

All the Wget Commands You Should Know



/ By Amit Agarwal / Published on 2014-12-09

PUBLISHED IN: **LINUX**

How do I download an entire website for offline viewing? How do I save all the MP3s from a website to a folder on my computer? How do I download files that are behind a login page? How do I build a mini-version of Google?

Wget is a free utility - available for Mac, Windows and Linux (included) - that can help you accomplish all this and more. What makes it different from most download managers is that wget can follow the HTML links on a web page and recursively download the files. It is the same tool that a soldier had used to download thousands of secret documents from the US army's Intranet that were later published on the Wikileaks website.

You mirror an entire website with wget

Spider Websites with Wget – 20 Practical Examples

Wget is extremely powerful, but like with most other command line programs, the plethora of options it supports can be intimidating to new users. Thus what we have here are a collection of wget commands that you can use to

accomplish common tasks from downloading single files to mirroring entire websites. It will help if you can read through the [wget manual](#) but for the busy souls, these commands are ready to execute.

1. Download a single file from the Internet

```
wget http://example.com/file.iso
```

2. Download a file but save it locally under a different name

```
wget --output-document=filename.html example.com
```

3. Download a file and save it in a specific folder

```
wget --directory-prefix=folder/subfolder example.com
```

4. Resume an interrupted download previously started by wget itself

```
wget --continue example.com/big.file.iso
```

5. Download a file but only if the version on server is newer than your local copy

```
wget --continue --timestamping wordpress.org/latest.zip
```

6. Download multiple URLs with wget. Put the list of URLs in another text file on separate lines and pass it to wget.

```
wget --input list-of-file-urls.txt
```

7. Download a list of sequentially numbered files from a server

```
wget http://example.com/images/{1..20}.jpg
```

8. Download a web page with all assets - like stylesheets and inline images - that are required to properly display the web page offline.

```
wget --page-requisites --span-hosts --convert-links --adjust-extend
```

Mirror websites with Wget

9. Download an entire website including all the linked pages and files

```
wget --execute robots=off --recursive --no-parent --continue --no-
```

10. Download all the MP3 files from a sub directory

```
wget --level=1 --recursive --no-parent --accept mp3,MP3 http://exa
```

11. Download all images from a website in a common folder

```
wget --directory-prefix=files/pictures --no-directories --recursiv
```

12. Download the PDF documents from a website through recursion but stay within specific domains.

```
wget --mirror --domains=abc.com,files.abc.com,docs.abc.com --accept
```

13. Download all files from a website but exclude a few directories.

```
wget --recursive --no-clobber --no-parent --exclude-directories /f
```

Wget for Downloading Restricted Content

Wget can be used for downloading content from sites that are behind a login screen or ones that check for the HTTP referer and the User Agent strings of the bot to prevent screen scraping.

14. Download files from websites that check the User Agent and the HTTP Referer

```
wget --refer=http://google.com --user-agent="Mozilla/5.0 Firefox/4
```

15. Download files from a password protected sites

```
wget --http-user=labnol --http-password=hello123 http://example.cc
```

16. Fetch pages that are behind a login page. You need to replace user and password with the actual form fields while the URL should point to the Form Submit (action) page.

```
wget --cookies=on --save-cookies cookies.txt --keep-session-cookie
```

Retrieve File Details with wget

17. Find the size of a file without downloading it (look for Content Length in the response, the size is in bytes)

```
wget --spider --server-response http://example.com/file.iso
```

18. Download a file and display the content on screen without saving it locally.

```
wget --output-document - --quiet google.com/humans.txt
```

19. Know the last modified date of a web page (check the Last Modified tag in the HTTP header).

```
wget --server-response --spider http://www.labnol.org/
```

20. Check the links on your website to ensure that they are working. The spider option will not save the pages locally.

```
wget --output-file=logfile.txt --recursive --spider http://example
```

Also see: [Essential Linux Commands](#)

Wget – How to be nice to the server?

The wget tool is essentially a spider that scrapes / leeches web pages but some web hosts may block these spiders with the robots.txt files. Also, wget will not follow links on web pages that use the [rel=nofollow](#) attribute.

You can however force wget to ignore the robots.txt and the nofollow directives by adding the switch **--execute robots=off** to all your wget commands. If a web host is blocking wget requests by looking at the User Agent string, you can always fake that with the **--user-agent=Mozilla** switch.

The wget command will put additional strain on the site's server because it will continuously traverse the links and download files. A good scraper would therefore limit the retrieval rate and also include a wait period between consecutive fetch requests to reduce the server load.

```
wget --limit-rate=20k --wait=60 --random-wait --mirror example.com
```

In the above example, we have limited the download bandwidth rate to 20 KB/s and the wget utility will wait anywhere between 30s and 90 seconds before retrieving the next resource.

Finally, a little quiz. What do you think this wget command will do?

```
wget --span-hosts --level=inf --recursive dmoz.org
```

PUBLISHED IN: **LINUX**

SHARE ON:



You'll also like:

- [Use an Old Linux Computer to Put your Baby to Sleep](#)
- [How to Install Linux on your Windows Computer](#)
- [Google Chrome for Linux and Mac](#)
- [Print Files on Linux Remotely using Dropbox](#)
- [Free Virtualization Software for your Windows PC](#)
- [Which is the Best Linux Distribution for your Desktop?](#)
- [How to Get Free Ubuntu CDs and DVDs in India](#)
- [Why Windows Users Buy Laptops Preloaded with Free DOS?](#)
- [Two Users Can Work On The Same Computer At The Same Time](#)
- [Create Screencasts in Linux with Screenr](#)

Looking for something? Find here!

Meet the Author

Web Geek, Tech Columnist



[Amit Agarwal](#) is a Google Developer Expert in GSuite and Google Apps Script. He holds an engineering degree in Computer Science (I.I.T.) and is the first professional blogger in India. Read more on [Lifehacker](#) and [YourStory](#).

Get in touch

- Send an email at amit@labnol.org
- Follow on [@twitter](#) and [@youtube](#).
- Subscribe to our [email newsletter](#) and never miss an update.

OVERVIEW

Digital Inspiration, established in 2004, helps businesses automate processes and improve productivity with Google services.

© 2004 — 2020 Amit Agarwal. All rights reserved.

RESOURCES

[About Us](#)

[Products](#)

[Code Snippets, Topics](#)

[Tech Support](#)

CONTACT

Amit Agarwal

phone: +91-562-4004563

email: amit@labnol.org



digital inspiration