Measurements on the Web: Where do they Lead ?

Brian O. Cusack,

School of Computing
Auckland University of Technology
Auckland, New Zealand
brian.cusack@aut.ac.nz

ABSTRACT

eBusiness has placed an increasing transaction load on the WWW and a demand for measurement models with forecasting potential.

This paper provides a literature review of various methodologies for measuring business activity on the Web, including hit counts, stress testing, customer profiles and innovative approaches. An assessment of the relative methodologies is also made to value the relative strengths and weaknesses of each approach. A framework is then elaborated in which business performance may be tested, and forecasting potentials assessed. The context of this paper is an on-going research project being undertaken by the author into eBusiness performance and forecasting methodologies on the Web.

KEY WORDS

Web Measurement, eBusiness Performance, Forecasting

1. INTRODUCTION

Measurement on the Web has traditionally been concerned with site and systems loadings, and judgements based on counts of targeted occurrences. The growth of eBusiness has placed demands on Web managers to become involved with Web analytics for business purposes. A recent study by PricewaterhouseCoopers (2000) showed that every dot.com eBusiness they examined inadvertently over or under-stated its fundamental site metric, that of visitor numbers. The consequence of "bad" measurement in this case, was linked to an erosion of business confidence in eBusiness activity, and a reduction in eBusiness transactions. Another researcher emphasized the importance of measurement (Dugan, 2000:1).

Too many Web businesses neglect their metrics and operate in the dark. Measuring certain things about your site, such as visitor traffic, page views, and return visits, is like breathing: If you are not doing it, you will not make it for long.

In this paper some innovative Web measures from the eBusiness perspective are reviewed to

juxtaposition more traditional Web metrics from computer systems and business studies. The growth of Web Analytics is also visited to report various methodologies being employed to bridge the gap between site management metrics and eBusiness metrics. The concluding section then turns to look at the implications of developments in Web metrics for forecasting eBusiness performance.

2. INNOVATIVE METHODOLOGIES

The best way to know whether a Web site is achieving its goals is to gather extensive traffic data — not just how many hits are made, but which pages are popular, who's visiting the site, when do they visit, and a host of other data that can give a clearer idea of what's going on. Metrics built on data can provide crucial insights into the health of a business. Measuring certain things about a site, such as visitor traffic, page views, and return visits, is fundamental to knowing the customer. Over and above fundamental measures eBusiness requires more sophisticated measures to, for example, guide business decision-making, and to retain control of volatile and ill conditioned activity. The following seven measures are examples of innovative and useful Web measures from a eBusiness perspective (Dugan, 2000:1-2).

2.1 Measure 1: The Front Page Stumble

Look at how much time your customers spend on the front page and correlate that with when they leave. In other words, find out if your home page drives away customers. You have 10 seconds to convince a customer to stay on your site.

2.2 Measurement 2: Dropouts

At exactly what point are customers leaving your site? Identify the pages that are driving traffic away from your site. Decide how valuable that information really is and remove the reason for leaving.

2.3 Measurement 3. The Clutter-to-Content Ratio

In kilobytes, what's the ratio of site content to clutter? Look at it from the customer point of view and be ruthless - assume advertising is considered clutter. The closer this ratio is to 1:1, the more likely customers will go elsewhere to find a simpler interface.

2.4 Measurement 4: Best Purchasers

Look at characteristics of top purchasers, those in the top 20 percent. How are they different from other customers? Consider how they got to your site, how much time they spend on it, and what tools and pages they use.

2.5 Measurement 5: Cross-Sales failure

Getting customers to buy additional products is a cross-selling strategy. Alienation can occur with irrelevant offers and customers leave. Don't destroy a relationship with a cross-selling strategies.

2.6 Measurement 6: Fulfillment

Measure the average time it takes to deliver a product from when the customers click on the Web site until delivery. Correlate the customers who never come back with this metric.

2.7 Measurement 7: Customer Loyalty

To measure loyalty compare the average customer to the customer who's received a promotional offer to draw them back. Customers who constantly need deals to entice them back would prefer another experience - change the baseline customer experience on the site to include the incentives offered.

3. MORE TRADITIONAL MEAS-

URES

3.1 Hits

Organizations that are able to capture and use accurate customer data online will have a distinct edge in retaining customers, increasing sales per customer and growing the business. Hits are the fundamental count of site activity, and the qualification of hits into types of request is the first layer of analysis. (See the Selected Reference List, “Hits”, below for 15 useful explanations.)

