Web Metrics Bibliography
Version 2

Compiled By
Anita Coleman and Chris Neuhaus

Educational Impact and Evaluation Standing Committee

National Science Digital Library

Table of Contents

Introduction

A review of the literature reveals that web metrics is a complex topic that can be found under many different terms and phrases: e-metrics, web site traffic measurement, web usage, web mining, online consumer/usage behavior, are just a few. Data mining, web analytics, knowledge data discovery, informetrics (bibliometrics and web-o-metrics) and business analytics are also relevant. 'Metrics' are measures and 'analytics' are measurements of information, processes and data analysis from processes but web analytics is also becoming synonymous for e-metrics and web metrics. 'Verticalization' is one of the newest trends in business web analytics/metrics; this means not just web traffic logging and analysis but has been broadened to include understanding and predicting customer behavior and customer relationship management. 'Personalization' is considered fundamental to improving customer motivation, satisfaction and retention.

What is the potential of web metrics for understanding educational use of the NSDL and measuring educational impact? As Einstein said, "Not everything that can be counted counts, and not everything that counts can be counted." What do we want to count for the NSDL how, and why? These are the questions that have motivated the creation of this bibliography. We hope it will be a useful starting point and reference document as we begin framing a plan of action for the NSDL in this important area.

Status: This bibliography is a work in progress. When it is completed (target date: 08/30/04) it will be a selective, annotated bibliography on web metrics. Currently, the abstracts in this bibliography are often taken directly from the source articles or websites and are not annotations. Books and journals dealing with this topic are not yet included (with one exception); we plan to include at least other texts and journals in the final version.

Acknowledgments: Chris Neuhaus jumpstarted this bibliography and Anita conducted a literature search in databases such as the ACM Digital Library besides editing it. In addition, we found the statements of the Webmetrics Workshop participants most helpful in preparing this bibliography and some of the references in the statements have been included here. We also acknowledge the labor of Shawn Nelson, SIRLS, University of Arizona, in locating the abstracts and articles/items listed in this bibliography. Your feedback and comments (especially critical comments and reader annotations about any of the items) will help to improve this selective, annotative bibliography and are greatly encouraged.

Version: This is version 2 of the bibliography prepared by volunteers of the Educational Impact and Evaluation Standing Committee for the NSDL Webmetrics Workshop (a joint workshop of the Technology Standing Committee and the EIESC), Aug. 2-3, Costa Mesa, CA. This version adds two tools mentioned at the workshop and includes citations to two papers that were distributed at the workshop as well. Version 1 of the bibliography, also a volunteer effort, was distributed as a paper copy to the 26 participants of the workshop.

For details about the workshop, visit http://webmetrics.comm.nsdl.org/.

This bibliography is being made available through DLIST, http://dlist.sir.arizona.edu/.

The Context – Digital Libraries, Educational Impact and Evaluation

Astin, A. and others. 9 Principles of Good Practice for Assessing Student Learning. *AAHE*. URL: http://www.aahe.org/assessment/principl.htm

ACRL Task Force on Academic Library Outcomes Assessment Report, (1998). Association of College and Research Libraries. June. URL: http://www.ala.org/ala/acrl/acrlpubs/whitepapers/taskforceacademic.htm Viewed: 07/08/04

Abstract: Report from the Association of College and Research Libraries (ACRL) Task Force on Academic Library Outcomes Assessment, which includes suggestions for incorporating teaching and learning outcomes assessment into ACRL standards, as well as for using them in other contexts.

Bertot, John C, (2004). Libraries and Networked Information Services: Issues and Considerations in Measurement. *Performance Measurement and Metrics, Vol. 5*(1), 11-19.

Abstract: The Internet is an integral part of library service that can take many forms - an extension of library collections and resources through licensed and/or digitized content, a gateway service through public access workstations, or as a means through which customers can interact with the library through such services as digital reference. Along with these evolving forms of electronic library services, there is a need to examine our ability to engage in a multi-faceted assessment of network-based information services and resources that includes input/output evaluation approaches as well as those grounded in service quality, outcomes, and other frameworks as determined by the data needs of the library. Information professionals, and those relying on information professionals, face a number of challenges in the networked information resources and services environment. Meeting these challenges requires a variety of issues and strategies for libraries to consider, particularly when engaging in evaluation activities.

Blixrud, Julia, (2003). Mainstreaming New Measures. *ARL Bimonthly Report* 230/231, October/December. URL: http://www.arl.org/newsltr/230/mainstreaming.html

Abstract: In the 1990s, escalating economic pressures and demands for accountability from governing boards, university administrators, and government agencies strongly encouraged ARL member libraries and many other libraries to find new ways to illustrate the contribution their collections and services provided to their constituencies. Traditional, input measures (i.e., size of collections, budget, and staff) were not meeting institutional accountability needs. Although efforts were in place to measure some forms of output--e.g., activities such as circulation, interlibrary borrowing, and group presentations--these, too, did not always provide a good or complete indication of library performance and the cost effectiveness of their services. In the 1990s, the ARL Statistics and Measurement Committee began discussing the need to move beyond the traditional measures, building on earlier work in organizational development conducted by ARL's Office of Management Services in the late 1980s and throughout the 1990s.[1] In early 1999, the ARL New Measures Initiative officially began to develop measures that would better demonstrate libraries' institutional value and describe their current operations. Now after four years of intensive work, the

New Measures Initiative is no longer new. ARL member libraries have integrated some of the measures into the mainstream of library operations and the Statistics and Measurement program has an active agenda for research, experimentation, and demonstration projects.

Fraser, Bruce T. and others, (2002). Toward a Framework for Assessing Library and Institutional Outcomes. *portal: Libraries and the Academy* 2 (4), October, 505-528. URL: http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v002/2.4fraser.pdf [subscription needed]

Abstract: The Association of Research Libraries (ARL) E-Metrics project sought to understand how academic libraries might specify, produce, and assess institutional outcomes. This paper reviews the findings from a discussion forum, site visits, an analysis and review of accreditation standards, and a survey of ARL member directors, and offers a framework for approaching the outcomes assessment process. The paper concludes by suggesting that much work remains to integrate outcomes assessment successfully in a university setting. Moreover, multiple approaches to assessment, of which outcomes assessment is but one, are still needed for a comprehensive assessment of libraries in the broader university and societal context.

Troll, Denise A., (2002). How and Why Libraries are Changing: What We Know and What We Need to Know. *portal: Libraries and the Academy* 2 (1), January, 99-123. URL: http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v002/2.1troll.pdf [subscription needed]

Abstract: In November 2000, the Digital Library Federation (DLF) and the Council on Library and Information Resources (CLIR) commissioned a white paper to initiate discussion of how and why libraries are changing. Eight academic library directors met with representatives from the DLF and CLIR in March 2001 to discuss the issues. The outcome of this meeting was a proposal to conduct research that will begin to fill significant gaps in our understanding and to better position libraries to meet the needs and expectations of academic users and university and college administrators. The white paper and the forthcoming research should be of interest to anyone trying to cope with or make sense of transformational change in academic libraries. To reach a broader audience, the white paper has been revised and updated for publication in portal: Libraries and the Academy. The article examines the importance of understanding how and why libraries are changing, analyzes the limitations and difficulties of traditional library performance measures, and explores environmental factors that may help account for why library use is changing. It concludes with an overview of research designed to develop an understanding of how user behavior and preferences affect demand for and use of library collections, services, and facilities, and a call to contribute conscientiously to the legacy of academic libraries and librarianship.

