SAWMILL

Page Tagging vs. Log Analysis

An Executive White Paper

April 2008

FlowerFire

Introduction

We often get the question of what is the best method for collecting web analytics data. A web analytics solution that you will use with a web server, should provide you with the ability to do either log file analysis or use page tagging on each page.

There have been concerns over the accuracy of log file analysis, because of the ability to cache the information. Thus web counters were used to track the number of visits to each page. Also cookies were introduced to assign a unique cookie to a visitor. Page tagging also allowed third party companies to analyze data and produce reports for their clients.

We would like to highlight the disadvantages of page tagging vs. log analysis:

1. Third Party Cookies.
   Cookies are delivered by the analytics site, which makes them third party to the website being analyzed. Some browsers disable third party cookies, therefore creating inaccurate statistics.

2. Non-page Tracking.
   Page tagging cannot report on non-pages including:
   > Image file hits, PDFs, media files, downloaded files (ZIP, .EXE, etc.)

3. Bandwidth and Technical Tracking.
   Page tagging cannot report bandwidth or other technical metrics, like server response time.

4. Confidentiality
   Page tagging is implemented by putting the website statistics in a third party database. As mentioned above, this means that another company has access to your web statistics, which could be a privacy and confidentiality concern.

5. Historical Analysis
   Page tagging cannot report statistics prior to the implementation of the page tagging software on the server.
   Log analysis software can report as far back as you have log files, which is typically back to the inception of the server.

6. Non-web Traffic
   Page tagging cannot report anything but website traffic; log analysis can analyze firewalls, mail servers, media servers, network devices, or anything else that generates logs.

7. Cost
   The analysis engine has to cope with a huge volume of transactions, one for each page hit on each site analyzed. This makes the cost of data capture much higher than log analysis, where the logging cost is minimal (writing lines to an open file on the server is practically free). This translates to higher overall prices for page tagging solutions versus log analysis solutions.

Summary
There are many advantages to doing log file analysis. The economic factors are that the analysis can be done in house without hiring a third party company. Page tagging often requires a monthly fee, or at least someone on staff trained to keep the page tagging up to date. The log files are already on the server, ready to be analyzed. The web server will record accurately every transaction it makes. Page tagging relies on the visitor's browser being willing to accept a cookie or have Java-script turned on, which many do not.
Historical data can always be analyzed, from the beginning of the server's life, and it is in standard format, not a third party format. Spiders can be tracked using log file analysis, this can be important when you want to optimize your website for particular search engines. Failed requests are also tracked; page tagging records the event, but not that it failed.