

Navigation

- [index](#)
- [modules](#) |
- [next](#) |
- [previous](#) |
- [Python v2.6.4 documentation](#) »
- [The Python Standard Library](#) »
- [14. File Formats](#) »

14.3. robotparser — Parser for robots.txt

Note

The `robotparser` module has been renamed `urllib.robotparser` in Python 3.0. The [2to3](#) tool will automatically adapt imports when converting your sources to 3.0.

This module provides a single class, [RobotFileParser](#), which answers questions about whether or not a particular user agent can fetch a URL on the Web site that published the `robots.txt` file. For more details on the structure of `robots.txt` files, see <http://www.robotstxt.org/orig.html>.

```
class robotparser.RobotFileParser:
```

This class provides a set of methods to read, parse and answer questions about a single `robots.txt` file.

```
set_url(url)
```

Sets the URL referring to a `robots.txt` file.

```
read()
```

Reads the `robots.txt` URL and feeds it to the parser.

```
parse(lines)
```

Parses the `lines` argument.

```
can_fetch(useragent, url)
```

Returns `True` if the `useragent` is allowed to fetch the `url` according to the rules contained in the parsed `robots.txt` file.

```
mtime()
```

Returns the time the `robots.txt` file was last fetched. This is useful for long-running web spiders that need to check for new `robots.txt` files periodically.

```
modified()
```

Sets the time the `robots.txt` file was last fetched to the current time.

The following example demonstrates basic use of the `RobotFileParser` class.

```
>>> import robotparser
>>> rp = robotparser.RobotFileParser()
>>> rp.set_url("http://www.musi-cal.com/robots.txt")
>>> rp.read()
>>> rp.can_fetch("*", "http://www.musi-cal.com/cgi-bin/search?city=San+Francisco")
False
>>> rp.can_fetch("*", "http://www.musi-cal.com/")
True
```

Previous topic

[14.2. ConfigParser — Configuration file parser](#)

Next topic

[14.4. netrc — netrc file processing](#)

This Page

- [Show Source](#)

Navigation

- [index](#)
- [modules](#) |
- [next](#) |
- [previous](#) |
- [Python v2.6.4 documentation](#) »

- [The Python Standard Library](#) »
- [14. File Formats](#) »

© [Copyright](#) 1990-2010, Python Software Foundation.

The Python Software Foundation is a non-profit corporation. [Please donate.](#)

Last updated on Feb 26, 2010. Created using [Sphinx](#) 0.6.3.