this site?" "which Pages are being accessed most frequently?", This requires the analysis of the structure of hyperlinks and the contents of the pages.

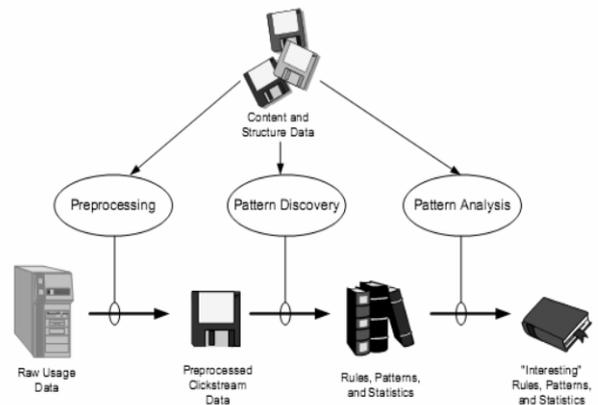


Fig. 1. Process of web usage mining

The data preparation process is often the most time consuming and computationally intensive step in the Web usage mining process, and often requires the use of special algorithms and heuristics. Understanding of these special algorithms can be a very hard nut to crack for normal administrators. They can avoid these complications by simply adopting software's present in the market which will internally do all these data preparation task.

4. DISCUSSION ON SOFTWARE

In this section various software are analysed and brief description are given along with their URL, advantages and drawbacks as shown in Table II. This discussion can be very favourable for the marketers to get the initial knowledge about the software.

5. METHODOLOGY

Our analysis of the Software's available are measured through scores given to various factors related to efficient mining of data from various weblog files as shown in Table I. The tests are done on log data received from a research institution Figure 3. The ranking of the software's are done through total score out of hundred. These factors along with their total score points are given as:

Table-1: Scores

Platforms =20	Ease of use=3 0	Cost= 10	Output formats =10	Availability =10	Report quality =20
--------------------------	------------------------------------	---------------------	-----------------------------------	-----------------------------	-----------------------------------

A. Platform

This factor is related to on which platform the software can be executed. The various platforms we are considering are: Windows 2000, XP, 2003, Vista, 2008, 7(x86, x64), Windows Server 2003/2008, GNU/Linux systems, FreeBSD & Mac OSX Version.

The scoring is done based on assigning five score points to each of these platforms.

- i. Windows 2000, XP, 2003, Vista, 2008, 7(x86, x64) - Score 5
- ii. Windows Server 2003/2008 - Score 5
- iii. GNU/Linux systems - Score 5
- iv. FreeBSD & Mac OSX Version - Score 5
- v. If platform independent – score 20

B. Ease of use

This factor is based on how much user friendly the software is. The total score given to this factor is 30 which are based on Graphical user interface – the user can interact with the software through GUI or is it Code based, Input type – what are the different files types that can deal as an input to the software e.g. IIS log files, apache log files, ZIP, GZ and BZ2 compressed logs, and Output type – whether the output type is statistical based, visual based (Pie/Bar Charts) or both.

The scoring is done based on assigning ten score points to each of these elements.

- i. Graphical user interface – Score 10 (If GUI available than 10 score otherwise 0 score)
- ii. Input type – Score 10 (IIS log files – 3.33 score, Apache log files – 3.33 score and compressed log files – 3.33 score)
- iii. Output type – Score 10 (Statistical based – 5 score and Visual based (Pie/Bar Charts) – 5 score

C. Cost

This factor is based on how much a user has to expense. The scoring is based on 10 score to the minimum free software and 1 score to maximum \$200. Briefly:

- 0 – 20 \$ - 10 score, 20 – 40 \$ - 9 score,
- 40 – 60 \$ - 8 score, 60 – 80 \$ - 7 score,
- 80 – 100 \$ - 6 score, 100 – 120 \$ - 5 score,
- 120 – 140 \$ - 4 score, 140 – 160 \$ - 3 score,
- 160 – 180 \$ - 2 score, 180 – above \$ - 1 score.

D. Output format

This factor is based on what are the different formats the output can be generated from the software. These formats can be of HTM, HTML, PDF, CSV, TXT format.

The scoring is done based on assigning two score points to each of these elements.

