



ON COMPUTING WEB ANALYTICS FOR GENERATING USAGE STATISTICS

JKR Sastry, Y. Nithin, M. L. Vyshnavi, N. Sai Keerthi and A. Lakshmi Anvitha

KL University, Vaddeswaram, Guntur District, Andhra Pradesh, India

E-Mail: drsastry@kluniversity.in

ABSTRACT

Many web sites are in vogue, which are hosted by many business establishments' world-over. Many users browse through the web site for their information requirements and understanding. Some of the users frequently visit the WEB site for their daily information requirements. Users loses lots of time and energy in looking for the information that they need when the WEB site are loosely designed and takes quite a time in navigating through different hyperlinks so as to arrive at the page which contains information needed by them. The WEB sites have to be designed and developed in such way that browsing can be done in easiest manner. It is necessary to analyse the way the users wants to use the WEB site and use the analysis for improving the design and implementation of the WEB sites. Efficient WEB analytics are to be identified and computed that reflect the effectiveness of the WEB sites and also provide foundations and insights into improving of the WEB sites. This paper presents efficient WEB analytics and the related computational aspects that determine the WEB usage. A model is also presented that help determining the improvements in the design and implementation of the WEB sites. The WEB analytics are computed for an existing WEB site and has been shown in the paper, the way the WEB analytics have been improved after the WEB site has been modified based on previously computed analytical values.

Keywords: WEB analytics, performance of WEB sites.

1. INTRODUCTION

Web analytics are the measurement, collection, analysis and reporting of web data for purposes of understanding and optimizing web usage. Web analytics is not just a process for measuring web traffic but can be used as a tool for business and market research, and to assess and improve the effectiveness of a website. Web analytics applications can also help companies measure the results of traditional print or broadcast advertising campaigns. It helps one to estimate how changes traffic to a website after the launch of a new advertising campaign. Web analytics provides information about the number of visitors to a website and the number of page views. It

helps to gauge traffic and popularity trends which are useful for market research.

Web Analytics are useful for measuring and maximizes ROI (return on investment), targeting marketing efforts, and converting visitors to the WEB site as customers, subscribers or members, provide online support leading to reduction in the offline costs, determine trends on the moment of products and determine the products that lead the market etc. Many analytics are being used to compute efficiency of a website. The analytics however differs from website to website. Some of the factors that are used to compute analytics of a website are shown in the Table-1.

Table-1. Frequently used WEB analytics.

Serial number	Type of WEB analytic	Description of the WEB analytics
1	HIT	A request for a file from the web server. Available only in log analysis. The number of hits received by a website is frequently cited to assert its popularity, but this number is extremely misleading and dramatically overestimates popularity. A single web-page typically consists of multiple of discrete files, each of which is counted as a hit as the page is downloaded, so the number of hits is really an arbitrary number more reflective of the complexity of individual pages on the website than the website's actual popularity. The total number of visits or page views provides a more realistic and accurate assessment of popularity.
2	Page View	A request for a file, or sometimes an event such as a mouse click, that is defined as a page in the setup of the web analytics tool. An occurrence of the script being run in page tagging. In log analysis, a single page view may generate multiple hits as all the resources required to view the page (images, .js and .css files) are also requested from the web server.
3	Event	A discrete action or class of actions that occurs on a website. A page view is a type of event. Events also encapsulate clicks, form submissions, keypress events, and other client-side user actions



