Understanding and Elevating the Web Practice through Web Analytics

Dr. Gurdev Singh
Gurukul Vidyapeeth Institute of Engineering & Technology Faridkot, Punjab-151203, India

Swati Chawla
Adesh Institute of Engineering & Technology Faridkot Punjab-151203, India

ABSTRACT

Web analytics is the measurement, collection, analysis and summarizing of internet data for purposes of understanding and optimizing web usage. To check the web analytics the website IJCT is considered. This paper is the first step to analyze the availability and performability of the site. Availability depicts that for how much time the system is up i.e. check the utilization of the system on which that site works. Performability is the method to analyze each and every module separately and clearly in an efficient manner. Availability and performability collectively check the reliability of the system to ensure the performance measurement of that system. Web analytics is not just a tool for measuring and depicting website traffic but can be used as a tool for business research and market research, as well as a means to check and improve upon the effectiveness and utilization of a web site .The tool used for the web analytics for this site is AWStats. AWStats is an open source Web analytics reporting tool, which used for analyzing the amount of data from Internet services such as web, streaming media, mail and FTP servers .This tool check the performance of the IJCT separately by dividing one month module to days in month, week in month ,hours in month ,contribution of various countries, traffic generated etc. This paper deal with the securitization of IJCT for January month to indicate how many events and functionality occurs during surfing on this journal by various users on the Web. This critique helps us to identify the throughput of IJCT. Here it is clear that users do not actually read: instead, they download and upload the data. A study of different online journals defines that this website scored massive response instead of the short time span. It completely and efficiently elaborate the usage of the IJCT by stating user visited, number of visits, number of hits ,bandwidth and many more.

KEYWORDS

Unique Visitor, Hits, Bandwidth, Performance, Analysis.

1. INTRODUCTION

Web analysis is the building block of every website because the existence of any site wholly depends upon it. As we know there are certain software development phases to find out an accurate and desired product to maintain the performability of that product. Similarly for website analysis certain set of tools are necessary to analyze the performability of site. Each and every website goes through this analysis phase before finding out a feasible and optimal solution. For this purpose Web analytic software is used. Web analytics (WA) is that software that you can use to track and analyze how people use your web sites and applications [1]. WA deals with the reliability, availability and throughput of a particular system on which that website works.

These three factors construct a system a real system which responds to each and every action occurs on that site. It mainly works with website to set and store up to five custom properties on all tracking events (user visits, number of visits, pages, hits, action events, etc.) on definite bandwidth. Custom variables can be set for a single tracking request, across all requests that make up a single session/visit, or across all visits for a single visitor [1].Web analysis is required because there is too much uncertainty without it. Marketing without metrics is lazy and probably ineffective. Metrics deals with the measurement of each and every part in a well efficient manner. It is a valuation technique in which a recently design asset is used to determine the value of a similar asset by comparing its all modules. It saves you from wasteful and unnecessary time spending in the long run. It enables you to explore multiple facets of internet surfing with confidence. Along with this, it provides you the ability to understand site utilization and help you to achieve your goals faster.

A full and complete log analysis is done by using AWStats tool which shows you the following information in a brief manner.

1. Number of visits: Number of visits and number of unique visitors are considered which describes the type of users.

2. Visits duration and last visits: Visits duration and last visits which indicate the time consumption by all users.

3. Authenticated users, and last authenticated visits: Authenticated users, and last authenticated visits which define the approval of type of visitors visits the website.

4. Days of week and rush hours: Days of week and rush hours (pages, hits, KB/MB for each hour and day of week) fully analyze the site by stating number of pages visited and action performed.

5 .Domains/countries of hosts visitors: Domains/countries of hosts visitors (pages, hits, KB, 453 domains/countries detected) to show the involvement of users around worldwide.

6. Hosts list, last visits and unresolved IP addresses: Hosts list, last visits and unresolved IP addresses list are shown here to states the popularity.

7. Most viewed, entry and exit pages: Most viewed, entry and exit pages are shown which declares the start and end point of each and every user while undergoes resource utilization of IJCT.

8. Files type: Files type is mentioned here to discuss the format of every downloaded and uploaded file.

9. Web compression statistics: Web compression statistics takes place by using AWStats web analytic tool.

10. Operating System (OS): OS used (pages, hits, KB/MB for each OS, 10 OS detected) indicates the type of interface used between user and hardware.

