

Mid-Term Research Paper
Web Analytics

Todd J.

Barry University – Prof: Michel Sily
IT-338 – Web Design Authoring & Publishing
Wednesday, September 23rd 2009

Web Analytics is one of the most comprehensive tools available on the internet to gauge the quality and level of traffic to a website. This article will discuss the history to present of web analytics as well as the basic functioning and technical aspect of its operation. It will primarily focus on Google Analytics as the primary means of discussion.

Web Analytics basically describes the research and analysis of internet traffic. It provides a means for the web site owner or host to have a real clear picture as to the effectiveness of their site. Specifically, it allows the customer to get a better idea of exactly what it is that draws customers to their site, and which pages get the most amount of traffic. While this information might normally be an otherwise obvious piece of information to the site owner, it can be invaluable to websites that maintain a lot of similar or like information. Normally, an e-Commerce site (a site which sells goods and services) would know what products are most popular by sales numbers, and therefore know which products are less popular. For websites that don't provide a sellable product, but instead indirectly provide a service, this information is very important.

Before the onset of web analytics as we know it today, most sites had little, if any way to tell if their site was popular. The earliest method of determining this was with the use of a visitor counter. The website counter was nothing more than a simple numeric counter much like you would find when walking into a stadium after first entering through the gates. Each person walking through the counter would cause the counter to increase by one and thereby adding to the number of people in attendance. The "Website Visit" counter as it was known, worked on the same principle. Some were eventually designed to count only "unique visits", while others were programmed to count every time the page was loaded. Unique visits gave the site owner a clearer picture as it only counted visitors based on the specific IP address used at the time. These counters worked great, but they never really gave the host a long-term understanding of their website goals. Certainly a large number would show the site was popular, but what would it say about the number of visits per week, or even per day? This is where true web analytics began.

Early web analytics companies were divided on the best methods of obtaining analytic data. There were really two schools of thought when it came to obtaining the “numbers”. The first method was through the use of browser tags, and the other was through the use of web logs. As much as this is true today as it was in the past, the idea of web analytics wasn’t simply for the benefit of the customer or owner of the website, it was also, but primarily for the benefit of the search engines. The search engines used this information to determine priority and relevance in the search criteria. As such, this information was equally critical to them as it was for the customer of the site. Each method had its own importance to the search engines.

A *Browser Tag*, is the technical description for the encapsulating syntax used in HTML code. This syntax would be used to wrap certain pieces within a web page to either aid in formatting, or for definition. In the case of web analytics, the tags analyzed most frequently were the header tags. In the early progression stages of the internet in the public’s hands, there were a very basic set of standard formatting tags. These tags were used to define certain criteria on the web page. In this example, the <h1> tag was almost always used as the primary header of a web page. The <h1> tag defaulted automatically with a stronger, bolder font, with more size than any of the other default tags. This tag would undoubtedly be used primarily as the title of the webpage, or to expand on what was contained in the <title> tag. Each successive <h#> tag would have a progressively decreased level of importance. The search engines felt this information was critical for determining the relevance of its searches and indexing them appropriately within the database. It was through the use of tags that the search engine could remotely determine the exact content of the web site.

The second most critical piece to the customer and the search engine companies, was the use of web logs. The web log is a general term used to define rudimentary information that’s stored either locally on the host machine, or on the individual search engine providers. The data would be similar, but used for two entirely different things. The search engine would build its own web logs based on the recorded usage of its search engine. For example, if a person is searching for websites all about cats, the search engine would return a series of available websites. Every time an internet “surfer” clicks on one of the links from that search engine, it records a single tick, or “+1” on the counter for that website. Depending on which sites get the most clicks, it helps the website determine the resulting priority for displaying the list. If a link that was previously near the bottom became

more popular, it would likely receive more clicks. A result of this is that based on the formula used, the website would return at a higher position each time for that search result. This data is of course equally important to the website owner because it lets him/her know the ranking their website takes in a particular category, as well as the number of hits that it's received. For the website host, it gives a better indication of how much to charge the customer for the web hosting.