Serial number	Type of WEB analytic	Description of the WEB analytics
4	Visit / Session	A visit or session is defined as a series of page requests or, in the case of tags, image requests from the same uniquely identified client. A unique client is commonly identified by an IP address or a unique ID that is placed in the browser cookie. A visit is considered ended when no requests have been recorded in some number of elapsed minutes. A 30 minute limit ("time out") is used by many analytics tools but can, in some tools (such as Google Analytics), be changed to another number of minutes. Analytics data collectors and analysis tools have no reliable way of knowing if a visitor has looked at other sites between page views; a visit is considered one visit as long as the events (page views, clicks, whatever is being recorded) are 30 minutes or less closer together. Note that a visit can consist of one page view, or thousands. A unique visit's session can also be extended if the time between page loads indicates that a visitor has been viewing the pages continuously.
5	First Visit / First Session	(Also called 'Absolute Unique Visitor' in some tools). A visit from a uniquely identified client that has theoretically not made any previous visits. Since the only way of knowing whether the uniquely identified client has been to the site before is the presence of a persistent cookie or via digital fingerprinting that had been received on a previous visit, the First Visit label is not reliable if the site's cookies have been deleted since their previous visit.
6	Visitor / Unique Visitor / Unique User	The uniquely identified client that is generating page views or hits within a defined time period (e.g. day, week or month). A uniquely identified client is usually a combination of a machine (one's desktop computer at work for example) and a browser (Firefox on that machine). The identification is usually via a persistent cookie that has been placed on the computer by the site page code. An older method, used in log file analysis, is the unique combination of the computer's IP address and the User Agent (browser) information provided to the web server by the browser. It is important to understand that the "Visitor" is not the same as the human being sitting at the computer at the time of the visit, since an individual human can use different computers or, on the same computer, can use different browsers, and will be seen as a different visitor in each circumstance. Increasingly, but still somewhat rarely, visitors are uniquely identified by Flash LSO's (Local Shared Object), which are less susceptible to privacy enforcement
7	Repeat Visitor	A visitor that has made at least one previous visit. The period between the last and current visit is called visitor decency and is measured in days.
8	Return Visitor	A Unique visitor with activity consisting of a visit to a site during a reporting period and where the Unique visitor visited the site prior to the reporting period. The individual is counted only once during the reporting period
9	New Visitor	A visitor that has not made any previous visits. This definition creates a certain amount of confusion (see common confusions below), and is sometimes substituted with analysis of first visits
10	Impression	The most common definition of "Impression" is an instance of an advertisement appearing on a viewed page. Note that an advertisement can be displayed on a viewed page below the area actually displayed on the screen, so most measures of impressions do not necessarily mean an advertisement has been viewable
11	Single Page Visit / Singleton	A visit in which only a single page is viewed (a 'bounce').
12	Bounce Rate	The percentage of visits that are single page visits
13	Exit Rate / % Exit	A statistic applied to an individual page, not a web site. The percentage of visits seeing a page where that page is the final page viewed in the visit.
14	Page Time Viewed / Page Visibility Time / Page View Duration	The time a single page (or a blog, Ad Banner...) is on the screen, measured as the calculated difference between the time of the request for that page and the time of the next recorded request. If there is no next recorded request, then the viewing time of that instance of that page is not included in reports.
15	Session Duration / Visit Duration	Average amount of time those visitors spend on the site each time they visit. This metric can be complicated by the fact that analytics programs cannot measure the length of the final page view.
16	Average Page View Duration	Average amount of time that visitors spend on an average page of the site.
17	Active Time / Engagement Time	Average amount of time that visitors spend actually interacting with content on a web page, based on mouse moves, clicks, hovers and scrolls. Unlike Session Duration and Page View Duration / Time on Page, this metric can accurately measure the length of engagement in the final page view, but it is not available in many analytics tools or data collection methods
18	Average Page Depth / Page Views per Average Session	Page Depth is the approximate "size" of an average visit, calculated by dividing total number of page views by total number of visits.



Serial number	Type of WEB analytic	Description of the WEB analytics
19	Frequency / Session per Unique	Frequency measures how often visitors come to a website in a given time period. It is calculated by dividing the total number of sessions (or visits) by the total number of unique visitors during a specified time period, such as a month or year. Sometimes it is used interchangeable with the term "loyalty."
20	Click path	The chronological sequence of page views within a visit or session
21	Click	Refers to a single instance of a user following a hyperlink from one page in a site to another

Web analytics can be classified as off-site and on-site web analytics. Off-site web analytics refers to web measurement and analysis regardless of whether you own or maintain a website. It includes the measurement of a website's potential audience (opportunity), share of voice (visibility), and buzz (comments) that is happening on the Internet as a whole. **On-site** web analytics measure a visitor's behaviour while navigating the website. This includes its drivers and conversions such as the degree to which different pages are associated with transactions such as online purchases. On-site web analytics measures the performance of a website in a commercial context. This WEB analytics data is compared against key performance indicators for performance, and used to improve a website or marketing campaign's audience response. Google Analytics is the most widely used on-site web analytics service; although new tools are emerging that provide additional layers of information, including heat maps and session replay.

