

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.

Company Name

What company report you would like to check?

Report amount

How many reports would you like to take?

1

5

25

☒ Download the CSV report?

Download

Some reports files may be heavy so downloading may take a while

☒ PDF

☐ HTML

Date (optional)

From which day you would like to check reports? (dd-mm-yyyy). If empty all dates will be taken.

Report type

What type of report would you like to take?

☒ current

☒ semi-annual

☒ quarterly

☒ Interim

☒ annual

Report category

What category of report would you like to take?

☒ EBI

☒ ESPI

Company Name: not specified

Report amount: 20

Download report types: not specified

Report date: not specified

Report type: current semi-annual quarterly interim annual

Report category: EBI ESPI

Scraped data

date	title	report type	report category	exchange rate
12-01-2025 09:56:51	Zawarcie umowy kredytu	Bieżący	ESPI	3.62
10-01-2025 23:29:02	Projekty uchwał na Nadzwyczajne Walne Zgromadzenie "MERCOR" S.A. zwołane na dzień 6 lutego 2025 roku	Bieżący	ESPI	-5.61
10-01-2025 23:18:36	Zwołanie Nadzwyczajnego Walnego Zgromadzenia "MERCOR" S.A. na dzień 6 lutego 2025 roku	Bieżący	ESPI	-5.61
10-01-2025 22:31:42	Oddalenie wniosku o zabezpieczenie roszczenia	Bieżący	ESPI	3.5
10-01-2025 21:49:10	Otrzymanie zawiadomienia w trybie art. 69 Ustawy o ofercie publicznej	Bieżący	ESPI	-0.22

Run

Output

SUCCESS! downloaded 3 files | CSV file saved
files saved in:
E:\PyCharm Projekty\Motokomando_800_czw\REPORTS\

report 2 file 1 Projekty Uchwał NWZA_2025-02-06.pdf

liczba stron: 18

Oto odpowiedź na Twoje pytanie w języku polskim:

§ 4. Liczba akcji podlegających Programowi oraz cena nabycia akcji

1. Programem objęte będą w pełni opłacone akcje Spółki, będące przedmiotem obrotu i notowane na rynku regulowanym, prowadzonym przez Giełdę Papierów Wartościowych w Warszawie S.A. z siedzibą w Warszawie (dalej: „GPW”), oznaczone kodem ISIN: PLMRCOR00016.

2. Łączna kwota przeznaczona na nabycie akcji własnych wynosi: 10 000 000,00 zł (słownie: dziesięć milionów złotych 00/100).

3. Nabycie akcji własnych będzie finansowane ze środków własnych Spółki, z wcześniej utworzonego kapitału rezerwowego z kwoty, która zgodnie z art. 348 § 1 k.s.h. może być przeznaczona do podziału między Akcjonariuszami.

4. Przedmiotem odpłatnego nabycia w ramach Programu będzie nie więcej niż: 778 463 akcji (siedemset siedemdziesiąt osiem tysięcy czterysta sześćdziesiąt trzy), o wartości nominalnej: 0,25 zł (zero złotych 25/100) każda, stanowiących nie więcej niż 5,00 % kapitału.

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 comand

```
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