Young, Sheila and Blixrud, Julia. Research Library Involvement in Learning Outcomes Assessment Programs. *ARL Bimonthly Report* 230/231, October/December 2003. URL: http://www.arl.org/newsltr/230/learnout.html

Abstract: This article is a brief report of a working group that explored the question: How could research libraries participate in the assessment of learning outcomes at the institutional level and determine how to measure their contribution to the learning process? This is the question that was pursued by the ARL Learning Outcomes Working Group in 2003.

Definition, Origins, History, and Scope of Webmetrics

Blixrud, Julia. (2002, August 13-25). Measures for Electronic Use: The ARL E-Metrics Project. Presented at the Statistics in Practice IFLA Preconference. Loughborough, UK. URL: http://www.lboro.ac.uk/departments/dils/lisu/downloads/statsinpractice-pdfs/blixrud.pdf Viewed: 07/05/04

Abstract: Libraries are spending a larger proportion of their materials budget on electronic resources. In 2001, members of the Association of Research Libraries (ARL) spent more than $132 million on electronic resources. In order to determine if their dollars are being well spent, 24 members of ARL self-funded a research project to develop measures for describing the resources, expenditures, and usage of electronic resources. The project resulted in a set of recommended measures and a data collection process that will continue to be refined and tested by ARL members.

Coffey, Steve. (2002). Internet Metrics: The Loyal Audience. *The OPA White Papers, 1*(1). URL: http://www.online-publishers.org/opa_loyal_audience.pdf Viewed: 07/07/04

Abstract: The growing body of advertising research clearly demonstrates that online advertising is effective. It raises awareness, conveys advertisers' messages, and enhances brand perceptions. Additional research illustrates that increasing exposure frequency results in even greater advertising impact. However, despite all of the research proving the efficacy of the medium, the metrics to support buying decisions are still not firmly in place. Total Reach (monthly cumulative site audience or "unique visitors per month") has been the dominant comparison metric for the past five years. Unfortunately, it does not meaningfully inform advertising planning, buying or selling decisions, and may even mislead planners if used improperly. This paper uses custom analyses to make better use of existing measurement data for advertising planning purposes and provides some guidelines for thinking about the advertising planning process. As you will see, alternative statistics calculated from the existing audience measurement data offer a more appropriate understanding of the relative strengths of competing media packages.

Dhyani, Devanshu, Wee Keong Ng, and Sourav S. Bhowmick. (2002). A Survey of Web Metrics. *ACM Computing Surveys (CSUR),* 34(4), 469-503. URL: http://doi.acm.org/10.1145/592642.592645

Abstract: The unabated growth and increasing significance of the World Wide Web has resulted in a flurry of research activity to improve its capacity for serving information more effectively. But at the heart of these efforts lie implicit assumptions about "quality" and "usefulness" of Web resources and services. This observation points towards measurements

and models that quantify various attributes of web sites. The science of measuring all aspects of information, especially its storage and retrieval or *informetrics* has interested information scientists for decades before the existence of the Web. Is Web informetrics any different, or is it just an application of classical informetrics to a new medium? In this article, we examine this issue by classifying and discussing a wide ranging set of Web metrics. We present the origins, measurement functions, formulations and comparisons of well-known Web metrics for quantifying *Web graph properties*, *Web page significance*, *Web page similarity*, *search and retrieval*, *usage characterization* and *information theoretic properties*.

Interactive Advertising Bureau. (2002, January). Interactive Audience Measurement and Advertising Campaign Reporting and Audit Guidelines. *Managing the Digital Enterprise.* URL: http://www.iab.net/standards/measure_guide.pdf

Abstract: This document establishes detailed definitions for several key metrics used in Internet measurement and provides certain guidelines for Internet advertising sellers (herein referred to as "media companies" or "sites") and ad serving organizations for establishing consistent and accurate measurements.
Additionally, this document is intended to provide guidance to users of Internet measurements for understanding the origin of key metrics and a roadmap for evaluating the quality of procedures applied by media companies and/or ad serving organizations. The definitions and guidelines contained in this document originated from Phase 1 of the Interactive Advertising Bureau's (IAB's) Ad Campaign Measurement Project, conducted from May through December 2001, as explained below. Phase 2 of the project, quantifying the numeric differences in counts arising from various measurement options, is in process.

Kohavi, R. and others, (2002). Evolving Data Mining into Solutions for Insights: Emerging Trends in Business Analytics. *Communications of the ACM*, 45 (8): 45-48.

Abstract: The goal is business effectiveness through 'verticalization,' usability, and integration with operational systems.

Kohavi, R., (2001). Mining E-commerce data: the good, the bad, and the ugly. *Conference on Knowledge Discovery in Data***:** *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.* San Francisco, California, p. 8 - 13

Abstract: Organizations conducting Electronic Commerce (e-commerce) can greatly benefit from the insight that data mining of transactional and clickstream data provides. Such insight helps not only to improve the electronic channel (e.g., a web site), but it is also a learning vehicle for the bigger organization conducting business at brick-and-mortar stores. The e-commerce site serves as an early alert system for emerging patterns and a laboratory for experimentation. For successful data mining, several ingredients are needed and e-commerce provides all the right ones (the Good). Web server logs, which are commonly used as the source of data for mining e-commerce data, were designed to debug web servers, and the data they provide is insufficient, requiring the use of heuristics to reconstruct events. Moreover, many events are never logged in web server logs, limiting the source of data for mining (the Bad). Many of the problems of dealing with web server log data can be resolved by properly architecting the e-commerce sites to generate data needed for mining. Even with

a good architecture, however, there are challenging problems that remain hard to solve (the Ugly). Lessons and metrics based on mining real e-commerce data are presented.

Patton, Susannah. (2002). Web Metrics that Matter. *Computer World.* URL: http://www.computerworld.com/databasetopics/data/story/0,10801,76002,00.html Viewed: 07/07/04

Abstract: During the dotcom heyday, companies slapped sites on the Web and waited for traffic to pour in. They counted "eyeballs" and measured their site's "stickiness" as a way to convey the online real estate's value to advertisers. When the Internet bubble burst, "sticky eyeballs" seemed suddenly worthless. Now, as the Web has moved from being a technology pipe to a sales channel, companies such as Saab Cars USA need to update their Web measurement strategy with new metrics and analysis tools that can help them analyze customer behavior and improve their site's business success. But while hits were once the metric du jour, the new metrics are not so clear-cut. "There is no standard metric that a company can rely on for its Web site," says Randy Souza, an analyst at Cambridge, Mass.-based Forrester Research. "Metrics will be different from company to company." Where a retail site might be focused on conversion rate (the number of online shoppers who actually buy something), a business-to-business site might value site reliability and speed above all. In short, the most valuable metrics will depend on what you are trying to do with your site. Once that is determined, large enterprises should consider buying software to help analyze Web data, while midsize and smaller companies should consider a hosted service. "Web metrics should absolutely be the concern of the CIO even if it's housed in another part of the organization," says Gary Massel, vice president and CIO at Boise Office Solutions in Itasca, Ill. "If you're concerned about your customer, then you have to be concerned about your Web site's performance."

