Web Site Management

Counting Clicks: Auditing Your Web Site Activity

Jack Powers
Director, Graphics Research Laboratory
Publisher, www.electric-pages.com

Web Site Management

2

Web Site Management

Web Site Auditing

- Management
  — Financial analysis
- Editorial Issues
  — Art and Design
- Marketing Issues
  — Site market research
  — Advertisers' links

3

Web Site Management

Internet World Agenda

- Web Server Overview
- The Web Log
- Statistical Analysis Software
- What to do with the data

Web Site Management

Web Server Overview

- http://www.electric-pages.com
- HTTP: HyperText Transmission Protocol
- HTML: Hyper Text Markup Language
- IP: Internet Protocol Address
  — 122.22.23 > ny27.interramp.com
  — permanent or dynamic IPs

Web Site Management

File Types

- HTML (Page Views)
- GIFs, JPGs

- MOVs, AUs, WAVs, PPTs, RAMs
- PDFs, CMXs, WRLs
- CGI Scripts

6

Web Site Management

Counting Clicks

- HIT: A single file sent to a browser
- VISIT: A single visit from an IP address
- UNIQUE VISIT: Count of different IPs
  - Dynamic IPs reset every 20 minutes
- REGISTERED USER: Subscriber
- COOKIE: Netscape secret number

Web Site Management

The Web Log

- NCSA/CERN standards
  - Access log
  - Error Log
  - Referer Log
  - Agent Log
- Netscape, Open Market, Microsoft, others
  - Extended Log Formats

8

Web Site Management

Many log variations

Common log file format (CERN) Intersé market focus 1 database Intersé market focus 2 database EMWACS log file format Intersé extended log file format Microsoft IIS standard logging format Microsoft IIS extended logging format Microsoft IIS hyperextended logging format Microsoft IIS ODBC log database NCSA combined log file format NCSA with servername log file format NCSA combined with servername log file format Netscape flexible log file format (Netscape Enterprise servers) Netscape Proxy extended logging format Open Market extended log file format O'Reilly multihome common log file format O'Reilly Windows log file format Real Audio log file format SiteTrack log file format Spry web server ASCII log file format Spry web server ODBC log database Universal log file format UUNET extended log file format WebFacts audit log file format WebStar log file format Zeus common log file format WU Archive ftp log file format

— from http://www.interse.com/ourproducts/faq.html#1

9

Web Site Management

The Web Log

```
ip24.umich.edu - -
[24/Jan/1996:12:27:13 -0500]
"GET /index.html HTTP/1.0"
200 7114
```

10

Web Site Management

Log Analysis Software

- Unix freeware
  - Getstats
  - wwwstats
- Commercial packages
  - I/Audit from I/PRO & Nielsen
  - Intersé, Web Audit
- Home brew analysis

Web Site Management

Stats Examples

GETSTATS
http://www.electric-pages.com/jan96.htm

WEBAUDIT (ACNielsen, Australia)
http://www.nielsen.com.au/webaudit/-#Objectives

I/PRO Third Party Auditng
http://www.ipro.com

12

Web Site Management

The Web Log

- Host IP address
- Username
- Start date and time of transaction
- Request and Filename sent
- Size of file sent
- Type of browser
- Refering URL
- http status code

Web Site Management

IP address

- 150.180.13.54
- ny25.interramp.com
- ab23.aol.com
- ip24.f-uberlin.de
- Registered domains
  — Organization
  — Location

Web Site Management

Invisible readers

- Internet gateways
  — ISPs or corporate networks
  — Allocate dynamic Ips
  — Cache pages
- Browser cache
  — Intermittently updated
- Off-line browsers

Web Site Management

Time of transaction

- Starting time is logged
- Ending time can be computed
- Local time can be guessed
- Time stats show
  — Event effects
  — Peak usage
  — Weekend vs weekday activity
  — Working day vs lunch hour and home time

Web Site Management

Request and file name

- Which pages are hot
- Which pages are not
- Sorting by directory
- Screening out GIFs and JPEGs
  — Measuring the reader's patience and the art director's skill

17

Web Site Management

Referrer URL

- Where readers came from
- Which ad placements work best
- What search targets brought the most traffic

18

Web Site Management

What to do with the data

- Management
  - Financial analysis
- Editorial Issues
  - Art and Design
- Marketing Issues
  - Site market research
  - Advertisers' links

Web Site Management

Better measurement

- Logging client usage with cookies
- Registering users/subscribers
- Writing profiles
  - Server-side
  - Client-side
  - Sharing cookies
- No two web pages will ever be the same

20

Web Site Management

What do things cost?

- Hosting service stats
- Shareware
  - Server-based Unix programs
  - PC-based dedicated systems
- Commercial products
  - PC-based

21

Web Site Management

Ad Pricing Options

- Standard space ad deal
  - Bottom-up TFN CPMs
- Basic rate plus per inquiry
- Per inquiry only
- Percentage of transactions
- Confusing combinations
- CASIE standards versus smart buyers

Web Site Management

Clickstream trends

- Basic site research
  - Management
  - Editorial research
  - Advertising
- Customization/personalization
  - HTML-on-the-fly
  - Tuning the editorial and ads to the reader

Web Site Management

Web publishing futures

- Privacy concerns
  - Browser aliasing
  - Cookie warping
- Constant editorial reengineering
- Algorithmic advertising
- Combining electronic with conventional research