- i. HTM – Score 2
- ii. HTML – Score 2
- iii. PDF – Score 2
- iv. CSV – Score 2
- v. TXT – Score 2

E. Report quality

This factor depends on quality of the report generated by the software, which is based on evaluating the software on various aspects as questions .These questions are:-

- i. What are the top ten downloads?
- ii. Maximum time duration of a session?
- iii. Maximum hits on which time of a day?
- iv. Maximum hits on which page on a website?
- v. What are the pages with least hit counts?
- vi. Maximum hits on which days of a week?
- vii. Website is viewed most in which browser?
- viii. Errors occur during a client request.
- ix. Maximum traffic experienced on which part of a year?
- x. What are the top visiting countries?
- xi. Ability to select multiple log files.
- xii. Website is viewed in which operating system?
- xiii. What are the top five ip accesses?

The scoring is done based on assigning 1.54 score points to each of these questions, out of total score of 20 assigned to the Report quality factor.

F. Availability

This factor is depending on availability of the software in the market for the user. The availability is characterized through easily available and hardly available.

The scoring is done based on assigning five score points to each of these elements.

- i. Easily available – Score 5
- ii. Hardly available – Score 5

6. EVALUATION OF THE SOFTWARE

In this section we present the evaluation of the software's based on its executable platforms, its ease of use i.e. GUI based or code based, its input type, output type, cost, output formats and its availability. The results of the evaluation are present in Table III. The software's are also analyzed through various queries by which the marketer

can inspect what type of information he can get from the software. The results are present in Table IV.

7. SCORING OF WEBLOG SOFTWARE'S

In this section after analyzing all the weblog software with respect to the above methodology scoring has been done on various factors as shown in Table V. The scoring has been done on total of hundred as shown in Table V

Table-2: Software Description

S.No	Software	Description	url	Advantage	Drawbacks
1	Download Analyzer	Download Analyzer is web log analysis software specialized in the analysis of file downloads (audio, video, and graphics files, and software distribution kits). It creates special interactive reports on downloads, refers, search phrases and individual search words[10].	http://www.downloadanalyzer.com/	<ol style="list-style-type: none"> 1.Runs on all window versions 2. Commercial software product is distributed as try-before-you-buy. 3.User friendly GUI 4. Easily downloadable from website. 	<ol style="list-style-type: none"> 1.Did not Run on other platforms. 2.Only statistical reports, no graphs(Pie/Bar Charts) 3. Personal License (\$49), Business License - \$99(US).
2	AlterWind Log Analyzer Professional	AlterWind Log Analyzer Professional is unique web log analysis software. New unprecedented reports for web site search engine optimization (SEO), web site promotion, and pay-per-click programs will help administrators save hundreds of hours and thousands of dollars. Administrators can increase the traffic to the web site with a minimal effort [11].	http://www.alterwind.com/	<ol style="list-style-type: none"> 1. Runs on all window versions. 2. Trial version is available free of cost. 3. User friendly GUI. 4. Reports comes in both statistical and graphical format. 5. Easily downloadable from website. 	<ol style="list-style-type: none"> 1. Did not Run on other platforms. 2. Costs \$125 US dollars. 3. Reports can only be get in HTML format.
3	Nihuo Web Log Analyzer	Nihuo Web Log Analyzer is a fast and powerful web access log analyzer for small and medium size websites. By telling administrators where their web site visitors come from, which pages are most popular, and which search engine phrases brought visitors to the web site, you can make your site a more effective sales machine [12].	http://www.loganalyzer.net/	<ol style="list-style-type: none"> 1. Runs on all window versions and Linux, FreeBSD & Mac OSX Version. 2. Trial version is available free of cost. 3. Supports all types of compressed log files. 4. User friendly GUI. 5. Reports comes in both statistical and graphical format. 6. Easily downloadable from website. 	<ol style="list-style-type: none"> 1. Reports can only be get in HTML format. 2. Costs Individual License (10 projects)-\$99.
4	Weblog Expert	Weblog Expert is a fast and powerful access log Analyzer. It will give administrators information about their site's visitors: activity statistics, accessed files, paths through the site, information about referring pages, search engines, browsers, operating systems, and more. The program produces easy-to-read reports that include both text information (tables) and charts [13].	http://www.weblogexpert.com/	<ol style="list-style-type: none"> 1. Runs on all window versions. 2. Trial version is available free of cost. 3. Supports IIS and Apache logs, ZIP, GZ and BZ2 compressed logs. 4. User friendly GUI. 5. Reports comes in both statistical and graphical format. 6. Reports can be get in HTML,PDF and CSV format. 7. Easily downloadable from website. 	<ol style="list-style-type: none"> 1. Did not Run on other platforms. 2.Costs WebLog Expert Standard - \$99.00