Rappa, Michael, *Webmetrics*. URL: http://digitalenterprise.org/metrics/metrics.html (Last updated: March 2004) Viewed: 07/08/04

Abstract: Digital technologies open the door to new possibilities in data collection and the analysis of business processes. As you travel the digital world, everywhere you go and everything you do leaves a trail of bits -- a virtual calling card, if you will -- that records your activities. Depending upon how you are connected to the Internet, those bits can be more or less easy to track back to the user. The fact that activity on the web is "measureable" is alone not enough: measurements need to be meaningful. Much work is needed to develop analytical approaches to the data that yield meaningful statistics.

Sen, Shahana, Balaji Padmanabhan, Alexander Tuzhilin, Norman White and Roger Stein. (2004). *On the Analysis of Web Site Usage Data: How Much Can We Learn about the Consumer from Web Logfiles?* Forthcoming in the European Journal of Marketing: Special Issue on Marketing in Cyberspace. URL: http://opim-sun.wharton.upenn.edu/~balaji/ejm.pdf Viewed: 07/07/04

Abstract: In this paper, we investigate the information needs of marketers on the Web for consumer analysis purposes and examine how these can be met. Grounding our work in the existing bases of segmentation in the literature, we propose new Web-specific segmentation variables, by operationalizing and extending the existing variables for consumer analysis and

segmentation. We discuss how marketers' information needs could be met from the different information sources available to a marketer, including Web logfiles, secondary sources, and information elicited from the consumer. We conclude that the readily available data in Web logfiles satisfies a significant part of the behavior segmentation information needs of marketers, while most of the demographic, geographic, psychographic and benefit segmentation information needs require information to be elicited from the visitors.

Sterne, J. (2002). Web Metrics: Proven Methods for Measuring Web Site Success. New York: J. Wiley and Sons, Inc.

Abstract: This book is bout measuring website success with customers. This book looks at the Web as a business tool for communicating with your customers. Sterne has been in sales and marketing for 20 years and a marketing and customer service consultant for the past nine of them. Therefore, this book is about measuring those places where the customer and the company come together: Advertising: Raise awareness; Marketing: Educate and persuade; Sales: Complete the transaction; Customer Service: Answer questions and solve problems. This is a business book for business people. This book takes a commonsense approach to what can be measured and business perspective about what should be measured.

Approaches, Projects, Methods, and Methodologies

Academic Web Link Database Project. URL: http://dlist.sir.arizona.edu/archive/00000099/01/index.html (Last updated: not stated) Viewed: 07/08/04

Abstract: This project was created in response to the need for research into web links: including web link mining, and the creation of link metrics. It is aimed at providing the raw data and software for researchers to analyse link structures without having to rely upon commercial search engines, and without having to run their own web crawler. This site will contain all of the following: Complete databases of link structures of collections of academic web sites; Files of summary statistics about the link databases; Software tools for researchers to extract the information that they are particularly interested in; Descriptions of the methodologies used to crawl the web so that the information provided can be critically evaluated; Files of information used in the web crawling process.

Ali, K. and Kethcpal, S. 2003. Golden Path Analyzer: Using Divide-and-Conquer to Cluster Web Clickstream Analysis. Conference on Knowledge Discovery in Data: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., 349 – 358.

Abstract: This paper describes a novel algorithm and deployed system Golden Path Analyzer (GPA) that analyzes clickstreams of people trying to complete the same task on a website. It finds the shortest, successful paths taken by users - 'golden paths' - and uses these as seeds for clickstream clusters. Other users are assigned to a cluster if their clickstream is a supersequence of the golden path. The advantages of this approach are that the resulting

clusters are easily comprehended, they are few in number, correspond to semantically different strategies used by the users, and jointly partition all the clickstreams. GPA's key contribution over prior work in process funnels is that by not excluding users that make diversions from the golden path, GPA is able to assign more users to fewer clusters. Another key contribution is to use actual full clickstreams as cluster seeds to which supersequences of other users are added. Golden paths correspond to complete clickstreams that are based on actual user page transitions. GPA is particularly useful for site designers to improve processes such as shopping, returns and registration. Its analyses identify which web pages cause many users to deviate from a golden path, which links distract users and the percentage of users taking each golden path. GPA has demonstrated value on more than twenty client projects in diverse industries.

Blixrud, J. and Kyrillidou, M. (2003) E-Metrics: Next Steps for Measuring Electronic Resources, *ARL Bimonthly Report* 230/231, October/December, 11-13. URL: http://www.arl.org/newsltr/230/emetrics.html
Viewed: 07/29/04

Abstract: Responding to user demand, libraries have steadily been shifting the focus of their collection development to the acquisition and licensing of electronic content, much of it via consortia. In the last decade, the average percentage of acquisition dollars that ARL member libraries directed to electronic resources rose from 3% to 20%. In 2002, 110 ARL university libraries reported spending more than $171 million on electronic resources, and 48 ARL libraries reported another $20 million expended on their behalf through centrally funded consortia.[1] In spite of considerable efforts, aggregate data documenting the quantity and use of these e-resources was proving to be elusive. The ARL E-Metrics work was established in 2000 to develop standard definitions for measures that libraries could use to describe: (a) the e-resources they make accessible, (b) the use made of the resources, and (c) the level of library expenditures.

Bollen, J. and Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. *D-Lib Magazine*, 8(6). URL: http://www.dlib.org/dlib/june02/bollen/06bollen.html

Abstract: At present, digital library (DL) policy is largely informed by management intuition and coarse measures of user satisfaction. Most DLs, however, maintain extensive server logs of user retrieval requests that contain a wealth of information on user preferences and the structure of user retrieval patterns. We propose a quantitative approach to DL evaluation that analyzes the retrieval habits of users to assess the impact of a collection of documents and to determine the structure of a given DL user community. We discuss a system that we have developed to automatically generate extensive journal and document networks from an efficient and simple analysis of user retrieval sequences registered in a particular DL's server logs.

Bollen, J. and others, (2003). Usage Analysis for the Identification of Research Trends in Digital Libraries. *D-Lib Magazine* 9 (5), May. URL: http://www.dlib.org/dlib/may03/bollen/05bollen.html

Abstract: The analysis of user logs from large-scale digital libraries offers new opportunities to assess research trends in an institution's user communities. We describe the application of a methodology to derive weighted journal relationship networks from reader logs at the Los Alamos National Laboratory's Research Library during 1998 and 2001. A journal impact metric is defined that derives journal impact from the structural features of the generated journal relationship networks, much in the same manner Google's PageRank evaluates the impact of web pages for a given subject on the basis of its context of hyperlinks to other pages. A comparison of this reader impact metric to the ISI Impact Factor values for the same journals in 1998 and 2001 allows us to detect and interpret community-specific research trends where the LANL community deviates from more general trends as indicated by changes in the Institute for Scientific Indexing (ISI) Impact Factors during those same years. Such analysis yields information to aid digital library managers to improve the evaluation of not only which parts of the collection are most highly valued by their local community, but it also detects research trends in user communities as they evolve over time.

