I have completed an enhanced robots.txt file.

Initial Website XML Maps I was making were returning just under 500 results. The Sitemap now returns 45 results, one of which is an error, that I have filed a separate bug report for. No pages are duplicated, all pages with relevant content are crawled.

Of note is that I have re-enabled the file gallery being crawled. It was returning large number of files and was disabled to prevent pressure on the web server. Results can now be indexed without the massive duplication that existed before.

There is a section in the file that can be commented out to prevent duplicated pages when SEF URL's is enabled. Applied automatically, it will prevent the entire website from being crawled.
Of importance is also that I have only tested this for the feature set that I am using. It WILL create a better crawling and therefor better Search Engine ranking with other features, but may not take it as far as one could be achieved.

Using this robots.txt file not only decreases server load, but also leads to better search engine rankings.

Demonstrate Bug

Please demonstrate your bug on show2.tikiwiki.org

Version: trunk ▼ Create show2.tikiwiki.org instance

Ticket ID
5223

Created
Thursday 01 May, 2014 03:58:25 GMT-0000

LastModif
Thursday 19 May, 2016 20:08:28 GMT-0000

Comments

Attachments

<table>
<thead>
<tr>
<th>filename</th>
<th>created</th>
<th>hits</th>
<th>comment</th>
<th>version</th>
<th>filetype</th>
</tr>
</thead>
<tbody>
<tr>
<td>robots.txt</td>
<td>01 May 14 04:00 GMT-0000</td>
<td>331</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The original document is available at http://dev.tiki.org/item5223-Robots-txt-Enhancements