



## Downloading an Entire Web Site with wget

Sep 05, 2008 By [Dashamir Hoxha \(/user/1002681\)](#)  
in

If you ever need to download an entire Web site, perhaps for off-line viewing, wget can do the job—for example:

```
$ wget \
  --recursive \
  --no-clobber \
  --page-requisites \
  --html-extension \
  --convert-links \
  --restrict-file-names=windows \
  --domains website.org \
  --no-parent \
  www.website.org/tutorials/html/ (http://www.website.org/tutorials/html/)
```

This command downloads the Web site [www.website.org/tutorials/html/](http://www.website.org/tutorials/html/) (<http://www.website.org/tutorials/html/>).

The options are:

- --recursive: download the entire Web site.
  - --domains website.org: don't follow links outside website.org.
  - --no-parent: don't follow links outside the directory tutorials/html/.
  - --page-requisites: get all the elements that compose the page (images, CSS and so on).
  - --html-extension: save files with the .html extension.
  - --convert-links: convert links so that they work locally, off-line.
  - --restrict-file-names=windows: modify filenames so that they will work in Windows as well.
  - --no-clobber: don't overwrite any existing files (used in case the download is interrupted and resumed).
-

## Comments

### Comment viewing options

Threaded list - expanded  Date - newest first  50 comments per page

Select your preferred way to display the comments and click "Save settings" to activate your changes.

#### [its nice to read cool post](#) (/content/downloading-entire-web-site-wget#comment-356811)

Submitted by [quran online](#) (<http://www.learningquranonline.com>) (not verified) on Tue, 10/12/2010 - 13:29.

wile reading this article i feel more touched the way you explained and made it easy from the readers to read its quite resourceful keep the good work tack care  
[read quran online](#)



#### [How to view as a web page](#) (/content/downloading-entire-web-site-wget#comment-356486)

Submitted by Slagathor (not verified) on Fri, 10/01/2010 - 12:20.

I am now downloading a site, using the "wget -m" option, and there are all this files within a folder. How do I view it in a web browser? :)



#### [Thanks](#) (/content/downloading-entire-web-site-wget#comment-356216)

Submitted by herif (not verified) on Thu, 09/23/2010 - 23:58.

Thank you, for these helpful information  
than you,



#### [wget](#) (/content/downloading-entire-web-site-wget#comment-355991)

Submitted by Anonymous (not verified) on Sun, 09/19/2010 - 04:26.

I had make a script to dowload a html file from the website. But it takes all in the sense header , body footer like that .i want only certain text to be copied. can u help me



#### [wGet script](#) (/content/downloading-entire-web-site-wget#comment-354671)

Submitted by [s3nzo](#) (<http://konvict.mobi>) (not verified) on Mon, 08/09/2010 - 04:07.

Can you please post the full code



#### [Thanks!](#) (/content/downloading-entire-web-site-wget#comment-349901)

Submitted by [Mike T.](#) (<http://www.mikestechblog.com>) (not verified) on Thu, 03/25/2010 - 07:33.

Thanks a lot for posting this. I needed to get a backup of a website quickly and easily.  
Wget did the trick!



#### [wget -m http://website.com](#) (/content/downloading-entire-web-site-wget#comment-347724)

Submitted by Anonymous (not verified) on Thu, 01/21/2010 - 17:52.

More easy:

wget -m <http://website.com> (<http://website.com>)



#### [Mirroring website with -m Flag](#) (/content/downloading-entire-web-site-wget#comment-355045)

Submitted by [Ahamed Bauani](#) (<http://basic-linux.bauani.org/>) (not verified) on Thu, 08/19/2010 - 03:03.

Yes, weget has it's own built-in signal flag '-m' or '--mirror' which is easy to use.

wget -m <http://basic-linux.bauani.org/> (<http://basic-linux.bauani.org/>)

will give you the full website on your HDD.

But the script above his more powerful, as I can control the downloading speed. More important thing is if you are going to use -m option to a well maintained, you most probably got BAN. As



you are ignoring rootbs.txt , downloading files without any delay. Like if you found someone is downloading ALL of your files with speed, making high load on your web server, will you allow it?

Regards

Ahamed Bauani

[Bauani's Technology Related Blog \(http://blog.bauani.org\)](http://blog.bauani.org)

**[Thank You for the Script](#) (/content/downloading-entire-web-site-wget#comment-355046)**

Submitted by [Ahamed Bauani](#) (<http://basic-linux.bauani.org/>) (not verified) on Thu, 08/19/2010 - 03:07.

Ho, the main thing, Thanks to the writer of script, it is very useful to me or probably others.

