

Web Log Analysis: A Study of Instructor Evaluations Done Online

Kenneth J. Klassen
Brock University, St. Catharines,
ON, Canada

kklassen@brocku.ca

Wayne Smith
California State University,
Northridge, CA, USA

wayne.smith@csun.edu

Executive Summary

This paper focuses on developing a relatively simple method for analyzing web-logs. It also explores the challenges and benefits of web-log analysis. The study of student behavior on this site provides insights into website design and the effectiveness of this site in particular. Another benefit realized from the paper is the ease with which these concepts can be discussed with students. The purpose and context of the website used are easily understood by instructors and students, providing the basis for some rich discussions regarding: web-logs in general, web-log analysis and its' challenges, the use of data for decision-making purposes, and other aspects.

Web-logs record every page that users access during their visit to the site, thus providing a picture of their behavior based on a number of things, including the order in which they hit pages (i.e., the "traversal path"), the length of time spent on each page, whether they go to pages in error, and whether they return more than once to a page. This is a vast data resource which many companies have difficulty using in an efficient and effective way. Using the on-line instructor evaluation site provided a finite database to study; in this case raw data of approximately 35,000 records represented 3,368 student-class evaluations done over a two-week period (a small web-log). Open-source software was used to calculate a number of statistics, and it was found that students do access the evaluation site from home, taking advantage of the 24-hour availability. However, many students fill out evaluations on weekdays between 9 am and 3 pm, suggesting that a fair proportion do the evaluations in between classes when they are on campus.

The main portion of this work involved identifying and analyzing the traversal paths of the students through the site. There were a huge number of unique paths, highlighting the difficulty in isolating common user behaviors. It was found that students are able to complete the entire process (signing in, filling out one or more evaluations, etc.) in just over 8 minutes (on average). Students spend on average 82 seconds answering 12 multiple choice questions and 110 seconds filling out qualitative comments, suggesting that they spend a sufficient amount of time to ade-

Material published as part of this journal, either on-line or in print, is copyrighted by the publisher of the Journal of Information Technology Education. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Editor@JITE.org to request redistribution permission.

quately evaluate the course. In fact, students provide more qualitative comments on the online evaluations than they do on the in-class paper evaluations. Further path analysis helped to show why there are so many unique paths; for instance, if a user hits "reload", the page will register again in the web-log, creating a different path than a user who only loads the page once even though these two

students are identical from a user-need standpoint. Also, the use of the “Back” button in the web browser and the ease with which users can jump from page to page around the site also creates many different paths.

This study found that the vast majority of users to the site were able to complete evaluations, and to do so in an efficient, time-effective manner. Thus, this is a well designed site that helps and directs users to the appropriate pages. It was determined that future analyses would benefit from some prior planning on the part of the web-log administrator, including some additional information in the form of “markers” in the logs to track user behavior, which would reduce the complexity of the analyses required.

The main reason for evaluating courses online is that it can drastically reduce the administrative cost and time required. The results here suggest that this is also an effective and efficient method for both students and instructors. Finally, as mentioned above, the information presented here can provide the basis for interesting class discussion of these issues.

Keywords: Web-log analysis, instructor evaluations, analyzing data with spreadsheets

Introduction

The rapid development of Internet-based technology and an increased desire to reduce the cost and time required to administer and record students’ evaluations of faculty members has prompted some schools to move towards the use of online instructor evaluations. The ability to authenticate a user and to securely transmit and store data has helped make this possible. Additional benefits (beyond cost savings) include the ability to effectively administer evaluations in a wide range of learning contexts and locations (e.g., distance learning and other online learning environments) and the ease of compiling summary data by instructor, by course, by department, by question, or in any other way desired.

Initial experimentation has been done at the College of Business and Economics (COBAE) at California State University, Northridge (CSUN). As with any change that can have major implications for many people (including promotion and tenure decisions), there are a number of questions, concerns, and related issues with which to deal. This research focuses on some of the technical aspects of implementation, but other concerns are also discussed below.

In addition to the interest in instructor evaluations, this research also considers implications for analysis of web-logs in general. Use of the World Wide Web is becoming an important addition to traditional business and organizational service encounters (e.g., Haksever, Render, Russell, & Murdick, 2000; Turban, Lee, King, Warkentin, & Chung, 2000). Accordingly, a number of strategic advantages for improving service through technology are possible (Berry, 1996). While the sage advice “you can’t manage what you can’t measure” (Drucker, 1993) is still very much true, in an e-Business environment it is also true that “you can’t measure what you can’t monitor”. Inspection, productivity, or quality control techniques in traditional service environments may be inadequate, inappropriate, or simply impossible in an electronic service environment. Learning about precise user behavior on web sites is of significant interest to both academics (Spiliopoulou, 2000; Joshi, Joshi, Yesha, & Krishnapuram, 1999) and practitioners (Davenport, 1999; Kimball, 2000).

Despite the abundance of available data that are related to the use of web sites, few organizations have yet developed the knowledge and skills necessary to study these data and transform them into information for the purposes of improving operations. This is partly because the Internet environment is relatively complex and data about transactions and events are generated so quickly; each time a user accesses (“hits”) a web page, a web-log entry is generated. Since typical users jump from page to page quite quickly, changes happen quickly and a plethora of information is

generated. Thus, “web-log analysis to improve web page content and design is not an easy task” (Drott, 1998, p. 50). Recording web hits on even a relatively small web server can result in log files with hundreds of thousands of lines of data - or more.

Because individual customers cannot be physically observed on a web site, studying user behavior on a web site is different than studying customers in a traditional service. On the Internet, customers move quickly, and many customers can be on a site at any given point in time, making it impossible to observe individual behavior in real time. However, it is possible to study the web-logs that were generated by the users. Web servers can be set up to maintain access logs that are recorded passively (i.e., without user disruption or intervention). It may be tempting to treat these logs as accesses originating from “machines” (i.e., computers), utilizing concepts from manufacturing operations. However, this approach is less than ideal since there are many personal characteristics involved in the way individuals use websites.

The COBAE instructor evaluation site captures and records individual student responses regarding course mission, curricular content and faculty evaluation. Although progress has been made in implementing and streamlining this process and each semester some instructors use the online method, the application is still in active development and refinement. Understanding students’ behavior in filling out these online forms should enable the technological and operational aspects of the process to be improved for future use. Another benefit of web-log analysis is that it provides an “arm’s length” quasi-audit feature of the application. The web-log only records meta-data about the transaction, that is, data about the student evaluation data. The authors of this paper see pure transactions, not individual names or content. (Note that there is no indication that the students do not trust the confidentiality of their individual responses.) The COBAE student evaluation application was designed and developed primarily by Professor Ken Chapman (2000). The web server for the online evaluation application uses the Microsoft Internet Information Server (IIS). For each “hit”, IIS records a number of fields of interest, including: Date, Time, Client IP Address, User Name, Service Name, Server Name, Server IP Address, Server Port, Method, URI System, URI Query, Protocol Status, Win32 Status, Bytes Sent, Bytes Received, Time Taken, Protocol Version, Host, User Agent, Cookie, and Referrer.

Next, the paper reviews relevant literature that provide the basis for the research issues and methodology. The COBAE evaluation website and its’ web logs will be introduced and various analyses carried out on the web-logs. Finally, implications of the research and how this study can be used in teaching IT will be discussed.

