# A Model of Online Instructional Design Analytics

Kenneth W. Fansler
Director, Technology Services
College of Education
Illinois State University

Rodney P. Riegle, Ph.D.
Professor of Education
Illinois State University

As an instructional medium, the Internet offers an unprecedented level of statistical information to assist in making effective, informed decisions about instructional design. As more educational organizations view the Web as an integral part of their operations, interest in the measurement and evaluation of instructional web site usage will increase. Log file analysis is one of several such mechanisms. Server logs can be used to glean a certain amount of quantitative usage information. Compiled and interpreted properly, log information provides a baseline of statistics that indicate use levels and patterns and/or growth comparisons among parts of a site or over time. The objectives of this paper are 1) to provide a data-driven example of instructional design evaluation for online course designers and administrators, and 2) to provide a theoretical framework of instructional design analytics that results in a deeper understanding of online teaching and learning.

## Overview of Log Files

A log file is a computer file that records the activity on a Web server. Almost every Web server has some sort of system that stores information about which pages, images, and files are requested, who requests them, and how many bytes are transferred. All of this information is dumped into a log file that is stored in a specific location on the server. These log files can be explored by simply opening the log file in an ordinary text editor and reading the raw data. For a more user-friendly view of the information, the log file can be analyzed by one of any number of log analysis software packages that display the data as charts, graphs, or tables that clearly illustrate users' activities. For example, WebTrends (popular log file analysis software) records online traffic patterns and behaviors.

From log files, instructors can determine what web browser and operating system students are using, where they come from when they visit an online course, the path they use as they navigate through the course, how long they stay on each page, how often each page is visited, from what page they exited the course, and when (date and time) and from where (IP address) all these events happen. This information has the potential to be used to evaluate the effectiveness of instructional design (e.g., are web pages tailored for the most popular browser, is a particular web page unnecessary, are students finding essential web pages, are students spending the right amount of time on various web pages, do successful students follow a certain pattern through the pages, etc.). Log file analysis can also be used to track changes in online traffic patterns and behavior over time. This information can be used to evaluate the effectiveness of instructional design within a semester or from semester to semester.

Advantages of using this method to examine instructional design include: 1) it is quantifiable, 2) it is unobtrusive, and 3) it is scalable (the method can be applied to any number of students or courses). This paper proposes a model of online instructional design analytics based on the use of log file analysis.

## The Limits of Log File Analysis

There are certain things that log files cannot track. Unless the Web visitor is required to login to the site with a username and password, no user demographics are recorded at all. On educational Web sites that do not required a login, it is difficult, if not impossible, to study the behavioral patterns of an individual. Even on courseware sites (WebCT™, BlackBoard™, etc.) that do require logins, the log files record only the frequency of logins by a single user. The analysis of these log files may show statistics such as the top ten authenticated users or the average daily time spent on the site by a single user. It does not show how a single human may have traversed through the site.

Log files record only "request" transactions. This happens when the client (user's Web browsing software) requests a page from the Web server. There are other, more sophisticated tools being used on Web servers that are not recorded on the log files. For example, Macromedia Flash™ is used by some educators to provide a multimedia-rich interactive experience for their students. Flash files are loaded and run on the user's local computer instead of the Web server. As a result, this activity is not recorded in the server log files. Java applets, common gateway interface (cgi) applications, and other server-residing programs that produce dynamic information (content that is not contained in static or permanent files) share a similar fate. As Web-based instructional tools such as these become more sophisticated, it is conceivable that less user activity will be logged unless changes are made to how activity is recorded.

Log file analysis can be misleading if improperly interpreted. Following are just some of the possible misconceptions that can be construed from the data reported in log files:

1. **Site hits are an indicator of the number of visits**
   The analysis software measures both "site hits" and "page hits". The "site hits" statistic measures not only the Web page that was hit, but also the total number of graphics and other multimedia files that are associated with the page. For example, if a Web page has three images, the number of hits recorded in the log files for a visit to that page would be at least four. This does not mean there were four visits to the page.