5	123LogAnalyzer	123LogAnalyzer is the fastest, most powerful and detailed solution for seeing who your site visitors are, where they come from, where they go, and more! Features include high-speed processing, low disk space requirements, built-in IP mapping and 123LogAnalyzer's unique filtering capabilities. Very simple to use, with detailed tables and charts produced in an easy to understand, professional-quality report. Fast, powerful and professional [14].	http://download.cnet.com/123LogAnalyzer	<ol style="list-style-type: none"> 1. Runs on all window versions. 2. Trial version is available free of cost. 3. Supports IIS and Apache logs, ZIP, GZ and BZ2 compressed logs. 4. User friendly GUI. 5. Reports comes in both statistical and graphical format. 6. Easily downloadable from website. 	<ol style="list-style-type: none"> 1. Did not Run on other platforms. 2. Costs \$129.95 3. Reports can only be get in HTML format.
6	Apache Logs Viewer	Apache Logs Viewer (ALV) is a free and powerful tool which lets administrators monitor, view and analyze Apache/IIS logs with more ease. It offers search and filter functionality for the log file, highlighting the various http requests based on their status code [15].	http://www.apacheviewer.com/	<ol style="list-style-type: none"> 1. Runs on all window versions. 2. Trial version is available free of cost. 3. Supports IIS and Apache logs, 4. User friendly GUI. 5. Reports comes in both statistical and graphical format. 6. Easily downloadable from website. 	<ol style="list-style-type: none"> 1. Did not Run on other platforms. 2. Costs €40 3. Reports can only be get in TXT and CSV format.
7	Indihiang Web Log Analyzer	Indihiang Project is a web log analyzing tool. The tool analyzes IIS and Apache Web servers in comprehensive graphs and reports [16].	http://indihiang.com/deplex.com/	<ol style="list-style-type: none"> 1. Runs on all window versions. 2. Free of cost. 3. User friendly GUI. 4. Reports comes in both statistical and graphical format. 5. Easily downloadable from website. 	<ol style="list-style-type: none"> 1. Did not Run on other platforms. 2. Reports are not in HTML, PDF, TXT or CSV format. 3. Only take IIS log files as input.
8	A1WebStats	Identify companies visiting your website and what they are interested in. A1WebStats can show you the visitors: Company name, Phone number, Website address, Keywords typed, Pages visited and whether the visit came from an Organic search or by Pay-Per-Click [17].	http://www.a1webstats.com/	<ol style="list-style-type: none"> 1. Does not depend on platform (Code based). 2. Easily downloadable from website. 3. Trial version is available free of cost. 	<ol style="list-style-type: none"> 1. No GUI. 2. Costs month to month: £49 + VAT per month. 3. Reports are provided through e-mail. 4. Only statistical reports, no graphs(Pie/Bar Charts)
9	htminer	Support analysis of web logs (including unique visitors, sessions, and transactions); organizes the data in a PostgreSQL data warehouse [18].	http://www.htminer.org/en/	<ol style="list-style-type: none"> 1. Code based runs on Linux system. 2. Easily downloadable from website. 3. Open source software. 4. Supports IIS and Apache log files. 	<ol style="list-style-type: none"> 1. No GUI. 2. Does not run on other platforms.

10	Extreme tracking	A code based web traffic analyzer with two versions ProTracker and FreeTracker having some basic features like Daily, Weekly and Monthly visits. Referrers, Search Engine Queries, Keywords, Geographic's, OS's, Browsers, Display Resolutions and much more, and advanced features like very detailed information per visitor. Including; IP address + resolved, geographic location, organization, system information, net speed, referrer, search engine, keywords up to the exact path the visitor travelled through your site [19].	http://extremetracking.com/	<ol style="list-style-type: none"> 1. Code based 2. Easily downloadable from website. 3. Two versions present ProTracker and FreeTracker. 4. Internally process the website no manual process. 	<ol style="list-style-type: none"> 1. No GUI. 2. Yearly plan: € 4.50/month.
11	Deep log analyzer	Deep Log Analyzer is advanced and affordable web analytics solution for small and medium size websites. One can analyze web site visitors' behavior and get complete website usage statistics in several easy steps! With this website statistics and web analytics software administrators will know exactly where the customers came from, how they moved through the site and where they left it [20].	http://www.deep-software.com/	<ol style="list-style-type: none"> 1. Runs on all window versions. 2. Trial version is available free of cost. 3. User friendly GUI. 4. Reports comes in both statistical and graphical format. 5. Easily downloadable from website. 6. Output can be achieved in browser and excel format. 	<ol style="list-style-type: none"> 1. Did not Run on other platforms. 2. Costs \$299.95 For single computer license.