Calore, Michael, *Log File Lowdown*. URL: http://hotwired.lycos.com/webmonkey/templates/print_template.htmlt?meta=/webmonkey/01/24/index4a_meta.html (Last updated: not stated) Viewed: 07/08/04

Abstract: Most users like to travel the Web incognito. They come to your site, poke around a few files, download a PDF or two, and then -- poof -- disappear, leaving nothing but questions in their wake: Where did they come from? Which browsers are they using? Are they experiencing any errors? Short of asking outright, there are ways to find the answers to these questions. The most celebrated method of tracking users is by planting cookies, which some folks consider rude or invasive, and, oh yeah, you need to know how to program them. Not to worry -- there is an option that requires very little technical know-how, comes at no (or nominal) cost, and may already be a part of your site's backend. I'm talking about logs! No, not those fancy browser-based publishing automation systems with a cult-like following. I speak of log files.

Cooley, R. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. In *ACM Transactions on Internet Technology* 3 (2): 93-116.

Abstract: The discipline of Web Usage Mining has grown rapidly in the past few years, despite the crash of the e-commerce boom of the late 1990s. Web Usage Mining is the application of data mining techniques to Web clickstream data in order to extract usage patterns. Yet, with all of the resources put into the problem, claims of success have been limited and are often tied to specific Web site properties that are not found in general. One reason for the limited success has been a component of Web Usage Mining that is often overlooked---the need to understand the content and structure of a Web site. The processing and quantification of a Web sites content and structure for all but completely static and single frame Web sites is arguably one of the most difficult tasks to automate in the Web Usage Mining process. This article shows that, not only is the Web Usage Mining process enhanced by content and structure, it cannot be completed without it. The results of experiments run on data from a large e-commerce site are presented to show that proper preprocessing cannot be completed without the use of Web site content and structure, and that the effectiveness of pattern analysis is greatly enhanced.

Crane, Allen S., *Actionable E-Metrics*. URL: http://www.intelligententerprise.com/030201/603feat2_1.jhtml?_requestid=643238 (Last updated: February 1, 2003) Viewed: 07/08/04

Abstract: In this article, I won't attempt to redefine the industry of e-commerce measurement tools, nor will I claim to offer a silver-bullet cure for struggling e-businesses. Rather, I'll present an original framework for how to measure and analyze real-world, relevant, e-commerce metrics, and how to present them in an actionable, executable format. Ultimately, this framework gives you the ability to: Rapidly diagnose site problems; Understand the elements of site conversion; Understand the relationships among units, revenue, and margin, and the behaviors that drive them.

Darmoni, Stefan J., Francis Roussel, Jacques Benichou, Benoit Thirion, and Nicole Pinhas. (2002). Reading Factor: A New Bibliometric Criterion for Managing Digital Libraries. *Journal of the Medical Library Association, 90*(3), 323-327. URL: http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=116406&action=stream&blobtype=pdf Viewed: 07/05/04

Abstract: This paper describes a way to measure the electronic consultation rate that the authors have termed the reading factor (RF), in direct correlation to the impact factor. RF measures the interest in a journal within the limit of a given readership. It is a direct measurement of journal use that is related to both clinical medicine and library or information science. Results of the observed distribution of RF for the year 1998 and IF for 1997 are shown in Table 1. The Rouen University Hospital (RUH) medical digital library was created in 1997. This library allows all 304 RUH senior physicians to access MEDLINE, forty-six

Downie, J. Stephen. (2002, July 13-17). *Worskshop on the Creation of Standardized Test Collections, Tasks and Metrics for Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation*. JCDL 2002 - Joint Conference on Digital Libraries. Portland, Oregon. URL: http://www.jcdl.info/jcdl02/menu.cgi?opt=sked-ws&F=1#W4 Viewed: 07/05/04

Abstract: There is a very strong feeling among the MIR and MDL research communities that we must somehow get our act together and acquire standardized collections of music (in multiple representations), standardized retrieval tasks (i.e., query sets, etc.), and standardized retrieval metrics. At the heart of all of this is very real concern that we are currently only "playing" at being objective and scientific insofar as our work is represented in our literature. One model being bandied about is the formation of a TREC-like set of collections and evaluation criteria. These collections and criteria will have to be designed specifically to address the wide variety of music-specific problems facing the MIR and MDL research communities. Establishing the ideal framework for developing these collections and criteria is the open question to be addressed at the Workshop. As it stands now, notwithstanding everyone's best efforts, MIR and MDL research is being conducted and reported upon in an ad hoc fashion with no mechanism for meaningful comparisons between proposed systems (i.e., each team has no standard way of comparing their results to the results of others). For example, we have absolutely no way right now to evaluate the relative merits of the various approaches when, for example, one group of researchers has a collection of MP3 rock songs, another 9354 folks in MIDI format, and yet another team uses 200 classical piano pieces in

GUIDO format. Add to this problematic situation, the fact that each team reports its results in whatever metric with which they might be most familiar. Since this is the case, MIR and MDL research and development is falling short of being truly scientific in those things where science--that is, the traditional scientific method--can help advance the state-of-the-art.

French, James C. and Allison L. Powell. (2000). Metrics for Evaluating Database Selection Techniques. *World Wide Web, (3),* 153-163. URL: none

Abstract: The increasing availability of online databases and other information resources in digital libraries and on the World Wide Web has created the need for efficient and effective algorithms for selecting databases to search. A number of techniques have been proposed for query routing or database selection. We have developed a methodology and metrics that can be used to directly compare competing techniques. They can also be used to isolate factors that influence the performance of these techniques so that we can better understand performance issues. In this paper we describe the methodology we have used to examine the performance of database selection algorithms such as gGlOSS and CORI. In addition we develop the theory behind a "random" database selection algorithm and show how it can be used to help analyze the behavior of realistic database selection algorithms.

Gonçalves, M., Panchanathan, G., Ravindranathan, U., Krowne, A., Fox, E.A., Jagodzinski, E., and Cassel, L. The XML Log Standard for Digital Libraries: Analysis, Evolution and Deployment. URL: http://www.dlib.vt.edu/projects/DLLogging/

Abstract: Log analysis can be a primary source of knowledge about how digital library patrons actually use DL systems and services and how systems behave while trying to support user information seeking activities. Log recording and analysis allow evaluation assessment, and open opportunities to improvements and enhanced new services. In this project, we propose an XML-based digital library log format standard that captures a rich, detailed set of system and user behaviors supported by current digital library services. The format is implemented in a generic log component tool, which can be plugged into any digital library system. The focus of the work is on interoperability, reusability, and completeness. Specifications, implementation details, and examples of use, are given below.
Griffiths, Jillian R., R.J. Hartley and Jonathan P. Willson. (2002). An improved method of studying user-system interaction by combining transaction log analysis and protocol analysis. *Information Research, 7*(4). URL: http://informationr.net/ir/7-4/paper139.html

Abstract: The paper reports a novel approach to studying user-system interaction that captures a complete record of the searcher's actions, the system responses and synchronised talk-aloud comments from the searcher. The data is recorded unobtrusively and is available for later analysis. The approach is set in context by a discussion of transaction logging and protocol analysis and examples of the search logging in operation are presented.

