

Analytics For Content Based Sites – A Comparative Study

Dr Veena Dhari
Department of Computer Application,
AISECT University
Bhopal , Madhya Pradesh -400071, India

Meenakshi Garg
Department of MCA,
Vivekanand Institute Of Technology
Mumbai, Maharashtra 400071, India

Abstract— *Examination of innovations that mine vast measure of organized and unstructured information to pick up bits of knowledge are turning out to be progressively essential to organizations. Web investigation for the most part concentrate on e-trade locales, where the visits need to unite in a buy however the conduct of e-learning situations clients is driven by data gaining. In this paper we are going to break down a content based site avatto.com utilizing two investigative apparatuses Google analytics and statcounter.*

Keywords: *Web Analytics, Google Analytics, analytics for online education, statcounter, Online Analytical Tools*

I. INTRODUCTION

Every time an asset facilitated on a web server is asked for on the web [1], a record is composed in the web server log. An asset can be a HTML document which is ordinarily rendered on the client's web program as a page. Then again it can be a picture, a sight and sound document or other media that can be gotten to by a web program or some other web customer, for example, a crawler. For every asset asked for, the web server records the IP location of the web customer, date and time of the solicitation, the HTTP convention utilized, the name of the asset, the HTTP convention status of the server reaction, the measure of the reaction, the ID of the web customer and subtle elements of treats utilized. There are two methods for breaking down the movement to a site; either by utilizing web server logs or page labeling.

The main technique [6] utilizes a product called a web server log analyzer or simply log analyzer, to prepare the web server logs and give an account of different attributes of the activity, for example, the IP location of the web customer, pages asked for, recurrence of solicitations, and so forth. The second technique includes including a few lines of JavaScript code to "tag" all pages that we need to examine. Every time a labeled website page is gotten to, the JavaScript code sends data [8] about the entrance to a remote server which totals the information and provides details regarding key parameters. For the analysis of content based websites in this paper second technique is used. Although Google Analytics visitor counter is invisible to the users of site where as Statcounter visitor counter may be shown to the visitors.

II. BACKGROUND

Web Analytics According to the Web Analytics Association, Web Analytics (WA) [4] is the estimation, accumulation, investigation and reporting of Internet information for the reasons of comprehension and upgrading web utilization the estimation of site activity is finished by utilizing web investigation measurements. The principle bodies who have info in characterizing web investigation measurements are Jicwebs (Industry Committee for Web Standards)/ABCe (Auditing Bureau of Circulations electronic, UK and Europe), The WAA [7] (Web Analytics Association, US) and to a lesser degree the IAB (Interactive Advertising Bureau). Both the ABCe and the WAA give records (i.e., a Report and a Draft for Public Comment) that contain arrangements of such definitions.

Google Analytics (GA) [3] is a free web examination instrument offered by Google, being the most generally utilized web investigation instrument.

Google Analytics is utilized by including a piece of JavaScript code [3] that the client includes into each page of site. These page labels are utilized to gather the guest information and send it back to Google information accumulation servers for handling. Additionally, GA sets first gathering treats on every guest's PC so as to have the capacity to figure out whether the guest has been to the site some time recently.

The principle GA [7] confinements are trademark to web examination instruments that gather nearby guest information utilizing page labeling:

- Blocking JavaScript code. This keeps movement and clients from being followed, and prompts uncollected information.
- Deleting or blocking treats. Such activities lead to error as returning guests can't be followed.

StatCounter [2] is a web movement investigation apparatus. Access to fundamental administrations is free and propelled administrations can cost between US\$5 to US\$119 a month The organization is situated in Dublin. The measurements from StatCounter are utilized to register web utilization offer for

instance. StatCounter is utilized on 3% of all sites. StatCounter measurements are specifically gotten from hits (not interesting guests) from 3 million destinations utilizing. StatCounter totaling more than 15 billion hits for each month the organization was established by “Aodhan Cullen” at 16 the age of years. He remains the CEO of StatCounter. Cullen got the “Web Hero” grant at the 2008 Eircom Golden Spider Awards. He was likewise named the 2007 Business Week Young European Entrepreneur of the Year. In 2008, the organization surpassed two million individuals all around the globe.

II. MAIN CHALLENGES

Information preparing in web examination begins with deciding interesting guests and visits.

A. 3.1 Unique User Identification

One noteworthy test in web investigation is to recognize unique users[7]. One strategy is to recognize them in light of their IP addresses and the User Agent. An option is to utilize treats. Hence, Google Analytics and other web examination instruments use them to decide novel guests. Treats are utilized on the grounds that IP locations are not generally extraordinary to clients and might be shared by expansive gatherings or intermediaries. On the other hand, there are different circumstances in which both of these (i.e., IP + User Agent and treats) routines are incorrect.

