

This Free E-Book is brought to you by [Natural-Aging.com](http://Natural-Aging.com).

**100% Effective Natural Hormone Treatment**  
**Menopause, Andropause And Other Hormone Imbalances**  
**Impair Healthy Healing In People Over The Age Of 30!**

## Creating Robots.txt File and its Importance

By San Christopher

### Creating Robots.txt File and its Importance by San Christopher

If you are thinking you have developed a truly great keyword-rich-unique-content fully optimized website for the search engines and an attracting site for the visitors – that's fine, but do you know you are missing something? A robots.txt file. Did you include it? By the way do you know what's the importance of a robots.txt file?

Success of big companies lies in keeping their confidential data a secret, hidden from all. They tell the world something and do something. This enables them to execute their future course of action easily and change plans according to the situation. Job of robots.txt file is the same. It can or cannot allow a search engine to visit some or all of your web pages. Of course a human visitor is free to visit these pages. That being the case, for the search engines your website may be different than what a visitor is seeing. If you think one or some of the pages/files aren't good enough to be visited by a particular search engine or engines you can do it. Although this is not recommended – your website should be made in such a way it should not shy away from the search engines. Nevertheless its always better to know the basics of writing robots.txt file. It will help you. We will discuss farther down – robots.txt file is important. I repeat again – don't make pages you think should be hidden from the search engines. If any search engine think you are up to some tricks, it may panelize your site causing a no-rank – in the worst case for ever!

Every search engine has a "robot" (a software program) that does the job of visiting a website. Their purpose is to "know" the website, what it is all about, gather all information about it etc. Search engine robots gather this information and bring them back to their databases to show them in their search results. So, if your site is not there in their database it never shows up in the search results.

Web Robots are sometimes referred to as Web Crawlers, or Spiders. Therefore the process of a robot visiting your website is called "Spidering" or "Crawling". When somebody says "the search engines have spidered my website," it means the search engine robots have visited their website. This robot is known by a name and has an independent IP address. This IP address is of no importance to us, but knowing their names will help since this name will be used when we create a robots.txt file. This is why the file is called "robots.txt." Given below is the list of the robots of some of the very popular search

## Creating Robots.txt File and its Importance

engines:

Search Engine – Robot

Alexa.com – ia\_archiver

Altavista.com – Scooter (Bought by Yahoo)

UK.Altavista.com – AltaVista–Intranet (Bought by Yahoo)

Alltheweb.com – FAST–WebCrawler (Bought by Yahoo)

Excite.com – ArchitextSpider

Euroseek.net – Arachnoidea

Gendoor.com (Genealogical Search Engine) – GenCrawler

Google.com – Googlebot (<http://www.google.com/bot.html>)

Hotbot.com (uses Inktomi's robot) – Slurp

Inktomi.com Slurp – ([slurp@inktomi.com](mailto:slurp@inktomi.com)) (Bought by Yahoo)

Infoseek.com – UltraSeek

Looksmart.com – MantraAgent

Lycos.com – Lycos\_Spider\_(T–Rex)

Northernlight.com – Gulliver

Nationaldirectory.com – NationalDirectory–SuperSpider

UKSearcher.co.uk – UK Searcher Spider

Writing Robots.txt:

Let's learn to write robots command. Note that there are two ways to write robots command. One is to include all the commands in a text file called "robots.txt" and another is to write robots command in the meta tag.

We will learn both ways of writing robots command.

Writing robots command in Meta tag:

There are 4 things you can tell a search engine robot when it requests (visits) your page:

- 1) Do not index this page – the search engines will not index the page.
- 2) Do not follow any links on this page – the search engines will not follow the links included in the page, i.e. they will not index any page that this page links to.
- 3) Do index this page – the search engines will index the page.
- 4) Do follow the links – the search engines will index the pages that this page links to.

Note that "index" is different than "spider". A search engine first spiders a page and then indexes it. Indexing is giving a certain importance to the page on the basis of its content, information, meta tags, link popularity with respect to the searched keyword. All this is decided at run time. When you tell search engines not to index a page, it means they know that "certain" page exists but do not rank them. That is, a no–index page will never be shown in their search results. This in any case does not mean a no–index page will not get visitors, it might get visitors indirectly from a page which links to it.