Literature Review and Related Work

Although there has been a plethora of research dealing with instructor evaluations, the vast majority has dealt with the (important) psychological aspects of the validity and reliability of their use as measurement instruments. For instance, Marsh (1987) conducted a meta-analysis where he reviewed approximately 200 prior studies of instructor evaluations. Many of the articles he quoted were themselves meta-analyses of other research. Marsh considered a large number of constructs, concluding that “student ratings are multidimensional, quite reliable, reasonably valid, reasonably uncontaminated by many variables often seen as sources or potential bias” (p. 369), and that students, faculty and administrators view the evaluations as useful tools. However, he also concluded that they may suffer from some degree of halo effect (e.g., instructor’s personality having an effect on teaching effectiveness scores) and that there is always uncertainty involved in any sort of psychological measurement.

Similarly, McKeachie (1997) reviews a number of studies in an effort to summarize their validity as tools for promotion and tenure decisions. He suggests that most potential biases (such as giving higher grades) are quite limited when it comes to their effect on scores. For instance, if an instruc-

tor teaches poorly but gives good grades in an effort to obtain high evaluations, students will likely see through this and correctly evaluate the instructor lower. He also suggests the halo effect on evaluations may actually increase their validity. For instance, if an instructor conveys to students that he/she cares about their learning, this likely has a positive effect on learning, justifying higher evaluation scores. McKeachie concludes that “student ratings are the single most valid source of data on teaching effectiveness” (p. 1219) and that they are more valid than many personnel committees give them credit for. However, it is not appropriate to simply use the average or median score to make tenure and promotion decisions; scores should be interpreted by the instructor and by the committee. The distribution of student ratings is important, with consideration given to the proportion of students who rated the instructor good or excellent, and the proportion who were dissatisfied.

The above studies were not focused on the *process* of carrying out evaluations. In fact, in the large Marsh (1987) study, there was only a very brief mention of administrative aspects (anonymity of the students, purpose of the ratings and the point during the term when the evaluations are given), and these do not address the process issues discussed in this research.

Online Evaluations

There is very little published regarding the study of online instructor evaluations, which is likely due to their relative lack of use to date. Likely the most comprehensive treatment is given by Reid (2001), where he summarizes the need for and usefulness of online evaluations, discusses characteristics that make such a system valuable (such as confidentiality, ease of use, etc.) and reports on a case study where such a system was used. The purpose of the system was to provide feedback to improve instruction and monitor whether a course was improving over time. Thus, it was not used for promotion and tenure decisions. In this environment, the system effectively allowed instructors to customize the questions for their classes, and thus they were given access to the actual program for this purpose. Reid also points out that the system is very cost-effective, but that any such system will need to remain flexible and stay up-to-date in terms of technology and other changes in the learning environment.

Yin-Sum & Tak-Wing (2002) report using online evaluations to evaluate a new, entirely online course. The focus here was experimentation with 100% online delivery of a course, not on the evaluations. They concluded that online course delivery can be as effective as traditional classroom delivery. However, the class used for the study had only eight students, which may have made it easier for the instructor to respond more quickly to students online than if the class was larger. Thus, in the future the results should be verified with a larger class. They make no conclusions about the effectiveness of the online evaluations or the process of administering them.

Comparing Scores for Online and Paper Evaluations

Other researchers have compared average scores and student response rates between online evaluations and traditional paper responses (Baum, Chapman, Dommeyer, & Hanna, 2001; Dommeyer, Baum, & Hanna, 2002). Among other things, they found that average scores are the same or slightly higher when online evaluations are used. However, response rates are significantly lower when evaluations are done online. On the other hand, a different study reported “satisfactory” response rates when courses have significant online components (Reid, 2001), suggesting that if students are accustomed to going online for other aspects of course delivery, they are more inclined to also carry out evaluations online. In the Dommeyer & Baum studies above, one factor tested was the attempt to increase response rates by providing a grade incentive for students who filled out the evaluations. This was quite successful in improving response rates to a level similar to the in-class paper method (i.e., no statistical difference in response rates), even

when the grade incentive was very small (e.g., $\frac{1}{4}$ % of the final grade). Also, students seem to write more comprehensive qualitative comments online than they do on in-class paper evaluations, which is very encouraging.

Not all the news is positive, as some question the ethics behind providing bonus points for evaluations. On the other hand, sometimes the smaller sample of students can result in biased results. Despite these challenges, there is little doubt that eventually online evaluations will be fully implemented at COBAE. Barriers to implementation will be dealt with because the cost-effectiveness and administrative time saved are so attractive.

This current research does not consider actual scores or ethical considerations, but looks at student behavior on the site from a technical standpoint. As such, the remaining literature review focuses on other technical studies.

Web-log Analysis

Spiliopoulou (2000, p. 128) writes, “The only information left behind by many users visiting a web site is the trace through the pages they have accessed.” In this environment, traditional service quality evaluation techniques can be implemented, but they come at a high price. These techniques may include finding a group of representative test persons to use the site and establishing an experimental environment for them that will promote similar behavior to what would occur when they are in other, more typical life settings. Very few companies have the resources to launch this process each time they want to monitor a site’s quality. Thus, different methodological approaches may be needed for online, virtual service environments.

Turner (2003) discusses the Analog program, which is one of the most popular open source web-log analysis programs currently available. Analog is a general-purpose reporting tool that generates statistics (counts and proportions) on fields in the log file such as the number of successful page requests, monthly, weekly, daily and hourly usage patterns, referring IP address, organization from which the user logs in, etc. A listing of the report options available from Analog is given in Appendix 1. A simple Internet search reveals a number of other programs available for analyzing web-log files that include various functionalities. For instance, the major statistical software vendors (SAS and SPSS) are developing sophisticated new tools to help researchers and web site designers analyze web-logs (Whiting, 1999).

However, it is often necessary to understand much more than the basic frequency distributions generated by these types of programs. A web page designer may be interested in the actual traversal paths users take. This is a large methodological jump from generating web-log statistics, and we know of no open source software that performs detailed “clickstream” analysis on web-logs. Clickstream analysis refers to the suite of techniques used to understand and predict individual traversal patterns within one or more web sites (Monticino, 1998).

In some cases, the “mental model” for how the web page designer believes users traverse the site is very different from the actual behavior. Therefore, accurate and relevant traversal path information can be useful in improving the web application delivery environment. Finally, an adaptive site (one that is capable of being customized to each individual) benefits from (or even “requires”) knowledge of prior web site traversal behavior (Perkowitz & Etzioni, 1997). Often, just knowing how users *leave* the web site system is important information (Andersen et al., 2000).

Even before analysis starts, a number of basic data management tasks confront the web data-mining researcher. Both Spiliopoulou (2000) and Joshi et al. (1999) indicate that web server log files need extensive preprocessing. Web-logs can be missing records due to client-side caching, server-side caching, intermediate network caching or errors. Two of the most difficult problems are: distinguishing what is a unique, single transaction “session” (Andersen et al., 2000), and

uniquely identifying users through one or more sessions (Spiliopoulou, 2000). These are not intended to invade privacy in any way, but rather to link clicks together in an accurate and anonymous manner.

A number of data mining approaches for analyzing web-logs have been offered, including traditional relational database star schemas where data are held centrally and can be accessed for different purposes using various types of queries (Kimball & Merz, 2000; Joshi et al., 1999). In this case, “fact” tables are generated to extend the schema models and aggregate groups of small or very small sequences of clicks (Andersen et al., 2000). This is an exhaustive approach that generates all possible traversal paths, even if they are never used. Joshi et al. (1999) takes a different approach by suggesting that “typical user sessions” are easier to identify than an exhaustive set of mouse-click associations. Probably the most sophisticated open source clickstream analysis program was an experimental program called “SpeedTracer” created by the researchers at IBM (Wu, Yu, & Ballman, 1998), but this program is no longer available for download. Speedtracer works by finding the “maximum forward traversal paths” (p. 95). Most promising, by combining the results of a “clickstream analysis” with other information about a commercial transaction, spatial diagrams (“information visualization”) can be generated to help managers understand the entire service or purchase cycle (Lee & Podlaseck, 2000).

