Student Work   Vol.1 No.1 June 1999

Web site usage monitoring –
Does it contribute to the bottom line?

O.G. Garrett
Lexinfo and Fairbridge Ardene & Lawton
library2@fairbridges.co.za
Postgraduate Diploma in Information Management
Rand Afrikaans University

Contents

---

1. Introduction

One of the strategies that is employed to evaluate a Web presence is the monitoring of site traffic, the number of (successful) visits to a site and, more particularly, the number of return visits. It is possible to monitor Web usage because the Internet server on which a site is hosted tracks all requests for pages. Sophisticated software is available to process this traffic information so that the site manager can analyse and interpret the data, enabling him to evaluate the site from both technical and content perspectives. This exercise will influence the way in which the site is developed and modified; it may also have a direct bearing on the operation of the enterprise itself.

The advantages of monitoring Web site usage, and the problems experienced with current methods of traffic analysis, are discussed. During the course of this discussion the following matters will be addressed:
- Why a Web site in the first place?
- What information can be established by monitoring a web site?
- What are the current methods of monitoring a Web site?
- What are the shortcomings of these methods?
- How does monitoring assist an evaluation of the effectiveness of a Web site?

These considered, the author then determines the central issue, how can monitoring of Web site usage impact on, or contribute to, the overall success or profitability of an enterprise, that is, how does it contribute to the infamous 'bottom line'?

2. Why a Web site in the first place?

The decisions regarding site monitoring depend largely on the nature of the site itself. To illustrate the complexity of the problem the author considers case studies of two business enterprises that range in purpose and consequently also in the approach to Web site construction. These differences give rise to different expectations that may in turn impact on the type of monitoring that is indicated.

Case 1: Wits Law School Site (http://www.law.wits.ac.za)

This is a non-commercial site that posts information of public interest. Specifically it publishes the judgments of the Constitutional, Labour and Land Claims Courts as well as reports of the South African Law Commission and other human rights bodies. It provides links to a number of prominent local and international legal sites. The site and its links are well maintained and it is the first port of call for many legal queries. The Wits Law Site clearly draws its visitors to locate case material that is posted by the site itself, and it leads the user to other appropriate sites by judicious maintenance of suitable links.

Since it is non-commercial, it does not carry advertising banners and it is not responsible to advertisers. As such its continued existence is not dependent on the number of visits and there is no need to publish its logfile data. Monitoring the site traffic is however helpful in determining client confidence with the services that are offered and eliciting information about potential services. The alerting service whereby a subscriber is informed electronically about new judgments, for example, provides the site manager with valuable market research data regarding its serious users.

Case 2: Amazon.com (http://www.amazon.com)

The virtual bookstore known throughout the Web world as Amazon.com is one of the great success stories of book supplier ventures. The Web site reflects the commercial nature of its operation by clearly advertising the latest offerings and by recommending specific book and/or music titles to users. The level of sophistication both in the marketing and distribution of titles has made it a household name. Web usage is carefully monitored and user profiles compiled. The key to access this information is stored in a cookie on the client machine that is automatically activated when the user revisits the Amazon.com domain. Personal information is stored in the cookie together with an ID. This information is transmitted to the server, which uses the ID to retrieve the user's profile and display customised information. This is mass customisation (Schonfeld 1998) at its most effective. The success of Amazon.com is partly due to the fact that its sales force neither slumbers nor sleeps: its service is available 24 hours a day, every day from anywhere that is connected. It is this effectiveness and efficiency that gives electronic commerce the edge over traditional commercial exchanges.

These two case studies illustrate the range from non-commercial through a commercial service-based venture to a strictly sales-based operation. In each case the Web site mirrors and supports the nature of the enterprise. An evaluation of the site is a de facto evaluation of the enterprise as a whole and is of strategic importance to an enterprise.