11. Browsers: Browsers used (pages, hits, KB /MB for each browser, each version) defines the list of various browsers. A web browser is a software application for retrieving, presenting,

and traversing information resources on the World Wide Web [2].

12. Visits of robots: Visits of robots detected the number of viewed or not viewed by visitors.

13. Search engines: Search engines, key phrases and keywords used to find your site. The search engine depicts certain set of data. Some of them which are detected are cloud, traffic, technology etc.

This tool is used to calculate the utility rate of the website IJCT. The International Journal of Computers & Technology (IJCT) has been running online since December 2011. It accosts original research papers addressing theoretical and practical implementations in information system applications. Along with this its complete performance is analyze with the help of various browser applications, operating system, through navigation etc by number of user's response [3].

The above written factors construct the performance measurement of the website IJCT. These are the fundamental considered for this site behind the metrics to detect the performance.

2. RESULT AND ANALYSIS

By using the Online web analytic tool AWStats, the whole performance and functionality is managed and analyzed effectively and specifically. The whole analysis of IJCT is based upon type of data used; number of visitors visit the site, type of visitors, navigation behavior etc. They explained the site in a very crucial manner. It is clearly stated and shown below:

Antiquity of IJCT

Antiquity deals with the type of data and various set of changes involved in the past month January.

This part is analyzed by defining five factors:

1. Unique Visitors (UV): UV relies on user entries. This portion focuses on quantity. More the number of UV more will be the performance criteria**.**

2. Number of visits (NV): NV indicates the participation of users depending upon the visitors. It is define as

UV=Visits/Visitors

3. Pages: This dictates the number of pages visit by each and every user. It judges the involvement of visitor on every click.

4. Hits: The action takes place on the site while surfing is considered as hit.

5. Bandwidth (BW): BW deals with the data rate over the network. The bandwidth is usually expressed in bits (of data) per second (bps).Here the bigger unit MB (Mega Byte) is considered to define the range.

1 Byte=8 Bits

MB= 10^6 Bits

2.1.1 Monthly

This shows the performance and usage of IJCT for 30 days in a combine module. It shows the utilization rate (UR) by already mentioned factors. UR checks the utility of action performed by all users. The bandwidth is used in MB.

2.1.2 Days in month

During this month the value of NV visited in this site as per the day is mentioned below. As far as this detail is concerned it shows that the number of visitors increases day by day. This increases the whole scenario. This gives the per formability of

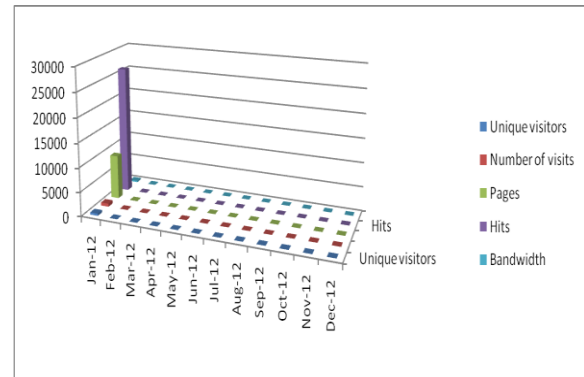website day wise in a month to determine the deep understanding of the IJCT.

Figure.1-Monthly history

The bandwidth has MB units. It completely shown in the graph mentioned below:

These all features collectively generate an output to shows the variation in every field even it may be UV, hits generates or bandwidth. It is shown in the graph mention in Fig. 2

2.1.3 Days in week

The deep understanding of the performance analysis is checked out by mentioning feasibility study of a week. Feasibility study means to say proper testing of numb of hits takes place on how much number of pages and under what bandwidth. Bandwidth is considered to show the proper range of signals utilize during the completion of various set of processes. It shows the week wise analysis of IJCT which defines the involvement of each visitor with the website. The used bandwidth is in MB.
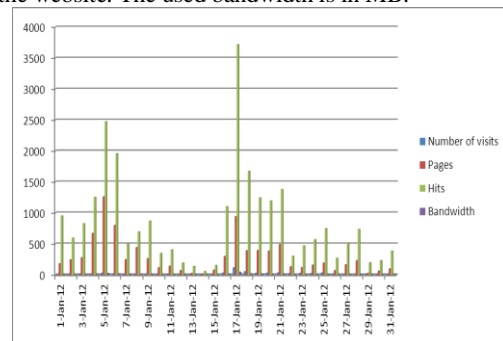

Figure.2- Days in month

2.1.4 Hours

This implies the performance measurement on the behalf of hourly rate of this respective month. It indicates the functionality of the website per hour rate to verify the ability as well as resource utilization of IJCT for deep review. This is as same as we are already discussed in above modules i.e. it also utilize UV, Hits and Bandwidth as a major portion.