The logs of a commercial web server, particularly one involved in robust e-commerce, can be very large. A nationwide, online book reseller generates a 300 GB log file each day (Dutta, VanderMeer, Datta, & Ramamritham, 2001). And although data mining techniques such as “sequences” and “associations” have been routinely performed in the last two decades (especially for “market basket” analysis of shopper behavior), the algorithms for evaluating traversal paths are not well known. The study of identifying, enumerating, and understanding traversal paths is a complex mathematical undertaking in itself (Chen, Park, & Yu, 1998). In general, few empirical studies to understand web behavior exist, especially for web searching (Jansen & Spink, 2000). Given the degree of processing power needed, the secondary storage requirements, and the application of sub-optimal algorithms, it is little wonder that few researchers have undertaken comprehensive clickstream analysis.

The Current Study

For this research, the site developed in the COBAE is studied both to learn about students’ behavior and as an example site to glean insights for broader implications. Nearly all of the empirical research in this area attempts to formulate analytic techniques, methods, and algorithms that can be widely generalizable within the context of a large, unstructured web site (Fong & Wong, 2002; Montgomery, 2001; Yang, Huang, & Ng, 2003). This “needle in the haystack” approach might appear appealing, but the results are difficult to come by and even more difficult to diffuse. Our research, on the other hand, uses a case study to rigorously analyze each specific path of a known, time-bound, relatively small web system. Then, in future research it may be possible to incrementally build upon that knowledge to develop efficient analytical algorithms for larger systems. Basic statistics and frequency distributions are studied, and student traversal paths (the order in which they hit various pages) through the site are studied as well. Although there are commercial programs that could handle the traversal path analysis more quickly (discussed below), one of the goals of this research is to gain an in-depth understanding of web-logs (instead of just getting the results from a “black box” approach).

The COBAE Evaluation Site

The COBAE instructor evaluation site is “open” for a full two-week period during the last two weeks of the academic term. This site has a number of pages, as shown in Figure 1. On the site-

entry page (rootfolder/ in Figure 1), students enter their student ID and password for the site. If it is their first time on the site, they go to new.asp. Those that have been before but forgot their password go to forgot.asp. The evaluations are accessed from ticket.asp and filled in under survey.asp. Finish.asp is accessed twice; once to record and save the answers to the multiple-choice portion, and once for input and recording of any qualitative comments.

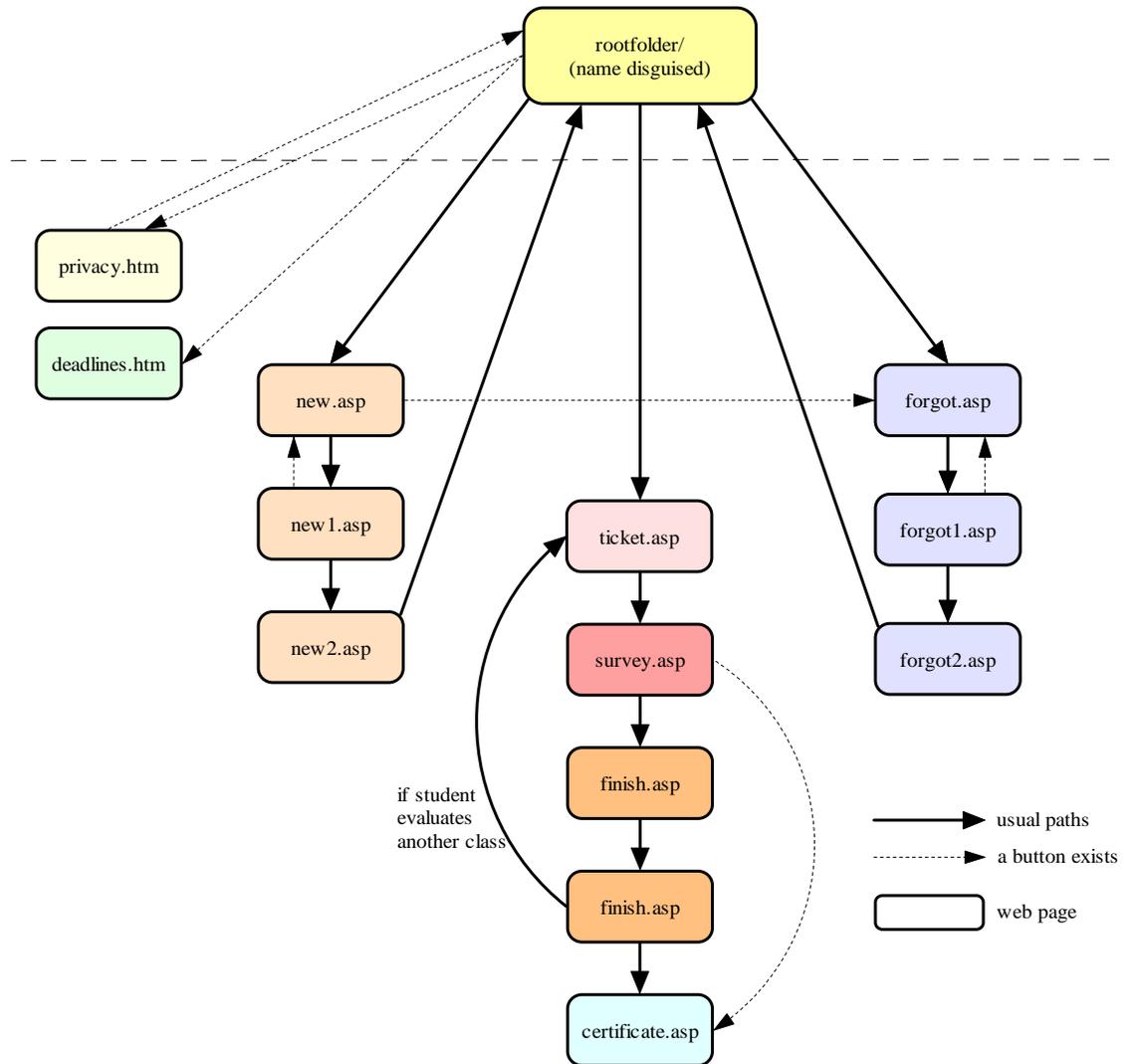


Figure 1: The Student Evaluation Site

Students have the option to print out a certificate proving they completed the evaluation (the certificate does not list the scores). This certificate can be given to the instructor if bonus points are given for evaluation. Students will usually follow the path designated by the arrows, but they can jump anywhere on the site by typing the URL of the page they want, using an earlier bookmark, or by using the “Back” button in their browser. Note that if students type a URL directly or use a bookmark, site security restricts them from actually filling out evaluations or getting a certificate until after they have signed in with their password.

Most of the pages are identical regardless of the specific course the student is evaluating, but two of the pages are customized for each student; `ticket.asp` only lists the courses the student is

currently enrolled for, and `survey.asp` is unique by department. Although each student has different courses and each department has a unique set of questions for their own courses, the pages show up in the web-logs as shown in Figure 1 regardless of the class being evaluated. Also, if a student makes an error on a page, an error message with the same page address may be generated.

The web-logs include a summary of every individual access of the evaluation site and all the pages accessed during the visit. Appendix 2 shows a few records from the COBAE logs. The logs do not include students' responses to questions or identify individual students, but rather the pages accessed, the time taken on each, and the order in which they were used. The data also include the numeric IP address of the student's Internet-connected PC when they log on, the domain (e.g., EarthLink, AOL, a specific educational institution, etc.) and the top-level domain (e.g., .net, .edu, .com, .gov, etc.) the request came from, the filename that is requested, whether or not the page was successfully downloaded, the number of bytes sent, and other data related to the web page request from the user's browser. The web-log includes one entry for each web page accessed (occasionally there is more than one entry per page if the page includes a graphic (e.g., .jpg, .gif) that is loaded with the page), with a column for each type of data recorded. Thus, if a student accesses 10 pages (i.e., 10 files) to fill out evaluations for various courses, there will be 10 records for that student, plus additional records for any graphics. Note that students are restricted to analyzing each course only once – the system checks their student ID number and once they have completed evaluating a course, they are restricted from accessing that course again.