5	123LogAnalyzer	Windows XP/2003/Vista/2008/7	GUI based, Input - Supports IIS and Apache logs, .log, .zip, .gz, or .txt format log files, Output-Graphical representation of report generated	\$129.95	HTM	Easily available
6	Apache Logs Viewer	Windows 2000, XP, 2003, Vista, 2008, 7, 8/8.1, 2012 (x86, x64), Microsoft .NET Framework 3.5	GUI based, Input- apache or IIS log files. Output-Statistical, Visual Reports (Pie/Bar Charts).	€40	Txt,CSV	Easily available
7	Indihiang Web Log Analyzer	Windows 2000, XP, 2003, Vista, 2008, 7(x86, x64), Windows Server 2003/2008, Microsoft .NET Framework 3.5	GUI based, Input- IIS Log files. Graphical representation of report generated	Free	Table format	Easily available
8	AIWebStats	JavaScript code on html page	Code based	Month to month: £49 + VAT per month.		Easily available
9	htminer	GNU/Linux systems	Code based, Input- IIS Log files, Apache log files,	open-source software		Easily available
10	Extreme tracking	To get the tracker up and running it takes a few minutes signing up and placing the tracker code on web site. Immediately after tracker will be activated automatically	Code based, Statistical report	Yearly plan: € 4.50/month.	Table ,CSV	Easily available

11	Deep log analyzer	Windows 2000, XP, 2003, Vista, 2008, 7(x86, x64), Windows Server 2003/2008.	GUI based, Input - IIS and Apache log files, Output – Table format, Graphical representation of report generated ,	Costs \$299.95 For single computer license	HTML, HTM,C SV,Table	Easily available
----	-------------------	---	--	--	----------------------	------------------

Table-4: Software Report Quality

S.No	Software	Top 10 Down loads	Max time duration of a session	Max hits on which time of a day	Max hit on which page on a website	What are the pages with least hit counts	Max hits on which days of a week	Website is viewed most in which browser	Errors occur during a client request	Max traffic experienced on which part of the year	Top Visiting countries
1.	Download Analyzer	no	no	no	yes	yes	no	no	no	no	no
2.	AlterWind Log Analyzer Professional	yes	yes	yes	yes	yes	yes	yes	yes	no	yes
3.	Nihuo Web Log Analyzer	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
4.	Weblog Expert	yes	yes	yes	yes	yes	no	yes	yes	no	no
5.	123LogAnalyzer	yes	yes	yes	yes	yes	yes	yes	yes	no	yes
6.	Apache Logs Viewer	no	yes	yes	yes	yes	yes	yes	no	no	yes
7.	Indihiang Web Log Analyzer	no	yes	yes	yes	yes	yes	yes	no	no	no
8.	A1WebStats	-	-	-	-	-	-	-	-	-	-
9.	htminer	-	-	-	-	-	-	-	-	-	-
10.	Extreme tracking	-	yes	yes	yes	yes	yes	yes	yes	no	yes
11.	Deep log analyzer	yes	yes	yes	yes	yes	yes	yes	yes	no	yes

Ability to Select Multiple log files	Viewed In which Operating System	Top 5 ip access
yes	No	no
yes	Yes	yes
yes	Yes	yes
no	Yes	yes
yes	Yes	yes
yes	Yes	yes
yes	Yes	yes
-	-	-

-	-	-
-	Yes	yes
yes	Yes	no

Table-5: Software Scoring

S.No	Software	Platforms	Ease of use	cost	Output formats	Availability	Report quality
1.	Download Analyzer	5	21.5	8	2	10	3
2.	AlterWind Log Analyzer Professional	5	26.5	4	2	10	18
3.	Nihuo Web Log Analyzer	15	30	6	2	10	20
4.	Weblog Expert	5	30	6	6	10	14
5.	123LogAnalyzer	5	30	4	2	10	18
6.	Apache Logs Viewer	5	26.5	8	4	10	15
7.	Indihiang Web Log Analyzer	10	23.33	10	0	10	14
8.	A1WebStats	20	-	6	-	10	-
9.	htminer	5	-	10	-	10	-
10.	Extreme tracking	20	20	7	8	10	15
11.	Deep log analyzer	10	27	1	8	10	17

