

WGet and Downloading an entire remote directory

Sometimes you need to retrieve a remote url (directory) with everything inside it, when you miss a live presentation or a forum you often find material published on the web in a certain site; you'd like to get the entire presentation (usually several html pages and links) or read it online.

When no "download all" button is available or when you don't have spare time to read it immediately you wish to grab all the directory content and read it offline later, I usually download material for my personal digital library, I always think: *"maybe I don't need that tip now but in the future it will help.."*

Everybody knows **wget** and how to use it, it's one of my favorite tools especially when I need to download an ISO or a single file, using wget with recurse on an entire site is not a big problem but when you need to download only a specified directory it could cause headaches when dealing with different options

Here's what I do when I need to download a specific directory located on a remote site (an html presentation for example)

```
wget -r -level=0 -E -ignore-length -x -k -p -erobots=off -np -N  
http://www.remote.com/remote/presentation/dir
```

Here are the options:

- r** : Recursive retrieving (important)
- level=0**: Specify recursion maximum depth level (0 for no limit), very important
- E**: append ".html" extension to every document declared as "**application/html**"
useful when you deal with dirs (that are not dirs but index.html files)
- ignore-length**: Ignore "Content-length" http headers, sometimes useful when dealing with bugged CGI programs
- x**: Force dirs, create an hierarchy of directories even if one would not been created otherwise
- k**: here's one of the most useful options, it converts remote links to local for best viewing
- p**: download ll the files that are necessary for proper display of the page
(not so reliable when dealing with JS code but useful)
- erobots=off**: turn off http robots.txt usage

-np: no parent, do not ascend to parent dir when retrieving recursively,
one of the most useful function I've seen

Please keep in mind this is focused on downloading only a specific dir and for html presentations or materials you can even view offline, when dealing with flash or other kind of contents it could be a mess, for my basic needs I've just created a script and I pass the url to it