Webalizer Quick Help

- **Main Headings**

  Hits represent the total number of requests made to the server during the given time period (month, day, hour etc..).

  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.

  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.

  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**.

  A **KByte** (KB) is 1024 bytes (1 Kilobyte). Used to show the amount of data that was transfered between the server and the remote machine, based on the data found in the server log.

- **Common Definitions**

  A **Site** is a remote machine that makes requests to your server, and is based on the remote machines IP Address/Hostname.

  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 accessable to the remote user, or results in an error (ie: 404 - Not found). URLs can be of any type (HTML, Audio, Graphics, etc...).

  Referrers are those 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.

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.

Note: Only available if that information is contained in the server logs.

User Agents are a fancy name 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.

Note: Only available if that information is contained in the server logs.

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.

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 Isreal, 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 (RFC 2068; See Chapter 10). These codes are generated by the web server and indicate the completion status of each request made to it.