Historically, web analytics has been used to refer to on-site visitor measurement. However, this meaning has been changed mainly because vendors are producing tools that provide on-site web analytics.

There are two main technical ways of collecting the data. The first and traditional method is the server log file analysis. The WEB server logs data about the surfing done by the users. The second method is called page tagging which is achieved through embedding JavaScript within a WEB page through which a request is made to third party server for collecting data about the WEB usage and store the data in a specially designed database. The third party servers can be enquired for getting reports related to WEB analytics based on the data that server could collect through invocation of a request initiated by a potential user

The Web analytics shown in Table-1 while measure to certain extent the effectiveness of the use of WEB site, it does not reflect any thing about the mission content for which the user is looking for. No insights have been thrown related to the kind of improvements that must be made in order to increase either the WEB analytic that express the positives of the site or decrease in the value of WEB analytic which measures the negative aspect of a WEB site.

Many WEB analytics are to be added and the measurements of the same are to be carried so that true usage of the WEB site is assessed. Some of the WEB analytics required include number of downloads, number of times access control is set over a time, extent of the content searched, number of times the WEB site is self-configured etc.

Every WEB site has its own features. Not all WEB analytics are suitable for every WEB site. It becomes necessary to assess the applicability of WEB analytics given the features of a WEB site so that the effectiveness of a WEB site can be assessed and the changes to be carried to the WEB site for improving the WEB analytics can be identified and the WEB sites can be improved by making those changes.

Web analytics are the best source to compute the efficiency and effectiveness of a given website. Number of web analytics has been proposed in the literature that can be generally considered for evaluating the effectiveness of a website. However not all web analytics proposed in the literature are relevant for every type of website. Therefore there should be an analytical approach using which most relevant and applicable web analytics are selected which are suitable for a given website and the same are used for evaluating effectiveness of the websites.

WEB analytics must be exhaustive so that one can select most appropriate WEB analytics suitable to compute the effectiveness of a WEB site. The changes that must be made to a WEB site based on the value of the analytics must be determined so that the changes can be carried and the improvements made to the WEB sites are measured yet again using the chosen WEB analytics that are related to the very WEB site which is being evaluated.

2. RELATED WORK

Catherine Dwyer [1] applied Web analytics to a research oriented virtual community and measure member's usage characteristics and interaction with the site. Web analytics uses data mining techniques to uncover browsing patterns by examining the content of Web server logs. It has been successfully applied in usability studies to highlight Web design problems, and has also been used in e-commerce to predict and/or influence shopper's choices. They have presented three basic steps for processing Web analytics using WEB logs. The steps include Pre-processing that deal with conversion of WEB log entries into abstractions, Pattern discovery through application of statistical analysis, association, clustering, and classification and Pattern analysis that focus on meaningful analysis.

Daniel Waisberg And Avinash Kaushik [2] have expressed that web analytics can be effectively used for improving the customer experience using statistics, data mining techniques and special processes that can be specially designed and implemented for the purpose. Web analytics can be used for improving the website such that the customer satisfaction is improved. The customer satisfaction can be expressed in terms of creativity, quality of images and content. The extent to which the



improvements can be made are expressed in terms of WEB analytics and improved levels of customer satisfaction. The WEB analytics can be related to key performance indicators (KPI). KPIs can be determined for each of the WEB site and mapped to different types of analytics. The computed values of the WEB analytics reveal the state of KPI. Improvement in WEB analytics can be used to indicate the extent to which the KPIs have been achieved. The WEB use can be collected by capturing the clicks made by the user for navigating through the WEB site

Mohammad Amin Omidvar *et al.*, [3] have presented a flexible method for analysing the use of WEB sites through regression analysis which can be carried through Google's WEB analytic tool. They have presented that grouping the WEB users based on the territory will give better insights into use of WEB sites.