Methodology

First, some basic user statistics were calculated using the open-source software *Analog* (Turner, 2003). This software was able to analyze the un-edited logs, and we had it calculate various statistics such as which days and hours during the week students used it, where they logged in from, etc. It was apparent that analyzing every single page would have been cumbersome, confusing and would not add much value. As such, statistics were calculated only on successful uses of `finish.asp`.

Secondly, traversal paths were examined. However, before paths could be determined, unique student visits had to be identified. This required extensive work to prepare the logs. MS Excel was used for this analysis in order to facilitate an understandable and appealing visual format, have the ability to easily implement changes to the model, and facilitate basic statistical analysis. Many database functions in Excel were used; future work will likely utilize a more advanced programming language to achieve more efficiency.

One way to determine traversal paths in a spreadsheet / database environment is to group all hits from each user session together. (The logs are not initially organized this way because if two or more users are on the site at the same time, the log records hits strictly in time sequence.) However, before hits can be sorted by session, it helps if non-student hits and other unnecessary data are deleted. Thus, rows that had administrative information other than hits to the site were deleted. Next, since graphics on a page show up in the logs as a separate record from the actual page they appear on, all .gifs and .jpgs were removed. The data were then analyzed to identify and delete records that were obviously not student hits. For instance, the following were deleted: hits from administrators (using the Client User column), incomplete IP addresses, page requests from robots and spiders, unsuccessful attempts to download pages (based on the HTTP status), records with undefined cookies, and other instances where something went wrong with the use of the site.

After this, there were 27,807 records remaining (a small web-log). One of the benefits of working with a fairly small site that is only open a few weeks each year is the ability to work with all hits

in a comprehensive fashion. It is hoped this will help provide a basis of understanding for working with larger web-logs.

Next, a program was written to format some of the data. For instance, the times were changed to a 24 hour clock, and the root page reference and .asp extension were truncated from the page request column. Then, all unnecessary columns were deleted, leaving only the date, time, client IP address, page request, and cookie. Since IP addresses may be the same even for different users, the cookies were used to identify unique users. The data were sorted by cookie, by day, and by time (start time on the page). This ensured that hits from unique visits were listed together, in order of their occurrence.

With the data in this format, it was then possible to calculate the time spent by each student on each page by subtracting the start time of each page from the start time of the following page. The only exception to this occurred for the last page a student hit. This page had no following page (at least no following page on this site), so the time on this page could not be calculated.

After this, a program was written that searched through each unique cookie and built the path for each unique visit, using the “page requested” field. A qualifier was added such that if the time on a page was more than 30 minutes, the session was considered terminated. Although the precise time for this might vary, the use of 30 minutes as a cutoff is relatively common in web-log analysis (e.g., Anderson et al., 2000). Thus, at times hits with the same cookie were divided into more than one session if the delay on one page was more than 30 minutes. In addition, total time on the site for each user was calculated.

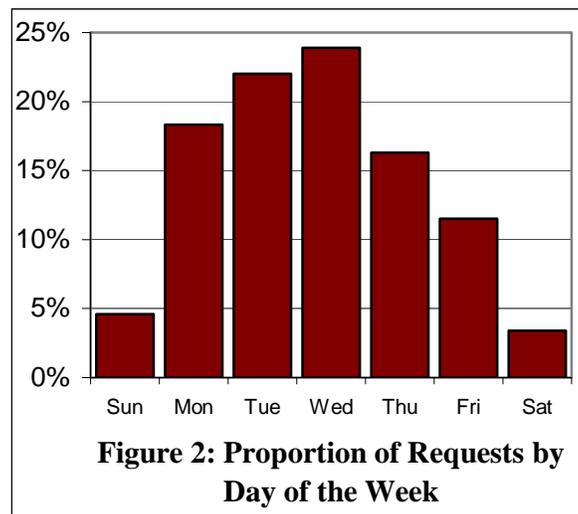
Finally, the most common paths were identified. This is particularly difficult because, while a computer program can easily tell if two values or text strings are the same or not, it is very difficult to look for things that are “almost” the same. Thus, using all of the final paths determined above, various analyses (using Excel’s Sort, Count and CountIf functions) were done to shed some insight on the paths. This was a relatively manual process compared to the programs written above.

Results and Analysis

User Statistics

During the two-week period, there were 6,736 successful uses of `finish.asp`. Since each student actually accesses this page twice (explained above), this means 3,368 student-classes were evaluated (this number is low because many professors in the College did paper evaluations this particular term). Figures 2 and 3 show what day of the week and time of the day students did evaluations, and Table 1 shows the frequency of requests from various domains.

Not surprisingly, most students did evaluations on weekdays between 8 am and midnight. Table 1 shows that a little less than half the students logged on



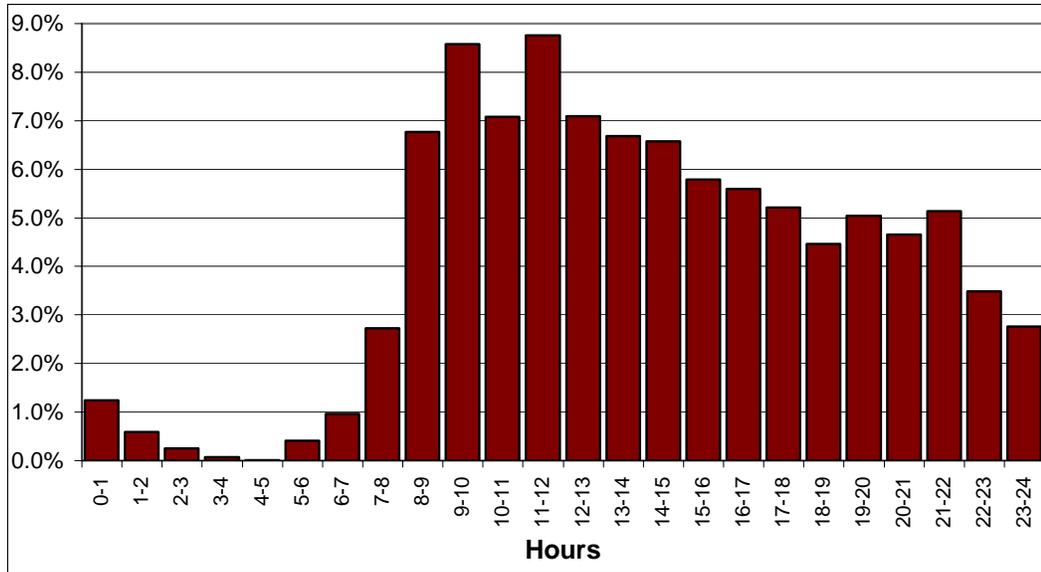


Figure 3: Proportion of Requests by Hour

from educational institutions (about 99% of these are from CSUN). Thus, as expected, many students either use the school computers or log in from home using the university server to get on the Internet.

There was, however, one possibly surprising result. Before this study was carried out it was felt many students would take advantage of the 24-hour availability, as well as the accessibility from home. Many did this (as evidenced by the number of evaluations done in the evenings), but even more did evaluations during the time of day when they have classes. The period between 9 am and 3 pm is the busiest time on campus (Klassen, Kumar, & Trybus, 2002), and this is when more evaluations were done than any other time (Figure 3). Thus, by deduction, it appears that many students do the evaluations on campus between class meeting times.

