Robots.txt

The robots.txt file is a web standard that gives web robots instructions about what pages they should and shouldn't crawl when visiting a website. In order to use a robots.txt (which is recommended even if all pages are allowed) you need to upload it to the root of your website. This will mean that it will then load in the default location www.yourdomain.com/robots.txt and web robots will find it.

Basics:

Disallow all robots from crawling your site:

User-agent: *

Disallow: /

Allow all robots full access to your site:

User-agent: *

Disallow:

Exclude certain parts of your site:

User-agent: *

Disallow: /directory/

To exclude a specific robot from your server:

User-agent: (BotNameHere)

Disallow: /

Actinic/Sellerdeck Example

User-agent: *

Disallow: /acatalog/test/

Disallow: /cgi-bin/

Disallow: /account/

User-agent: googlebot

Disallow:

User-agent: googlebot-image

Disallow:

Sitemap: http://domainname.com/sitemap.xml

Magento Example

The following is a well-tested version of Robots.txt file which you can use; just edit the lines not applicable for your store's set up. If you are not certain on some elements then consult with your developer.

Robots.txt for Magento Community and Enterprise

GENERAL SETTINGS

 # Enables robots.txt rules for all crawlers

 User-agent: *

 # # Crawl-delay parameter: the number of seconds you want to wait between successful requests to the same server.

Set a crawl rate, if your server's traffic problems. Please note that Google ignore crawl-delay setting in Robots.txt. You can set up this in Google Webmaster tool

Crawl-delay: 30

 # # Magento sitemap: URL to your sitemap file in Magento

Sitemap: http://www.mystore.com/sitemap/sitemap.xml

 # # Settings that relate to the UNDER CONSTRUCTION

 # # Do not allow indexing files and folders that are required during development: CVS, SVN directory and dump files

Disallow: / CVS

Disallow: / *. Svn $

Disallow: / *. Idea $

Disallow: / *. Sql $

Disallow: / *. Tgz $

 # # GENERAL SETTINGS For MAGENTO

 # # Do not index the page Magento admin

Disallow: / admin /

 # # Do not index the general technical Magento directory

Disallow: / app /

Disallow: / downloader /

Disallow: / errors /

Disallow: / includes /

Disallow: / lib /

Disallow: / pkginfo /

Disallow: / shell /

Disallow: / var /

Do not index the shared files Magento

Disallow: / api.php

Disallow: / cron.php

Disallow: / cron.sh

Disallow: / error_log

Disallow: / get.php

Disallow: / install.php

Disallow: / LICENSE.html

Disallow: / LICENSE.txt

Disallow: / LICENSE_AFL.txt

Disallow: / README.txt

Disallow: / RELEASE_NOTES.txt

MAGENTO SEA IMPROVEMENT

Do not index the page subcategories that are sorted or filtered.

Disallow: / *? Dir *

Disallow: / *? Dir = desc

Disallow: / *? Dir = asc

Disallow: / *? Limit = all

Disallow: / *? Mode *

Do not index the second copy of the home page (example.com / index.php /). Un-comment only if you have activated Magento SEO URLs.

Disallow: / index.php /

Do not index the link from the session ID

Disallow: / *? SID =

 # # Do not index the page checkout and user account

Disallow: / checkout /

Disallow: / onestepcheckout /

Disallow: / customer /

Disallow: / customer / account /

Disallow: / customer / account / login /

 # # Do not index the search page and CEO, non-optimized link categories

Disallow: / catalogsearch /

Disallow: / catalog / product_compare /

Disallow: / catalog / category / view /

Disallow: / catalog / product / view /

 # # Server Settings

 # # Do not index the general technical directories and files on a server

Disallow: / cgi-bin /

Disallow: / cleanup.php

Disallow: / apc.php

Disallow: / memcache.php

Disallow: / phpinfo.php

 # # SETTINGS Image indexing

 # # Optional: If you do not want to Google and Bing to index your images

User-agent: Googlebot-Image

Disallow: /

User-agent: msnbot-media

Disallow: /