3.2 Customer Profiling

When a customer places an order over the Internet, it is far simpler and easier to have the customer complete just one or two simple questions, such as name and address, and have all the other relevant data - such as delivery preferences, buying history and other demographic and psychographic profile information automatically updated and retrieved. A complete and best fit data profile of the customer can be made available online and in real-time for relationship building and personal customer interaction. The ability to populate customer records with online data enables e-businesses to automate customer relationship management and minimise anonymous cash-based transactions. The customer profile hence provides a basic unit for analysis and metric construction. (See the Selected Reference List, “Customer Profiles”, below for 13 useful explanations.)

3.3 Stress Testing

Stress testing involves subjecting a system to unreasonable loads, and denying the system the necessary resources to process that load. Unit measures may be obtained for systems performance through stress testing. On the Web the sites may under go full-scale stress testing identify a site’s capacity constraints and pinpoint performance bottlenecks. Using test scripts and run tests to emulate a real-life load situation on an organisation’s Web site, definitive measures may be found for factors affecting Service Level Agreements or other performance thresholds. (See the Selected Reference List, “Stress Testing”, below for 9 useful explanations.)

4. WEB ANALYTICS

Web analytics is the monitoring and reporting of Web site usage. The analysis can provide a better understanding of the interactions between Web visitor and what the Web site offers, as well as optimizing the site for increased customer loyalty and business benefits. Web Analytics may provide answers to the success (or otherwise) of a promotion campaign if site traffic had made a simple count increase, or to answer more complex questions requiring the use of data mining or OLAP tools. A further development at the innovative end of the spectrum is the building of comprehensive models (math, data, heuristic, and AI) to provide confidence to forecast futures on the Web.

4.1 Data Collection

Three fruitful approaches to data collection for Web Analytics are network monitors, single-pixel solutions, and HTTP server log analysis.

4.2 Network Monitors

Network monitors perform packet sniffing through an application that reads each package of data called into view by the operating system. A network monitor could be installed on each Web server or a single one on traditional Ethernet LANs. A single network monitor could report on every HTTP event but workload capacities need to be balanced in order to maintain server performance. A network monitor can see everything, including client requests, server responses, cookies, and HTML files. It can also track stop requests issued from the browser, making it possible to list the pages that are taking too long to generate. It can measure the Web server’s response time to different requests. Some network monitors can report on content-related HTML tags and capture form data transmitted via a POST request when the visitor hits a submit button.

4.3 Single Pixel Solutions

One of the newer methods for Web site data collection is single-pixel technology which enables online data collection of page view information. It originates from the transmission of data during the request for a one-pixel image placed somewhere on a Web page. The data collected includes basic information similar to that logged by an HTTP server, as well as

client-side behavior. The information is collected when the enabled page is loaded in a browser, and then sent back to the analysis server for the analyzing and reporting processes. Single-pixel information collects information by page views, rather than by "hits," as recorded in HTTP server logs. It collects the following data:

- time it takes the page to load;
- whether or not the page loads with errors;
- whether or not the page load is aborted by the user;
- referring page of the loaded page (if there was one);
- link destination clicked to leave page (if there was one);
- usage information about any forms on the page;
- session state (for example, prior visits, duration, number of page views, and so on);
- traversal path; and,
- user defined data.

4.4 HTTP Server Logs

This form of Web analytics involves the analysis of log files produced by the HTTP servers in a Web server environment. Logging capabilities can be controlled from within server software but a typical HTTP logging configuration would result in a detailed log entry for each HTTP request or hit to the server. Web analytics software can parse and process these logs files, in batch, to combine information from each request to give a view of the Web site's traffic, including basic metrics such as the number of hits, visitors, visitor duration, visitor origin (subdomain, referral link), visitor IP address, browser type and version, platform, and cookies, as well as more advanced metrics derived from the manipulation of data through techniques such as categorization and aggregation.