The reasons that can therefore be advanced for establishing a Web presence include client expectations, where clients are themselves Web literate, promotion of the enterprise and of its services as an adjunct to traditional advertising, expansion of the client base, more effective market penetration and greater operating effectiveness and efficiency. To these obvious advantages can be added the development of new markets and services, by-products of an interactive marketing strategy that electronic commerce facilitates. The net gain in increased sales of products and/or services is achieved at little or no additional expense. Details of a product or service are posted once and the bulletin board can advertise the product to thousands of potential clients over an indefinite period of time. No copy-based advertising can achieve this degree of exposure, and at negligible marginal cost. All of the above contribute to enterprise profitability and a healthy bottom line.

3. **What information can be established by monitoring a Web site?**

Web sites are constructed for a variety of reasons, but in a business environment the decision to establish a Web presence centres on increasing sales of products and/or services. Failure to monitor a site is a failure to monitor market response to the site, the product and the enterprise itself. Commercial enterprises expend a great deal of creative energy and money on market research; careful and judicious monitoring of Web traffic can yield equally valuable findings. Not only are existing services evaluated – either directly by the questionnaire method or indirectly by recording the number of times a particular file is requested – but untapped markets and new products can be determined as the user profile emerges.

Mariva Aviram (1998), writing for Builder.com, identifies three motivations for analysing Web traffic. These incorporate many of the reasons earlier advanced for creating and developing a Web presence, and are grouped under the headings of business development, marketing and advertising sales, and technical resource and capacity planning.

3.1 Business development

Business development yields important information about a client base, the satisfying of present demand and the potential for future development. Market demand for a product or service can be determined by monitoring site traffic, particularly the ability to track how a user came to the site, information obtained from the so-called referrer logfile, which pages he/she accessed, and the last page visited. Interactive capability to process suggestions that are sent by electronic mail or submission forms can also assist business development.

When a visitor accesses the LegiSmart Web site, he/she is invited to register in order to proceed through the pages of the site. While the literature indicates some market resistance to registration (Aviram 1998; Winett 1998), many of the reputable commercial sites now require registration, and it appears that a paradigm shift is in progress (Burger, D.J.I. 1998. Personal communication – Lex-Info, lexinfo@iafrica.co.za). Registration typically requires name, e-mail address, occupation (market sector) and position. These data are a valuable source of information regarding the client base (real and potential) and the market (real and potential). Analysis of the data and, more important, translation of these into strategic planning promote business development.

3.2 Marketing and advertising sales

Fully commercial sites that are supported by sponsorship and advertising have obligations to their sponsors and advertisers to give accurate information regarding distribution figures, or the equivalent in electronic parlance. Most sites are able to quote logfile statistics regarding the number of hits, or visits, and it is these that are considered by advertisers when they agree to put up banners. But as we shall see later these figures are not altogether reliable and need to be approached with caution. They are, however, useful to both advertisers and sales personnel, especially as a comparative tool.

3.3 Technical resource and capacity planning

When businessmen speak of the bottom line they use the term to denote the profitability of an enterprise. All business expenses must be monitored and evaluated, and the decision to continue paying for something is subject to its cost effectiveness. Web site development is no exception. The expenses in terms of man-hours and software and hardware expenses must be audited in the light of actual returns and the costs justified. The advantage of doing business on the Internet is the potential increase in sales at little or no increase in fixed costs. When additional copy is added to a site, it costs very little more than the fixed maintenance costs, but the information that is posted is available at all times to any potential client and represents an omnipresent salesman (Burger, D.J.I. 1998. Personal communication – Lex-Info, lexinfo@iafrica.co.za).

4. **What are the current methods of monitoring a Web site?**

Web usage monitoring falls into two categories: statistical information, known as basic metrics (Aviram 1998), and demographic information that yields information about individual and group expectations, and market trends. The first reveals trends on a broad scale; the latter more personal, in-depth information.

Basic metrics are available as standard logfiles generated by the server (Aviram 1998) and can include the IP address, date, time, files accessed, browser information and referrer. Logs include varying amounts of data about the client, but the most useful information that is collected records the number of visits to a Web site, the path taken by users to arrive on the site navigation within a site, files accessed, and the final page that was requested.

