Mirroring a site using Wget

Wget is the GNU/FSF alternative to CURL used to retrieve files from the network via command line. Gettting a single file is easy but trying to get an entire directory it's not so easy.

I've used a similar version to this command in the past but it turned into huge downloads that would take forever to complete.

I found this version in Guy Rutenberg's site and liked it after I used it to mirror a site.

The command is:

```
wget --mirror \
--convert-links \
--adjust-extension \
--page-requisites
--no-parent http://example.org
```

These are flags we're using:

- **–mirror** – Mirrors a site by making the download recursive and enabling additional flags
- **–convert-links** – After the download is complete, convert the links in the document to make them suitable for local viewing. This affects not only the visible hyperlinks, but any part of the document that links to external content
- **–adjust-extension** – If a file of type `application/xhtml+xml` or `text/html` is downloaded and the URL does not end with the regexp `\.[Hh][Tt][Mm][Ll]?`, this option will cause the suffix '.html' to be appended to the local filename.

  This is useful, for instance, when you're mirroring a remote site that uses '.asp' pages, but you want the mirrored pages to be viewable on your stock Apache server

- **–page-requisites** – Download things like CSS style-sheets and images required to properly display the page offline.
- **–no-parent** – Do not go to the parent directory. It useful for restricting the download to only a portion of the site.

Alternatively, the command above may be shortened:

```
wget -mkEpnp http://example.org
```

If the server's robots.txt file is not configured to forbid it, you can run this command to mirror the specified directory and prepare it for offlline viewing by making some changes to the content and downloading all the needed materials to make the page viewable offlline or in a different server.