At some point through the next couple of years, a new breed of web analytics companies realized the need for incorporating both features formerly created by companies such as WebTrends and WebSideStory. This new generation of web analytics companies brought in a whole host of new technologies and web analysis. Gone were the days of simple visit counters. Websites now had complete information at their fingertips that would provide them a full spectrum of information on their site visits.

This new technology expanded upon individual visits. Most analytics companies required that scripting code be added into the website in order to properly facilitate the information without having direct access to the hosting company's visit logs. As is the case with Google Analytics, a code snippet needs to be added:

```
<script type="text/javascript">
  try { var pageTracker = _gat._getTracker("UA-7791171-1");
        pageTracker._trackPageview(); }
    catch(err) {}
</script>
```

This script gets embedded into every single page on the website. Having this script on each page now provides additional information like, visits per page, unique visits, number of additional pages viewed by unique viewer, location of visitor, source of visitor and bounce rate. With this information, even further calculations can be made to provide even more information, like time spent on page, percentage of new visitors per week versus existing visitors, and total time spent on site per visitor. As our society becomes more and more global, and our consumers become more and more internet savvy, it's critically important for most corporations to be able to reach their customers in this new marketplace. The information provided by these web analytics companies provides exactly that.

Part of what makes the information so critical is the old adage, “know your market”. Google Analytics, as well as other providers, have the ability to tell the customer exactly where their viewers are coming from. By storing and logging each IP address that accesses the website, the script can determine exactly from what country, what state, what city, and even what provider the visitor comes from. Also known as *Location Targeting*, this location information provides the marketing team with what they need to know about their customer so they can focus primarily most of their advertising resources.

This same theory is equally important for the marketing team when it comes to what site or referral sites have provided the customers to their own site. With Google Analytics, the site owner and marketing team can determine what sites provided the most exposure to their link. This of course is important because they’ll know what sites and search engines they need to provide and focus most of their advertising resources to.

Another important aspect of modern web analytics are with respect to the text used in the search engines. This text or phrases are known as “keywords”. Each search engine has its own formula for determining query results, but several things primarily remain the same. Each search engine will index every page it gets exposed to and record much of the text from them as well. This data and text gets stored in a database. Depending upon the tags used or the placement of the text, it will use this text to compare against keywords when displaying the query results.

With the advent of this information, marketing research gurus have been able to take it to the next level. Using what marketing researchers call “psychographics”, the data analyst is able to determine the shopping habits, and even the demographics of the person visiting a web site. It might be typical to think that a person who’s IP hails from the New England area might be more apt to purchase a snow-blower when visiting a site such as Home Depot or Hechingers. Likewise, a person in Canada doing a search for vacations might definitely be interested in information regarding travel to Miami. This demographic and psychological information (again, known as *psychographics*), allows companies to better suit their webpage content and advertisements for the specific type of customer visiting their site.

Marketing research and web statistics have come a long way in the past 20 years of the internet boom. . The industry alone employs hundreds of thousands of people for web analytics jobs. As technology continues to progress, more and more potential information is becoming available from the customer and to the website company. There is quite literally no end in sight for how much information stands to be gained by this industry.

REFERENCES

1. Alan K'necht - Digital Web Magazine
(2005, October) "*Dollars & Sense of Web Analytics*"
Retrieved September 12th, 2009,
From http://www.digital-web.com/articles/dollars_and_sense_of_web_analytics/
2. Ballardvale Research - Counsel for Online Best Practices
(2004, June) "*Market Trends - Web Analytics: History and Future*"
Retrieved September 19th, 2009,
From <http://www.ballardvale.com/free/WAHistory.htm>
3. Get Smarter With Google Analytics - Education
Retrieved September 19th, 2009,
From: <http://www.google.com/analytics/education.html>