Table 1: Frequencies for the Most Common Domains

Extension	Domain	Frequency
.edu	USA Educational	44%
.com	Commercial	29%
.net	Network	20%
Unres	unresolved numerical addresses	5%
None	domain not given	1%
.us	United States	<1%
.org	Non-Profit Organizations	<1%
.mil	USA Military	<1%
.ch	Switzerland	<1%
.gov	USA Government	<1%

Traversal Path Analysis

Information from traversal path analysis sheds light on whether the site is efficient in helping students complete the evaluations in a timely fashion. It is also interesting to see if students complete evaluations once they start or if they jump around the site without completing them.

The 27,807 records (see Methodology) resulted in 3,154 final paths; thus, this is the total number of visits to be analyzed. Of these, there was a surprisingly high proportion of unique paths (1,192). This shows how differently users choose to use a website. Even on this simple website,

where the goal of each visitor is very similar, there are numerous different ways of working toward that goal.

Time on individual pages

Table 2 summarizes the time spent on individual pages. It does not include pages that were the last page accessed by a student; as explained above, it was not possible to calculate times for these. The page most affected by this is the certificate page; many students, naturally, end on this page. When considering the time to fill out the surveys, note that the surveys had on average 12 questions (this varied slightly for different departments). As is typical on an instructor evaluation, these questions were all short and easily understandable.

The time spent by students on each page varied significantly. It is likely that the minimum times in Table 2 are students who quickly clicked “through” the page without entering any data. These may be students just checking the site for later use, or times when they knew exactly what to do

Table 2: Time Spent on Individual Pages

	Page (*.asp)													
	root-folder	dead-lines	privacy	new	new1	new2	forgot	forgot1	forgot2	ticket	survey	finish	certif.-.icate	
Ave (sec.)	29.0	47.3	43.9	23.3	34.1	7.5	17.0	25.6	11.3	20.0	81.7	69.7	79.6	
Max (sec.)	1594.0	324.0	475.0	1303.0	1548.0	466.0	1399.0	762.0	506.0	1796.0	1630.0	1666.0	1714.0	
Max(min.)	26.6	5.4	7.9	21.7	25.8	7.8	23.3	12.7	8.4	29.9	27.2	27.8	28.6	
Min (sec.)	0.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
Count	2430	43	23	1791	1684	1172	801	751	711	4928	3745	5439	1135	

and thus were very fast. Not surprisingly, the longest average times are for the survey and finish pages. Students take on average 82 seconds to fill in a survey. This is a sufficient amount of time to answer approximately 12 multiple choice questions. As instructors, it would be nice to know that students spend considerable time thinking about their answers, but it is apparent that most students desire to get it done quickly. The maximum times shown suggest that some students do spend more time.

However, Figure 4 demonstrates how quickly most students fill out the evaluation. In fact, 63% of students take 75 seconds or less, 53% take 50 seconds or less, and 32% take 25 seconds or less. The evaluations vary in length, but most have between 10 and 20 questions. If we assume it takes 2-3 seconds to read and comprehend each question, this suggests that students taking less than 25 seconds may not be reading the questions carefully, and instead using an overall feeling for the course to quickly evaluate it. As mentioned above, this “halo effect” may actually have positive implications on validity (McKeachie, 1997).

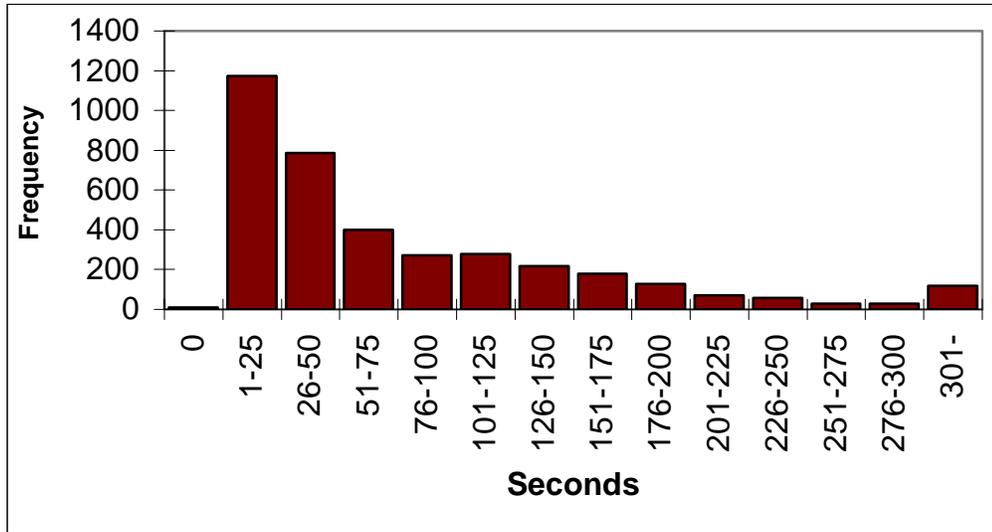


Figure 4: Time Spent by Students to Fill out the Evaluation (`survey.asp`)

Now turning to the `finish.asp` page, the average time spent (69.7 seconds) is somewhat misleading, since the finish page is actually accessed twice by each student; once for the computer to record the quantitative scores and once for them to fill in qualitative comments. If we assume that the computer will record the quantitative scores very quickly, and filling in the qualitative comments takes longer, it may be valid to divide these data at the median, with the shorter half representing the computer processing times. The average time for the shorter half of the data is 5 seconds, suggesting that this is the amount of time it took the server to process the information. For the second half, all times over 10 minutes were omitted as anomalous, with the average of the remaining times being 110 seconds (out of 6604 recorded `finish.asp` times, 131 were over 10 minutes, with the longest being 21 hours). One minute and 50 seconds seems to be a sufficient amount of time to provide useful qualitative feedback. This was also evidenced by the number of comments made by students – in two independent studies, Dommeyer et al. (2002) and Baum et al. (2001) found that there were actually more comments made on the online forms than on the in-class evaluations.

Students also spent a great deal of time on the certificate page, although this was probably not time working on the page. More likely, it was the time they waited for the printer to print the certificate, or they simply left their browser on this page when they were finished evaluating a particular class without immediately closing the site.

Identifying common paths

With 1,192 unique paths, the challenge becomes identifying the most common paths. As mentioned above, this is particularly difficult because, while a computer program can easily tell if two values or text strings are the same or not, it is very difficult to look for things that are “almost” the same. Thus, using the original 3,154 final paths, various analyses were done to shed some insight on the paths. The most common paths are shown in Table 3.

Table 3: Most Common Paths

	Path	# of occurrences
1	rootfolder/new/new1/new2/ticket/survey/finish/finish/certificate	306
2	rootfolder/new	233
3	rootfolder/ticket/survey/finish/finish/certificate	217
4	rootfolder/	125
5	rootfolder/ticket/survey/certificate	111
6	rootfolder/ticket/new/new1/new2/ticket/survey/finish/finish/certificate	92
7	rootfolder/ticket	71
8	rootfolder/new/forgot	58
9	rootfolder/forgot	53
10	rootfolder/new/new1/new2/ticket/survey/finish/finish/certificate/ticket/survey/finish/finish/certificate	48
11	rootfolder/new/new1/forgot/forgot1/forgot2/ticket/survey/finish/finish/certificate	44
12	rootfolder/ticket/ticket/new/new1/new2/ticket/survey/finish/finish/certificate	38

For complete understanding, Table 3 should be analyzed in conjunction with Figure 1. Paths 1, 3, 6, 10, 11, and 12 are paths where students fill out evaluations (as denoted by the /ticket/survey/finish/finish sequence in each). Path 3 is a typical, efficient visit, where a student signs in and evaluates one course. Path 1 is a typical visit for a first time user, first choosing a password and then evaluating one course. Path 10 represents new users who evaluated two courses before logging off. Paths 6 and 12 are also new users, except the students first went to `ticket.asp` before realizing they had to go to `new.asp` to get a password. Paths 6 and 12 represent why it is so difficult to identify similar paths; these two are identical, except in path 12 the students hit `ticket.asp` twice (possibly reloading the page) before going to `new.asp`. Path 11 represents students who forgot they had used the system in the past. They go to `new.asp` trying to enter a “new” password, but the site informs them they already have one and directs them to `forgot.asp`.