4.5 Data Treatment

Data warehousing, data-mining, and OLAP are key techniques in the data treatment process for Web Analytics. The large amount of data collected on eBusiness activity is stored in a Data warehouse. The aggregation of different data types provides a data-mining context, and the slicing & dicing of data over different dimensions gives the OLAP context. Effective data treatment, consequently requires the concurrent exercise of the three key techniques. For example, OLAP systems can report only on directly observed and easily correlated information. They rely on the user to discover patterns and decide what to do with them. To solve this problem, marketers and business analysts use data-mining techniques. These are machine learning algorithms that find patterns buried in databases and report or act on those findings. When Web data is combined with other enterprise data (such as customer and product profiles, order and fulfillment system data), it makes for a powerful business intelligence solution.

4.6 Strengths and Weaknesses

To determine which Web metrics analysis method (or combination of methods) is appropriate to a given situation the strengths and weaknesses of each need to be weighed. Issues such as cost (capacities and financial) and operational requirements influence decision-making. For example, log analysis implementations are generally less expensive than implementing network monitors or real-time analysis systems. Log analysis solutions give robust historical trending information, as well as the ability to filter, categorize, sessionize, and aggregate the results. However, there are management problems, including set up, resource consumption, and its not real-time. Alternatively, packet-sniffing solutions provide real-time capabilities but bring higher setup and maintenance costs, and could introduce risks to the production Web-serving environment. A combination of these methods is a plausible compromise.

5. EBUSINESS PERFORMANCE

EBusiness site performance can be measured by load test emulations of real customers. Tests may be run to find real-time online measures for response times, for transactions, and infra-structure component stress. The testing process brings together the trader, the customer, and the technical expert to optimize performance. Bottlenecks to optimal performance and meeting customer expectation can be pinpointed, their causes identified, and fine-tuning for best possible user experience completed. By taking this approach the major stakeholders in an eBusiness work together to create a working and winning situation. Typically optimization occurs on the five interfaces of:

- Data base tuning;

- Network bottlenecks;
- Load balancing;
- Server configurations; and,
- The Human Computer Interface.

Performance testing is an integral part of designing, building, and maintaining eBusiness applications. Validating the system architecture in the early stages of development is just as important as continuously monitoring its performance in the service environment. Considering scalability in the design phase validates that the planned architecture will handle expected traffic volumes. In the most basic terms, the final goal for any eBusiness application for high-volume use is for business users to consistently have the positive experience they expect. E-business applications demand non-disruptive growth, continuous availability, and consistent response times, even during peak usage times. There are five phases required for eBusiness performance to this standard (Shea, 2000:1):

- architecture validation;
- performance benchmarking;
- performance regression;
- performance tuning and acceptance; and,
- continuous performance monitoring.

6. CONCLUSION

Web metrics are important for the informational, operational, and marketing aspects of eBusiness. A business Web presence does not guarantee competitive advantage over conventional competitors, but Web analytics can create competitive advantage, and speed the evolution of a company. For example, recognizing that Web site behavior has a direct effect on business success and customer loyalty, acknowledges that a better understanding of users can make more responsive and interesting sites, that are easier to navigate and to make purchases. Optimal performance may be found in sites which are fit for purpose, and present trustworthy metrics.

Note

The help by Research Assistant Andrew Zimmer in compiling a comprehensive set of readings which support this paper, is acknowledged with gratitude.

REFERENCES

Dugan, S. 2000. Learn, know, and live them: The seven metrics of highly successful Web sites. InfoWorld. 22:26. 150.

Pricewaterhousecoopers(2000). Pricewaterhouse Coopers calls for industry standards for Internet 'Metrics'. M2 Presswire - 30 November 2000. 1-2.

READINGS

GENERAL

Making Sense of all that Data - http://www.zdnet.com/ecommerce/stories/main/0,10475,2671987,00.html

Tuning RED for Web Traffic – (PDF file named IEEE-ToN-01.pdf) http://citeseer.nj.nec.com/christiansen00tuning.html

New Metrics for New Media: Toward the Development of Web Measurement Standards http://www2000.ogsm.vanderbilt.edu/novak/web.standards/web-stand.html

Analyzing Web Site Traffic http://linz1.net/wdev/wdevArt38.html

Things to Consider When Selecting a Web Traffic Analysis Tool http://dcb.sun.com/practices/how-tos/web_traffic.jsp

Data's Dandy. But Do a Reality Check, Too http://clickz.com/article/cz.3726.html

When Readers Write... White Papers http://clickz.com/article/cz.3597.html

A Performance Study of Internet Web Servers (1996) http://citeseer.nj.nec.com/arlitt96performance.html