2.1.5 Traffic

Traffic indicates the performance of a website over the network. Network traffic is data in the network for each and every website and the data is encapsulated in network packets. These packets contain whole information regarding the data which is under consideration.
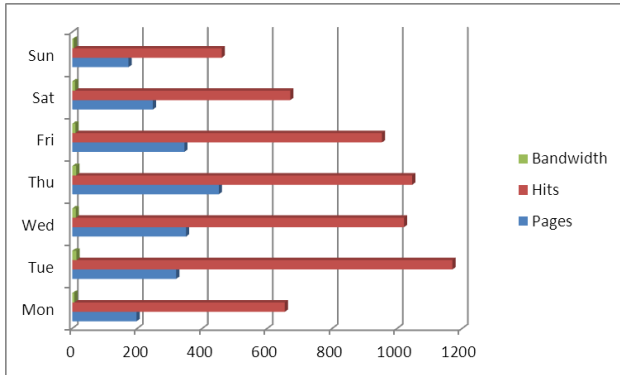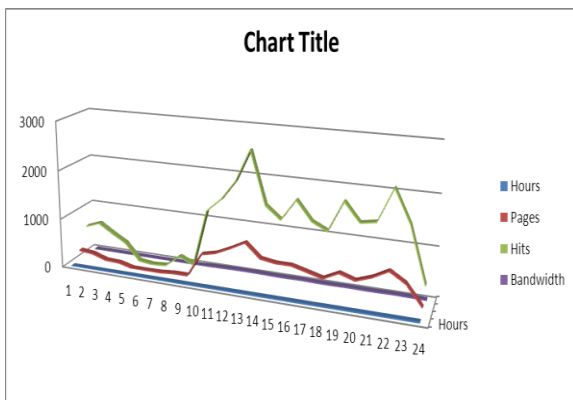
Figure 3-Days in week

Figure 4-Hours

This complete process is done with the response of users. Traffic generates or terminate will effects the performability of the IJCT. More will be the traffic more will be the popularity and more will be time consumption for resource management and analysis is difficult but the comparison is easy and vice-versa. Here the classification of users takes place by indicating number of visits as shown in Table 1 which is viewed or not viewed.

	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Viewed traffic *	453	649 (1.43 visits)	9,042 (13.93 Pages)	26,193 (40.35 Hits)	285.43MB (450.36 KB/Visit)
Not viewed traffic *			1,317	8,673	19.22 MB

Table1. Traffic

Not viewed traffic includes traffic generated by robots, worms, or replies with special HTTP status codes.

In this traffic is calculated by indicating the number of page visits by various users and their responses under particular region of bandwidth i.e. 285.43 MB.

2.2 Retort Observation
2.2.1 over the world
This website compares with other websites on the basis of the performance recorded by AWStats web analytic tool. This information is based on the actions of web users all over the world (unlike the hit-wise data often quoted in the WWW), but the web users whose browsing produces the rankings are a self-selecting group, so this site becomes popular among number of countries all over the world. Some websites are probably more popular than their web analytics tool rankings which indicate traffic likely comes because of web users. However, it should also be noted that IJCT becomes more popular than other websites due to its performability, availability and throughput. These performance measures generate a report which defines the accurate and exact performance of the site. Some of the countries which undergo takes part in the popularity of the IJCT are India, United States, Canada, Germany Norway, Poland, Japan etc.
2.1.2 Host
In operating systems, the term terminal host traditionally denotes a multi-user computer or software providing services to computer terminals, or a computer that provides services to smaller or less capable devices [4]. A host can be a node on the Internet. A computer that serves pages for one or more Web sites totally depends upon the host functionality. Along with this the throughput and reliability of a website relies on the number of visitors visit that site and what type of utility is performed by them. Host indicates the various set of IP addresses from where the user visits the site IJCT. It describes resolved as well as unresolved users /visitors. Along with this it also implies the various pages involved during the searching of any content or while downloading or uploading the data with the detail regarding last visit visited by visitor and number of actions performed by them under certain range of bandwidth.
2.1.3 Authenticated Users (AU)
Authenticated users are those users whose visit is recorded during the occurrence of any event in the website. The set of pages consider under particular bandwidth and respective responses are considered in AU. It discuss the set of various logins from where the visitor login. AU effectively mentions the popularity among resolved users whose complete functionality and performance measurement has to be calculated. The vital response is shown in Table 2