There were many close variations of paths 1, 3, 6, 10, 11, and 12. For instance, a number of students generated path 1 with the addition of “/ticket” on the end. This suggests that they were a typical new user, except that once they were done, they went back to `ticket.asp` either by mistake, or to see if they had any other courses to evaluate. At times other pages are also “added on” at the end of sessions. Also, many times a page shows up twice in a row (e.g., /survey/survey). This can occur if a student reloads the page, but can also occur in the normal execution of the program; there are some instances in the program where certain actions cause the same page reference to come up with slightly different information for the student. All of these cases could be considered the same path as 1, 3, 6, 10, 11, or 12, but don’t get recorded as such. Thus, there are likely many more instances of these paths than denoted above. This is explored further later.

Now considering the other paths in Table 3, path 5 occurs if a student has already evaluated the course but did not print a certificate proving they were done. In this case they can return later,

Table 4: Statistics for Individual Pages in the Traversal Paths

	Page (*.asp)					
	forgot	new	ticket	survey	finish	certificate
Number of paths with the page	786	1852	2549	2320	1929	2071
Proportion of paths with the page	24.9%	58.7%	80.8%	73.6%	61.2%	65.7%
Number of paths that start with the page	194	1277	1511	4	9	0
Proportion of paths that start with the page	6.2%	40.5%	47.9%	0.1%	0.3%	0.0%

sign in, choose the course in `ticket.asp`, go to `survey.asp`, but then jump to the certificate in order to print it. Paths 2, 4, 7, 8, 9 represent visits where students did not fill out evaluations. It is assumed that many of these are students who are going to the site to verify that they can find it (and possibly book marking it) for later use. For path 7, it is possible that a student signs on, gets to `ticket.asp`, but finds that there are no courses to evaluate online (or no remaining courses). This could occur because some instructors still use paper evaluations. Students may not recall which courses are online and which are paper, and may try to log on to evaluate a course that is not online.

Table 4 shows the number and proportion of paths that have at least one instance of the page noted. Note that these do not match statistics in Table 2 above, because in Table 2 the last page of each session was left out, and also, in Table 4 a page may occur twice in a session but would only be counted once for that path.

Table 4 should be interpreted remembering that these are any occurrence of the page, without regard to what order they appeared. Thus, every use of each page may not have been successful due to being accessed in an incorrect sequence, and interpretation of these data is somewhat ambiguous. However, some inferences can be made. The proportions suggest that 61.2% of visitors to the site finish the evaluation. (It is unknown how many visitors actually wanted to fill out an evaluation; some may have been book marking the site for future reference.) The reason `certificate.asp` is higher than `finish.asp` is likely because, as stated above, a student returning to just get a certificate (having filled out the evaluation earlier but choosing not to print a certificate from `certificate.asp` immediately) will go directly from `survey` to `certificate`, skipping the `finish` pages. Interestingly, it appears that 58.7% of users are new, and 24.9% of students forgot their password (although recall that some students that forgot their password went first to `new.asp`). Thus, out of all returning users (41.3% of users), approximately 60% forgot their password (24.9/41.3). This is not surprising, since for most students the last time they used it would have been 5 months earlier in the prior term. In any case, the site is set up to accurately identify students whether or not they remember their password. Finally, considering the page that students visit first, it is apparent that most new users correctly go to `new.asp`, but that many who have forgotten their password try a different password (bringing them to `ticket.asp`) or try entering a new password. Thus, the site seems fairly efficient in directing new users and returning users that know their password to the right location.

Table 5: Statistics for Key Page Combinations in the Traversal Paths

	Page Combinations	
	ticket-survey-finish	ticket-survey-finish-certificate
Number of paths with the pages noted	1917	1772
Proportion of paths with the pages noted	60.8%	56.2%
Average time on site	8.2 min.	8.3 min.

Considering page combinations can provide more accurate information on those that actually filled out questionnaires. Table 5 shows some additional statistics for the commonly expected partial-paths for those students that actually fill out an evaluation.

It is not surprising that all students did not get certificates, since some instructors did not give bonus points, and thus the certificate was not required. The data suggest that approximately 61% of visits resulted in completed evaluations, which is virtually the same as the figure above denoting how many reached the `finish.asp` page. Also, we see that most students spent just over 8 minutes on the site. This includes signing in, getting a new password or getting help with an old one, choosing courses to evaluate, finishing the evaluation(s) (including qualitative comments), and getting the certificate(s) printed out. This average includes single and multiple evaluations, but a perusal of the logs suggests that the majority of users do only one evaluation per visit. Thus, it seems that evaluating a

course takes approximately 8 minutes. This is typically less time than that allowed for in-class evaluations. Some students do spend a considerable amount of time, as the maximum time spent was 70.2 minutes. This student was a new user who evaluated three courses, and apparently unsuccessfully tried to evaluate two others (as evidenced by a slightly convoluted traversal path). Figure 5 shows the breakdown of time spent by users that completed evaluations.

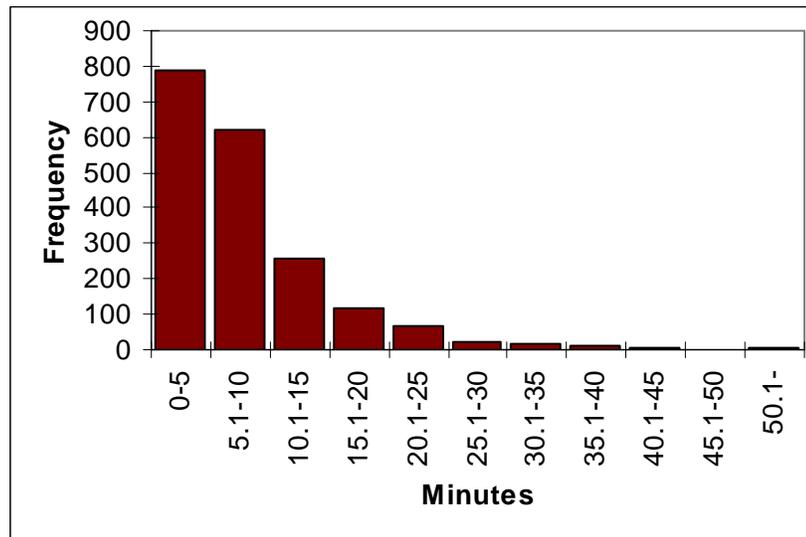


Figure 5: Total Time Spent on the Site by Students who Completed the Evaluation

Figure 5 demonstrates that most students do not need an extended time period on the site. In fact, 74% take 10 min or less from start to finish, and 42% take 5 minutes or less.

Implications and Conclusions

There were four main goals in this exploratory research; to learn about student behavior when filling out online instructor evaluations, to learn about analyzing web-logs in general, to provide the basis for discussion in class, and to promote further thinking and research in this little-known area.

