

## I, Robots.txt - They Came, They Saw, They Cataloged!

The robots.txt file is a file placed in your web server's root directory (meaning it should be accessible by typing [www.yoursite.com/robots.txt](http://www.yoursite.com/robots.txt)) that contains specific details about your site, making a search engine's job much easier, as well as telling it what NOT to index. This is called the 'Robot Exclusion Standard'.

The format for the robots.txt file is special. It consists of records. Each record consists of two fields : a User-agent line and one or more Disallow: lines. The format is:

```
[field] ":" [value]
```

The following tags are allowed in the robots.txt file, and examples are given for their usage:

### User-agent:

The User-agent line specifies the robot. For example, to disallow ALL robots:

```
User-agent: googlebot OR User-agent: *
```

You can find user agent names in your own logs by checking for requests to robots.txt. Most major search engines have short names for their spiders.

### Disallow:

The second part of a record consists of Disallow: directive lines. These lines specify files and/or directories. For example:

```
Disallow: email.htm OR Disallow: /cgi-bin/
```

If you leave the Disallow line blank, it indicates that ALL files may be retrieved. At least one disallow line must be present for each User-agent directive to be correct. A completely empty Robots.txt file is the same as if it were not present.

Any line in the robots.txt that begins with # is considered to be a comment only. The standard allows for comments at the end of directive lines, but is considered poor style:

```
Disallow: bob #comment
```

### EXAMPLE ROBOTS.TXT FILE:

```
-----  
#Allowing all robots everywhere:
```

```
User-agent: *
```

```
Disallow:
```

```
#This one keeps all those nosy robots out:
```

```
User-agent: *
```

```
Disallow: /
```

```
#The next one bars all robots from the illegal_documents and invoices  
directories:
```

```
User-agent: *
```

```
Disallow: /illegal_documents/
```

```
Disallow: /invoices/
```

```
#This one bans Google from poking around:
```

```
User-agent: Google
```

```
Disallow: /
```

```
#This one keeps googlebot from indexing "secret.html":  
User-agent: googlebot  
Disallow: secret.html
```

-----  
Once you are finished banning and allowing robots, run your file through the [Robots.txt file validator](#). Let us know [how you did!](#)

**..:About the Author**

*William Kinirons is the president of BMK Media, a web and graphic design company based in Coconut Creek that also offers in-home computer hardware & software support For more information on web design, onsite computer software support, or managed web hosting call BMK Media at (954) 818-2010.*