Demographic user statistics on the other hand are tied to a unique user ID rather than an (anonymous) IP address. This enables a unique user profile to be assembled, recorded and stored, usually in a database on the server machine. Demographic information can be obtained by conducting random surveys (Aviram 1998) and by registering all visitors. These may in turn become subscribers to a free (Wits Law) or a fee-paying (LegiSmart) service. They may also become customers (Amazon.com). The information obtained from users on repeat visits is of far greater value to a site manager than that obtained on a casual basis from random surveys. Users are routinely invited to comment on the site or the services offered by the enterprise, and this becomes the key to future product development.

It has been seen that basic metrics and profiles stored and retrieved via cookies are concerned with logging user visits, but the definition of visit is not clear. We need to examine this and other words that are associated with site usage.

4.1 Visits

A visit is generally regarded as a session at a Web site made by one user in a continuous period of time (Meeker 1996). When the user requests pages, the unique session ID in the cookie is sent with the request, and the server logs this information. By looking for all requests with the same session ID, a report specific for that user can be created (Garratt, N. 1998. Personal communication – AGC Information Systems, neil@agc.co.za). When the user is inactive for longer than a set period of time – anywhere between ten and 30 minutes – the log regards the session as terminated. When the user reactivates the flow of information, he/she can be logged on with a new IP address and a second visit is recorded. This phenomenon can exaggerate the number of visits and confuse the statistical analysis.

4.2 Hits

A hit is any request for a file that is received by a server (Winett 1998). This may be in response to a search conducted on the user's behalf by a search engine, or it may be as the result of following links. It provides useful statistical information about the most popular files posted on a Web site. This information can be saved in database format so that the marketing department can analyse the data and use it to help map market trends. This information is useful, but is not always reliable (Winett 1998).

4.3 Pageview

This method of logging a session counts the hits (requests for files) from one page (or site) as a pageview (Winette 1998). This occurs when a user requests multiple files (images on a page, for example) from one page or site. Rather than log each request separately, the entire page request is logged as one pageview.

4.4 Logfiles

Data relating to visits, hits and pageview requests are counted (or logged) and sent to a logfile. Standard logfiles contain a modest amount of statistical information - IP address, date, time and names of files requested - and take up a modest amount of space. They run quickly and are not costly to manage (Garratt, N. 1998. Personal communication – AGC Information Systems, neil@agc.co.za). Other extended logfiles such as the NCSA and W3C extended logs can gather and store a great deal more information, but process time is extended and is consequently more expensive. Not all servers can handle extended logs (Garratt, N. 1998.

Personal communication – AGC Information Systems, neil@agc.co.za).

4.5 Referrer log

Most servers keep a referrer log that contains the page that the user last viewed. It tracks the movement of a user between linked pages so that it can inform the site manager of the link that was used to bring a user to the homepage.

4.6 IP Address

The primary tool for tracking web traffic begins with logging the IP address. When a user goes online, a unique IP (Internet Protocol) address is allocated to him for the duration of the session. Once the session is terminated, the IP address can be re-allocated to another user within the organisation. This means that if a user reconnects to the Internet later he may have a different IP address, and the server would log this as a new user session (Garratt, N. 1998. Personal communication – AGC Information Systems, neil@agc.co.za).

Logfile data regarding the visit, referrer log and navigation within the site itself are logged by the server and can be retrieved and packaged by one of the many software options that is available commercially, or on shareware or on the browser itself (Aviram 1998; Garratt, N. 1998. Personal communication – AGC Information Systems, neil@agc.co.za).

The more sophisticated of these products save information to a searchable database (Aviram 1998). Examples quoted in the literature are Accrue Insight, Andromedia Aria and I/PRO's Netline. Netline can send reports in graphical or textual form to be read in commercial software such as Microsoft's Excel (Aviram 1998). Other features that a Web manager might consider in the choice of software are user-friendliness, flexibility - the ability to modify both form and content of the report, and the capability to combine traffic data with other business information. An example of this feature is the facility to determine that a user is responding to a special offer on a product or service (Aviram 1998). This has important implications for the marketing department of a commercial enterprise. MB Interactive, for example, conduct random surveys among users to confirm market trends or determine untapped markets. Software like Accrue Insight can track a user click by click and can detect when a file transfer is stopped at the request of the user (Aviram 1998). When analysed on a broad scale this can be an invaluable diagnostic tool in the detection of problem areas, such as irrelevant or outdated material, or technical problems such as graphics taking too long to download.