User Behavior

The site is effective in helping students quickly carry out evaluations. Even though the majority of users needed a password or had forgotten their existing password (up to 83%), the average time to completion was just over eight minutes, and many (42%) required less than five minutes. Also, it seems most students spend enough time to conscientiously fill in the answers. The majority spend more than 25 seconds on the multiple choice, and almost two minutes (110 seconds on average) filling out the qualitative comments. It is apparent that some students probably fill in the multiple choice answers without reading the questions carefully, but this apparently happens for in-class evaluations as well (Marsh, 1987; McKeachie, 1997). This observation supports the conclusion (from other studies) that a halo effect exists (i.e., that some students give an instructor a similar score on all questions without thinking about each question individually). Also, it appears that most visitors are able to accomplish their goal of completing the evaluation. Sixty percent of users do complete evaluations, and if those “just looking” (in preparation for future use) were deleted (this figure is unknown), we would expect that the proportion of successful completions would increase. In fact, there were very few complaints from students regarding trouble with the website.

Thus, we can conclude that this was a well-developed web site for its purpose. It is not a “fancy” or “flashy” website, but it has the necessary links for students, and sends them to the appropriate pages if they go in the wrong direction (i.e., if they try to fill out a survey without signing in, or if they forgot their password). In addition, because the data are stored in a well-designed database, it has made it extremely easy to calculate statistics by instructor, by department, by course, or any other categorization desired. Also, as Reid (2001) found with his on-line evaluations (reviewed above), this system saves an incredible amount of paperwork and time for administrators, and provides department chairs and instructors with more comparative information than the prior in-class system.

The purpose of this research was not to evaluate the effectiveness of online evaluations compared to paper evaluations. However, as noted earlier in the paper, others have independently addressed this question (e.g., Baum et al. 2001 and Dommeyer et al., 2002), and found that scores are similar between the two methods, the key being the need to get a high enough response rate that appropriate statistics can be generated.

Web-log Analysis

Now considering web-log analysis in general, it is apparent that the spreadsheet analysis provided significant insights because it allows for many intermediate “checkpoints” in the analysis and is very flexible allowing output of whatever is of interest. However, a more advanced programming language would possibly be more efficient and may be easier to adjust for handling larger web-logs. It is desirable to develop a general program that could analyze a wider range of types of web-logs; for this, a spreadsheet approach would be quite inefficient. As stated by Chen et al. (1998), a major challenge is to develop a program to identify similar (but not necessarily identical) paths. Although insights from a spreadsheet approach can provide guidance, some form of artificial intelligence or genetic algorithm will be required to accomplish this in a reliable and

flexible manner. This remains a difficult challenge, as evidenced by the lack of open source software available to determine traversal paths. Thus, a major research area is the development of flexible and efficient analytical algorithms that can determine traversal paths in a wider range of web-logs.

Future analysis would also benefit from some prior planning on the part of the web-log administrator. In this research, the user sessions were identified solely by the cookies generated from the user's computer. It would be easier to program the logs so that markers would be included that would explicitly and uniquely identify individual sessions. Having markers will also reduce the number of conditions and steps it takes to generate the final paths. This final point highlights the fact that web-log administrators may want to include markers in all logs to make identification of individual sessions easier.

Pedagogical Uses

The information in this paper can also be used for an in-class discussion in a business IT or IS class regarding web-logs. The question could be posed: "How do users use a website?" Since the context and purpose of the site used in this research is already understood by students and instructors, and because this is a fairly simple site, the material in this paper could provide the basis for some rich discussion regarding the benefits and challenges in analyzing web-logs, how to use a simple spreadsheet to analyze them, benefits and drawbacks of using a spreadsheet, other methods to analyze web-logs, and other methods to analyze user behavior.

The discussion could be made even more relevant by giving students an assignment to analyze a small web-log (10,000-30,000 hits). They could be asked to duplicate some of the analyses in this paper. An introductory class could download and use the Analog program, and a more advanced class (especially a programming class) could be asked to write a program for traversal path analysis (either in Excel or using another language).

Besides learning about web-logs, this type of work could help students: realize the complexity inherent in web-sites and their administration, understand the use and impact of robots on the web (students need to sift through those hits), and understand how the various pages of a site work together. This could be a real eye-opener, since many students are only familiar with the easy-to-use point-and-click interface most web pages have today and do not understand what actually happens when a "click" is made on a website.

In future pedagogical research, it would be interesting to see if this perspective on web-logs does actually increase student's understanding of the complexity of how the Internet works. Studies could be done that test "before and after" teaching this module, or by using a control group that does not get a module on web-log analysis.

Finally, this article can also provide the basis to discuss how to access and use data in a more general sense – there is so much data available to businesses today, but very little is analyzed or used effectively for decision-making purposes.

References

- Andersen, J., Larsen, R., Giversen, A., Pedersen, T., Jensen, A., & Skyt, J. (2000). Analyzing clickstreams using subsessions. *Proceedings of the ACM third international workshop on Data warehousing and OLAP*, McLean, VA, 25-32.
- Baum, P., Chapman, K., Dommeyer, C., Hanna, R. (2001). Online versus in-class student evaluations of Faculty. *Proceedings of the Hawaii Conference on Business*.

- Berry, L. (1996). *On great service: A framework for action*, New York, NY: The Free Press.
- Chen, M., Park, J., & Yu, P. (1998). Efficient data mining for path traversal patterns. *IEEE Transactions on Knowledge and Data Engineering*, 10 (2), 209-221.
- Davenport, T. (1999, September). The eyes have it. *CIO Magazine*.
- Dommeyer, C., Baum, P., & Hanna, R. (2002). College students' attitudes toward two methods of collecting teaching evaluations: In-class vs. online. *Journal of Education for Business*.
- Drott, M. (1998). Using web server logs to improve site design. Proceedings of the *Sixteenth Annual International Conference on Computer Documentation*, 43-50.
- Drucker, P. (1993). *The practice of management* (Reissue edition). New York, NY: HarperBusiness.
- Dutta, K., VanderMeer, D., Datta, A., & Ramamritham, K. (2001). Discovering critical edge sequences in e-commerce catalogs. Proceedings of the *3rd ACM conference on Electronic Commerce*, (McLean, VA), 65-74.
- Fong, J. & H. K. Wong. (2002). Online analytical mining of path traversal patterns for web measurement. *Journal of Database Management*, 13 (4), 39-61.
- Haksever, C., Render, B., Russell, R., & Murdick, R. (2000). *Service management and operations*. New Jersey: Prentice-Hall.
- Jansen, B. & Spink, A. (2000). Methodological approach in discovering user search patterns through web-log analysis. *Bulletin of the American Society of Information Science*, 15-17.
- Joshi, K., Joshi, A., Yesha, Y., & Krishnapuram, R. (1999, November). Warehousing and mining web-logs. Proceedings of the *Second International Workshop on Web Information and Data Management*, 63-68.
- Kimball, W. (2000). The special dimensions of the clickstream. *Intelligent Enterprise Magazine*, 3 (2).
- Kimball, R. & Merz, R. (2000). *The data webhouse toolkit: Building the web-enabled data warehouse*. New York, NY: John Wiley and Sons.
- Klassen, K., Kumar, A., & Trybus, E. (2002). Parking needs analysis for multiple facilities using multiple lots. *The Service Industries Journal*, 22 (4), 71-92.
- Lee, J., & Podlaseck, M. (2000, October). Using a starfield visualization for analyzing product performance of online stores. Proceedings of the *2nd ACM conference on Electronic Commerce*, Minneapolis, MN, 168-175.
- Marsh, H. (1987). Students' evaluations of university teaching: Research findings, methodological issues and directions for future research. *International Journal of Educational Research*, 11 (3), 255-388.
- McKeachie, W. (1997). Student ratings: the validity of use. *American Psychologist*, 52 (11), 1218-1225.
- Monticino, M. (1998). Web-analysis: Stripping away the hype. *Computer*, 31(12), 130-132.
- Montgomery, A. (2001). Applying quantitative marketing techniques to the Internet. *Interfaces*, 31 (2), 90-108.
- Perkowitz, M. & Etzioni, O. (1997). Adaptive sites: automatically learning from user access patterns. Proceedings of the *Sixth International World Wide Web Conference*.

