

# PETTT Technical Report

## PETTT-02-CT-01



## Evaluating Your SimpleSite with Log File Analysis

How do your students use your site? Is your site effective?

Kristen Shuyler & Eric Maddox  
August, 2002

University of Washington  
Program for Educational Transformation Through Technology  
Box 353080  
Seattle, Washington 98195-3080

<http://depts.washington.edu/pettt/home.html>

To cite this document, use the format:

Shuyler, K. and Maddox, E. (2002). Evaluating Your SimpleSite with Log File Analysis. Technical Report PETTT-02-CT-01. Program for Educational Transformation Through Technology. University of Washington, Seattle, Washington.

*Notice: PETTT has made its working papers, reports, and other documents available for your free use, but only if you (a) reference us in your bibliography and (b) tell us about it by emailing us at [pettt@u.washington.edu](mailto:pettt@u.washington.edu).*

## TABLE OF CONTENTS

|  |           |
|--|-----------|
| <b>Introduction</b>  | <b>1</b>  |
| <b>What are log files?</b>   | <b>1</b>  |
| <b>Why should I investigate how my students use my web site?</b>   | <b>3</b>  |
| <b>Why is log file analysis a good way to discover how my students use my site?</b>  | <b>3</b>  |
| <b>What exactly can I learn from log file analysis?</b>  | <b>4</b>  |
| <b>How can I improve my web site's instructional and informational materials with the data I get from log file analysis?</b> | <b>8</b>  |
| <b>Conclusion</b>  | <b>9</b>  |
| <b>Bibliography and further reading</b>  | <b>10</b> |

## **Introduction**

If you have created a SimpleSite or other web site for your classes, you might be curious whether your students use it, how they use it, and what they think about it. After all, you have spent your time and energy on it, so you might want to be reassured that your students have found it helpful. Asking your students directly (through surveys or conversations) is one way to find out what they think about your site. However, if you want to examine how your students actually use your site, you will have to go beyond their perceptions, attitudes, and preferences—you will need to observe their actions.

Log file analysis is a fast and easy method for examining actual student uses of your site. By providing you with a dynamic picture of the patterns of use (Ramey 2000: 398), log file analysis is a powerful way to learn about your site. Moreover, it is not as technical or difficult as it may sound. Below you will find some explanations of "log files" and other techno-terms, what you might learn from log file analysis, how you might use it to efficiently improve your site, and links to step-by-step instructions and further reading.

## **What are log files?**

Log files are the records of activity on a web site (Ramey 2000: 398). These records require no special tools to collect them (Aviram 1998). Your SimpleSite is stored on a UW-owned computer called a "web server." When someone types your web site address into a web browser (like Internet Explorer), his or her computer contacts the UW web server and requests to view the files (web pages, images, etc.) that comprise your web site. These requests are also called "transactions." Most web servers automatically record every transaction between the server and another computer (Ingram 1999: 138). These records are called "log files." Analysis of these web server records is called "log file analysis."

Log files usually include information such as (Ramey 2000: 398-399):

- the date and time of the transaction,
- the IP address of the requesting computer (An IP address, or Internet Protocol address, is a numeric identifier unique to each computer on a network [[http://www.webopedia.com/TERM/I/IP\\_address.html](http://www.webopedia.com/TERM/I/IP_address.html)]),
- the file that was requested,
- the size of the file,
- the status of the request (e.g. was the request fulfilled, or did it end in an error?),
- the visitor's browser (e.g. Netscape or Explorer) and platform (e.g. Macintosh or PC),
- and the web page that the visitor was viewing before they viewed yours.

Log files may give you detailed information about how your web site is used, but the statistics derived from them are not always easy to interpret. Due to the way log files are collected, they have several inherent limitations, such as (Ramey 2000: 398):

- Log files track requests for files (images or web pages), **not** users.
- Log files recognize specific computers viewing your web site, **not** specific people.
- Technical aspects of web browsers, such as the cache (a file on the user's computer that temporarily stores web sites) may prevent the log files from recording every time a user looks at your pages.