Far from creating mass markets, electronic commerce is creating mass customisation. 'Companies with millions of customers are starting to build products designed just for you' (Schonfeld 1998). Amazon.com illustrates the principle perfectly. A user registers with the company, giving some personal information and expressing interest in particular books (or music CDs) or topics. He/she may order a title and authorises the company to charge it to his/her credit card. This done the customer awaits the delivery of the order. When he/she subsequently requests access to the Amazon.com site, he/she is greeted by name as a valued customer. Based on the titles of earlier orders, the customer is given a list of suggested new titles. If the customer does not know what to choose, he/she can participate in an online questionnaire that will elicit information about his/her taste in books (or music). Based on the responses, he/she is then given a list of recommendations. If he/she chooses, he/she may request a critical review on a particular title. This is marketing at its best. We see the perfect salesperson, who remembers everything material about the customer, who knows all the best titles to suggest, who is infinitely patient and who is at the customer's beck and call all day, every day.

The user profile that is the key to this service excellence is gathered to, stored in and accessed from a so-called cookie.

4.7 Cookies

Cookies expand the abilities of HTTP by allowing information to be stored on a user's hard drive. Any information can be stored in cookies, but routinely they contain sufficient data only to identify the user on future visits. The server tracks any information that has been obtained by means of a questionnaire, typically name, gender, age group, occupation and position held,

for example, and binds it to a user ID. This can be either session or persistent ID, depending on the way in which the logging process is set up (Garratt, 1998). Other technical data that such as computer platform and browser is also routinely captured on the HTTP header.

Client suspicion about cookies is largely unfounded. Cookies are not a security threat since they do not contain the information but rather hold the key to information that is stored on the server.

5. What are the shortcomings of these methods?

Web usage monitoring is important, and the logfile results in particular are frequently posted on the homepage. It is, however, unfortunate that there is still a lack of consensus on what constitutes a visit. Statistics that are not qualified are therefore treated with some suspicion.

Visits have traditionally been regarded as the means for monitoring Web site usage. But this logging of visits is by no means straightforward. A user may, for example download material, disconnect while he reads the page, and then return online to access a further file or page. Depending on the length of time he is disconnected, the return online can be logged as a new session.

There are other situations where the statistics are not logged accurately. The logfile that records every hit or file request as a visit will distort the figures upward when one compares statistics with a site that logs visits in terms of pageviews. Where a page for example comprises a number of graphical images and frames each one that is downloaded can in some instances be logged as a hit. Then again when the proxy server caches pages a request for these pages by multiple users will be logged by the server as originating from the one proxy server (Garratt, N. 1998. Personal communication – AGC Information Systems, neil@agc.co.za). Reallocation of an IP address to another user can also distort the log as the server may regard this as a continuation of the earlier session.

The inaccuracy of logging statistics is widely accepted by information technologists. Businessmen therefore have to look to alternative strategies to monitor site usage and usefulness. It is perhaps less important to consider the technical shortcomings than it is to understand the uselessness of statistical information without (strategic) action (Burger, D.J.I. 1998. Personal communications – Lex-Info, lexinfo@iafrica.co.za).

6. How does monitoring assist an evaluation of the effectiveness of a Web site?

Earlier it was stated that an evaluation of a Web site is in effect an evaluation of the enterprise as a whole, and is of strategic importance to the enterprise. Failure to monitor the site represents a failed opportunity to evaluate both the site and the enterprise. An enterprise may establish a Web presence at some considerable expense and effort. Unless it is able to evaluate it objectively there is no way in which the expense can be quantified and then justified.

Site monitoring is an objective evaluation of the site. A standard logfile logs the number of visitors to a site on any given day as well as the file/s he/she requests. These statistics indicate the degree of visibility of a Web site and, by recording which files were requested in each session, they reveal which services provided by an enterprise are considered to be most attractive to a visitor.