## Creating Robots.txt File and its Importance

Yes, no direct visitors from the search engines.

Suppose you want the search engines to index and also index (follow) its linked pages then include the following command in the Meta Tag:

Suppose you want the search engines to index a page but not follow its links then include the following command in the Meta Tag:

Suppose you do not want the search engines to index a page but follow its links then include the following command in the Meta Tag:

Suppose you do not want the search engines to either index or follow links of a particular page then include the following command in the Meta Tag:

Note:

Google makes a "Cached" of every file it spiders. It's a small snap shot of the page. Want to stop Google from doing so? Include the following Meta Tag:

Like any meta tag the above written tags should be placed in the HEAD section of an HTML page:

your title

Creating robots.txt file:

A robots.txt file is an independent file and should be written in a plain text editor like Notepad. Do not use MS-Word or any other text editor to create robots.txt. The bottom line is this file should have the extension ".txt" else it will be useless.

Let's begin. Open Notepad (it comes free with Microsoft Windows) and save the file with the name "robots.txt". Make sure that the extension is .txt.

By the way, did you note we did not use name of any robot in the meta tag! What does it indicate? Simple – by using meta you direct all the search engines to do something or not do something on a page. You do not have control over any one search engine. The solution is robots.txt.

It can always happen you do not want a particular search engine to index a page for certain reasons. In that case using a robots.txt file will help. Even though I do not recommend such a thing. The search engines get you traffic, why hate them. Stop them from doing their job and they hate you. I again repeat keep your pages smart for the search engines and welcome them. Fine, then why take the trouble to learn robots.txt? Why should you include a robots.txt file at all?

Let's suppose yours is a dynamic database site containing information of your newsletter subscribers, customers, their address, phone numbers etc. All these confidential information is kept in a separate

## Creating Robots.txt File and its Importance

directory called "admin". (It is recommended to keep such information in a separate directory. Handling data will be easier for you and so will be easy to keep the search engines away. We will just know how.) I am sure you would never want any unauthorized person to visit this area leave alone the search engines. It does not help the search engines either since they have nothing to do with the data or files there. Here comes the role of a robots.txt file. Write the following in the robots.txt file: (Ignore the horizontal row – they are included only to separate the commands from rest of the text.)

---

```
User-agent: *  
Disallow: /admin/
```

---

This does not allow the spiders to index anything in the admin directory also including sub-directories if any.

The asterisk (\*) mark indicates all the search engines. How do you stop a particular search engine from spidering your files or directory?

Suppose you want to stop Excite from spidering this directory:

---

```
User-agent: ArchitextSpider  
Disallow: /admin/
```

---

Suppose you want to stop Excite and Google from spidering this directory:

---

```
User-agent: ArchitextSpider  
Disallow: /admin/
```

```
User-agent: Googlebot  
Disallow: /admin/
```

---

Files are no different. Suppose you want a file datafile.html not to be spidered by Excite:

---

## Creating Robots.txt File and its Importance

User-Agent: ArchitextSpider  
Disallow: /datafile.html

---

Similarly, you do not want it to be spidered by Google too:

---

User-agent: ArchitextSpider  
Disallow: /datafile.html

User-agent: Googlebot  
Disallow: /datafile.html

---

Suppose you want two files datafile1.html and datafile2.html not to be spidered by Excite:

---

User-Agent: ArchitextSpider  
Disallow: /datafile1.html  
Disallow: /datafile2.html

---

Can you guess what does the following mean?

---

User-agent: ArchitextSpider  
Disallow: /datafile1.html  
Disallow: /datafile2.html

User-agent: Googlebot  
Disallow: /datafile1.html

---

Excite will not spider datafile1.html and datafile2.html, but Google will not spider only datafile1.html. It will spider datafile2.html and the rest of the files in the directory.

Imagine you have a file kept in a sub-directory that you wouldn't like to be spidered. What do you do? Lets suppose the sub-directory is "official" and the file is "confidential.html".

```
User-agent: *  
Disallow: /official/confidential.html
```

---

I hope that's enough. A little practice is of course required. If the syntax of your robots.txt file is not written correctly, the search engines will ignore that particular command. Before uploading the robots.txt file double check for any possible errors. You should upload robots.txt file in the ROOT Directory of your server. The search engines look for robots.txt file only in the root directory else they totally ignore it. Mostly root directory is the directory where the index page is kept. In that case keep the robots.txt file in the same directory as the index file.

