Block bad bots from accessing your website

⚠ You can learn the general details of robots.txt in our Knowledge Base at this link.

What is a Bad Bot?
They can be thought of as the bots or spiders that do more harm than good to your website.

An example of a bad bot would be an email harvester which scans your web page code for email addresses that can then be used to send spam to. Another example is an unwanted bot which consumes too much bandwidth, or causes the load to go up on your server, causing it to go slow or at worst, completely offline due to overload.

While the worst of the "Bad Bots" will ignore your robots.txt directives completely, there are some bots that are not necessarily intending to be a Bad Bot, but may they may be unwanted by you. For the bots that ignore your robots.txt file, they would need to be blocked by using user-agent directives in your . htaccess file, but that topic is beyond the scope of this simple guide.

For the bots that are not intending to be malicious, but sometimes are, we can take care of them in your robots.txt file. For example, if you have a site based in the USA, you may not want bots from Non-English speaking countries coming in and eating up your bandwidth or other resources. Many bots will follow your rules, and this simple guide can help you to control the bots which access your site.

How to block Bad Bots
Follow these steps to block the bad bots and spiders from accessing your website.

Step 1:
Open your favorite text editor and create a file called **robots.txt**.

Step 2:
Place the following code in this file:

```
# Deny all robots that we do not specifically want to allow
User-agent: *
Disallow: /

# Allow these robots only
User-agent: googlebot
Allow: /
```

The code above will block all bots from accessing your website, with the exception of Google (googlebot).

ⓘ See the end of this post for more search engines / robots that are safe to add to your robots.txt file.

Step 3:
Save the file and upload it to your **public_html** directory. You can upload it via FTP or through the cPanel file manager.

More Good Bots to allow
The example above only uses Googlebot. There are others that you may want to add to your robots.txt file. Here are a few.

- Googlebot-News - Google News
- Googlebot-Image - Google Images
- Googlebot-Mobile - Google Mobile
- MSNBot - Microsoft MSN
- Teoma - Teoma Search
- bingbot - Bing Search
- Slurp - yahoo! Search
- Scooter - AltaVista Search
- Scrubby - Scrub the Web

You can add these bots into the robots.txt file in the following format:

```
User-agent: BOTNAME
Allow: /
```

Where **BOTNAME** is the name of the bot listed above.

So one example of a robots.txt file which bans all robots except yahoo, bing, and google might look like this:

```
# Deny all robots that we do not specifically want to allow
User-agent: *
Disallow: /

# Allow these robots only
User-agent: slurp
Allow: /

User-agent: bingbot
Allow: /

User-agent: googlebot
Allow: /
```

But if robots.txt doesn't help, you may block bots in your .htaccess file. First of all, we need to find out how to identify a bot. You will need to check your raw access logs using appropriate option in your Cpanel. The "User Agent" string in the logs is the one we need. For example, in the line below you may see **YandexBot** string:

```
HTTP/1.1" 200 927 "-" "Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)"
```

This is what we need. In order to block Yandex bot, you need to add the following into your .htaccess:

```
BrowserMatchNoCase YandexBot bad_bot
Order Deny,Allow
Deny from env=bad_bot
```

The other bots can be blocked by adding BrowserMatchNoCase directive in the same way.

More .htaccess Related

- .htaccess Robots, Redirects, Rewrites and Password Protected Directories