B. Multiple IP locations - Single Visitor

A person that gets to the site from distinctive areas/gadgets will have [7] diverse IP addresses (separately diverse treat ID) from visit to visit and consequently will be numbered more than once. This makes following rehash visits from the same client troublesome.

C. Multiple User Agents - Single Visitor

A client that uses more than one browser, even on the same machine[7], will show up as numerous clients. Besides, cookies can be erased or blocked. The most precise method is to use registered user account data keeping in mind the end goal to recognize people, particularly for e-learning as most online instructive frameworks use client verification. This arrangement is the most reasonable one and can be actualized just in an incorporated framework.

D. Visit/Session recognizable proof

Distinguishing exact visits is not a minor errand[7]. That is basically on the grounds that HTTP convention is stateless and connectionless. In this way, it is for all intents and purposes difficult to decide when a client is counseling the webpage or going by different destinations or if really leaves the site. Additionally, some ISPs or protection apparatuses arbitrarily allot every solicitation from a client to one of a few IP addresses. Albeit uncommon, in these cases, a solitary server session can have different IP addresses. There are three primary heuristics that are by and large used to decide the visit end:

- Temporal heuristics that limits the length of time of the whole visit to a predefined upper bound (generally acknowledged as 30 minutes)
- Heuristic taking into account the navigational examples where all pages inside of a session must be connected specifically or in a roundabout way (i.e., the page has a referral that is a page got to beforehand in the present visit).

The initial two techniques, in spite of the fact that they may deal with ecommerce destinations, [8] where the customer makes the buy and leaves the site, it is not material to e-learning as the learning process requires some investment.

III. OBJECTIVE

Aim of this paper is to mainly focus on following areas

- A. To compare number of visitors and page view over a certain period of time
- B. To compare unique visits and returning visits using web analysis tools.
- C. To compare different browsers using which website is browsed
- D. To study the impact of National and public holidays on content based sites.

We intend to analyze tools on many areas but we will focus mainly on the above areas.

V. WEBSITE ANALYSIS

This paper displays the outcomes got by deciphering the same web use information by two diverse web examination instruments. On one side there is an examination instrument, called Google Analytics [3] and the other one is Statcounter. Both tools require simple JavaScript to be inserted in web pages.

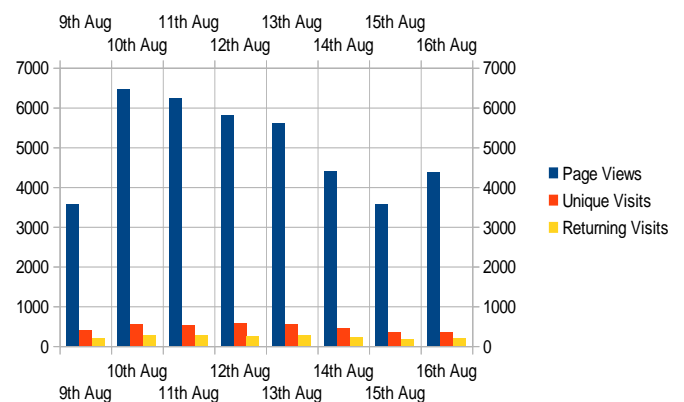


Fig 1 Statistics from Google Analytics showing page views

For analysis purpose one week data from 9th Aug 15 to 16th Aug 15 is being considered. Fig 1 shows number of unique page views, page views, and returning users starting from top.

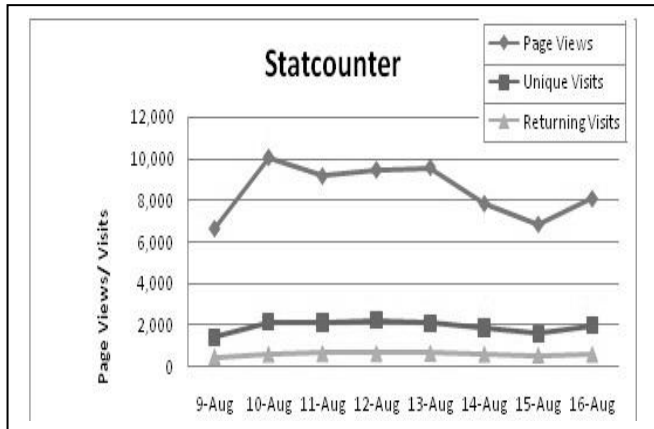


Fig 2 Statcounter Analysis

Period from 9th August to 16th August was chosen as we intend to analyze the impact of National and public Holidays on content based sites.

IV. OBSERVATIONS

It can be seen that there is sharp decrease in website traffic on national and public holidays. One can observe that there is 90% increase in number of page views in Google Analytics from 9th Aug 15 to 10th Aug 15 and there is an increase of 33% in Statcounter.

