Exercises in static scraping using Scrapy

Anna Lewczuk & Przemysław Kurek

Exercise 1

We will work with a wikipedia web page which contains the list of musicians.

`https://en.wikipedia.org/wiki/Lists_of_musicians`

- Extract links to web pages containing information about musicians specialising in music genre beginning with "A".
- Extract links to artists' web pages for the first of the links from the previous step.
- From the artist pages extract: the name of the band and years active.

Submissions

- Create three Scrapy spiders named `"s_1.py"`, `"s_2.py"`, `"s_3.py"`.
- For spider names use: `link_lists`, `links`, `musicians`.
- Each of those spiders must contain full spider, which after being copied to `spiders` folder can be run without errors with `scrapy crawl spider_name -o output_name.csv` command (test it in command line).
- As output files (which will be used in the next steps) use: `link_lists.csv`, `links.csv`, `musicians.csv`.
- Commit three spider files to your Github Classroom repository into `"06"` folder. Do not commit anything else.
- The deadline is: nearest Sunday, 23:59.
- You can try using Jupyter Notebook.