April 28, 2017

Using AWStats to Analyze Logs from EZProxy and from the Public OPAC Logs

Stacey Knight-Davis

Available at: https://works.bepress.com/stacey_knight-davis/62/
Using AWStats to Analyze Logs from EZProxy and from the Public OPAC Logs

Stacey Knight-Davis
Eastern Illinois University
Log Analysis

With Google Analytics Query String Added Extras!
Log only exists when turned on!
Why might the access log be Off?

• It’s big
• It grows bigger every day
• Balancing patron privacy and data driven decision making

ILA best practice is to delete old logs or anonymize them
What might be in a log?

• Images served
• Pages served
• Search terms, if configured to do so
• Date
• Time
• User’s Browser
  • User’s IP
  • User’s login

Personally Identifying Information – Protect it!
Where can I get logs?

• CARLI search logs
• Pre-anonymized for your convivence
Stuff you Need

• Download Them All – Firefox Plugin

• 7zip
  http://www.7-zip.org/

• Log URL
  https://www.carli.illinois.edu/logs/
Without analysis, logs are about as useful as a chunk of wood.
Analyzing logs to make them Useful

- AWStats
- Unpleasant to install
- Requires full admin rights
- Requires IIS or Apache
- Requires Perl
- Can be run on Windows
- Can be run on a desktop
- [http://www.awstats.org/](http://www.awstats.org/)
What does AWStats do?

• Finds all the matching needles in the haystack

We need the needles that match our CARLI code
Most viewed records at EIU, 2/19/2017-4/19

<table>
<thead>
<tr>
<th>Pages-URL</th>
<th>Viewed</th>
<th>Average size</th>
</tr>
</thead>
<tbody>
<tr>
<td>/vf-eiu/Record/</td>
<td>74,192</td>
<td>575 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1309074</td>
<td>156</td>
<td>6.13 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_92052</td>
<td>129</td>
<td>6.04 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/ebl_1441586</td>
<td>43</td>
<td>1.30 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/ebl_1441789</td>
<td>43</td>
<td>1.27 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1330918</td>
<td>43</td>
<td>1.28 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1033553</td>
<td>41</td>
<td>1.05 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1073775/Holdings</td>
<td>41</td>
<td>1.03 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1361564</td>
<td>41</td>
<td>1.05 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_01112/AJAX</td>
<td>34</td>
<td>674 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1262785</td>
<td>31</td>
<td>6.13 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_7195621/AJAX</td>
<td>27</td>
<td>592 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/smgs.cgi</td>
<td>25</td>
<td>1.27 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_691160/AJAX</td>
<td>23</td>
<td>581 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1136915/Cite</td>
<td>23</td>
<td>6.00 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1354543/AJAX</td>
<td>22</td>
<td>509 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/nu_756493/TOC</td>
<td>20</td>
<td>5.89 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/nu_756493/AJAX</td>
<td>19</td>
<td>449 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_1038843</td>
<td>18</td>
<td>6.21 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_0872005</td>
<td>18</td>
<td>6.35 KB</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_77338/AJAX</td>
<td>18</td>
<td>496 Bytes</td>
</tr>
<tr>
<td>/vf-eiu/Record/eiu_060690</td>
<td>18</td>
<td>6.26 KB</td>
</tr>
</tbody>
</table>
Click the link to see what it is

• Encyclopedia of public relations (156)
• Business regulation in the Common Market nations (129)
• Hispanic-serving institutions: advancing research and transformative practice (43)
• Higher education in America (43 CARLI EBL)
• Higher education in the digital age (43 CARLI EBL)
Other Log Sources

• EZ Proxy Access Log
• This will not be sanitized unless you sanitize it yourself
• Contains IP address
EZ Proxy logs can tell us

• What is used off campus
• What time of day users are online
• Top downloaded articles
• Usage by provider (requires some Excel work or several config files)

• AW Stats Config info
  • LogType=W
  • LogFormat=4
Special Bonus Feature: Google Analytics

Google Analytics can capture search terms from a locally hosted search box.
Go to All Pages, Filter for query string
### Finding Search Terms from Article Searches

<table>
<thead>
<tr>
<th></th>
<th>Query</th>
<th>Hits</th>
<th>TF-IDF</th>
<th>Matches</th>
<th>Score</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>critical thinking</td>
<td>20</td>
<td>0.17%</td>
<td>0</td>
<td>75.00</td>
<td>$0.00</td>
</tr>
<tr>
<td>4.</td>
<td>arthritis</td>
<td>17</td>
<td>0.14%</td>
<td>0</td>
<td>52.94</td>
<td>$0.00</td>
</tr>
<tr>
<td>5.</td>
<td>social media</td>
<td>16</td>
<td>0.14%</td>
<td>2</td>
<td>100.00</td>
<td>$0.00</td>
</tr>
<tr>
<td>6.</td>
<td>leadership</td>
<td>15</td>
<td>0.13%</td>
<td>0</td>
<td>73.33</td>
<td>$0.00</td>
</tr>
<tr>
<td>7.</td>
<td>nurse to patient ratio</td>
<td>14</td>
<td>0.12%</td>
<td>8</td>
<td>62.50</td>
<td>$0.00</td>
</tr>
</tbody>
</table>
The end

• Questions?
• Comments?
• Presentation is posted on CARLI website