Due to these and other limitations, most log file analyses can give a good estimate of how your site is used, but not an exact description (e.g. the number of users will never be exact). In addition, your log files cannot tell you what each user does; they can only give the average statistic for whatever time period you are investigating. Lastly, **log files cannot tell you *why* students use your site the way they do.** Interpretations of the statistics are up to you. Often, log file analysis will be more useful and accurate if you combine it with other inquiry methods (Ingram 1999: 151), such as informal chatting with students, in-class shows of hands, surveys, or written feedback about your web site.

## Why should I investigate how my students use my web site?

"Formative evaluation" is the evaluation of an educational program while it is still in development, and with the purpose of continually improving the program (Maslowski et al 2000: 6). Examining how your students use your web site is one way you can evaluate your instructional design in a formative manner. It may help you improve your instructional materials based on the students' performance and attitudes with respect to the web site (Ingram 1999: 137).

Log files can be useful for other reasons. For instance, discovering what parts of your web site are actually used by students can help you spend your time and resources efficiently—by focusing on the most popular or useful parts of the site (Koyani and Mathews 2001: 131) and spending less time on the less popular aspects of it. Additionally, web sites can frustrate students when they have difficulty finding information (Koyani and Mathews 2001: 130). Assessing a site's usage may help you improve it, which may improve your students' opinions about your class.

## Why is log file analysis a good way to discover how my students use my site?

- **Log file analysis reveals what people are *actually* doing when they use a site.** Other methods, such as surveys, may reveal what students think they are doing. Surveys and similar methods rely on students to self-report their actions. However, log files show what they actually do—which is often not the same as what they say they do (Kushniruk, et al. 2000: 5).
- **Log file analysis is an unobtrusive method of collecting and analyzing data.** You will not need to ask your students to spend their time filling out a survey.
- **The log files are automatically collected for you.** The University's web servers collect the log files so you will not spend time collecting them. Why not take

advantage of the available information to make sure you are offering your students a useful site and spending your time most effectively on it?

- **Log file analysis does not preclude you from using other methods to examine your web site.** In fact, combining methods may be the most useful way to evaluate your web site (Ingram 1999: 151). For example, you could ask your classes to respond to questions about the web site during class (show of hands or discussion) or over e-mail. Or you could have them fill out a short survey (online, using WebQ from Catalyst, or on paper). This information, when combined with the primary data from log files, would present a more complete assessment of how your site is used.

## What exactly can I learn from log file analysis?

Analyzing log files can reveal certain characteristics of your site and its visitors. Below are some of the statistics that are most easily extracted from the log files, explanations of these statistics, and some notes on how you might interpret them.

**1. Number of visits to your web site.** A "visit" consists of one computer requesting one or several files from a server in a certain amount of time. A visit is not the same as a "hit," which is a single file sent to a requesting computer. For example, when a student views your web site, the log files may show one visit and eight hits (six images and two pages). Visits do not tell you how many people have visited the site—just how many times it has been visited. However, finding the number of visits is more informative than the number of "hits," since the number of hits can be artificially inflated by multiple images on each page (Ingram 1999: 142).

**2. How many people visited your site?** Your web site's log files collect the number of files requested, *not* the number of visitors (Ramey 2000: 398). However, you can estimate the number of visitors from the number of visits and IP addresses (Ingram 1999: 140; Aviram 1998). When students visit your site, their computer's IP address is recorded in the log files. The number of unique IP addresses provides an estimate of how

many people are using your site, if you assume that each IP address represents a single user (Ramey 2000: 400). However, the number of IP's is not the same as the number of visitors. Some Internet service providers (ISP's) assign IP addresses that change while the user is online, so the log files may incorrectly indicate more than one visitor. In addition, the number of visitors tells you the number of distinct computers that visited the site, not the number of people. You would not know if more than one person was looking at the computer, or if one computer was used to view your site by several people at different times (such as a computer in a lab on campus) (Ingram 1999: 142).