Testing Services http://dmoz.org/Computers/Software/Testing_Services/

Following is a list of resources for the software testing professional http://www.ondaweb.com/sti/stivend.htm

Network Performance - http://dmoz.org/Computers/Software/Networking/Network_Performance/

Samples of Practical eValid Applications http://www.soft.com/eValid/Applications/testing.html

Web Informant #180, 11 Decmber 1999: Measuring web performance http://www.strom.com/awards/180.html

Site Optimization - http://hotwired.lycos.com/webmonkey/templates/print_template.htmlt?meta=/

webmonkey/98/26/index3a_meta.html

Web site hosting and analysis - http://www.espion.ch/

Performance measuring products - http://www-svca.mercuryinteractive.com/

Five Steps to Managing the Performance of E-business Applications - http://www.stqemagazine.com/featured.asp?id=10

Web Site Test Tools and Site Management Tools - http://www.softwareqatest.com/qatweb1.html

Hits

C&W Adds Web Traffic Reporting Tool http://www.crmdaily.com/perl/story/?id=6154

Measuring Your Sales Success http://clickz.com/print.jsp?article=3416

Who's Counting? Ratings Data is Beginning to Stream In http://www.channelseven.com/newsbeat/2001features/news20010104.shtml#synd

http://www.stats4you.com/

Analyzing Website Traffic http://www.wilsonweb.com/webmarket/traffic.htm

Stuck in Traffic? http://clickz.com/article/cz.3158.html

Counting what counts http://clickz.com/article/cz.3436.html

Experience the Power of Conversion http://clickz.com/article/cz.3533.html

Not All Hits Are Created Equal: Cooperative Proxy Caching Over a Wide-Area Network (1998) http://citeseer.nj.nec.com/rabinovich98not.html

WebMeter - http://208.221.168.24/response.html

Counting the "Hits" A guide to measuring website traffic http://www.canadaone.com/technology/website-promotion7.html

Web Traffic Ratings Not There Yet http://www.tech-web.com/wire/story/TWB19980612S0022

Surfstats Log Analyzer - http://www.surfstats.com/

Hitometer - http://websitegarage.netscape.com/O=wsg/turbocharge/hitometer/index.html

Web-Site-Tools | SuperStats - http://www.web-site-tools.com/superst.htm

Customer Profiles

Measuring Customer Performance: What Counts? http://clickz.com/article/cz.3680.html

The Myth of the Average Customer http://clickz.com/article/cz.3699.html

Knowing Where to Go http://clickz.com/article/cz.3765.html

Links to academic papers:

Forecasting Repeat Sales at CDNOW: A Case Study - Peter Fader Bruce

Capturing Evolving Visit Behavior in Clickstream Data - Wendy Moe

Modeling the Evolution of Repeat Buying - Peter Fader Bruce

Discovery of Interesting Usage Patterns from Web Data - Robert Cooley Pang-Ning (1999)

Web Mining: Information and Pattern Discovery on the World Wide.. - R. Cooley, B.. (1997)

Using Data Mining Techniques on Web Access Logs to Dynamically.. - F. Masseglia, P.. (1999)

Hosted Web Performance Monitoring Service http://www-svca.mercuryinteractive.com/products/activewatch/

Nettracker - Who is coming to your web site? http://www.sane.com/ads/whoiscoming.html

Customer Profiling And Segmentation Services http://www.sms-direct.com/profilingsegmentation.html

Pegasus Solutions Debuts Web-Enabled Customer Relationship Management Tools - http://online.hitec.org/news/4005556.2000343.htm

Novantus Customer Profiling http://www.novantus.com/profiling.htm

SAS® e-Intelligence software earns Product of the Year award http://www.sas.com/news/preleases/011901/news1.html

St George - http://www.stgeorgeifs.com/solutions/profiling.html

Data Mining and Customer Profiling with good white page links http://www.vignette.com/CDA/Site/0,2097,1-1-1329-2075-1341-2235,00.html

Web Site Analysis Tools - http://www.zdnet.com/products/content/pcmg/1705/281511.html

Load Testing

Web Server Performance Analysis http://citeseer.nj.nec.com/398280.html

Measuring Web Performance in the Wide Area (1999) http://citeseer.nj.nec.com/barford99measuring.html

Load testing tools http://www.webperformanceinc.com/