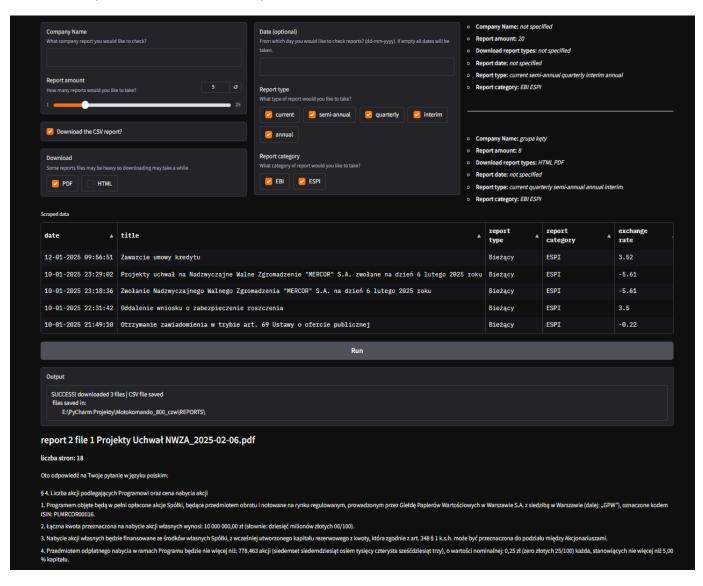# GPW Scraping Tool

- Jakub Żakowski
- Michał Kornatowski
- Filip Maksymiuk
- Piotr Kowalczyk

What is a GPW Scraping Tool?

GPW (stock exchange) Scraping Tool is a tool that allows you to download reports from the WSE stock exchange. The tool uses a scraping script that automatically downloads reports for selected companies, based on certain parameters.

Functionality

1. Downloading reports for selected companies from the Stock Exchange.

2. Filtering of reports based on date, report type and category.

3. Ability to download attachments in PDF or HTML format.

4. Exporting data to CSV file.

5. Exporting data to MYSQL database.

6. Summary downloaded reports

7. Displays history searches

Content

- user_interface.py: Gradio-based graphical user interface for convenient parameter entry and display of results.
- scrape_script.py: Script responsible for downloading reports and attachments from the stock exchange.
- summary.py: Part responsible for summarise using local model
- database_connection.py: Part responsible for database operations
- gpw_data.sql: MySQL database with structure and example data
- REPORTS: folder that will be created when you run the script. Reports and CSV data are saved here.

Requirements

- At least 10GB free on disk
- git
- Windows only (for now)
- Internet connection
- NVIDIA GPU (The better the GPU, the faster the summaries will show up)
- `Python 3.12` (Microsoft store)
- any database hosting like XAMPP (https://www.apachefriends.org/pl/download.html)
- ollama installed with `llama3.2:latest` model (https://ollama.com/)

Installation

Open CMD in desire directory and clone project using: ### Step 1 (clone repository)

```
git clone https://github.com/xHezuSx/Motokomando_800_czw
```

Step 2 (setup database)

1. run XAMPP with apache and MYSQL
2. click `Admin` button browser with admin panel will appear
3. select `New` and name database `gpw data` and click `Create`. **IMPORTANT name has to be `gpw data` otherwise it won't work**
4. select `Import` section and select `gpw_data.sql` file from cloned repository
5. scroll down and click `Import`

Step 3 (setup ollama)

1. Download ollama (https://ollama.com/)
2. Ollama PATH = `C:\Users\<USERNAME>\AppData\Local\Programs\Ollama` it's your username on Windows
3. open settings search for `environment variables for system`
4. click `environment variables` select `PATH` and `edit`
5. click `New` and paste your ollama PATH

Step 4 (download model)

1. Open CMD and run download AI model (2GB)

```
ollama pull llama3.2
```

2. after download use:

```
ollama serve
```

Step 5 (install dependencies)

On windows right click on folder where project is, select "open in terminal/CMD" option and run following command:

```
pip install -r requirements.txt
```

Usage

Step 1 (make sure ollama is running)

- open CMD and use comamnd

```
ollama serve
```

Step 2 (make sure XAMPP (or other MySQL engine) is running)

On windows right click on folder where project is, select "open in terminal/CMD" option and run following command:

Step 3 (Run app)

If you run app for the first time it will take more time

```
python .\user_interface.py
```

if everything is good you should see `Running on local URL:` http://127.0.0.1:7860 Simply open the link and voilà :)

Pssst… don't worry about warnings in CMD