Interactive capability extends the type of information that can be determined by a viewer. Surveys and sales records that reveal demographic information in respect of age, gender, occupation and interests can be stored in a server database and retrieved on subsequent visits by means of the cookies that reside on the client machine.

The ability of a Web site to continue to attract visitors is a measure of the success of the Web site. The ability to attract users on a return visit is a measure of the success of the enterprise itself. If the visitor to the LegiSmart site completes the registration form and then accesses the files dealing with current legislation, it suggests that the site is meeting a user need. If he/she then proceeds down the levels from the list of titles to the pages dealing with long title, date of implementation and citation, this confirms both the value of the Web site and the service. When he revisits the site one assumes that the (free) legislation service it provides is able to meet his expectations. When, however, he requests information about other fee-based services it can be said that the Web site has begun to assume strategic value.

Putting a Web site up without monitoring the traffic flow and providing the mechanism for interactive comment leaves the Web manager unaware how it is received and how to improve it. Of greater concern, however, is the failure to utilise information about the role of the Web site as the mirror of the enterprise.

There is, as we have seen above, some dissatisfaction with the unit of measurement that is used to log Web usage – hit or visit or pageview. Unless the terminology and its meaning can be standardised there is no basis for comparative analysis and ratings are useless. These shortcomings are debated in the literature and alternative strategies are being developed to compensate. The third party audit is growing in popularity as a means to measure the claims of site managers when rating their sites (Meeker 1996). One of the better known audit packages is I/PRO. The Nielsen rating system, well known in American television broadcasting circles, has now entered the computer online arena and is associated with I/PRO in a widely subscribed evaluation service. According to Aviram, I/PRO gives regular unbiased data. This it achieves by requesting logfiles, doing random site visits and validation of URLs (Aviram 1998).

7. How does site monitoring improve the bottom line?

The bottom line expresses the financial state of an enterprise. The goal of business operations is to increase income and/or reduce expenditure in order to improve profitability. A Web presence does not reduce expenditure, but the rationale behind its development is significantly increased sales at relatively low increase in either fixed or marginal costs. This is enabled by the creation of the omnipresent salesman (Burger, D.J.I. 1998. Personal communications – Lex-Info, lexinfo@iafrica.co.za) an agent who makes available all material information about designated products or services at any time of the day from any location. The greater potential for market development that the Web facilitates remains underexploited unless the response of the market to the product is evaluated.

Web usage can be monitored by standard logfiles that log the basic metrics of those who visit the site, where they come from, and which files they request. While there is some strategic value in assembling this data, in themselves the numbers have no meaning. It is when the Web site requires its users to register that really useful information emerges, such as how many repeat visits occur as a percentage of total visits, or which files are consulted in depth, also expressed as a percentage of total requests for a file. These are the statistics that assist the enterprise's management team in evaluating the success or failure of the Web site in marketing their products and services.

Surveys, registration forms and interactive capability creates profiles of users in terms of age, gender, income group, occupation, position, etc. This assists strategic planning with regard to the market, the client base and product development.

8. Conclusion

Web site usage monitoring helps to identify trends. When users return constantly to a particular page it confirms that there is a need for the service that is provided by the enterprise. Support for a service suggests secondly that there is scope for development. If LegiSmart's legislation service is well subscribed, there may be some justification for preparing it, for example, in hardcopy format. Thirdly, analysis of the site usage in conjunction with the

interactive comments of its users may indicate new ways of developing market penetration and expanding the client base.

9. **References**

Aviram, M.H. 1998. Analyze your Web site traffic. [Online]. Available WWW:http://www.builder.com/Servers/Traffic

Meeker, M. 1996. The Internet advertising report. Technology: Internet/new media. [Previously online at http://www.ms.com, but now no longer on the Web]

Schonfeld, E. 1998. The customized, digitized, have-it-your-way economy. Fortune, September 28, 1998:69-74.

Winett, B 1998. Tracking your visitors. Webmonkey/e-business. [Online]. Available WWW: http://www.hotwired.com/webmonkey/temp…eta=/webmonkey/98/16/index21meta.html

Disclaimer