webpage_scraper

cli app that given a URL, scraps a website: HTML (and its pandoc Markdown conversion), info JSON and images.

Also contains the `webpage2pdf` binary that only from the URL converts the webpage to PDF.
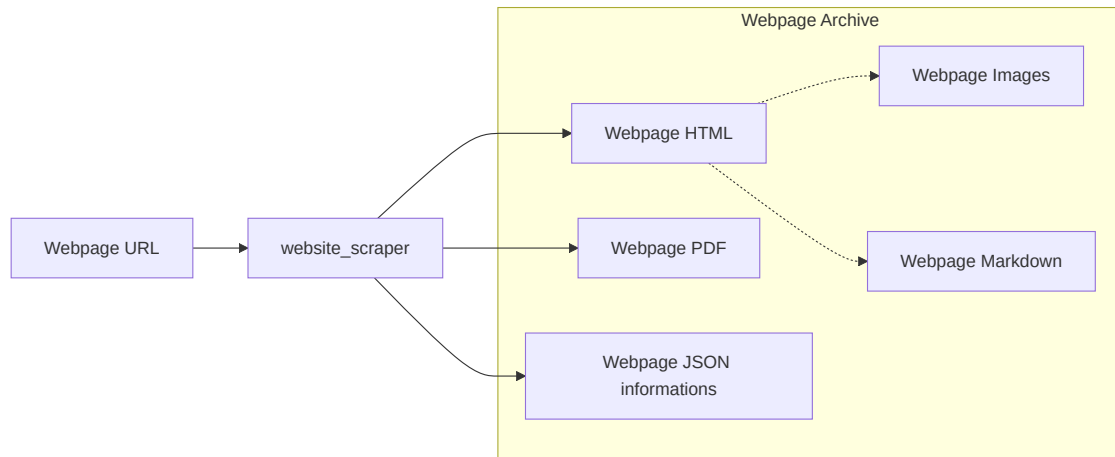

Figure 1: Overview of `webpage_scraper`

Dependencies

Both binaries need document converter [pandoc](#) installed.

Usage

```
Usage: webpage_scraper <URL> [OUTPUT_DIRECTORY]

Arguments:
  <URL>               URL of the webpage to be scraped
  [OUTPUT_DIRECTORY]  Name of the output_directory if not given, will use the name o

Options:
  -h, --help     Print help
  -V, --version  Print version
```

```
Usage: webpage2pdf <URL>

Arguments:
  <URL>  URL of the website to convert to PDF

Options:
  -h, --help     Print help
  -V, --version  Print version
```