

wget

General Usage

```
wget -Flags URL
```

Full site mirror

```
wget -mkprE -np -nc -l inf -e robots=off --restrict-file-names=windows -D example.com --no-check-certificate http://example.com
```

Flags

- `--mirror / -m` Turns on all options to make a mirror copy of a site (`--recursive` , `--timestamping` , `--level=inf` , `--no-remove-listing`)`
- `--convert-links / -k` Converts the links in the downloaded documents to make them suitable for offline viewing. This means that all the links will point to the local files instead of the original URLs.
- `adjust-extension / -E` Adjusts the file extension based on the MIME type. For example, if a file is served as HTML but doesn't have an .html extension, this flag will rename the file to include the .html extension
- `--page-requisites / -p` Downloads all the resources that a page requires to display properly, such as images, CSS, and JavaScript files.
- `--restrict-file-names=windows` Modifies filenames so that they are compatible with Windows. This avoids issues with characters that are not allowed in Windows filenames (e.g : or *).
`--domains example.com` : - Restricts the download to the specified domain (example.com in this case). wget will only download files from this domain and no other domains.`
- `--no-clobber / -nc` Prevents wget from overwriting existing files. If a file already exists, wget will skip the download of that file.
- `--no-check-certificate` Disables SSL/TLS certificate validation. This can be useful if the server's certificate is self-signed or otherwise not trusted by default.
- `-e robots=off` Ignores the robots.txt file on the server, allowing wget to download pages and resources that might otherwise be disallowed by the server's robots.txt file. Use this

option responsibly.

- `--recursive / -r` Enables recursive downloading. This means wget will follow links found in the downloaded files and download those files as well. This is necessary for downloading entire websites or sections of websites.
- `--level=inf` Specifies the maximum recursion depth. Setting it to inf (infinity) means wget will follow links at any depth, effectively downloading the entire site as long as links are found.
- `--no-parent / -np` Prevents wget from following links outside the specified directory. This means it won't ascend to parent directories but will stay within the directory structure of the specified URL.

Revision #4

Created 22 March 2025 09:38:28 by Conor

Updated 22 March 2025 09:45:46 by Conor