Kristin Glass, *et al.*, [4] have presented a method that explore analysing web sites from the perspectives of security relevant information discovery, Target awareness and Predictive analysis. They have presented three approaches for assessing the security of web sites that include Textual, Relation and temporal

Bhavana Beri *et al.*, [5] have expressed that the usability of a web site is very much dependent on the conversion rate which is nothing but the measurement that indicates the extent to which the WEB site is used by the customer.

They have presented that the web analytics are to be extended to find the extent to which the WEB sites are being used. They have presented a method to compute website usability. The method is based on user testing which is carried by making the user to accomplish some tasks. The testing is carried through implementation of a think-aloud protocol while the user is browsing through the website

They have expressed yet another method which is based on Usability-expert walkthroughs. Both the methods presented by them are useful for the contextual design of software meant for managing mass market where the user is unknown.

Gurpal Singh [6] has presented various operations of web analytics, Types of Web Analytics, advantages and disadvantages of the web analytics. A comparison of different types of WEB analytics is provided. The operation of Web Analytics collecting relevant data in terms of time during which the WEB site is used, number of clicks made etc. Hybrid methods analyses the WEB site through exploring both the files and page tagging. Some of the Hybrid Methods discussed include Geolocation of Visitors, Click Analytics and Customer Life Cycle Analytics.

Pawandeep Singh, *et al.*, [7] have expressed that online web analytics is a tool to understand the user behaviour and to take the decision accordingly. Information gathering, user views, conversion, usability are the parts of the web analytics application. The success of an application lies on the usability and the rate of conversion and web analytics will help improving the website usability and conversion rate.

Vishal Machewad and Sneha Agarwal [8] have expressed that Responsive web design (RWD) is the associate approach. It permits style and development to resort to the user behaviour and setting, supported screen size, platform and orientation that includes straight forward reading. RWD gained attention owing to the rise in the varieties of devices used for web site browsing which include mobile devices like sensible phones and tablets and enormous screen devices like sensible TV's. Responsive web site goals, KPIs and targets support the user in providing the content through various devices. Internet analytics is a platform to know and analyse the performance of a web site across devices and screen sizes.

Infosys white paper [9] explained a general scenario which deals with the key issues of presentation components where it includes collection of data, measuring, analysing and reporting the data. This paper presents the basic introduction of web analytics and its advantages and importance. It also provides the key business issues required for implementation of web analytics. The two different strategies to measure the effectiveness is discussed and also the performance management details of the relevancy of web analytics for emerging markets is presented

Fabio Rizzotto [10] gave clear view about the latest communication technology tools named "Web Analytics". He described the way to interpret analytic data and also to report it. He has shown how web analytics can be used for different purpose.

3. MORE WEB ANALYTICS

The list of web analytics that have been collected from the literature have been shown in Table 1. More of web analytics can be determined than what are presented in the literature that are quite frequently used for different purposes. Some of the additional WEB analytics required to make the WEB sites extensive are shown in the Table 2.

The way the WEB analytics can be measured also are described in the table. The need for additional web analytics have been investigated by working with many of the WEB sites that have different features and current day requirements as felt by many of the WEB sufferers.



Table-2. More WEB analytics.

S. No.	Type of WEB analytic	Description of the WEB analytics
1	Number of Downloads within a specified time	This is a measure of the need for referring to the documents related to businesses
2	Speed of down loading	The speed of down loading is a measure of rate of data transfer generally measured in terms of KB/s. If this measure is small the user will be dis-satisfied and look for some other method getting the required resources
S. No.	Type of WEB analytic	Description of the WEB analytics
3	Speed of up loading	The speed of up loading is a measure of rate of data transfer generally measured in terms of KB/s. If this measure is small the user will be dis-satisfied and look for some other method of uploading the required resources
4	Ease of data collection	Users wants to input as less input as possible in search of accessing the information through minimal processing to be undertaken on the WEB server side. This can be measured in terms of ratio of number of characters of data inputted to the total content that should have been accessed using the user interface.
5	Time to search	Amount of time taken to search for content based on the WEB snippets
6	Number of WEB pages searched	The minimal number of WEB pages that must be surfed to get the content required through searching
7	Complexity of configuring the WEB site in terms of fewer parametric entries	The ease with which the customer can specify the navigation and the pages that are frequently surfed can be measured in terms of number of parameters sued for configuration.
8	Speed of display of right and left menus measured in milli seconds from the time of either click or focus of the pointer	The acceptable time is 20 Milli seconds
9	URL redirection	Must be done in 20Milli Seconds
10	Auto refreshing of a page, the content of which is changed on the server side while the page is being navigated by the user	Must be done in 20Milli Seconds
11	Context level help	Must be provided on at least 80% of the objects displayed on a page within 20 Milli Seconds when pointer is focussed on the object
12.	Number of uploads	This is a measure of need for making information public for the perusal of the stake holders of the business
13	Ease of display of output	Number of characters displayed per second

