Get to Know Your Audience
Web Site Log Files Inform Design

By James Kalbach

Good design depends on knowing your audience. When rebuilding a Web site, there is significant information to be found in analyzing the visitors' paths through the site. The record of these paths is called the log file.

A log file is a text file in which the Web server records an entry for every item requested by a Web browser like: '123.123.123.11 - - [15/May/2000:17:56:04 -0400] "GET /index.html HTTP/1.1" 206 0 "-" "Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"'
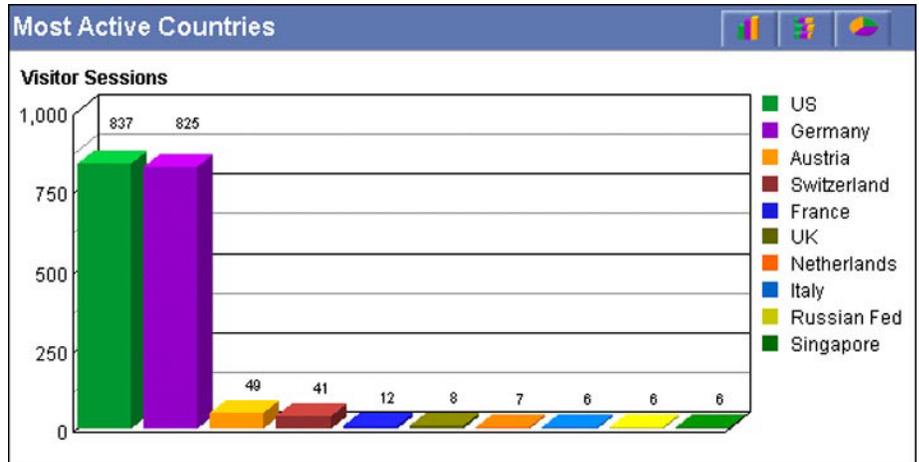
There is some interesting information in this line: the address of who requested it, when it was requested, what was requested, the type of browser, and what page they came from. It would be very difficult to read this file by eye, so a number of companies have written software to help analyze the raw information.

The analysis of log file data traditionally falls under the responsibility of a Web master or technical site manager. The resulting figures are typically large, round numbers that indicate total traffic, visitors or page views for a given period that are often used to cost justify

Most Active Countries

Analysis programs generate easy-to-read graphs. This example may suggest a German language version of the site. Source: WebTrends

james kalbach is an information architect at razorfish in hamburg.
<james.kalbach@razorfish.de>

the existence of a Web site to the client and marketing managers. But an inferential analysis of site traffic can provide valuable design feedback that can be used to optimize navigability and usability of a Web site. Log file analysis can also provide discussions with an objectivity that clients respond to. Qualitative ethnographic data, such as user surveys, should be supplemented with the log statistics.

Log analysis, however, is limited by the fact that the data are only approximate records of visitor behavior. User satisfaction is difficult to infer from the way a visitor goes through a site. For example, many hits from a single given visitor could indicate someone who really likes a site or someone who gets lost there. However, if servers and analysis software are properly configured, a wealth of usage data for an existing site can be gathered, often without heavy time or cost investments.

Basic Traffic Measurements

These basic whole-site measurements don't inform design directly, but are important to be aware of.

Hits A hit, sometimes called a request, refers to a file request made to a Web server by a browser. This most basic measurement is simply a count of the lines in the log file. Hits do not accurately measure the popularity of a site as one HTML page with ten images on it counts as eleven hits.

Page views Measuring page views is a much better way of tracking overall traffic to a site than counting hits. This is total number of times your site's pages are viewed, a good way to monitor overall traffic.

Unique Visitors The total number of individual visitors who viewed your site, the best measure of the user base.

The Most Useful Log Data Types

Most log analysis programs can provide detailed information on the use of each page. After identifying problem pages, one can look for their flaws and fix them.

Most popular pages Popular pages should be (and usually are) close to the surface. If popular pages are deep within the structure of the site, some

rearranging or adding of quick links should be done. Knowing the most popular pages can also prove or disprove hypotheses of design priority and can support data from market reports, questionnaires and interviews.

Least popular pages A low number of page views may indicate confusing navigation: Some pages may just be difficult to get to. Do they need to be more visible or do they need to be removed?

Entry pages Web sites are integrated into the Web fabric and are not always entered through the "front door" (homepage), but directly to sub-pages. Knowing the most frequent entry points can help you identify the purpose for each pages and provide navigation to other relevant areas of the site. "Sideways" visitors will need to immediately know where they are and where they can go.

Single access pages Pages that visitors enter and exit without viewing any other pages of the site. Analyzing these pages for layout, content and navigation might boost site traffic and stickiness, as well as identifying what the audience is looking for.

Exit pages These are the pages last seen by visitors before leaving the site. This is generally good indicator of the effectiveness of your content, site architecture and navigation. Most frequent drop-off pages should ordinarily not be navigational pages. Are people lost on these pages, are there technical malfunctions or are visitors just bored?

Page errors Visitors can't use pages they can't load. Be sure to notify developers and technologists about recurring technical failures.

Most trafficked paths A path is a succession of pages that unique visitors that have taken through a site. Paths

can indicate the effectiveness of your navigation: Do your visitors achieve their goals or do they leave your site prematurely? Placing a quick link to popular path destinations on the homepage, for example, can reduce the number of clicks it takes to get to these content

Log files have as much to tell a Web designer as a Web master

areas. Knowing popular paths can also help identify the most effective advertising or promotion space on your site.

Depth of visit This shows how deep users are drilling down into your site. Knowing where users are generally operating within your site can help you optimize the navigation to facilitate assumed user tasking in these areas.

Visitor's Country For sites that are intended to have an international market, knowing exactly where visitors come from is important. These metrics can justify translation efforts, help create more effective marketing campaigns, and aid in creating appropriate designs for a local market.

Referring Sites A key step in increasing traffic is identifying where existing traffic comes from. The most common sources of traffic are search engines, directories, and sites of related subjects. Targeting the primary referring site can help effectively promote your site. The better log analysis tools can determine not only referring search engines, but also what keywords people used to find your site. This is a direct indication of what the people are looking for.

Client side statistics Information about browser versions, OS, plug-ins and much more can easily be found from log

file data. This can assist in design decisions, such as whether to use DHTML.

All of these statistics, used in conjunction with user research and common sense can improve the quality of the redesign of the Web site. Failure to take the user's needs and behaviors into account is the simplest prescription for a failed project.

Popular Log Analysis Software
Don't underestimate the power of low-cost programs or the effort necessary to interpret the results of any program.

Accrue Insight (high-end / expensive)
www.accrue.com

WebTrends (mid-end / mid-price)
www.Webtrends.com

Wusage (low-end / cheap)
www.boutell.com/wusage

Webalizer (low-end / free)
www.mrunix.net/Webalizer

for more information

A practical discussion of the issues
"There's gold in them there log files!", 2000, Charlie Morris
www.wdvl.com/Internet/Management/index.html

An excellent academic-level essay
"Reading Reader Reaction: A Proposal for Inferential Analysis of Web Server Log Files", 1997, Terry Sullivan
www.pantos.org/ts/papers/rrr.html

A basic look at analysis functionality
"Software for analyzing your website's traffic.", 1998, Bruce Morris
webdevelopersjournal.com/columns/analysis.html

razorfish reports are published regularly by the razorfish science department for our colleagues and the interested public.