2. **All page or site hits are generated by students**
   Web search engines like Google™ use "spiders" or "robots" to scour the Web for documents. References to these documents are added to the search engines' databases. The spider hits are recorded in the same manner that "regular" hits (those recorded due to actual humans visiting your site). While some Web analysis software programs separate this information from other statistics, they may also aggregate the spider hits and the regular hits in a grand total number of hits. This information can be misleading. Also, page hits may reflect your own viewing of the site while making changes to the course content or layout. Finally, outside guests who may simply be browsing your site will also be recorded in the Web server logs.

3. **Hits from a single Internet Protocol (IP) address reflect visits from a single user**
   An IP address is a unique address that belongs to a single device (like a computer) that is connected to the Internet. While it's possible that multiple hits logged from a single IP address might reflect activity from one person sitting at a single computer, it might also reflect multiple users accessing the site from a single computer over a period of time. It should also be noted that a computer's IP address can change, depending on how the computer is configured. This means that one individual's activity on a Web site may be logged as multiple IP addresses over a number of days, and might therefore be incorrectly interpreted as multiple visitors. Finally, many Internet Service Providers (ISPs) such as America Online™ (AOL) and EarthLink™ channel their users through IP addresses on computers that may be in a geographic location other than that of the user. For example, an AOL user in Montana who visits your site may first be channeled through

an AOL-owned computer in Virginia. The Web server will log this IP visit as a connection from a computer in Virginia instead of Montana.

4. **The "average time viewed" statistic accurately measures the amount of time someone spent on the Web page**
   It cannot be assumed that a visitor is engaged in the content of a Web page for the entire time reflected in the log files. It is quite possible that a visitor may walk away from the computer and leave it running, or may have a second Web browser window open to visit multiple sites at the same time. A visitor may also stay on a page just long enough to print the document for later reading offline.

## Characteristics of a Useful Model of Online Instructional Design Analytics

Log file analysis software is driven by commercial interests, namely, the tracking of online customer behavior. There are some similarities between online customers and online students (e.g., they both search for information), but there are also some obvious differences (e.g., purchasing is a much simpler act than learning). These differences are ignored by the current model of log file analytics, and this model has had a chilling effect on both the use of log file analysis in online instructional design and the thinking about online instructional design analytics. Below are some of the characteristics that are necessary for log file analysis to become truly useful for online instructional design. These characteristics are not limited by the current capabilities of log file software or even the current common log file format. We believe that it is at least as important to identify what online instructional designers need as it is to identify what is currently possible. It is our hope that once a sound theoretical model is developed, software engineers will be interested in working with Web server software writers to create applications that realize the potential of the model. Log file analysis software (often referred to as web analytics) is currently a $600 million dollar market and it is projected to grow at a 32% annual growth rate through 2005 (Blumstein, 2003). Given the hundreds of thousands of educational organizations worldwide, instructional design analytics is a market capable of generating serious commercial revenue. Thus, our hope seems at least theoretically justified.

1. Representations of log file data must be able to be tied to the "shape" of the online course. An online course is a collection of pages and links that form patterns through their relationships between and among each other. Together these patterns and relationships constitute the course design or "shape." Currently, log file analysis software displays log file data in charts and graphs that do not represent these patterns or relationships, but rather are limited to student activity.

2. Elements of the online course (e.g., pages, links, page visitors, length of stay, etc.) must be able to be represented graphically. For example, if pages are represented by points and links are represented by lines between points, then the size of a point and the thickness of a line could represent the number of visits. Currently, log file analysis software displays such activity numerically.

3. Log file data must be able to be displayed in three-dimensional representation of the course being analyzed. Some particularly complex combinations of instructional web pages are best (perhaps only) viewed in 3 dimensions (e.g., cubes and spheres) rather than 2 dimensions (e.g., squares and circles). To get the most accurate view of student navigation within an online course, an instructor may need to view their activity in 3 dimensions. Thus, the charts and graphs output of current instructional log file analysis software must be supplemented with representations of the log file data organized by the three-dimensional "shape" of the online course.

4. The point of view (perspective) of the three-dimensional representation of log file data must be able to be manipulated by the instructor. That is, the instructor must be able to choose a front, rear, or side perspective as well as an interior or exterior perspective. In addition, the instructor must be able to zoom in or out on a particular part of the three-dimensional representation.