Every web site is domain specific. The look and feel of the web sites is primarily dependent on the arrangement of the content for access by the users with ease. Creativity of the WEB designer comes into play on designing the effectiveness web sites. The extent to which a web site is used is dependent on the features designed into the WEB and ease with which the features are supported for access by the users. One can sought out variety of features that one can support through study of different types of web sites which are in vogue as on today

and also more of innovations can be carried to invent newer features that can be supported on the WEB site.

Some of the features that have been collected through study of different web sites are shown in Table 3. The features have an association with different kind of WEB analytics. Any web site can be chosen for which the WEB analytics are to be computed and improve the WEB site based on the level of WEB analytics, The WEB analytics related to selected web sites can be determined based on the features of the selected web site

**Table-3.** Mapping features to WEB analytics.

Feature serial	Feature serial	Whether supported in ECM web site	Existing WEB analytics serial	Additional web analytics serial
1	Search Interface	X		5
				6
2	Pane based presentation	√	14	
3	Existence of a tool bar	√	16	
4	Tree based menu	X	16	
5	Drop down menu list	√	16	
6	Display navigation path	X	16	
7	Site map	X	11	
8	Tab based menu	X	16	
9	Use of horizontal scroll bars	√	21	
10	Use of vertical scroll bars	√	21	
11	Scrolling menu	X	14	
Feature serial	Feature serial	Whether supported in ECM web site	Existing WEB analytics serial	Additional web analytics serial
12	Tool bar based menu	√	14	
13	Sliding menu	X	14	
14	Use of forms for capturing input	√		4
15	Use of forms for displaying output	√		14
16	Use of tables	√	3	
17	Social connectivity	X		9
18	Quick links	X		9
19	Use of frames	√	19	
20	Button for viewing cart in e-commerce sites	X	15	
21	Use of photo slides	√	19	
22	Announcements		19	
23	Bulletin boards	√	19	
24	Video streaming	X	19	
25	Category based menu for e-commerce	X	18	
26	Uploading Features	X	22	
			23	
27	Downloading features	X		1
28				2
29	Auto refreshing	X		10
30	URL Redirection	X		9
31	Context Level Help	X		11
32	Left tree menu	X		8
33	Right tree Menu	X		8
34	Drop down Menu	√	11	
35	Self-Configuration	X		7

4. COMPUTING THE WEB ANALYTICS

Some of WEB analytics can be computed with the help of WEB log and some WEB analytics are to be computed while the WEB site is in running mode. The



values of the WEB analytics can be computed and displayed offline or through execution of a query initiated

from the WEB site itself. Figure-1 shows the computational approach.

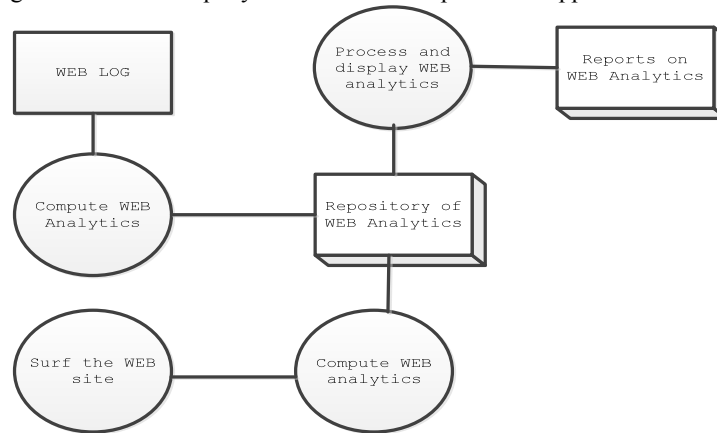


Figure-1. Computing and displaying WEB analytics.