Authenticated users : 0	Pages	Hits	Bandwidth	Last visit
Other logins (and/or anonymous users)	9,042	26,193	285.43 MB	

Table 2-Authenticated users

2.1.4 Non- Authenticated Users (NAU)
Non- Authenticated users are those users whose visit is not recorded during the occurrence of any event in the website. These NAU are considered as Robots/Spiders which show hiding of data. Some of the set of hits takes place under

particular bandwidth and respective responses along with their last visit are shown in the table mentioned below in Table 3

9 different robots*	Hits	Bandwidth	Last visit
Googlebot	1,254	17.89 MB	31 Jan 2012 - 22:24
Unknown robot (identified by 'bot*')	50	340.88 KB	30 Jan 2012 - 14:43
Unknown robot (identified by empty user agent string)	16	58.76 KB	21 Jan 2012 - 11:44
Unknown robot (identified by '*bot')	4	150.21 KB	14 Jan 2012 - 19:15
Unknown robot (identified by 'spider')	2	27.13 KB	28 Jan 2012 - 12:40
Unknown robot (identified by 'crawl')	2	0	27 Jan 2012 - 01:24
Alexa (IAArchiver)	1	0	08 Jan 2012 - 19:55

Table 3-Robots/Spiders visitors

2.5 Navigation

The importance of navigation structure cannot be over-emphasized. Without some sort of navigation, a site loses all sense of structure and organization [5]. There are lots of ways for presenting the navigation. Here we discuss the navigation by describing the time utilization by user during surfing on the website, the type or format of files involved in it, quantity of data downloaded from the site IJCT, the number of pages visited by the UV, type of operating system and last but not the least type of browser from which the site surfing starts. It is important to allow access to the rest of your site from anywhere. Ideally, you should be able to go to any page in a maximum of two or three clicks, with your main pages accessible in one [5]. Navigation deploys the behavior of UV as well as IJCT depending upon the below mentioned factors:

2.5.1 Visit Duration (VD)

VD discusses the time complexity i.e. it deals with the process of checking time function with effect of UV. Visit Duration's outcome depends upon the number of antiquity factors along with their percentage as shown in the Table 4

Number of visits: 649 - Average: 668 s	Number of visits	Percent
0s-30s	253	38.9 %
30s-2mn	79	12.1 %
2mn-5mn	86	13.2 %
5mn-15mn	91	14 %
15mn-30mn	47	7.2 %
30mn-1h	52	8 %
1h+	41	6.3 %

Table 4-Visits duration

2.5.2 File Type (FT)

File Type mentions the types of various data files involved during surfing on the site while uploading, downloading or any action performed by the visitor. Some of the list of files is mentioned below:
1. css (Cascading Style Sheet file)
2. js (JavaScript file)
3. jpg, gif, png
4. HTML, PHP, HTM
5. pdf doc
These all shows the number of hits performed by all users in their corresponding bandwidth range along with their respective percentage value.

2.5.2.1 Download Data (DD)

Files may be of any format or type but the vital role played here is downloading. DD shows the amount and quantity of data downloaded on IJCT. It evaluates the number of events fired during the site navigation. Along with this it also defined the size of each and every hit occurred with their respective bandwidth. Bandwidth shows the amount of data carries from one point to another in a given time period.

2.5.2.2 Viewed Pages(VP)

Viewed Pages deal with page URL (PU). PU deploys the number of various files involved during searching on IJCT for relevant output. It specifies the different and distinct entries through URL.
URL is Uniform Resource Locator/ Universal Resource Locator is a specific character string that constitutes a reference to an Internet resource [6]. By using URL it is quite easy to access any data easily and efficiently. VP mentions the number of pages viewed by visitors, their respective size along with the entry and exit information of each & every visitor. The counted number of viewed pages is nearly 700. It's a big figure which clarifies the popularity of IJCT.