Thanks again.  
Ahamed Bauani



**[Please use example.com.](#) (/content/downloading-entire-web-site-wget#comment-347689)**

Submitted by Anonymous (not verified) on Wed, 01/20/2010 - 12:10.

Thanks a lot! I needed to mirror a site on our local LAN, and this kept me from having to re-familiarize myself with the man page. But PLEASE use example.com for a placeholder domain name. It is reserved for exactly that purpose.



**[options that you should add to the main article](#) (/content/downloading-entire-web-site-wget#comment-342063)**

Submitted by [magi182](http://www.asomatic.net) (<http://www.asomatic.net>) (not verified) on Wed, 08/26/2009 - 09:51.

It would be a VERY good idea to add:

```
--wait=9 --limit-rate=10K
```

to your command so you don't kill the server you are trying to download from.

the --wait option introduces a number of seconds to wait between download attempts, the --limit-rate limits the amount of the servers bandwidth you are sucking up. Both good ideas if you don't want to be blacklisted by the servers admin.



**[Thanks](#) (/content/downloading-entire-web-site-wget#comment-339932)**

Submitted by Mini (not verified) on Fri, 07/24/2009 - 06:47.

What options of wget should I use to retrieve all the pages related to the links from a main search page ? I've been trying for days to achieve it, using Linux.

Faleminderit per postimin, ( tr - thanks for posting )



**[Good work](#) (/content/downloading-entire-web-site-wget#comment-339415)**

Submitted by [Arjun Pakrashi](http://phoxis.wordpress.com) (<http://phoxis.wordpress.com>) (not verified) on Thu, 07/09/2009 - 06:13.

I was always finding suggestions on appropriate switches to be used when downloading a complete website. This piece of document was very helpful. Thanks and keep up the good work.



**[Great](#) (/content/downloading-entire-web-site-wget#comment-326118)**

Submitted by [partero](#) ([/users/partero](#)) on Thu, 10/02/2008 - 16:52.

Very good instructions



([/users/partero](#))

**[My use](#) (/content/downloading-entire-web-site-wget#comment-325493)**

Submitted by Luigimax (not verified) on Mon, 09/15/2008 - 16:39.

Just to start, this post is most helpfull. Dashamir Hoxha thanks alot!

the reason for writing this is when downloading multiple sites in sequence will take much time. so to easily download multiple sites i set this up. and yes it would be more efficient to put the pipe command in the scrpt file.

what im using it for: dowload multiple websites (manga specifcly)

step 1: put the wget command in a script file (for ease of use)

```
#!/bin/bash
```

```
wget -r --page-requisites --convert-links --no-parent -I $2 -U Mozilla $1
```

ill call mine "meget"



run command: `chmod +x meget`  
is what i put in mine. how to use: `[script-name] [target website] [scan depth]`

step 2:

make a file with all the websites you want to download - one per line. ill call mine "zone"

step 3: run command:

```
cat zone | xargs -n1 -P 3 -i ./meget {} 1000
```

to increase the number of parallel downloads change the 3 to whatever number you need. keep in mind not to have a list of 300 sites and download them all at once - this may cause problems

be sure to also set the 1000 number to the depth you need. in my case to download a 1500 page manga i need to set it upto 1500 or more.

when it is running it will only show one download at a time. if still running it will always show something.

---

**[Just what I've been looking for](#)** [//content/downloading-entire-web-site-wget#comment-325280](#)

Submitted by TsueDesu (not verified) on Tue, 09/09/2008 - 08:34.

Just this week I needed to make a site available offline so I can reference to it while working at home. And YaY!! I have wget and love using it already. However, I advise taking note of how wget is saving the files, if it's a site with lots of PHP pages, then you'll have to change the reference in every .php to .php.html ... Not to fear though, your computer can already do the hard work for you. Just type

```
grep -rl .php *.html | xargs perl -pi~ -e 's/.php/.php.html/'
```

et voila Your pages will open and link with out a hitch...really interesting and marvelous this Linux thing.



---

**[I know that your post is](#)** [//content/downloading-entire-web-site-wget#comment-355050](#)

Submitted by Silvia (not verified) on Thu, 08/19/2010 - 04:34.

I know that your post is quite old, but I just wanted to add that wget can convert the references to renamed files in downloaded pages with option -k (--convert-links). This option is also very useful if you haven't downloaded all the referenced files, check out its magic in the manual.



---

**[I know that your post is](#)** [//content/downloading-entire-web-site-wget#comment-355049](#)

Submitted by Silvia (not verified) on Thu, 08/19/2010 - 04:34.

I know that your post is quite old, but I just wanted to add that wget can convert the references to renamed files in downloaded pages with option -k (--convert-links). This option is also very useful if you haven't downloaded all the referenced files, check out its magic in the manual.