5. DESIGNING TO COMPUTE WEB ANALYTICS

A sample web site has been developed and four analytics related to the web site have been computed through query execution initiated by the user from the

sample web site that has been modified to have the links required for executing the four types of Analytics. The details of the WEB analytics that have been programmed are placed in the Table-4.

Table-4. Details of WEB analytics computed.

S. No.	WEB analytic computed	WEB analytic serial	Computational query	Query results before improving the WEB site	Query results after improving the WEB site
1	Frequency of use of WEB site	1	SELECT url, count(url) from login_access group by url order by count(url) desc	32	237
2	Average Number of users accessing the WEB site	4	SELECT ceil(count(user_id)/count(distinct(date))) from login_access	4	12
3	Frequently used resource files	2	INSERT INTO cntmax(url,count) select url,count(url)from login_access group by url From the table created we are going to get the maximum value of the count using the following query SELECT url from cntmax where count=(select max(count) from cntmax)	Kluniversity.in/default.aspx Response (1 Secs)	Kluniversity.in/default.aspx Response (0.1 Secs)
4	Frequency of use of navigation path	3	SELECT user_id, url, date, time FROM login_access order by user_id,date,time asc	Kluniversity.in/faculty.aspx Extent of URL used 230 Minutes	Kluniversity.in/faculty.aspx Extent of URL used 990 Minutes

The results obtained out of executing the queries are shown in **Figure-2 to Figure-5**. From Figure-2 it can be seen that some of the URLs are used by very few users while some URLs are used most frequently. The most frequently used URLs are to be provided as quick links in the home page.

Figure-3 shows the number of users using the WEB site is 4 while after improving the navigation the average number of users accessing the WEB site has increased to 12. **Figure-4** shows that the <http://kluniversity.in/default.aspx> is quite frequently used and therefore needs to be cached both on the client side and



on the server side to improve the response from the WEB site. **Figure-5** shows the extent of use of navigation path.

The more the navigation path used, it becomes the case to place the URL right on to the home page.

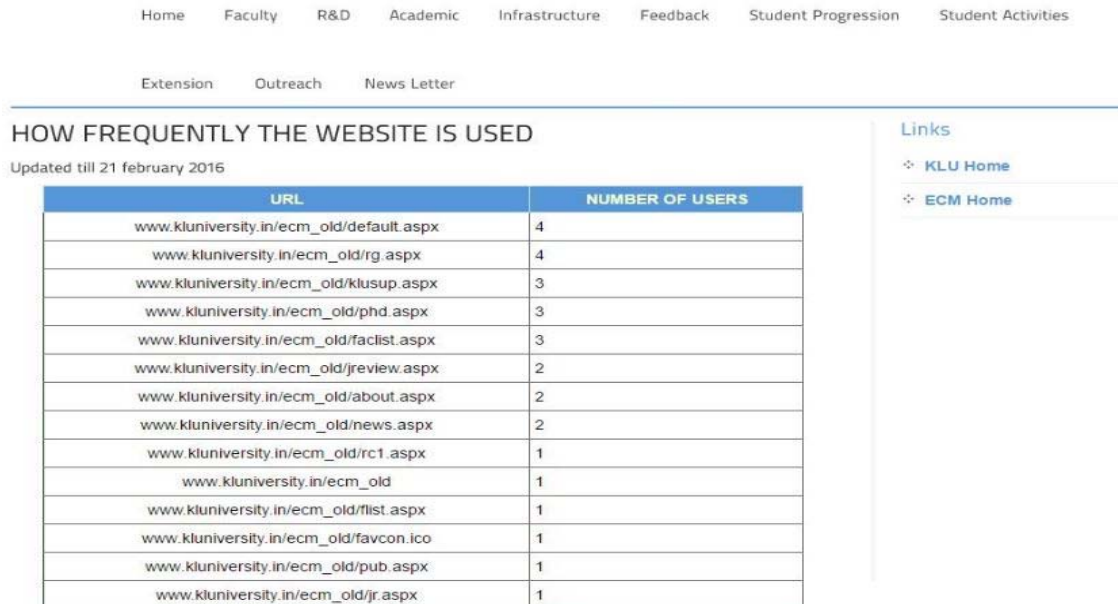


Figure-2. URL wise the frequency of WEB usage.



