

## Webalizer WebStats - Explanation of Terms

### Main Headings

**Hits:** represent the total number of requests made to the server during the given time period (month, day, hour etc..). Any request made to the server which is logged, is considered a \*hit\*. The requests can be for any type of file, e.g., HTML pages, graphic images, audio files, cgi scripts, etc... Each valid line in the server log is counted as a hit. This number represents the total number of requests that were made to the server during the specified report period.

**Files:** represent the total number of hits (requests) that actually resulted in something being sent back to the user. Not all hits will send data, such as 404-Not Found requests and requests for pages that are already in the browsers cache. Some requests made to the server require that the server then send some information back to the requesting client, such as an HTML page or graphic image. When this happens, it is considered a \*file\* and the files total count is incremented. The relationship between hits and files can be described as incoming requests and outgoing responses.

**Tip:** By looking at the difference between hits and files, you can get a rough indication of repeat visitors, as the greater the difference between the two, the more people are requesting pages they already have cached (have viewed already).

**Sites:** is the number of unique IP addresses/hostnames that made requests to the server. Care should be taken when using this metric for anything other than that. Many users can appear to come from a single site, and they can also appear to come from many ip addresses so it should be used simply as a rough gauge as to the number of visitors to your server.

**Visits:** occur when some remote site makes a request for a page on your server for the first time. As long as the same site keeps making requests within a given timeout period, they will all be considered part of the same Visit. If the site makes a request to your server, and the length of time since the last request is greater than the specified timeout period (default is 30 minutes), a new Visit is started and counted, and the sequence repeats. Since only pages will trigger a visit, remotes sites that link to graphic and other non- page URLs will not be counted in the visit totals, reducing the number of false visits.

Whenever a request is made to the server from a given IP address (site), the amount of time since a previous request by the same address is calculated (if any). If the time difference is greater than a preconfigured visit timeout value (or has never made a request before), it is considered a new visit, and this total is incremented (both for the site, and the IP address). The timeout value is set to 30 minutes, so if a user visits your site at 1:00 pm in the afternoon, and then returns at 3:00 pm, two visits would be registered. Technically due to the limitation of the HTTP protocol, log rotations and other factors, this number should not be taken as absolutely accurate, rather, it should be considered a pretty close "guess".

**Pages:** are those URLs that would be considered the actual page being requested, and not all of the individual items that make it up (such as graphics and audio clips). Some people call this metric page views or page impressions, and defaults to any URL that has an extension of .htm, .html or .cgi.

## Common Definitions

**A Site:** is a remote machine that makes requests to your server, and is based on the remote machines IP Address/Hostname. Each request made to the server comes from a unique \*site\*, which can be referenced by a name or ultimately, an IP address. The sites number shows how many unique IP addresses made requests to the server during the reporting time period. This DOES NOT mean the number of unique individual users (real people) that visited, which is impossible to determine using just logs and the HTTP protocol (however, this number might be about as close as you will get). This is because it is possible for multiple users to be hidden behind a single IP number - as in firewalls, etc.

**URL - Uniform Resource Locator:** All requests made to a web server need to request something. A URL is that something, and represents an object somewhere on your server, that is accessible to the remote user, or results in an error (i.e.: 404 - Not found). URLs can be of any type (HTML, Audio, Graphics, etc...).

**Referrers:** URLs that lead a user to your site or caused the browser to request something from your server. The vast majority of requests are made from your own URLs, since most HTML pages contain links to other objects such as graphics files. If one of your HTML pages contains links to 10 graphic images, then each request for the HTML page will produce 10 more hits with the referrer specified as the URL of your own HTML page. A \*Referrer\* is the URL of the last web page the user was on before coming to your web site. Therefore, this information reports the web pages that link to your site.

Technically, if a user types in your URL directly, the browser will not record a referrer. In addition, the referrer variable is a new feature that was not supported in all browser types and versions. The recent browsers support the referrer attribute.

**Search Strings:** are obtained from examining the referrer string and looking for known patterns from various search engines. The search engines and the patterns to look for can be specified by the user within a configuration file. The default will catch most of the major ones.

The **Search Keywords** is a list of keywords that were used in search engines to find your web site.

**User Agents:** are fancy names for browsers. Netscape, Opera, Konqueror, etc. are all User Agents, and each reports itself in a unique way to your server. Keep in mind however, that many browsers allow the user to change it's reported name, so you might see some obvious fake names in the listing.

The **User Agent** reports the number of browser types and versions used by users visiting your web site.

**Entry & Exit Pages:** are those pages that were the first requested in a visit (**Entry**), and the last requested (**Exit**). These pages are calculated using the Visits logic above. When a visit is first triggered, the requested page is counted as an Entry page, and whatever the last requested URL was, is counted as an Exit page.

The Top Entry and Exit Pages give a rough estimate of what URL's are used to enter your site, and the last pages viewed by a visitor. This information will give a good indication of the overall trend in where users come into your site - via bookmarks or search engines, and where users exit your site.

**Countries:** are determined based on the top level domain of the requesting site. This is somewhat questionable however; as there is no longer strong enforcement of domains as there was in the past. A .COM domain may reside in the US, or somewhere else. An .IL domain may actually be in Israel, however it may also be located in the US or elsewhere. The most common domains seen are .COM (US Commercial), .NET (Network), .ORG (Non-profit Organization) and .EDU (Educational). A large percentage may also be shown as Unresolved/Unknown, as a fairly large percentage of dialup and other customer access points do not resolve to a name and are left as an IP address.

**Response Codes:** are defined as part of the HTTP/1.1 protocol. These codes are generated by the web server and indicate the completion status of each request made to it.