

Bugs & Wish list

Robots.txt Disallow entries now must have leading slash | Tiki Wiki CMS Groupware :: Development
Robots.txt Disallow entries now must have leading slash

Status

✖ Closed

Subject

Robots.txt Disallow entries now must have leading slash

Version

1.8.x

1.9.x

2.x

Category

- Error

Feature

All / Undefined

Installer (profiles, upgrades and server-related issues)

Search engine optimization (SEO)

Resolution status

Fixed or Solved

Submitted by

John Hadjisky

Lastmod by

Marc Laporte

Rating

★ ★ ★ ★ ★ (0) ?

Description

Although the RFC (for example, <http://www.robotstxt.org/wc/norobots-rfc.html>) doesn't explicitly require a leading slash (/) before the page name, I have found that, as of late Oct, 2005, many 'bots, including Googlebots, have started requiring them.

For example, before the change,

Disallow: tiki-pagehistory.php

would prevent well-behaved 'bots from trying to index tiki-pagehistory.php. However, after the change, I had to have:

Disallow: /tiki-pagehistory.php

in robots.txt, or else all my page history would be indexed! I verified this using my server log, and also by doing google searches against my site for phrases that only appeared in page history. I have every reason to believe this is a problem for all other TikiWiki-based sites.

Others have noticed this. There is discussion in the forums at:

- Googlebot ignoring robots.txt — includes a graph of my bandwidth usage before and after.
- Adding <meta name="robots" content="noindex,nofollow"> to History pages
- Yahoo search indexes Print pages instead of Read pages

Solution

Putting a leading slash before all page references in robots.txt solved the problem. See <http://ihuck.com/robots.txt> (a TikiWiki site), compare to e.g. <http://dupli.tikiwiki.org/robots.txt>

1. Change the robots.txt in the CVS so that there is a leading slash before all page references (there is already a leading slash before all vdir references)
2. Post an article on tw.o urging existing users to change their robots.txt
3. Change the robots.txt on all sites *.tikiwiki.org that use the standard robots.txt, for example, dupli.tw.o, doc.tw.o, probably others.

I have many years of web dev experience, and two plus years experience with PHP and TikiWiki, but almost no CVS experience. I'm happy to learn CVS and implement this solution, but I am hoping first for some feedback from the community re have I overlooked any reason not to make these changes. Thanks.

Assign this back to me and I'll start working on the changes (except for the 3rd change which a *.tw.o admin will need to do).

Importance

5

Priority

25

Demonstrate Bug (Tiki 19+)

Please demonstrate your bug on show2.tiki.org

Version: trunk ▼

Demonstrate Bug (older Tiki versions)

Please demonstrate your bug on show.tikiwiki.org

Version: 18.x ▼

Ticket ID

442

Created

Wednesday 21 December, 2005 00:51:59 GMT-0000
by Unknown

LastModif

Tuesday 14 April, 2009 22:56:22 GMT-0000

Comments



fmathias 30 Jan 06 19:34 GMT-0000

have already been fixed

Attachments

filename	created	hits	comment	version	filetype
 robotsneedroot.png	21 Dec 05 13:28 GMT-0000	441		1	

The original document is available at
<https://dev.tiki.org/item442-Robots-txt-Disallow-entries-now-must-have-leading-slash>