Figure-3. Average number of users using the WEB site.



Figure-4. Frequently used resource files.



Home Faculty R&D Academic Infrastructure Feedback Student Progression Student Activities

Extension Outreach News Letter

FREQUENCY OF NAVIGATION PATH

Updated till 21 february 2016

USER ID	URL	DATE	TIME
5000	www.kluniversity.in/ecm_old/phd.aspx	2016-02-19	09:51:45
5000	www.kluniversity.in/ecm_old/faculist.aspx	2016-02-19	09:51:47
5000	www.kluniversity.in/ecm_old/default.aspx	2016-02-19	09:51:50
5000	www.kluniversity.in/ecm_old/jreview.aspx	2016-02-19	09:51:52
5001	www.kluniversity.in/ecm_old/default.aspx	2016-02-10	08:51:45
5001	www.kluniversity.in/ecm_old/jreview.aspx	2016-02-10	08:51:49
5001	www.kluniversity.in/ecm_old/phd.aspx	2016-02-10	08:51:52
5001	www.kluniversity.in/ecm_old/klusup.aspx	2016-02-10	08:51:55
5002	www.kluniversity.in/ecm_old/pub.aspx	2016-02-18	10:51:45

Links

❖ [KLU Home](#)

❖ [ECM Home](#)

Figure-5. Extent of use of navigation path.

6. CONCLUSIONS

Poorly designed web sites are seldom used. One of the main reasons for poor business performance is due to poorly designed WEB site. WEB analytics plays a major role in improving the response and performance of a web site leading to a better business. Not all the web sites have the same features. A web analytic can be related to a feature of web site. A set of WEB analytics that are related to a WEB site can be selected based on the feature of the WEB site. One can adapt different strategies to improve the performance of the web sites by looking at their WEB analytics as every web analytic gives an insight to improve the WEB site.

REFERENCES

- [1] Catherine Dwyer, Yi Zhang, Starr Roxanne Hiltz. Using Web Analytics to Measure the Activity in a Research-Oriented Online Community, Tenth Americas Conference on Information Systems.
- [2] Daniel Waisberg and Avinash Kaushik. 2009. Web Analytics 2.0: Empowering Customer Centricity. 2(1).
- [3] Mohammad Amin Omidvar Vahid Reza Mirabiand Narjes Shokry. 2011. Analysing the Impact of Visitors on Page Views with Google Analytics. International Journal of Web and Semantic Technology. 2(1).
- [4] Kristin Glass, Richard Colbaugh. 2011. Web Analytics for Security Informatics, intelligence and security informatics conference (EISIC).
- [5] Bhavna Beri. 2013. Web Analytics: Increasing Website's Usability and Conversion Rate, international Journal of Computer Applications. 72(6).
- [6] Gurpal Singh. 2014. Web analytics a survey of its various methods and types used in various fields. International Journal of advanced Engineering Technical. 5(2): 100-102
- [7] Pawandeep Singh, Ayushi, Neha Dogra, Dr. Bhawna Sharma, Ms. Sheetal Gandotra. 2015. Web Analytics: Taking Thoughtful Decisions, Based On User Information and Behaviour. International Journal of Innovative Research in Science, Engineering and Technology. 4(7).
- [8] <http://www.tcs.com/SiteCollectionDocuments/White%20Papers/Web-Analytics,-Measurement->



Responsive-Websites-0215-1.pdf, Date accessed: 20th
April 2016

- [9] <https://www.infosys.com/manufacturing/resource-center/Documents/insights-driven-computing.pdf>,
20th April 2016
- [10] https://www.domodomain.com/press/IDC_WHITEPAPE.pdf, 20th April 2016