

wget: Download entire websites easy

v1.0.4 (en)

Fooling sites to let wget crawl around

wget is a nice tool for downloading resources from the internet. The basic usage is `wget url`:

```
wget http://linuxreviews.org/
```

Therefore, wget ([manual page](#)) + less ([manual page](#)) is all you need to surf the internet. The power of wget is that you may download sites recursive, meaning you also get all pages (and images and other data) linked on the front page:

```
wget -r http://linuxreviews.org/
```

But many sites do not want you to download their entire site. To prevent this, they check how browsers identify. Many sites refuses you to connect or sends a blank page if they detect you are not using a web-browser. You might get a message like:

Sorry, but the download manager you are using to view this site is not supported. We do not support use of such download managers as flashget, go!zilla, or getright

Wget has a very handy `-U` option for sites like this. Use `-U My-browser` to tell the site you are using some commonly accepted browser:

```
wget -r -p -U Mozilla http://www.stupidsite.com/restricedplace.
```

The most important command line options are `--limit-rate=` and `--wait=`. You should add `--wait=20` to pause 20 seconds between retrievals, this makes sure you are not manually added to a blacklist. `--limit-rate` defaults to bytes, add `K` to set KB/s. Example:

```
wget --wait=20 --limit-rate=20K -r -p -U Mozilla
http://www.stupidsite.com/restricedplace.html
```

A web-site owner will probably get upset if you attempt to download his entire site using a simple `wget http://foo.bar` command. However, the web-site owner will not even notice you if you limit the download transfer rate and pause between fetching files.

Use `--no-parent`

`--no-parent` is a very handy option that guarantees `wget` will not download anything from the folders beneath the folder you want to acquire. Use this to make sure `wget` does not fetch more than it needs to if just want to download the files in a folder.

Copyright (c) 2000-2004 Øyvind Sæther. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

> [Linux Reviews](#) > [Tips and Tricks](#) >
wget: Download entire websites easy