robots.txt File

robots.txt File HTML-5.com is a great guide for web developers. [TV Series & Actors and Actresses](). Follow [TV Series]() and [HTML 5]() on Google+.

[HTML-5.com]() ➤ itemscopehttp://data-vocabulary.org/Breadcrumb[HTML 5]() ➤ itemscopehttp://data-vocabulary.org/Breadcrumb **robots.txt File**

---

robots.txt File

You can use robots.txt to disallow search engine crawling of specific directories or pages on your web site.

Patterns in the robots.txt file are matched against the request URI, which starts with a slash ("/"). Therefore, exclusions should also start with a "/":

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

The trailing slash ("/") should also be included on directories in order to avoid matches with the prefix of other names. For example, a URI such as "/oldies.html" does *not* match the pattern in the example above.

The importance of robots.txt should not be underestimated. For example, we learned from experience that Google Search favored the printer-friendly PDF versions of the pages on this site over the HTML documents in its search engine results, so it was important to disallow robots from indexing PDF files:

```
User-agent: *
Disallow: /old/
Disallow: *.pdf
```

Some robots may not recognize patterns with wildcards such as "*", so those exclusions should appear last in a User-agent group.

Alternatives to the robots.txt file include:

• the [<meta name="robots"/> tag]()

• the [`rel="nofollow"`]() attribute, which doesn't necessarily prevent a search engine from indexing a page but keeps it from discovering the page via the link on which the attribute appears

For more information, see the [Search Engine Optimization (SEO) Tutorial]().

[back to top]()

THE END