IBM Almaden Research Center, *Intelligent Information Systems.* URL: http://www.almaden.ibm.com/software/quest/ (Last updated: not stated) Viewed: 07/08/04

Abstract: The Intelligent Information Systems Research (aka. Quest) group, led by Dr. Rakesh Agrawal, IBM Fellow, is designing information systems that preserve the privacy and

ownership of data while not impeding the flow of information. Our work is motivated by the technical challenges posed by the emerging on demand world whose success is predicated on protecting the privacy, security, and integrity of interactions between individuals and enterprises as well as between enterprises.

Jones, Steve, Sally Jo Cunningham, Rodger J. McNab, Stefan Boddie. (2000). A Transaction Log Analysis of a Digital Library. *International Journal on Digital Libraries, 3*(2), 152-169. URL: http://www.cs.waikato.ac.nz/~stevej/Research/PAPERS/ijodllogs.pdf Viewed: 07/07/04

Abstract: As experimental digital library testbeds gain wider acceptance and develop significant user bases, it becomes important to investigate the ways in which users interact with the systems in practice. Transaction logs are one source of usage information, and the information on user behaviour can be culled from them both automatically (through calculation of summary statistics) and manually (by examining query strings for semantic clues on search motivations and searching strategy). We have conducted a transaction log analysis on user activity in the Computer Science Technical Reports Collection of the New Zealand Digital Library, and report insights gained and identify resulting search interface design issues. Specifically, we present the user demographics available with our library, discuss the use of operators and search options in qureies, and examine patterns in query construction and refinement. We also describe common mistakes in searching, and examine the distribution of query terms appearing in the logs.

Kass, Gordon, *End-to-End Web Performance: Ensuring a Positive Customer Experience.* URL: http://www.digitalenterprise.org/seminar/kass.html (Last updated: September 26, 2001) Viewed: 07/08/04

Abstract: The modern enterprise is becoming evermore "digital" in terms of what it is and what it does. This site deals exclusively with the digital aspect of the enterprise and covers such topics as: how to design virtual spaces that rival physical spaces in terms of ease of use; how to measure the performance and effectiveness of the digital enterprise; how the digital enterprise can alter or extend an organization's business model; how markets change as more consumers migrate online; how the automation of digital processes can change market dynamics; how trust can be fostered in computer-mediated intereaction; how to securely manage the digital enterprise; how to protect digital intellectual property.

Larsen, Ronald L. (2002). *The DLib Test Suite and Metrics Working Group: Harvesting the Experience from the Digital Library Initiative.* URL: http://www.dlib.org/metrics/public/papers/The_Dlib_Test_Suite_and_Metrics.pdf Viewed: 07/07/04

Abstract: The DLib test suite was conceived to address three needs: (1) lowering the barriers of entry for digital library researchers requiring access to large collections and information management services, (2) providing standard sets of data for quantitative and comparative research, and (3) supporting a distributed environment of heterogeneous resources organized to support interoperability experiments.

Lee, Juhnyoung, Robert Hoch, Mark Podlasek, Edith Schonberg, and Stephen Gomory. (2000). Analysis and Visualization of Metrics for Online Merchandizing. *Lecture Notes in*

Computer Science, 1836, 126. URL: http://www.ibm.com/iac/papers/lncs.pdf Viewed: 07/07/04

Abstract: While techniques and tools for Web marketing are being actively developed, there is much less available for Web merchandising. This paper contributes to the area of Web usage analysis for E-commerce merchandising. First, we categorize areas of analysis for Web merchandising such as product assortment, merchandising cues, shopping metaphors, and Web design features. Second, we define a new set of metrics for Web merchandising, which we call micro-conversion rates. The new metrics provide capabilities for examining data about merchandising in online stores, and detailed insight into the effectiveness of different Web merchandising efforts by answering related business questions. Third, we present a set of novel visualizations that explore patterns in micro-conversions in online stores reflecting in customer responses to various Web merchandising efforts. Through an empirical study using lookto-buy data from an online store, we demonstrate how the proposed starfield visualizations can be used to understand the shopping behavior in an online store and the effectiveness of various merchandising schemes it employs. Finally, we discuss the types of data required for this kind of visual analysis of online merchandising, and briefly describe how the data can be collected and integrated from an E-commerce site.

Malacinski, Andrei, Scott Dominick, and Tom Hatrick. (2001). *Measuring Web Traffic, Part 1*. URL: http://www-106.ibm.com/developerworks/web/library/wa-mwt1/

Abstract: There are several Web measurement approaches that have been adopted by the industry, such as network monitors, single-pixel solutions, and HTTP server log analysis. We begin this article with a discussion of why Web metrics are important and then provide an introduction to the various types of Web measurement approaches with site traffic metrics obtainable with these approaches. When choosing a Web measurement approach, decisions have to be made about the technology to employ, its cost, its setup, and coordinating it with existing enterprise data. This article looks at these and other decisions related to collecting Web measurements.

Malacinski, Andrei, Scott Dominick, and Tom Hatrick. (2001). *Measuring Web Traffic, Part 2*. URL: http://www-106.ibm.com/developerworks/web/library/wa-mwt2/

Abstract: In part two of this series, we concentrate on the approach of obtaining measurements from HTTP server logs. To help you better understand HTTP server log analysis, this article contains an explanation of HTTP logging (what it is, what the log formats are, what information is available, and so on) and reviews the metrics available from raw log data. Basic metrics include the number of hits, the number of visitors, visitor duration, visitor origin (subdomain, referral link), visitor IP address, browser type, and platform type; we also will discuss advanced metrics which are derived from the manipulation of data through techniques such as categorization and aggregation. Additionally, it is possible to expand the reaches of your solution by combining technologies such as data warehousing and data mining.

Menasce, Daniel A., *QoS Measurement and Control in Web and E-Commerce Services*. URL: http://www.digitalenterprise.org/seminar/menasce.html (Last updated: March 20, 2002) Viewed: 07/08/04

Abstract: As individuals and organizations increasingly rely on Web and E-commerce services for their day to day activities, it becomes crucial to guarantee that these services are reliable, scalable, robust, trustworthy, and secure. This talk addresses the Quality of Service (QoS) metrics that should be considered when assessing the quality of the service provided on the Web. Methods for workload characterization and benchmarking, including the TPC-W benchmark, are discussed. Finally, techniques for dynamic QoS control of e-commerce sites are presented.