2.5.3 Operating System (OS)

An **operating system** (**OS**) is a set of programs that manage computer hardware resources and provide common services for application software. The operating system is the most important type of system software in a computer system. A user cannot run an application program on the computer without an operating system, unless the application program is self booting [7].OS acts as interface between user and hardware. Here user is UV

and the hardware is on which UV works. A list of various OS is clearly shown in the graph which mentioned the popularity of distinct OS during resource utilization of IJCT in the form of %age by comparing with the number of hits performed by each UV.
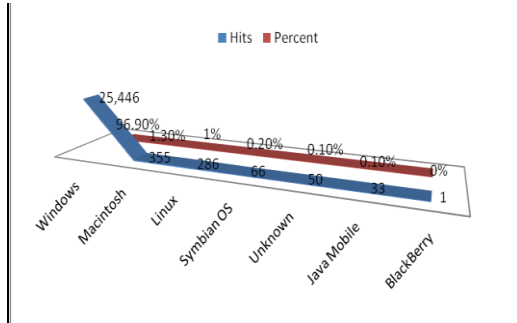

Figure.5-Operating Systems

2.5.4 Browsers

Browser acts as major tool for the web surfing. It deals with the processing of site which further depends upon the browser used. The below mentioned data show both authorized as well as unauthorized browsers. Along with this it generates the percentage of performance analysis for IJCT. Browsers indicate resource usage on the behalf of hits and their respective percentile. It used the word grabber to explain the utilization of various browsers. Grabber stated as if site had a visit from this browser which was listed then the value is positive means yes otherwise negative i.e. no. A list of non-grabber browsers are:
1. Google chrome, Firefox, MS Internet Explorer, Opera, Safari, Mozilla
2. Nokia Browser, Android Browser, Samsung Browser (These are PDA/Phone Browsers)

2.5.4.1 Connect to site

The **Internet** is a global system of interconnected computer networks that use the standard Internet protocol suite (TCP/IP) to serve billions of users worldwide [8]. It fully defines the utilization of bandwidth. Connection plays vital role in each and every task over the Web. Internet is a *network of networks* that consists of millions of private, public, academic, business, and government networks, of local to global scope, that are linked by a broad array of electronic, wireless and optical networking technologies [8].Connection indicates the interlinking of two or more units to generate one feasible output. Internet elaborates the connectivity through the network over the World Wide Web. This module acts as starter to IJCT as well as describes the linking from where every UV connected to the site. As well as it calculate the number of pages viewed and number of hits takes placed by UV in the percentile form.

2.5.4.2 Search keywords

It is defined as that type of search which looks for matching documents that contain one or more words specified by the user. The search engine is used to find out the desired data. Here search engine is used to search many words observed while surfing on IJCT. Some of the main words are
1. Cloud, Computing, Computer
2. Traffic, Technology
3. Journal
4. Website

It shows large number of search words whose approximate value is around 110

2.5.4.3 Search Key phrase

Words combine to make phrases, and phrases are one of the basic patterns out of which we build sentences [9]. This part conveys the various phrases involved during the utilization of the site. A phrase is an **expansion** of one of the words inside it. A phrase can be two words long [9]. Some of the searched phrases are cloud computing, web analytics, traffic metrics etc.

3. CONCLUSION

Web testing plays a vital role in the field of performance analysis. A performance analysis is generally called for when you want to improve a part of the organization (may be any website) or to fix a problem that someone has brought forth. Both are generally fixed in the similar manner. When performing an analysis phase, it is best to take a long term approach to ensure that the performance improvement initiative ties in with the organization's vision, mission, and values. This connects each need with a metric to ensure that it actually does what it is supposed to do. This is best accomplished by linking performance analysis needs with Web Analytics Tool AWStats which effectively checked the various modules via. Pages, UV, Hits, Bandwidth.

IJCT has been similar to most other Web usability work sites for journal publication and users have mainly looked at site architecture, navigation, page design, layout, search, graphic elements and style, and icons. Even so, this website has collected large number of user responses about the content during this hard and long series of competition and advancement. Responses can be calculated by undergoing performance analysis.

4. REFERENCES

[1] http://www.openwebanalytics.com

[2] http://en.wikipedia.org/wiki/Web_browser

[3] www.ijctonline.com

[4] http://en.wikipedia.org/wiki/Host_(network)

[5] http://en.wikipedia.org/wiki/Uniform_resource_locator

[6] http://en.wikipedia.org/wiki/Operating_system

[7] http://en.wikipedia.org/wiki/Internet

[8]http://www.phon.ucl.ac.uk/home/dick/tta/phrases/phrases.htm