

How To Use The Robots.txt File To Increase Your Web Ranking

Sometimes we rank well on one engine for a particular keyphrase and assume that all search engines will like our pages, and hence we will rank well for that keyphrase on a number of engines. Unfortunately this is rarely the case. All the major search engines differ somewhat, so what's get you ranked high on one engine may actually help to lower your ranking on another engine.

It is for this reason that some people like to optimize pages for each particular search engine. Usually these pages would only be slightly different but this slight difference could make all the difference when it comes to ranking high.

However because search engine spiders crawl through sites indexing every page it can find, it might come across your search engine specific optimizes pages and because they are very similar, the spider may think you are spamming it and will do one of two things, ban your site altogether or severely punish you in the form of lower rankings.

The solution in this case is to stop specific Search Engine spiders from indexing some of your web pages. This is done using a robots.txt file which resides on your webspace.

A Robots.txt file is a vital part of any webmasters battle against getting banned or punished by the search engines if he or she designs different pages for different search engine's.

The robots.txt file is just a simple text file as the file extension suggests. It's created using a simple text editor like notepad or WordPad, complicated word processors such as Microsoft Word will only corrupt the file.

You can insert certain code in this text file to make it work. This is how it can be done.

User-Agent: (Spider Name)

Disallow: (File Name)

The User-Agent is the name of the search engines spider and Disallow is the name of the file that you don't want that spider to index.

You have to start a new batch of code for each engine, but if you want to list multiply disallow files you can one under another. For example

User-Agent: Slurp (Inktomi's spider)

Disallow: xyz-gg.html

Disallow: xyz-al.html

Disallow: xxyyzz-gg.html

Disallow: xxyyzz-al.html

The above code disallows Inktomi to spider two pages optimized for Google (gg) and two pages optimized for AltaVista (al). If Inktomi were allowed to spider these pages as well as the pages specifically made for Inktomi, you may run the risk of being banned or penalized. Hence, it's always a good idea to use a robots.txt file.

The robots.txt file resides on your webspace, but where on your webspace? The root directory! If you upload your file to sub-directories it will not work. If you wanted to disallow all engines from indexing a file, you simply use the "*" character where the engines name would usually be. However beware that the "*" character won't work on the Disallow line.

Here are the names of a few of the big engines:

Excite - ArchitextSpider

AltaVista - Scooter

Lycos - Lycos_Spider_(T-Rex)

Google - Googlebot

Alltheweb - FAST-WebCrawler

Be sure to check over the file before uploading it, as you may have made a simple mistake, which could mean your pages are indexed by engines you don't want to index them, or even worse none of your pages might be indexed.

Source: <http://www.articlecircle.com>

About the Author

The SEO Company offers a complete Search Engine Optimization service and right now you can get a comprehensive ranking report for your website for free. Find out exactly where your website ranks in all three major search engines <http://www.the-seo-company.co.uk/free-ranking-report.html>