

robots.txt for confluence for SEO(Search Engine Optimization)

Below code seems pretty much helpful for crawler if it follows robots.txt

```
# Note: this files uses parameters specific to Google, parameters that are not in robots.txt standard
# http://www.google.com/support/webmasters/, http://www.robotstxt.org/wc/faq.html and http://en.wikipedia.org
/wiki/Robots_Exclusion_Standard were used to research said parameters
# some links shouldn't show to an anonymous browser such as GAS but are included for completeness

User-agent: * # match all bots. GSA is our primarily crawler but logs indicate there may be others on our
Intranet
Crawl-delay: 5 # per http://en.wikipedia.org/wiki/Robots.txt#Nonstandard_extensions, sets number of seconds to
wait between requests to 5 seconds. may not work

Disallow: /pages/ # this line to purge GSA of all old page entries, will be removed in next iteration so that
specific /pages/ lines below take effect
Disallow: /admin/ # administrator links
Disallow: /administrators.action? # remove any administrator links
Disallow: /createrssfeed.action? # remove internal RSS links
Disallow: /dashboard.action? # remove the dashboard, heavy resource hit
Allow: /display # ensure primary display pages are allowed
Disallow: /display/*&tasklist.complete= # remove tasklist links
Disallow: /display/*?decorator=printable # remove printable version links
Disallow: /display/*?focusedCommentId= # remove page comment focus links
Disallow: /display/*?refresh= # prevent crawler from clicking refresh button
Disallow: /display/*?replyToComment= # remove reply to comment links
Disallow: /display/*?rootCommentId= # remove news comment focus links
Disallow: /display/*?showComments=true&showCommentArea=true#addcomment # remove add comment links
Disallow: /doexportpage.action? # remove pdf export links
Disallow: /dopeopledirectorysearch.action # people search
Disallow: /dosearchsite.action? # remove specific site searches
Disallow: /exportword? # remove word export links
Disallow: /login.action? # remove the login page

# Allow: /pages/viewpage.action?* # allows indexing of pages with invalid titles for html (such as ?'s).
Unfortunately currently allows page history to sneak in
Disallow: /pages/ # this line to purge GSA of all old page entries, will be removed in next iteration so that
specific /pages/ lines below take effect
Disallow: /pages/copypage.action? # remove copy page links
Disallow: /pages/createblogpost.action? # remove add news links
Disallow: /pages/createpage.action? # remove add page links
Disallow: /pages/diffpages.action? # remove page comparison pages
Disallow: /pages/diffpagesbyversion.action? # remove page comparison links
Disallow: /pages/editblogpost.action? # remove edit news links
Disallow: /pages/editpage.action? # remove edit page links
Disallow: /pages/removepage.action? # remove the remove page links
Disallow: /pages/revertpagebacktoversion.action? # remove reversion links
Disallow: /pages/templates # remove template pages
Disallow: /pages/templates/ # block template indexes
Disallow: /pages/viewchangessincelastlogin.action? # remove page comparison pages
Disallow: /pages/viewpagesrc.action? # remove view page source links
Disallow: /pages/viewpreviouspageversions.action? # remove the link to previous versions
Disallow: /plugins/ # blocks plug-in calls
Disallow: /rpc/ # remove any RPC links
Disallow: /searchsite.action? # remove the wiki search engine pages
Disallow: /spaces/ # remove space action pages
Disallow: /themes/ # theme links
Disallow: /users/ # remove user action pages
Disallow: /x/ # remove tiny link urls

# End file
```

When you install confluence on Centos 7, it may not create a robots.txt, so you should create it manually at "<root>/confluence/robots.txt"