Table-6: Software Ranking

S.No	Software	Score
1	Download Analyzer	49.5
2	AlterWind Log Analyzer Professional	65.5
3	Nihuo Web Log Analyzer	77
4	Weblog Expert	71
5	123LogAnalyzer	69
6	Apache Logs Viewer	68.5
7	Indihiang Web Log Analyzer	67.33
8	A1WebStats	-
9	htminer	-
10	Extreme tracking	80
11	Deep log analyzer	73

8. CONCLUSION

Various category of information's are mined from log files are possible because of web usage mining. Enhancing the usability of website's, customer relation improvement are some of the major roles of web usage mining. And this is possible because of the various web mining software's present, each of them has some advantages and drawbacks. It has become laborious for the website developers and marketers to find the appropriate software with respect to there need. Towards this goal we discussed some of these software's available on various factors and score them accordingly with all its advantages and drawbacks, although research results may vary because of advanced versions of the software's or manual error. This reviews can be very useful for them to select relevant software.

REFERENCES

- [1] Rajni Pamnani and Pramila Chawan, "Web usage mining: a research area in web mining," Department of computer technology, VJTI University, Mumbai.
- [2] L.K. Joshila Grace, V.Maheswari and Dhinaharan Nagamalai, Analysis of weblogs and web user in web mining, Sathyabama University, Chennai, India, Wireilla Net Solutions PTY Ltd, Australia.
- [3] R.Cooley, B.Mobasher and J. Srivastava, "Web mining: information and pattern discovery on the world wide web," Department of computer Science and Engineering, University of Minnesota, Minneapolis. MN 55455, USA.
- [4] P.Perner and G.Fiss, "Intelligent e-marketing with web mining, personalization and user-adapted interfaces," Institute of Computer Vision and Applied Computer Sciences, Arno-Nitzsche-Str. 45, 04277 Leipzig.
- [5] Magdalini Eirinaki and Michalis Vazirgiannis, "Web mining for web personalization," Athens University of Economics and Buisness.
- [6] Andreas Hotho and Gerd Stumme, "Mining the world wide web – methods, applications, and perspectives," .
- [7] Renata Ivancsy and Istvan Vajk , "Frequent pattern mining in weblog data,". Mill Valley, Department of Automation and Applied Informatics and HAS-BUTE Control Research Group, Budapest University of Technology and Economics, Goldmann Gy. tér 3, H-1111 Budapest, Hungary.
- [8] Yan Wang, "Web mining and knowledge discovery of usage pattern,". CS 748T Project (part I).
- [9] Ajith Abraham, "Buisness intelligence from web usage mining,". Department of computer science, Oklahoma State University, 700 N Greenwood Avenue, Tulsa, Oklahoma 74106-0700, USA.
- [10] Downloadanalyzer.Internet: <http://www.downloadanalyzer.com/index.html>, date

[22/04/2014].

- [11] Alterwind: the software company. Internet: <http://www.alterwind.com/loganalyzer/log-analyzer-professional.html>, date updated* [23/04/2014].
- [12] Nihuo software. Internet: <http://www.loganalyzer.net/>, [23/04/2014].
- [13] Weblog expert. Internet: <http://www.weblogexpert.com/>, [24/04/2014].
- [14] 123Log analyzer. Internet: <http://www.123loganalyzer.com/>, [24/04/2014].
- [15] Apache log viewer. Internet: <http://www.apacheviewer.com/>, [24/04/2014].
- [16] Indihiang. Internet: <http://indihiang.codeplex.com/>, [24/04/2014].
- [17] A1webstats. Internet: <http://www.a1webstats.com/>, [25/04/2014].
- [18] ht://miner. Internet: <http://www.htminer.org/>, [25/04/2014].
- [19] eXTReMe tracking. Internet: <http://extremetracking.com/> [30/04/2014].
- [20] Deep log analyzer Internet: <http://www.deep-software.com/> [3/05/2014].

AUTHOR'S PROFILE

Sudip Kumar Suklabaidya has received his Master of Technology degree in Computer Science & Engineering from CDAC NOIDA, under Guru Gobind Singh Indraprastha University, New Delhi in the year 2013. His areas of interest are Programming, Database Management System, Design and Analysis of Algorithm, Data Structure, Web Based programming and Designing.

Pradeep Sharma has received his Master of Technology degree in Information Technology from Tezpur University, Tezpur, Assam in the year 2012. At present he is at Assam down town University working as an Assistant Professor in the department of Computer Science and Engineering, Guwahati, Assam. His areas of interests are Data Mining, Algorithm and Programming.