5. Instructional designers need to be able to understand not only how students move spatially through an online instructional environment but also how they move through it over time. This means that the output of instructional log file analysis software must go beyond the current static charts and graphs and into moving displays such as java applets or mpeg or flash movies.

6. The representation of the time dimension of log file data must be able to be manipulated by the instructor. That is, the instructor must be able to view data change forward or backward and at variable rates of speed in either direction.

7. Log file data must be tied to individuals or groups of students in order to analyze navigation patterns by such factors as final grades and course evaluations. Instructional designers could then attempt to determine whether different navigational patterns yield different final grades and/or course evaluations. This means either that log file formats must be extended beyond the current extremely limited types of data records (for explanations and examples of extended log formats, see Bacus Laboratories, 2001, Hallam-Baker, & Behlendorf, 1996, and Moeser, 2000) or that instructional log file analysis software must become integrated with cookies, spyware (tracking software installed surreptitiously on a computer, usually by clicking on an online ad, which reports usage data back to the installer), and other types of individual identification methods (for a discussion of this issue related to recent commercial trends. See Sullivan, 2001). Of course, this additional data must be closely guarded so as to avoid ethical problems.

8. The presentation of log file data, which is currently limited to .html pages and multimedia files (.gif, .mp3, .mov, etc.) must incorporate other kinds of interactive files such as java applets and flash movies. Clearly, student navigation patterns can be modified by information they receive from these files.

9. Log file data must be integrated with other kinds of student online experiences such as email and discussion boards. Clearly, student navigation patterns can be modified by information they receive by email or on a discussion board.

10. The information generated by the characteristics described above must be easily accessible in real time to students. Just as this information can be useful to instructional designers who are attempting to improve an online course, it can also be useful to students seeking to improve their performance while taking an online course.

## Conclusion

We do not claim that the model outlined above is perfect, or that it identifies every important feature of online instructional design. However, such a model is clearly superior to existing models. We believe that more discussion and research about models such as the one outlined above are crucial if progress is to be made in the field of online instructional design.

## References

Bacus Laboratories, Inc. (2001, December 7). Common log format: HTTP Web server log format. Retrieved May 28, 2004, from http://www.bacuslabs.com/WsvlCLF.html

Hallam-Baker, P.M., & Behlendorf, B. (1996, March 23). Extended log file format: W3C working draft WD-logfile-960323. Retrieved March 15, 2004, from http://www.w3.org/TR/WD-logfile.html

Moeser, E. (2000, November 15). Shop log file format. Retrieved April 3, 2004, from http://www.w3.org/TR/shoplogfileformat/

Sullivan, T. (2001, August 17). Web analytics move across the enterprise. *InfoWorld.* Retrieved April 21, 2004, from http://archive.infoworld.com/articles/fe/xml/01/08/20/010820feedge.xml

Blumstein, W. (2003, October).World Wide Web analytics software applications forecast and analysis, 2003-2007. Retrieved May 28, 2004, from http://idc.com/getdoc.jsp?containerId=30292

## Biographical Sketches

**Mr. Ken Fansler** is the Director of Technology Services for the College of Education at Illinois State University. He earned a B.M.E. from Shenandoah University and an M.M.E. from Illinois State University. He is currently a doctoral student in Educational Administration at Illinois State University. His dissertation focuses on the use of log file analysis to evaluate instructional design. He has made presentations to the Sloan-C International Conference on Asynchronous Learning Networks (2003) and the Nineteenth Annual Conference on Distance Teaching and Learning (2003).

Address: 331 DeGarmo Hall
Campus Box 5300
Illinois State University
Normal, IL 61790-5900
Phone: 309.438.7055
Fax: 309.438.3813

**Dr. Rod Riegle** has been a professor of education at Illinois State University since 1978.He earned a B.A. in education from Wittenberg University and an M.A. and Ph.D. in philosophy of education from Ohio State University. Dr. Riegle is the author of 4 books and 50 articles on a wide variety of topics on every level of education. He has made more than 50 national and international presentations to scholarly organizations and has served as a consultant on information technology to over 40 educational organizations.

Address: 331 DeGarmo Hall
Campus Box 5900
Illinois State University
Normal, IL 61790-5900
Phone: 309.454.2639
Fax: 309.438.8683