- Reid, I. (2001). Reflections on using the Internet for the evaluation of course delivery. *The Internet and Higher Education*, 4 (1), 61-75.
- Spiliopoulou, M. (2000). Web usage mining for web site evaluation: making a site better fit it's users. *Communications of the ACM*, 43 (8), 127-134.
- Turban, E., Lee, J., King, D., Warkentin, M., & Chung, M. (2000). *Electronic commerce: a managerial perspective*, Upper Saddle River, NJ: Prentice-Hall.
- Turner, S. (2003). *Analog 5.03*. Retrieved June 18, 2003, from <http://www.analog.cx>
- Whiting, R. (1999, November 29). SAS goes after the e-intelligence market. *Information Week*.
- Wu, K., Yu, P., & Ballman, A. (1998). SpeedTracer: A web usage mining and analysis tool. *IBM Systems Journal*, 37 (1), 89-105.
- Yang, Q, Huang, J., & Ng, M. (2003). A data cube model for prediction-based web prefetching", *Journal of Intelligent Information Systems*, 20 (1), 11-30.
- Yin-Sum, S. & Tak-Wing, S. (2002). Development of an online taxation course: From design to evaluation. *The International Tax Journal*, 28 (4), 23-51.

Biographies



Kenneth J. Klassen is an Associate Professor of Operations Management at Brock University. He received a Ph.D. in Operations Management at the University of Calgary. His teaching and research interests focus primarily on operations management in services. He has developed introductory and advanced courses in service operations, and also taught computer simulation and statistics courses. Research involves modeling in a number of areas, including outpatient appointment scheduling, demand and capacity management for services, planning campus parking and food services, and web log analysis.



Wayne Smith is the Director of Technology in the College of Business Administration and Economics at California State University, Northridge and is also a doctoral student in the School of Information Science at Claremont Graduate University. His research interests are in the areas of advanced TCP/IP internetworking, information visualization, and strategic management in higher education.

Appendix 1: Summary of the Statistics Available with Analog

<i>General Summary</i>	- <i>contains overall statistics.</i>
Monthly Report	- lists the activity in each month.
Weekly Report	- lists the activity in each week.
Daily Report	- lists the activity in each day.
Daily Summary	- lists the total activity for each day of the week, summed over all the weeks in the report.
Hourly Report	- lists the activity in each hour.
Hourly Summary	- lists the total activity for each hour of the day, summed over all the days in the report.
Quarter-Hour Report	- lists the activity in each quarter-hour period.
Five-Minute Report	- lists the activity in each five-minute period.
Domain Report	- lists the countries of the computers that requested files.
Organization Report	- lists the organizations of the computers that requested files.
Host Report	- lists the computers that requested files.
Referrer Report	- lists the referrers (where people followed links from, or pages which included this site's images).
Referring Site Report	- lists which servers people followed links from.
Browser Report	- lists the browsers used by visitors.
Browser Summary	- lists the vendors of visitors' browsers.
Operating System Report	- lists the operating systems used by visitors.
Virtual Host Report	- lists the activity on each of the virtual domains at the site.
Status Code Report	- lists the HTTP status codes of all requests.
Processing Time Report	- lists the times taken to process successful requests.
File Size Report	- lists the sizes of files.
File Type Report	- lists the extensions of requested files.
Directory Report	- lists the directories from which files were requested.
Request Report	- lists the files on the site.

Appendix 2: A Few Records from the COBAE Logs

DATE	TIME	CLIENT IP	SERVICE REQUEST	SERVER NAME*	SERVER IP*	NETWORK PORT
5/5/2001	10:53:46 AM	63.22.174.214	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	10:53:51 AM	63.22.174.214	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	10:54:05 AM	63.22.174.214	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	10:56:32 AM	63.22.174.214	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	10:56:41 AM	63.22.174.214	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	10:56:44 AM	63.22.174.214	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	12:27:55 PM	203.149.1.58	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	3:13:08 PM	152.163.189.135	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	3:13:17 PM	152.163.201.202	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	3:13:35 PM	152.163.189.66	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	3:13:42 PM	152.163.201.202	W3SVC1	COBAE	130.166.yy.zz	80
5/5/2001	3:14:07 PM	152.163.201.202	W3SVC1	COBAE	130.166.yy.zz	80

CONT...

* Surrogates have been used to disguise some actual data. This does not affect the interpretation of this table.

REQUEST TYPE	PAGE REQUEST*	HTTP STATUS	WIN32 STATUS	SERVER RESPONSE	CLIENT REQUEST	SERVER ELAP
POST	/ONLINE/default.asp	200	0	89	595	125
POST	/ONLINE/ticket.asp	200	0	89	625	422
POST	/ONLINE/survey.asp	200	0	89	593	375
POST	/ONLINE/finish.asp	200	0	89	886	157
POST	/ONLINE/finish.asp	200	0	89	1443	1453
POST	/ONLINE/certificate.asp	200	0	89	621	312
GET	/online/Default.asp	200	0	0	457	141
GET	/online/Default.asp	200	0	0	356	141
POST	/online/ticket.asp	200	0	0	571	282
GET	/online/new.asp	200	0	0	399	0
POST	/online/new1.asp	200	0	0	548	203
POST	/online/new2.asp	200	0	0	611	140

CONT...

HTTP VERSION	TARGET SERVER*	USER AGENT (browser)
HTTP/1.1	cobae.univ.edu	Mozilla/4.0+(compatible;+MSIE+5.0;+MSNIA;+Windows+98;+DigExt)
HTTP/1.1	cobae.univ.edu	Mozilla/4.0+(compatible;+MSIE+5.0;+Windows+98;+DigExt)
HTTP/1.0	cobae.univ.edu	Mozilla/4.0+(compatible;+MSIE+5.5;+AOL+5.0;+Windows+NT+5.0)

CONT...

* Surrogates have been used to disguise some actual data. This does not affect the interpretation of this table.

COOKIE**	REFERRER*	TEXT IP ADDR*
	http://cobae.univ.edu/ONLINE/forgot2.asp	1Cust214.tnt1.canoga-park.ca.da.uu.net
	http://cobae.univ.edu/ONLINE/default.asp	1Cust214.tnt1.canoga-park.ca.da.uu.net
	http://cobae.univ.edu/ONLINE/ticket.asp	1Cust214.tnt1.canoga-park.ca.da.uu.net
	http://cobae.univ.edu/ONLINE/survey.asp	1Cust214.tnt1.canoga-park.ca.da.uu.net
	http://cobae.univ.edu/ONLINE/finish.asp	1Cust214.tnt1.canoga-park.ca.da.uu.net
	http://cobae.univ.edu/ONLINE/finish.asp	1Cust214.tnt1.canoga-park.ca.da.uu.net
	http://www.univ.edu/~busecon/	*
	http://www.univ.edu/~rk33999	xxxx.proxy.aol.com
	http://cobae.univ.edu/online/	yyyy.proxy.aol.com
	http://cobae.univ.edu/online/	xxxx.proxy.aol.com
	http://cobae.univ.edu/online/new.asp	yyyy.proxy.aol.com
	http://cobae.univ.edu/online/new1.asp	yyyy.proxy.aol.com

* Surrogates have been used to disguise some actual data. This does not affect the interpretation of this table.

** Private information