Moe, Wendy W., and Peter S. Fader. (2001). Capturing Evolving Visit Behaviour in Clickstream Data. *Managing the Digital Enterprise.* URL: http://www-marketing.wharton.upenn.edu/ideas/pdf/00-003.pdf Viewed: 07/07/04

Abstract: Many online retailers monitor visitor traffic as a measure of their stores' success. However, summary measures such as the total number of visits per month provide little insight about individual-level shopping behavior. Additionally, behavior may evolve over time, especially in a changing environment like the Internet. Understanding the nature of this evolution provides valuable knowledge that can influence how a retail store is managed and marketed. This paper develops an individual-level model for store visiting behavior based on Internet clickstream data. We capture cross-sectional variation in store-visit behavior as well as changes over time as visitors gain experience with the store. That is, as someone makes more visits to a site, her latent rate of visit may increase, decrease, or remain unchanged as in the case of static, mature markets. So as the composition of the customer population changes (e.g., as customers mature or as large numbers of new and inexperienced Internet shoppers enter the market), the overall degree of visitor heterogeneity that each store faces may shift. We also examine the relationship between visiting frequency and purchasing propensity. Previous studies suggest that customers who shop frequently may be more likely to make a purchase on any given shopping occasion. As a result, frequent shoppers often comprise the preferred target segment. We find evidence supporting the fact that people who visit a store more frequently are more likely to buy. However, we also show that changes (i.e., evolution) in an individual's visit frequency over time provides further information regarding which customer segments are more likely to buy. Rather than simply targeting all frequent shoppers, our results suggest that a more refined segmentation approach that incorporates how much an individual's behavior is changing could more efficiently identify a profitable target segment.

Nicholson, Scott. (2003). The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology and Libraries, 22*(4), 146-151. URL: http://www.ala.org/ala/lita/litapublications/ital/2204nicholson.htm Viewed: 07/08/04

Abstract: Bibliomining, or data mining for libraries, is the application of data mining and bibliometric tools to data produced from library services. This article outlines the bibliomining process with emphasis on data warehousing issues. Methods for cleaning and anonymizing library data are presented with examples.

NIST WebMetrics Testbed. URL: http://zing.ncsl.nist.gov/WebTools/ Last updated: 29 January 2003. Viewed: 07/29/04

Abstract: The objective of the NIST Web Metrics Testbed is to explore the feasibility of a range of tools and techniques that support rapid, remote, and automated testing and evaluation of website usability. The prototypes are used to support the usability engineering research of the Visualization and Usability Group (VUG) at the National Institute of Standards and Technology (NIST)

Pitkow, James. 1997. In Search of Reliable Usage Data on the WWW. *Computer Networks and ISDN Systems* 29: 1343-1355. Also available online. URL: http://www2.parc.com/istl/projects/uir/pubs/items/UIR-1997-04-Pitkow-WWW6-UsageData.pdf Viewed: 08/03/04.

Abstract: The WWW is currently the hottest testbed for future interactive digital systems. While much is understood technically about how the WWW functions, substantially less is known about how this technology is used collectively and on an individual basis. This disparity of knowledge exists largely as a direct consequence of the decentralized nature of Web. Since each user of the Web is not uniquely identifiable across the system and the system employs various levels of caching, measurement of actual usage is problematic. This paper establishes terminology to frame the problem of reliably determining usage of WWW resources while reviewing current practice and their shortcomings. A review of the various metrics and analyses that can be performed to determine usage is then presented. This is followed by a discussion of the strengths and weaknesses of the hit-metering [8] proposal currently in consideration by the HTTP working group. Lastly, new proposals, based upon server-side sampling provides more reliable and useful usage data while requiring no change to the current HTTP protocol and enhancing user privacy.

Pitkow, James E. and Krishna A Bharat. (1994). *Webviz: A Tool For World-Wide Web Access Log Analysis*. In Proceedings of the First International WWW conference. URL: http://arbor.ee.ntu.edu.tw/~chyun/dmpaper/pitkwa94.pdf

Abstract: Various programs have emerged that provide statistical analysis of World-Wide Web (WWW) access logs. These programs typically detail the number of accesses for a file, the number of times a site has visited the database, and some programs even provide temporal analysis of requests1. However, these programs are not interactive nor do they provide visualizations of the local database. WebViz was developed with the intention of providing WWW database maintainers and designers with a graphical view of their local database and access patterns. That is, by incorporating the Web-Path paradigm into interactive software, users can see not only the documents (represented visually as nodes) in their database but also the hyperlinks travelled (represented visually as links) by users requesting documents from the database. WebViz further enables uses to selectively filter the access log (i.e. restrict the graphical view by specifying the desired domain names or DSN numbers,directory names, and start and stop times), control bindings to graph attributes (i.e. node size, border width and color as well as link width and color can be bound to frequency and recency information), play back the events in the access log (i.e. re-issue the logged sequence of requests), select a layout of nodes and links that best presents the database's structure, and examine the graph at any instant in time. Clearly, WebViz is a useful WWW database utility given that it can provide the user with graphical information about document accesses and the paths taken by users through the database. Such analyses can facilitate structural and contextual changes resulting in a more efficient use of the document

space. This paper details the implementation of WebViz and outlines possible future extensions.

Rappa, Michael, *Web Metrics Research Program.* URL: http://ecommerce.ncsu.edu/research/rappa.html (Last updated: October 06, 2002) Viewed: 07/08/04

Abstract: The newly established multi-disciplinary research program in Web Metrics focuses on a broad range of issues that connect the empirical measurement of web-based commercial activity with overall business process metrics. Work in this area includes: Log Analysis Metrics -- An evaluation of the strengths and limitations of existing metrics and associated tools for log analysis. Performance testing of existing commercial tools; User Surveys -- Studies of user preferences and satisfaction via a company's web presence, including evaluations of the comparative effectiveness of different customer channels. Also, research into the validity and reliability of web-based user survey methodologies; New Methods in Web Metrics -- Research into the refinement of existing web metrics and the development of new metrics, including those metrics that tie log data with business process and performance data; Business Performance Modelling -- This research aims to use web measurements to create predictive models of business performance.

Redalen, Aaron and Naomi Miller. (2000, January). Evaluating Website Modifications at the National Library of Medicine through Search Log Analysis. *D-Lib Magazine, 6*(1). URL: http://www.dlib.org/dlib/january00/redalen/01redalen.html Viewed: 07/07/04

Abstract: The effective design, presentation and functioning of search facilities are integral parts of website design, particularly for the sites of information-rich organizations such as libraries. Libraries face particular challenges because many of the traditional sources of bibliographic information, such as online databases, may not be integrated into the Web environment. As Clifford Lynch points out, "The Net is not a digital library." (Lynch, 1997). Since search functions are an important and heavily used facet of the National Library of Medicine website, and since NLM produces many databases that are not fully integrated into its website, we have undertaken a series of analyses and modifications to guide users in effective site navigation and use of the search functions. It is difficult to find much published literature on the use of World Wide Web search tools and their impact on users' navigation. However, in our review of the literature, we learned that many web users tend to immediately follow any available "search" links without first determining what other links are available, or what database or collection of documents is to be searched. In a 1997 article on search interfaces, web usability expert Jakob Nielsen observes that, "Our usability studies show that more that half of all users are search-dominant...The search-dominant users will usually go straight for the search button when they enter a website: they are not interested in looking around the site; they are task-focused and want to find specific information as fast as possible" (Nielsen, 1997). Despite the fact that many users gravitate towards search features in navigating websites, another study of searching behavior found that less experienced users frequently misunderstand the functioning of search engines (including what is being searched), and easily become frustrated when they are unable to locate relevant information (Pollock, 1997). These findings are consistent with our experiences in the work described below.

