

Redmine - Patch #3754

add some additional URL paths to robots.txt

2009-08-18 19:40 - mark burdett

Status: New	Start date: 2009-08-18
Priority: Normal	Due date:
Assignee:	% Done: 0%
Category:	Estimated time: 0.00 hour
Target version:	
Description	
<p>My apache logs show that some redmine URLs are being heavily indexed by robots, and it seems like it would be best to have them blocked by robots.txt:</p> <p>/issues /projects/*/time_entries /projects/N/wiki/* (where N is the numeric project id) /repositories/annotate/* /repositories/browse/* /repositories/changes/* /repositories/diff/* /repositories/entry/*</p>	
Related issues:	
Related to Redmine - Defect # 6734: robots.txt: disallow crawling issues list...	Closed 2010-10-24

History

#1 - 2009-08-18 21:35 - Mischa The Evil

See the [[Plugin_List#Bots-Filter-plugin|Bots Filter plugin]] which has some overlap (e.g. the repositories). Maybe you can modify it to adapt it to your precise requirements?

Regards,

Mischa.

#2 - 2009-08-21 11:48 - mark burdett

Or I can easily block these via my apache config. But I do think they should be added to robots.txt by default. I also wonder, how are Googlebot and others even finding some of these non-canonical paths? It could point to a bug elsewhere which is generating links to these paths?

#3 - 2009-09-23 11:18 - mark burdett

- *File robots.txt.patch added*

Here's a patch adding the additional problematic paths to the default robots.txt

#4 - 2009-10-01 00:01 - Eric Davis

I like having the robots crawl some of these pages, they even turn up when I'm searching for a bug that I've already fixed.

- wiki pages
- global issues list
- repositories

#5 - 2009-10-01 01:38 - mark burdett

- *File robots.txt.patch added*

The wiki pages that this patch blocks are not the canonical path, they use the numeric project id rather than project name.

I now realize that the initial version of this patch blocked the individual issue pages; I intended to only block /issues? -- i.e. the global issue search page.

#6 - 2009-11-11 20:44 - Jean-Philippe Lang

- *Tracker changed from Defect to Patch*

#7 - 2010-07-04 08:38 - Brad Schick

My site is also getting hammer on /repositories and /issues. Seems somewhat pointless to disallow access to these resources through /projects/... but not other urls.

#8 - 2013-03-11 21:55 - Antoine Beaupré

This patch has been ready for more than 3 years, why hasn't this been committed yet?

#9 - 2013-03-11 22:31 - Antoine Beaupré

- *File robots.txt-2.patch added*

Here's an updated patch for 1.4.

#10 - 2015-02-04 20:35 - Antoine Beaupré

Antoine Beaupré wrote:

| *Here's an updated patch for 1.4.*

and that was now two years ago, with the patch sitting here for 5 years. can we at least get feedback on what's wrong with the patch, if anything?

thanks.

Files

robots.txt.patch	587 Bytes	2009-09-23	mark burdett
robots.txt.patch	597 Bytes	2009-09-30	mark burdett
robots.txt-2.patch	554 Bytes	2013-03-11	Antoine Beaupré