Web Analytics
vs.
Log File Analysis

It is possible that analytics users have used different tools to audit and monitor the visits to their websites. Such processes can be carried out to audit the reports and test their accuracy. This normally leads to different results and the purpose of this document is to illustrate why this happens.

Why doesn't web analytics match with log files?

When comparing web analytics results to the log files it is important to know why certain discrepancies (sometimes really huge) come up in the reports. The reason for such discrepancies depends on the metrics the analytics are using which are different from the metrics being used to analyze the log files.

The main difference between the two is that Analytics uses client-side data to gather the information while for log files these contain server-side information. This important distinction leads to totally different results. In fact it is important to analyze the data with respect to how the data was gathered.

Visits

Analytics (most mainstream web analytics) rarely track 100% of visitors. There are various reasons for this like browsers that have JavaScript disabled and cookies that are deleted. Others can block access to analytics or connections to the analytics web server may be lost. Thus results that come out from web analytics have to be considered like a survey. Such losses would normally be valid for about 10% of websites visits.

On the other hand Analytics does not track bots and spiders. Log files will track all hits, and all files that are requested, regardless of who requests it. But one must note that if the user agent is logged in the log file than it is possible to filter out such requests. This is why it is important to use a good tool to analyze the log files.

Pageviews

Log files log requests to all files and media requested by the client. On the other hand analytics are only executed for html content. One must consider that browsers do keep a cached version of the page to their local caches. Thus sometimes when a browser reloads a page (like when the user presses the back button) the browser will most probably load the page from its local cache other that requesting it from the server. Hence for this the call does not get logged.

However analytics records even cached pageviews because the tracking cookie executes every time a page is displayed. This obviously leads to duplicate entries and additional page views. It is very common for somebody to refresh the page. Because of these variables, the number of total visits and pageviews will never match between analytics and log files.

Visitors

Analytics uses cookies to differentiate between a visitor and another. The cookie holds a unique visitor ID so that the analytics can tell whether a visitor has been to the site before. This also allows it to tie pageviews throughout a visit to a single visitor. But one must not forget the users that clean their cookie

cache when they close their browser. Such users will have each visit recorded as a new visit each time since they will be using a new tracking cookie.

Log file analyzers on the other hand cannot use such information and have to rely on the Internet Address and the browser.

Traffic Sources

Log files can determine the referring site for a visit most of the time. It is stored with the first page of their visit. This referrer can he further analysed to distinguish if it was a website, a paid advertisement or a search originating from a search engine. It is even possible to extract the search terms used. Together with this it is also possible to find the geographical country of origin by inspecting the Internet Address.

Analytics uses cookies to store the traffic source information. With this, the Internet Address and referrer they can also extract all the information listed above.

Privacy

Privacy is quite a debated topic; it is quite obvious that when a 3rd Party is involved (as with the case of analytics) there can be privacy concerns. Log files on the other hand if secured do not pose any of these risks. This is why some users delete their browsing history and cookie cache on browser exit.

Web Analytics never match Log Files Analysis

These are some of the most common reasons why analytics reports don't match up with log file reports.

Log files are literally raw files which need initial processing to be used for web site analytics. This is why it is important to use the correct tools when analyzing log files and not to treat them as simple click counters. Log files are files filled with important information, which if used correctly can help you improve and fine tune your website to be more efficient.