Sayal, Mehmet, Akhil Sahai, Vijay Machiraju, and Fabio Casati. (2003, March). Semantic Analysis of E-Business Operations. *Journal of Network and Systems Management, 11*(1).

Abstract: An e-business infrastructure comprises a large number of business processes that are exposed through web services. To manage such an infrastructure, it is necessary for business managers to be able to observe their e-business operations in greater detail. In particular, a framework, or tool, is required that allows business users to define qualitative and quantitative metrics, depending on their own (business or IT) goals. Once metrics have been defined, the tool should be able to measure them and support users in semantic analysis of the results, allowing them to quickly identify quality degradations as well as their causes.

Sevcik, Peter and Bartlett, J. (2001). Understanding Web Performance. *Business Communications Review,* October. URL: http://www.porivo.com/news/10-01BCRarticle.pdf Viewed: 07/07/04

Abstract: Is there an Internet performance problem? There are two views: The conventional wisdom held by 'Net insiders is that things are just fine. They point to the massive investment in bandwidth that has eliminated many congestion problems, and to the performance of the Keynote Business-40 Index, which has fallen from more than 12 seconds to less than 3 seconds in only four years. The other view is held by the vast majority of Internet users, who complain that the Web is an awfully slow way to do anything useful. To be sure, most connections are slow; no matter whose numbers you use, anywhere from 78 percent to 93 percent of the users in the U.S. use a dial-up modem to connect to the Internet. And most connections actually are at speeds closer to 30 kbps. Furthermore, most real users and network shoppers don't visit the Keynote Business-40 sites. Instead, they visit sites like MSN, AOL, Amazon, ICQ, ESPN and Disney, which are designed to be "interesting and cool" rather than optimized for performance; these sites are slow by design. In short, geography, demographics and interest all play important roles in determining what the "Internet experience" will be like

Shelman, Mark and Robert Warren, *The Web User Log: Its Value, Limits, Uses and Abuses.* URL: http://www.digitalenterprise.org/seminar/shelman.html (Last updated: November 28, 2001) Viewed: 07/08/04

Abstract: The typical web server captures information about every browser request that it receives, and records this information in web user logs. These logs are surprisingly valuable, but that value is not always easily unlocked. This talk will dissect the web user log, explaining the meaning and value of its various components. Through the use of case studies, Mr. Shelman and Dr. Warren will then discuss some of the business and technical issues - including web site performance, access patterns, and security - that web log analysis can address, and demonstrate some of the web log analysis tools that may be employed.

Smith, Melaney, *What Should You Measure? Part 1: E-Commerce Metrics.* URL: http://ecommerce.internet.com/how/biz/article/0,,10365_2197431,00.html (Last updated: April 28, 2003) Viewed: 07/08/04

Abstract: What should be measured for my Web business?" If you are wondering what to measure to evaluate your online business, the simple truth is it depends. The metrics that matter depend on your company's goals and strategies, your own role in the company, the time period under consideration, and more. To illustrate the variations in which measurements are considered "critical," we'll look at a few key industries and differences among businesses that would appear to have somewhat similar goals.

Smith, Melaney, *What Should You Measure? Part 2: Online Media.* URL: http://ecommerce.internet.com/how/customers/article/0,,10363_2203851,00.html (Last updated: May 9, 2003) Viewed: 07/08/04

Abstract: This is the second in a series aimed at that frequently asked question: What should I measure for my online business? My last column covered e-commerce metrics. We heard from folks in the trenches about which stats are mission-critical and why. This week, we move to online media, where answers are very different.

Smith, Melaney, *What Should You Measure? Part 3: Common Ground.* URL: http://ecommerce.internet.com/how/biz/article/0,,10365_2210941,00.html (Last updated: May 22, 2003) Viewed: 07/08/04

Abstract: This is the third column in a series aimed at that frequently asked question: What should I measure for my online business? Parts one and two focused on e-commerce, where revenue is unanimously considered the most critical metric, and online media, for which traffic is the primary metric of choice. But let's not oversimplify. There are too many variations in what constitutes "revenue" and "traffic" to assume one stat fits all. A recurring theme of talks I've had with executives is each company, and each person within the company, must determine which metrics are most relevant based on objectives.

Strupp, Paul. An Introduction to Web Traffic Measurement. URL: http://www.cs.colorado.edu/~sumner/cs6838_2001/Strupp/Strupp_web_traffic_lecture3.ppt

Abstract: The presentation discusses log file analysis (combined log format), presents some of the commercial tools (netGenesis, Wuusage, Accrue, NetAcumen, Hitbox, and Aria) and business/marketing applications, and highlights the issues and pitfalls.

Viles, Charles L. and James C. French. (1999). Content Locality in Distributed Digital Libraries. *Information Processing & Management, 35*, 317-336
Abstract: In this paper we introduce the notion of content locality in distributed document collections. Content locality is the degree to which content-similar documents are colocated in a distributed collection. We propose two metrics for measurement of content locality, one based on topic signatures and the other based on collection statistics. We provide derivations and analysis of both metrics and use them to measure the content locality in two kinds of document collections, the well-known TREC corpus and the Networked Computer Science Technical Report Library (NCSTRL), an operational digital library. We also show that content locality can be thought of temporally as well as spatially and provide evidence of its existence in temporally ordered document collections like news feeds.

Winett, Bill, *Tracking Your Visitors*. URL: http://hotwired.lycos.com/webmonkey/templates/print_template.htmlt?meta=/webmonkey/98/16/index2a_meta.html (Last updated: not stated)
Viewed: 07/08/04

Abstract: So you've created the ultimate Web site, and now you're sitting back watching your hit counter go wild. You may ask yourself, "I wonder how many pageviews my help page is getting?" or, "I wonder how many people are visiting my site?" Unfortunately, when most people start building a Web site, they don't consider they someday might want to track its traffic. It takes enough time just to design the site and create the content. Outlining what information they want to track is just more work that already overworked staffs tend to let slide. But when it comes down to it, we all quickly become bean counters on the Web. Once a site is up and running, we want to know how many people are looking at our pages and how many pages each of those people is looking at. That's usually when a lot of Web developers discover that had they spent more time thinking about setting up their site, they'd be able to track how it's being used much more easily. If you're in this situation right now, you've come to the right place. And if you haven't made your site public yet, you're lucky - you still have time to think about reporting before your design is set in stone. Don't miss out on this chance!