**3. Where are your visitors located?** Log files contain IP addresses, which offer information about the location of the computers that access your web site. Sometimes the log files reveal specific computers or computer labs (on the UW campus, for example), but more often they show the commercial Internet service provider (ISP) that the students use, or in which state they are located. You can also discover whether visitors are using computers in the educational (.edu), commercial (.com), government (.gov), or military (.mil) domains or foreign countries (.fr for France, .jp for Japan, etc.). This information may help you understand where students use your web site: on-campus accesses would be UW IP addresses in the .edu domain, whereas at-home accesses might be from dialing in to UW, or from connecting to AOL, AT&T, or another commercial service provider. In addition, you can discover if people other than your students are viewing your site. For example, if you notice a lot of foreign usage, your site might be popular among non-students around the world! (Ingram 1999: 146-147; Ramey 2000: 400)

**4. What pages do students view the most and least often?** By looking at the log files, you can determine the most commonly viewed resources on your site (Ingram 1999: 140). One way to do this would be to look at, for example, the top five (and bottom five) files requested and the frequency with which they were requested (Ramey 2000: 399). Knowing these numbers can give you an idea of the level of demand for the various types of content on your web page (Ramey 2000: 400).

**5. How much time do students spend at your web site?** Log files can tell you how many minutes students spend (on average) on your web site and each of its pages. This data can complement what you know about how often your pages are viewed and help you understand what parts of the site students use the most. It also tells you about the behavior of students when using the web site—do they spend, on average, a lot of time using your site? Or do they visit it more often for just a few moments? This statistic should be used cautiously, however, because it is difficult to interpret what these times actually mean (Ramey 2000: 401). For instance, a *long* amount of time spent on a page might indicate interest, but it might also signal confusion or difficulty in finding the desired content (Ramey 2000: 401). Moreover, a page that is "viewed" for several minutes may not have been attended to at all—the student could have left the page open while doing something else. On the other hand, a *short* amount of time on one page may indicate little interest in its content, but it might be that students print out pages that they do not want to read on the computer screen (Ingram 1999: 148-149). These statistics should not be ignored because they are difficult to interpret. If you use them critically, they could be very useful. For instance, one way to minimize the effects of disrupted user attention is to take an average over a long block of time such as an academic term (Ramey 2000: 401). In addition, if you compare the time data with other information about your web site, you could better understand the usage of your site. For instance, you could ask for a show of hands during class about printing pages from the web site, or ask students informally about it.

**6. When do students use your web site?** Log files can also tell you *when* your students use your web site. Do they tend to use it between midnight and 2 a.m., just before class meetings, or after class meetings? (Ingram 1999: 143). You can analyze the patterns of use during the day to find out. Weekly patterns can also be seen in the log files. You can determine if students tend to use the web site on weekends or during the week. Log files can also show if weekly usage peaks on the day of class meetings or if students put off visiting the site until the day before an exam (Ingram 1999: 143). In addition, it might be useful to know how students use a course web site during the academic term. For instance, if the site was heavily used at the beginning of the term, then use dropped to

almost nothing, it could be that many students printed up pages from the web site—or it could mean that students decided the site was not useful after using it once or twice (Ingram 1999: 145). Once you are aware of these three usage patterns (daily, weekly, and by the term), you might want to know the actual reasons behind them—this might be a good opportunity for an e-mail survey, WebQ, or in-class show of hands.

**7. What pages on your site do students look at first and last?** Log files can tell you what pages in your web site are viewed first and last. Do students enter on your main SimpleSite page, or do they begin with a different page that they find more useful? If students tend to enter on other pages, they could have bookmarked these pages, memorized the web addresses, or followed links in an e-mail or other web site. Do you care that they do not see the front page every time? Are they going to miss important announcements by skipping the first page? Log file analysis can also tell you the last pages that visitors view. Interpreting why students left after reading these pages can be difficult—it could be that these pages satisfied their information needs, or that they became frustrated by the time they arrived at these pages (Ingram 1999: 149).

**8. What sequences of pages do visitors follow?** Log files can be analyzed to show you in what order students view your web pages (Ingram 1999: 140). This kind of analysis is called "path analysis," and can show you how long they spent on each page as well. Path analysis is useful only for small groups because it is a time-intensive analysis (Ingram 1999: 151).

