farmdoc
UNIVERSITY OF ILLINOIS

The Use and Abuse of Web Traffic Statistics for Extension Program Evaluation

Mark A. Althouse and Scott H. Irwin



Agricultural and Consumer Economics

Sample of Raw Weblog File for the farmdoc Site, January 1, 2005

- #Software: Microsoft Internet Information Services 6.0
- #Version: 1.0
- #Date: 2005-01-01 00:01:21
- #Fields: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs(User-Agent) cs(Referer) sc-status sc-substatus sc-win32-status sc-bytes
- 2005-01-01 00:01:21 128.174.65.110 GET /images/finance_on.gif - 80 - 216.229.21.37 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+98;+Mozilla/4.0+(compatible;+MSIE+5.5;+Windows;+HEARTLAND+INET);+T312461) http://www.farmdoc.uiuc.edu/weatherprices/index.asp 304 0 0 211
- 2005-01-01 00:01:21 128.174.65.110 GET /images/marketing_on.gif - 80 - 216.229.21.37 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+98;+Mozilla/4.0+(compatible;+MSIE+5.5;+Windows;+HEARTLAND+INET);+T312461) http://www.farmdoc.uiuc.edu/weatherprices/index.asp 304 0 0 210
- ...
- 2005-01-01 00:11:48 128.174.65.110 GET /robots.txt - 80 - 207.68.146.55 msnbot/0.3+(+http://search.msn.com/msnbot.htm) - 404 0 2 1814
- 2005-01-01 00:11:48 128.174.65.110 GET /fasttools/register.html - 80 - 207.68.146.55 msnbot/0.3+(+http://search.msn.com/msnbot.htm) - 200 0 0 2524
- ...
- 2005-01-01 00:12:58 128.174.65.110 GET /ncrisk/grant_application/FinalReport.doc - 80 - 68.142.249.113 Mozilla/5.0+(compatible;+Yahoo!+Slurp;+http://help.yahoo.com/help/us/ysearch/slurp) - 200 0 0 21269
- 2005-01-01 00:14:02 128.174.65.110 GET /finance/FinancialCharacteristics/ratios.htm - 80 - 202.45.121.254 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;+SV1;+.NET+CLR+1.1.4322) http://www.google.com.au/search?q=%22servicing+ratio%22&hl=en&lr=&start=60&sa=N 200 0 0 11908
- ...

Single Log Entry

- 2005-01-01 09:56:33 128.174.65.110 GET /marketing/grainoutlook/0403bean/0403bean_text.html - 80 - 65.214.36.43 Mozilla/2.0+(compatible;+Ask+Jeeves/Teoma) - 200 0 0 36896
  - This file request occurred: *January 1, 2005 at 9:56AM*
  - The page request was a from IP (Internet Protocol) address *65.214.36.43* for the file */marketing/grainoutlook/0403bean/0403bean_text*.
  - The browser technology was *Mozilla/2.0* with a download of *36896kB*

Output from Commercial Web Traffic Software for the farmdoc site, January 2005

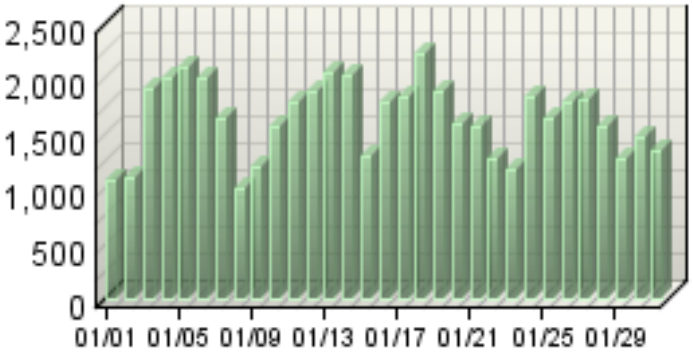
Table 1.a: Technical Summary	
Total Hits	2,252,356
Successful Hits	2,184,692
Successful Hits (as Percent)	97.00%
Failed Hits	67,664
Failed Hits (as Percent)	3.00%
Cached Hits	402,822
Cached Hits (as Percent)	17.88%

Table 1.b: Page View Summary	
Page Views	143,825
Average per Day	4,639
Average Page Views per Visit	2.83