I know a user-friendly software that will write robots command for you (the software is introduced at the

beginning of this article). It can make error-free robots.txt file very easily. This software RoboGen is a great tool. Never bother ever again to check the syntax of your robots.txt file or even write a robots.txt file yourself. RoboGen is a visual editor for Robot Exclusion Files and is easy to use. Just select files you want to be visited or not to be visited by the search engines, and it creates the robots.txt file. You can also select the search engines of your choice. RoboGen maintains a database of over 180 search engine user-agents, which are selectable from a drop down menu. It is the BEST and ONLY software on the Internet to write robots.txt file correctly and effectively. This great tool is cheaper than you expect. **CLICK HERE NOW** to know more!

Note: You should be able to see robots.txt file if you type the following in the address bar of your Internet browser.

```
http://www.your-domain.com/robots.txt
```

(Where your-domain is the domain name of your website. If yours is not a .com site, replace .com with the respective extension your website. For e.g. .net, .us, .org etc.)

You must be wondering whether to use Meta tag or Robots.txt or which of these is more effective!

A robots.txt correctly written is more effective than the meta tag. All search engines support robots.txt, but not all search engines support robots command written in the meta tags. I recommend that you use both so that you cover your site in both the scenarios. RoboGen will help you to write both!

One last thing – You can look in your web server log files to see what search engine robots have visited. They all leave signatures that can be detected. These signatures are nothing but name of their robots. For instance if Google has spidered your site it will leave a log file called Googlebot. This is how you know which search engine has spidered your pages and when!

Senior Manager – Internet Promotions <http://www.searchengineoptimizationpromotion.com>

## **The role of the robots.txt file to improve site ranking!**

**By Michael Kralj**

### **The role of the robots.txt file to improve site ranking! by Michael Kralj**

Not many web master take the time to use a robots.txt file for their website. For search engine spiders that use the robots.txt to see what directories to search through, the robots.txt file can be very helpful in keeping the spiders indexing your actual pages and not other information, such as looking through your stats!

The robots.txt file is useful in keeping your spiders from accessing parts folders and files in your hosting directory that are totally unrelated to your actual web site content. You can choose to have the spiders kept out of areas that contain programming that search engines cannot parse properly, and to keep them out of the web stats portion of your site.

Many search engines cannot view dynamically generated content properly, mainly created by programming languages, such as PHP or ASP. If you have an online store programmed in your hosting account, and it is in a separate directory, you would be wise to block out the spiders from this directory so it only finds relevant information.

The robots.txt file should be placed in the directory where your main files for your hosting are located. So you would be advised to create a blank text file, and save it as robots.txt, and then upload it to your web hosting to the same directory your index.htm file is located.

Here is examples of the use of the robots.txt file:

To block out a directory in a robots.txt file, such as a subdirectory for your online store called /store/ you would do the following:

```
Disallow: /store/
```

Another example to block out your stats directory:

```
Disallow: /stats/
```

You may also want to disallow individual files that you do not want searched by the search engines. For example you dont want search.php to be parsed by the Search Engines. To do this you type in the following on its own line:

```
Disallow: /search.php
```

Following the rules outlined and creating the robots.txt file, you will keep search engine spiders out of unwanted files and directories, and letting them go through the important files to see what your web site is all about!

Michael Kralj is owner of Emenki Web Solutions and Domains at Retail. Emenki Web Solutions are web site designers and programmers based in Hamilton, Ontario, providing businesses with an

informative and strategic approach to establishing an online presence on the web. Please visit Emenki Web Solutions on the web <http://www.emenki.com> Please visit Domains at Retail on the web: <http://www.domainsatretail.com>



This Free E-Book has been brought to you by [Natural-Aging.com](http://Natural-Aging.com).

**100% Effective Natural Hormone Treatment**  
**Menopause, Andropause And Other Hormone Imbalances**  
**Impair Healthy Healing In People Over The Age Of 30!**