Zucca, Joe. (2003). Traces in the Clickstream: Early Work on a Management Information Repository at the University of Pennsylvania. *Information Technology and Libraries, 22*(4), 175-178. URL: http://www.ala.org/ala/lita/litapublications/ital/2204zucca.htm

Abstract: For the past three years, the University of Pennsylvania (Penn) library has been building a data repository and developing computer functionality to support management information needs. This article traces the origin and evolution of Penn's evolving management information system (MIS) program, known as the data farm. It addresses problems pertaining to the collection, storage, anonymization, and normalization of data, and looks at current work on a database-driven model for future MIS functions.

Tools and Reviews of Tools

Tools for webmetrics can be categorized as tools for 1) collecting and gathering web traffic, 2) analyzing the data collected, and 3) reporting and visualizing the data collected and/or reported. A few of these are listed here along with a set of articles by Smith that review some of the common tools. Open source tools that have been used in NSDL Projects such as *Awstats, Wusage*, and *Webalyzer* are not yet included here. However, tools that were mentioned at the workshop, as the industry trend towards outsourcing webmetrics to third party tools (tools provided by application service providers) are included: *CoreMetrics, Omniture, WebSide Story (*see *Hitbox)* and *WebTrends On Demand (*see *WebTrends*).

Analog. URL: http://www.analog.cx/ Viewed: 07/29/04

Abstract: This is an open source logfile analysis tool and according to the website one of the most popular logfile analyzers.

BehaviorTracking. URL: http://www.behaviortracking.com/ Viewed: 07/29/04

Abstract: MondoSoft's BehaviorTracking is billed as the next-generation analytics software that Web marketers can use to quickly increase Web site sales and customer satisfaction while lowering costs. BehaviorTracking is indicative of an emerging new software category called Search Analytics, which will eclipse the existing Clickpath Analytics. BehaviorTracking provides reports-on-demand, not based on guesswork but on what visitors searched for and found, and sometimes more important - what they did not find. BehaviorTracking customize the data for each relevant team member and actually recommend what actions should be taken to adjust course, improve sales, cut costs or retain customers.

CoreMetrics. URL: http://www.coremetrics.com/ Viewed: 08/03/04

Abstract: Coremetrics provides a suite of powerful analytics products that address the most pressing business challenges facing eBusinesses today. Based upon the industry's only data platform storing LIVE (Lifetime Individual Visitor Experience) Profiles of all online visitor and customer interactions, Coremetrics solutions deliver the actionable insight necessary to help you measure and optimize all of their online customer interactions, while increasing the profitability of your online channel and boosting customer lifetime value. CoreMetrics offers four solutions: CoreMetrics Online Analytics 2004, CoreMetrics Data Service, CoreMetrics Strategic Services, and CoreMetrics LIVEApps.

Hitbox. URL: http://www.hitboxprofessional.com/ and http://www.hitbox.com/ Viewed: 07/29/04

Abstract: *HitBox* provides real-time Web site analysis to show you where site visitors come from, what content and products they look at, and which navigational paths they take through your site. HitBox Professional can also keep track of orders, customers and revenue, and pinpoint your best sources of customer acquisition. On-demand analytics and website optimization is also provided through an affiliated suite of tools, *WebSideStory*.

NetGenesis. URL: http://www.spss.com/netgenesis/ Viewed: 07/08/04

Abstract: *NetGenesis* Web analytics solutions transform high volumes of Web data into actionable e-metrics, the operational measurements of online business success. This proven technology provides decision makers with the e-metrics needed to manage your online business. With *NetGenesis*, a complete set of customizable e-metrics reports is ready for use out of the box. *NetGenesis* also supplies a platform for creating and customizing new e-metrics that are unique to your business.

NetTracker. URL: http://www.sane.com/products/NetTracker/ Viewed: 07/29/04

Abstract: NetTracker Web analytics solutions affordably put expert Web site traffic analysis in the hands of those who need it.

Omniture. URL: http://www.omniture.com/ Viewed: 08/03/04

Abstract: Omniture's SiteCatalyst™ empowers online marketing managers to measure success and drive ROI by providing timely, relevant reports about visitor interactions with their web site. Designed specifically for the needs of the enterprise, SiteCatalyst is a remotely hosted, subscription-based reporting solution for real-time web site analysis. By providing reports that contain powerful insights into the visitor interaction and customer behavior of your web site, SiteCatalyst provides you with the knowledge you need to make intelligent business decisions.

Smith, Melaney, *In the Sea of Analysis Tools, a Lifeboat?* URL: http://www.clickz.com/experts/crm/analyze_data/article.php/1477921 (Last updated: October 8, 2002) Viewed: 07/08/04

Abstract: We all seem to be unhappy with the reporting tools we have (or don't have) at our disposal. Clearly we're all looking for better ways to evaluate our businesses, whether it's from a high-level financial perspective or down at the individual email level. And many of you are willing to spend money to do it, if you could only find the right tool. So, I decided to turn this issue over to the readers. What tools do you use to access, manipulate, and interpret your data, so that it is useful for understanding your business and making decisions? I want to hear from people who use the tools, not those who sell the tools. What has (or hasn't) worked for you? What are the pros and cons of the tools at your disposal? Winners and losers? Bargain workhorses and expensive mistakes? Smith presents highlights.

Smith, Melaney, *Does a Perfect Web Metrics Tool Exist?, Part 1*. URL: http://www.clickz.com/experts/crm/analyze_data/article.php/1502661 (Last updated: November 19, 2002) Viewed: 07/08/04

Abstract: Experience with different analysis tools and reader feedback about the tools that have worked (or not worked) for them. Readers from all over the world: the U.S., England, Australia, and France, express a collective longing better tools, obviously a universal sentiment. Surprisingly, *Microsoft Excel* and it's add-on, *Spreadsheet Assistant*, emerged as the single most useful tool.

Smith, Melaney, *Does a Perfect Web Metrics Tool Exist?, Part 2*. URL: http://www.clickz.com/experts/crm/analyze_data/article.php/1548561 (Last updated: December 3, 2002) Viewed: 07/08/04

Abstract: Reader feedback in response to a previous column requesting input on the best, and worst, analysis tools. *WebTrends, ClickTracks,* and *RedSheriff*, are briefly described.

Smith, Melaney, *Does a Perfect Web Metrics Tool Exist?, Part 3*. URL: http://www.clickz.com/experts/crm/analyze_data/article.php/1556241 (Last updated: December 17, 2002) Viewed: 07/08/04

Abstract: In parts one and two, Smith shared reader feedback and recommendations on reporting tools, in response to a previous column. This week: more tools and more feedback. *Ariel* (which has changed names and become *Ecommstats*) is the log analysis tool to watch, Smith says, although she likes *NetTracker* best and a couple of readers gave *WebCEO* (promotion and analysis tool) and *HitBox* a nod.

WebTrends Log Analyzer. URL: http://www.netiq.com/webtrends/default.asp Viewed: 07/08/04

Abstract: The worldwide leader in web analytics, *WebTrends* is another commercial tool. There is also *WebTrends 7 On Demand*.

First created: 06/04
Last modified: 08/03/04