TABLE I. Browser Configuration in Google Analytics

| Google Analytics | |
|-------------------|------------|
| Browser | Percentage |
| Chrome | 55.79% |
| Firefox | 16.15% |
| Opera | 6.25% |
| Opera Mini | 5.71% |
| UC Browser | 5.38% |
| Internet Explorer | 3.88% |
| Safari | 3.54% |
| Android Browser | 2.61% |
| (not set) | 0.13% |
| Nokia Browser | 0.13% |
| Coc Coc | 0.12% |
| BlackBerry | 0.09% |
| Maxthon | 0.09% |
| YaBrowser | 0.03% |
| CoolNovo | 0.01% |

It can also be seen that number of unique visits seen by Google

Analytics are much lower than as seen by Statcounter. Similarly number of returning visits is also shown higher by Statcounter.

Table 1 shows data [9] from Google analytics and shows that more than 50% of users browse website using Google Chrome. It gives cumulative view about Google Chrome, but it does not talk about various versions of Chrome. Table 2 shows data [10] from various versions of browsers. It can be clearly seen that Chrome stands out in a crowd in this report too. Further it can also be observed that after Google Chrome Firefox takes place in both tools and Internet Explorer being at third place.

TABLE II. Browser Configuration in Statcounter

| StatCounter | |
|--------------------|------------|
| Browser | Percentage |
| Chrome 47.0 | 59.80% |
| IE 11.0 | 9.60% |
| Opera 12.1 | 7.80% |
| Firefox 43.0 | 7.20% |
| IE 10.0 | 3.20% |
| Firefox 11.0 | 2.40% |
| Opera 34.0 | 1.60% |
| Firefox 44.0 | 1.60% |
| UC Browser 8.7 | 0.80% |
| Safari 7.0 | 0.80% |
| Opera 0 | 0.80% |
| Firefox 41.0 | 0.80% |
| Firefox 39.0 | 0.80% |
| Chrome for Android | 0.80% |
| Chrome 48.0 | 0.80% |
| Chrome 46.0 | 0.80% |
| Firefox 34.0 | 0.40% |

V. CONCLUSION

Although Statcounter shows more number of page views, unique page views, and visitors, but both tools show a decline in traffic on national and public holidays. Both tools show Google Chrome as the first choice of users for accessing content-based sites. Further, Google Analytics shows data only for one version of browser, whereas Statcounter shows data for different versions of browsers, but Google Analytics maintains data for a wide variety of browsers.

This paper demonstrates that incorporated examination is more precise and accordingly more dependable than legacy investigation, which for this situation is spoken to by Google Analytics. Notwithstanding while utilizing the same web utilization log information, the contrasts between the two methods are noteworthy. These errors are significantly more intense as the investigation is done on an e-learning framework.

which is a substance based site, where the guidelines of e-business, for which Google Analytics is primarily utilized. Statcounter offers a more sensible in this manner solid elucidation by utilizing web data. As a parallel research, the creator examined in the contrasts in the middle of investigation that statcounter recognizes guests by taking into account the IP and client operators. The outcomes acquired by this instrument, albeit most exceedingly awful than Google Analytics. As a future work, the creator needs to concentrate on the unique user identification based on substance and hence measure the adequacy of individual pages. Additionally creator notices by looking at traffic trend and browser specific navigation both tools produce similar traffic trend and similar browser analysis.

VIII. REFERENCES

- [1] J.J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, “Web usage mining: discovery and applications of usage patterns from web data,” ACM SIGKDD Explorations Newsletter, vol. 1, no. 2, pp. 12–23, January 2000.
- [2] Wikipedia, “<https://en.wikipedia.org/wiki/StatCounter>”
- [3] Wikipedia, “https://en.wikipedia.org/wiki/Google_Analytics”
- [4] D. Waisberg and A. Kaushik, “Web analytics 2.0: Empowering customer centricity,” SEMJ.org, vol. 2, no. 1, 2009.
- [5] C. Romero and S. Ventura, “Educational data mining: A survey from 1995 to 2005,” Expert Systems with Applications, vol. 33, no. 1, p. 135146, July 2007.
- [6] N. K. Tyagi, A. K. Solanki, and S. Tyagi, “An algorithmic approach to data preprocessing in web usage mining,” International Journal of Information Technology and Knowledge Management, vol. 2, no. 2, pp. 279–283, July-December 2010.
- [7] Web Analytics Association, “Web analytics definitions,” Web Analytics Association, Draft for Public Comment, 2008.
- [8] The Joint Industry Committee for Web Standards (JICWEBS), “Reporting standards. website traffic.” Auditing Bureau of Circulations electronic (ABCe), Report 1, 2011.
- [9] Google, “Google analytics,” <http://www.google.com/analytics/>, viewed on August 2015
- [10] Statcounter “Statcounter, <http://www.statcounter.com> viewed on August 16