



Search Engine Tricks: What's a Robots.txt file?

Help potential customers locate your ecommerce site

Search engines look at millions of Web pages to generate search results. They do this with what we call "search engine spiders." This makes sense - spiders crawling around on the Web. But another word for them is "robots" because they are simply unmanned programs gathering data automatically.

In the beginning, these robots would spider every page, and every file attached to the Web. This caused problems for both the search engines and the people using them. Pages that really aren't worth looking at, such as header files to be

included in all pages on a site, were being spidered and appearing in search results. Have you ever searched on Google and gotten a partial page as a result? Or, worse, a forum post from four years before?

The solution was for Google and other search engines to begin looking for a "robots.txt file". The robots.txt file is placed in the root folder, meaning directly where the home page file is located to determine what should and shouldn't be searched. This is named "The Robots Exclusion Standard." It's not a standard that

is backed by any organization, it's simply a format agreed upon by search engine companies. You can learn more at robotstxt.org/wc/robots.html. This simple text file, created with Notepad or other simple text editor gives you complete control by telling the robots not to spider, or access, certain folders in your site. The result is happier visitors who come to your site from search engines and get only full pages that you want them to see, not partial, test or script pages you don't want them to see.

Look at the examples below to get started.

robots.txt file example

It is important that the file is a simple text file – do not use Microsoft Word to create it because it adds formatting that can make it unreadable by the search engine spiders. And be careful of how you type – it must look exactly like these examples, with caps only for the first letter, just the right spacing, etc. A poorly done robots.txt file could harm your site more than help it. To test that you've created your robots.txt file properly. You can use a cool online robots.txt file validator at searchengineworld.com/cgi-bin/robotcheck.cgi.