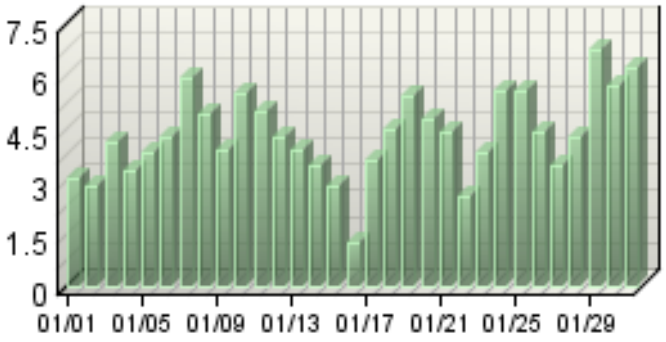
Table 1.c: Visit Summary	
Visits	50,776
Average per Day	1,637
Average Visit Duration	00:16:17
Median Visit Duration	00:02:59
International Visits	0.00%
Visits of Unknown Origin	100.00%
Visits from Your Country: United States (US)	0.00%

Table 1.d: Visitor Summary	
Visitors	24,619
Visitors Who Visited Once	19,411
Visitors Who Visited More Than Once	5,208
Average Visits per Visitor	2.06

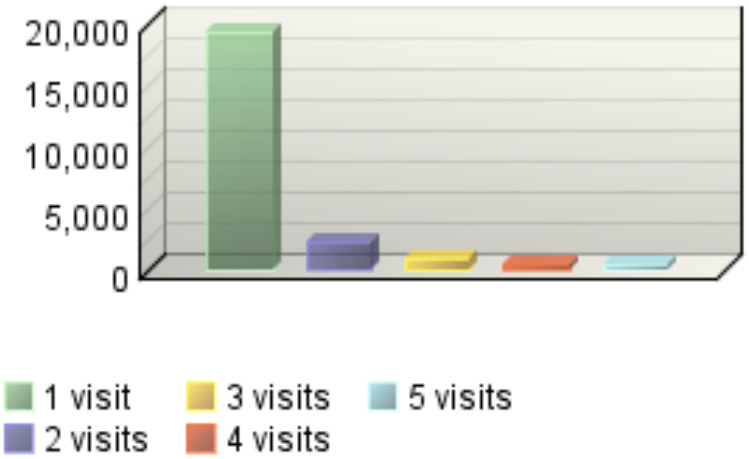
Active Visits

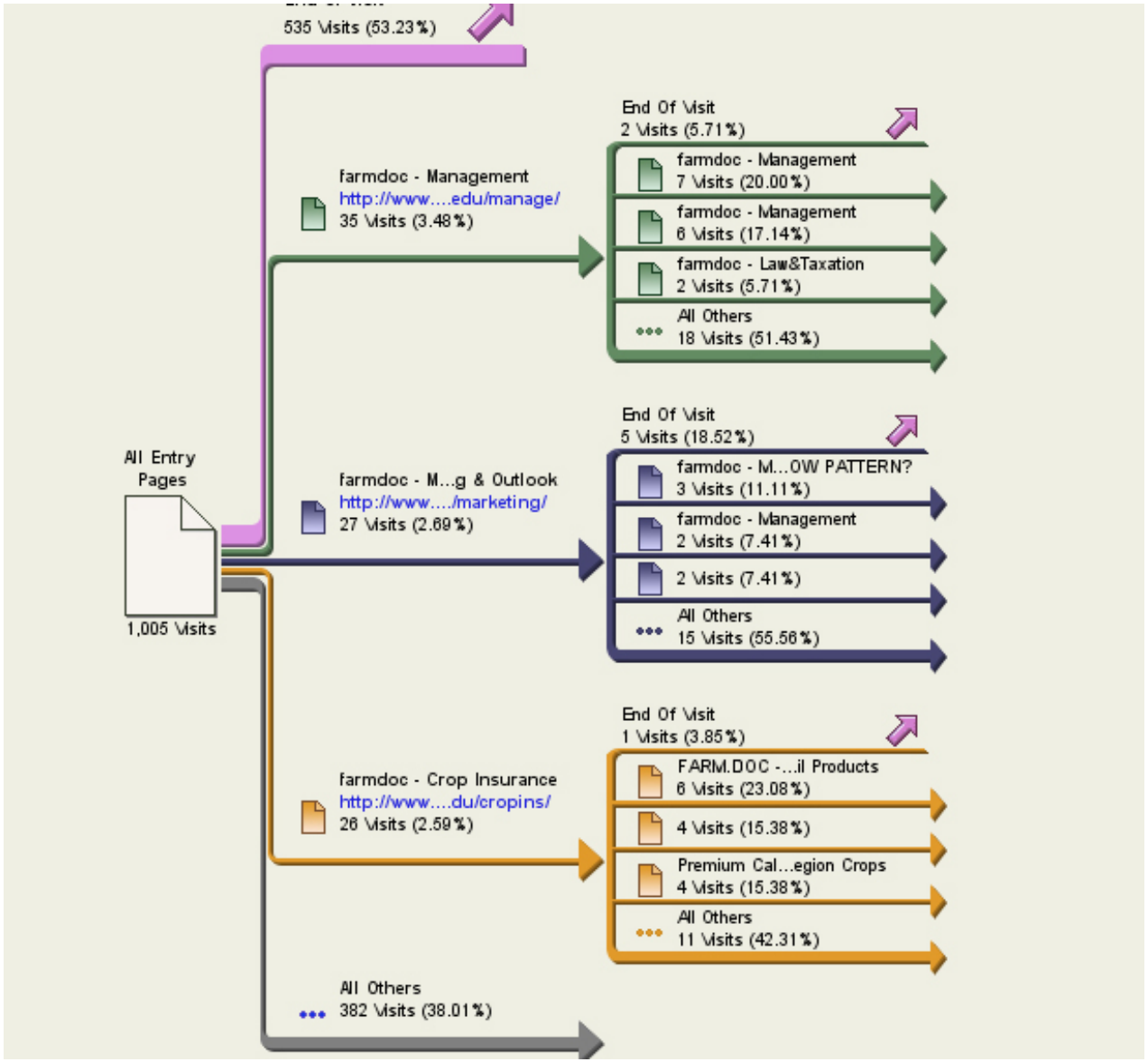


Average Visit Duration



Visitors



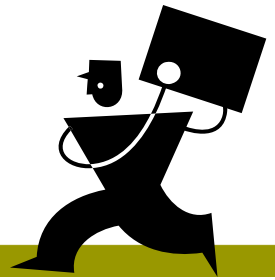

535 Visits (53.23%)
End Of Visit
2 Visits (5.71%)
farmdoc - Management
http://www....edu/manage/
35 Visits (3.48%)
farmdoc - Management
7 Visits (20.00%)
farmdoc - Management
6 Visits (17.14%)
farmdoc - Law&Taxation
2 Visits (5.71%)
All Others
18 Visits (51.43%)
All Entry Pages
1,005 Visits
End Of Visit
5 Visits (18.52%)
farmdoc - M...g & Outlook
http://www..../marketing/
27 Visits (2.69%)
farmdoc - M...OW PATTERN?
3 Visits (11.11%)
farmdoc - Management
2 Visits (7.41%)
2 Visits (7.41%)
All Others
15 Visits (55.56%)
End Of Visit
1 Visits (3.85%)
farmdoc - Crop Insurance
http://www....du/cropins/
26 Visits (2.59%)
FARM.DOC -...il Products
6 Visits (23.08%)
4 Visits (15.38%)
Premium Cal...egion Crops
4 Visits (15.38%)
All Others
11 Visits (42.31%)
All Others
382 Visits (38.01%)

Abuse #1: User Identity cannot be Determined

- User IP address is available but this is not equivalent to user identity
- Many large Internet service providers (ISPs) use the same public 'proxy' server IP address for large pools of users
- Many servers now use dynamically assigned public IP addresses that are reassigned on a regular schedule
- Information on the user IP address and referring site cannot be extrapolated to uniquely identify *users* nor their exact *geographic location* without explicit use of password, stored user data or other identity collection protocol

Abuse #2 The Number of Unique Visitors and Visits is Unknown

- The number of distinct hosts making requests is often used to define unique visitors; however, such visitor estimates unreliable
- Just as in the case of determining user identity, the estimate of unique visitors is thwarted by the practice of multiple users connecting from a proxy host and dynamic public IP numbers
- Commercial software incorporate “black box” models to estimate visitors and visits

