



Gately's

A Premier Internet Shopping Site

A Case Study Using Sawmill

July 2006



Summary

Gately's is a chain of web stores, selling affordable products with high-quality, service and integrity. Their primary objective was to obtain numerical data on the return on investment of each of their internet advertisements, especially their Google Adwords advertisements, to determine which ads were costing more than they brought in. A secondary objective was to create a system which could report which items were ordered through the website, and how often, and how much was spent on them.

Implementation

The core of the implementation was a Sawmill log format plug-in that was based on the standard Apache plug-in included in Sawmill. In addition to tracking all web log fields, this plug-in also tracked order information, conversion information, and numerical fields such as dollars spent.

Ideally, every paid advertisement would point to a URL that contains the search engine and search phrase of the advertisement. In this case, Sawmill can be configured to extract search engine and search phrase information from the referrer field, but it is not as reliable as query-based extraction; in particular, it is generally not effective for content-based AdWords. Nevertheless, a reasonable percentage of search phrases were extracted from referrers, using the standard search engines rules built in to Sawmill, supplemented with custom rules implemented through log filters. Gately's eventually implemented search engine and search phrase information to every advertisement, at which point the referrer-based rules were made secondary, so query-based information was used when available, with referrer-based information as a fallback.

On the Gately's websites, each incoming visitor was assigned a unique OrderID, which appeared in the log data as a cookie. A SQL database stored information on each order in an Orders table, including the OrderID items in the order (items purchased and quantity of each), tax, shipping, and customer information. Items were referred to by ItemID in the order record, and a separate Items table in the SQL database contained information about each item, including an English description of the item, the vendor, and the price. The Orders table and Items table existed prior to the use of

Sawmill, and nothing in the SQL database had to be changed to use it with Sawmill.

To integrate the SQL database with Sawmill's web log analysis, Gately's created a script which exported the contents of the Orders and Items tables in Sawmill's CFG (configuration group) format, a very simple, highly compact, highly legible, textual format. Once data is in a CFG file, it typically requires only one line in a log filter for Sawmill to access the file. This logic was added to the log format plug-in, to extract the item information, order information, and numerical values (dollars spent) from the Orders and Items file, and to add them as standard database fields, and standard reports, in the web log analysis.

To track conversions, it was necessary to associate the search engines and search phrases of the original hit (the one which resulted from the advertisement click) with the final order. This was done by keeping the search engine and search phrase information internally in a node in Sawmill, using the built-in Salang language of Sawmill, and recovering the search engine and search phrase information at the time the order hit occurred in the web log data. This was very simple, due to Salang's natural support for hierarchical hash-based nodes, which can be persistent across multiple invocations; for instance, if the advertisement hit occurs on one day, and the order occurs on a later day, the information will still be carried across properly.

Result

The resulting reports showed all of the standard web log analysis reports (date/time reports, source IPs and countries/regions/cities, search engines, search phrases, pages and directories, and much more), but integrated the order information naturally, through new numerical columns and new standard reports. Each report included a "purchase amount" column that indicates the dollar value of purchases associated with that column. This provided not only such useful information as sales by country and sales by domain (each store is in its own domain, so this is sales by store), but also the primary objective: the standard Search Engines, Search Phrases, and Search Phrases by Search Engine reports included in the column also, so under Google Adwords, each search phrase appeared as a

separate row, with the total sales dollars generated by the visitors referred by that keyword.

Secondarily, the profile included custom Orders and Items reports, which showed the sales of each item, in quantity and dollars. The reports were configured to zoom naturally; so the Domains report showed the sales by store, and zooming in on a particular store showed the orders for that store, and zooming in on an order showed the items for the order. Sawmill's unlimited zooming capabilities allowed any combination of filters or zooms to be applied, so for instance, it is possible to zoom on a particular date range, and then zoom on a particular store, and then zoom on a particular country, and finally click Items to see the items sold (including dollar amounts) for that date range, in the store, from buyers in that country. This easily fulfilled the order analysis requirement, and provided new ways of examining the data, which had not been in the original proposal.

Finally, the analysis was a full-featured web log analysis. Even ignoring the order information, the reports provided all the information of a web traffic analysis, from unique IPs to sessions (entry/exit pages, paths through the site, session durations, repeat visitors, and more), to visitor demographics (geographic, domains, IP addresses, web browsers and operating systems), to content information (which pages and directories were hit and which file types), to technical information (bandwidth usage, server response times, screen dimensions, server response codes, broken links, and other detailed metrics).

Cost

The software license cost under \$1000, and the customization work, which was contracted to Flowerfire by Gately's, took less than 40 hours. The total project cost was under \$10,000, and was completed in under six weeks. Work began in mid-November with a goal of having numbers for the holiday season, and the goal was achieved. It should be noted that though Flowerfire did the work on a professional services basis, none of the work required new development on Sawmill--everything in the project was done using the existing Sawmill infrastructure, and could have been done by any expert Sawmill user without assistance from Flowerfire.

Quote from Gately's

Pete Scudamore, who represented Gately's and was the primary point of contact at Gately's through the entire project, said, "I spent a long time looking for a solution which could provide actual return on investment for our pay-per-click advertisements, by connecting our existing SQL data with the web logs. I spoke with many companies that seemed, at first glance, to be offering this type of analysis. But when it came down to it, none of them had the flexibility to do what we needed; the solution simply was not available through any product other than Sawmill. Sawmill was the only solution I found which could provide the actual sales generated by each advertisement, from the information we had available. Flowerfire worked with me through the whole process, and completed the project on time and under budget."