**9. Other information.** Log file analysis can also show you what browsers (e.g. Internet Explorer or Netscape Communicator) and what operating systems (e.g. Macintosh or PC) your users employ, if that is of interest to you (Ingram 1999: 146). You can also find out how many visitors look at only one page before leaving (called "single page visits") (Ingram 1999: 148).

## **How can I improve my web site's instructional and informational materials with the data I get from log file analysis?**

Evaluating a web site requires knowledge of what the site's goals are. To begin, review the goals of your web site. Did you create it to provide:

- the syllabus,
- supplementary instructional material or required instructional material,
- "standalone" instruction,
- links to material that you did not create,
- highly interactive educational experiences, or
- course and administrative updates? (Ingram 1999: 154)

Once you have clarified the web site's goals, you can analyze statistics that will help you evaluate whether these goals are being achieved. For instance, you can examine how popular certain pages are, how long students spend on certain pages, what pages are largely ignored, and other topics discussed elsewhere in this report.

You might want to compare your initial assumptions about your web site with what you find through log file analysis. If you discover a difference, consider what you might do to make the usage of your web site meet your expectations, or how you might change your web site to meet your students' expectations instead (Ramey 2000: 400).

### *Some specific suggestions*

If you look at the data for the most and least visited pages in your site, you might ask yourself: Do you see unexpected patterns of use? Are your students avoiding content that you expected to be important? How could you redirect their attention to the parts of the site that you think are important? (Ramey 2000: 401). You might study the data about the most and least popular pages and then change the ordering and grouping of topics in your menu to reflect users' usual preferences (Ramey 2000: 402).

Your log files can show if students actually use media files, such as still images or streaming media. If they do not, then any instruction or information in these files is probably not reaching your students (Ingram 1999: 138) and you may want to reduce the amount of time you spend developing such resources.

For more discussion on how to design a web site based on findings from log file analysis, please see Ingram 1999 and Ramey 2000.

## **Conclusion**

Log file analysis can tell you how your students actually use your web site. Although log files are incomplete sources of information, they can offer insights about your web site that you cannot get anywhere else (Drott 1998: 50). If combined with other sources of information, log files can be interpreted and used to improve your web site.

For instructions on how to access, use, and analyze log files, please see "Tracking Web Site Usage" from UW Computing and Communications. This document is available at: <http://www.washington.edu/computing/web/publishing/usage.html>

## Bibliography and further reading

- Aviram, Mariva H. 1998. "Analyze your Web Site Traffic."  
<http://www.builder.com/Servers/Traffic> Accessed 8 August 2002.
- Drott, M. Carl. 1998. "Using Web Server Logs to Improve Site Design." Annual ACM Conference on Systems Documentation. Proceedings of the 16th Annual International Conference on Computer Documentation. ACM Press, page 43-50.
- Ingram, Albert L. 1999. "Using Web Server Logs in Evaluating Instructional Web Sites." *Journal of Educational Technology Systems* 28 (2): 137-157.
- Koyani, Sanjay, and James Mathews. 2001. "Cutting to the Chase: Delivering Usable and Useful On-Line Health Information Web Sites Through Usability Engineering." *Health Promotion Practice* 2 (2): 130-132.
- Kushniruk, Andre W., et al. 2000. "Evaluation of Web-Based Patient Information Resources: Application in the Assessment of a Patient Clinical Information System." Proceedings of the Fall 2000 American Medical Informatics Association Symposium. Available <http://www.amia.org/pubs/symposia/D200732.PDF>
- Maslowski, Ralf, Adrie J. Visscher, Betty Collis, and Paul P. M. Bloemen. 2000. "The Formative Evaluation of a Web-Based Course-Management System Within a University Setting." *Educational Technology* May-June 2000: 5-19.
- Ramey, Judith. 2000. "Guidelines for Web Data Collection: Understanding and Interacting with Your Users." *Technical Communication* Third quarter 2000: 397-410.
- Shelton, Brett, and Jennifer Turns. 1999. "Exploratory Arthritis Source Log Analysis, Part 1 and Part 2." Unpublished paper, University of Washington Program for Educational Transformation Through Technology. Available at <http://depts.washington.edu/pettt/papers>