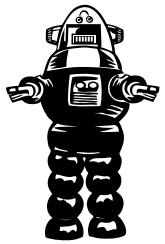


Abuse #3: User Paths within the Site cannot be Reconstructed

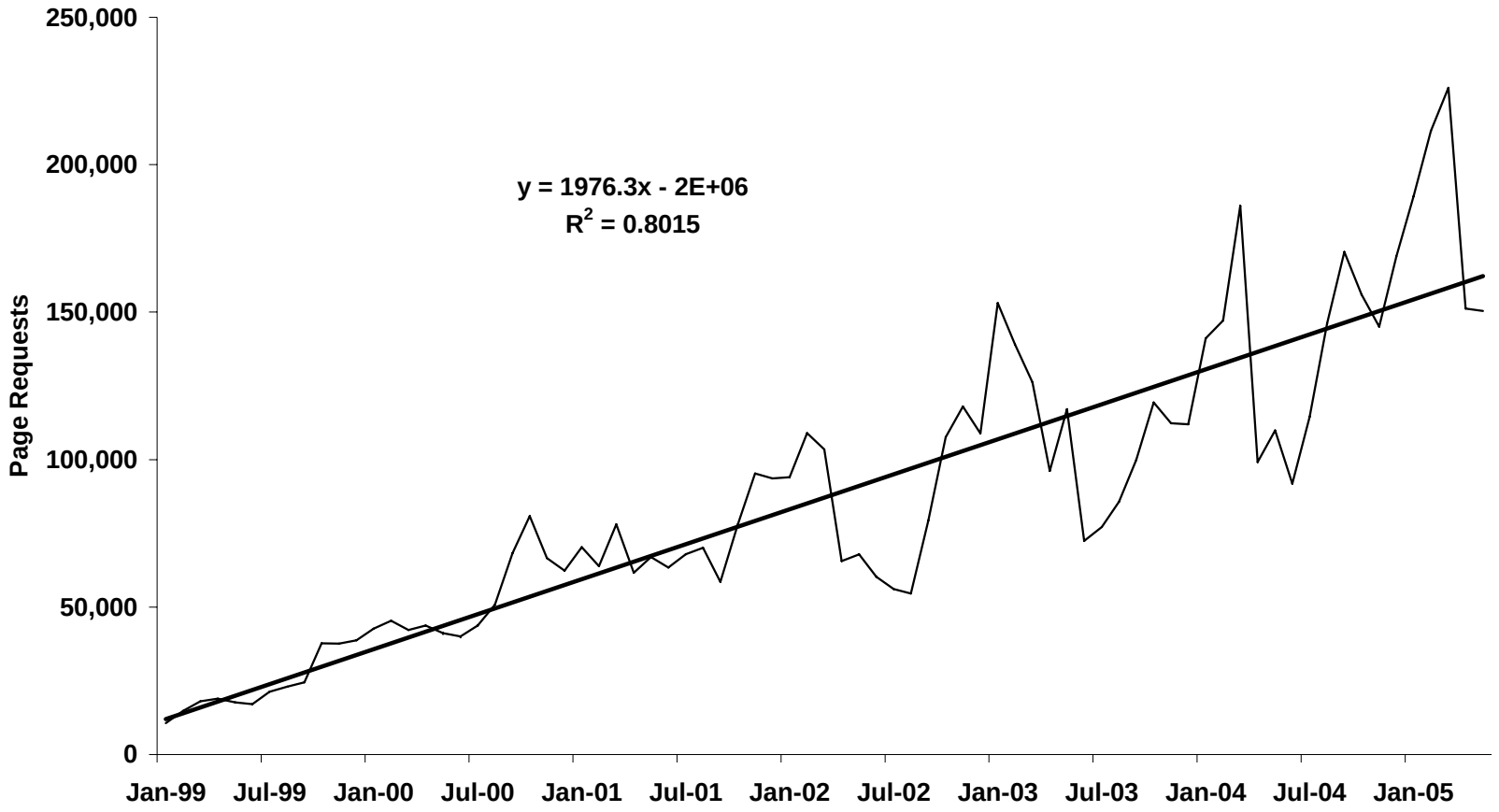
- The current practice of file caching, especially at the server level, means that accurate and complete weblog information on a unique visitor path is not available
- Exit pages cannot be determined since links to external websites are not recorded in the weblog
- Commercial software must apply a user path model to estimate unique visit paths and select the last requested page in that sequence as the exit page

Best Practice: Counting Page Requests

- The number of web pages viewed by visitors to a site
  - Roughly the number of “clicks” a user makes at a site
  - Standard measure for determining advertising rates on the web
- Measurement issues
  - Browser and server caching may cause actual use to be underestimated by page requests
  - Must filter out page requests from search engine “robots” and “crawlers” or effective use will be overestimated



Usage of the farmdoc Website: Entire Site



Usage of the farmdoc Website: Requests Growth Rate for the Entire Site (compared to same month in previous year)