

**Uniwersytet Przyrodniczo – Humanistyczny
w Siedlcach**

Kierunek Informatyka

Cezary Borkowski

Temat zadania:

Program pobierający dane z Biuletynu Informacji Publicznej

Praca wykonana pod kierunkiem
Dr Piotra Świtalskiego

Siedlce, 2022r.

Spis treści

| | |
|--|----------|
| Opis programu | 3 |
| Opis rozwiązania | 3 |
| Możliwości uruchamiania | 3 |
| Możliwości uruchamiania – z poziomu CMD | 3 |

Opis programu

Cele programu jest pobranie danych dotyczących aktów prawnych ze strony Biuletynu Informacji publicznej UPH oraz zapisanie ich w odpowiedniej postaci (JSON).

Opis rozwiązania

Program został napisany w języku Python, z wykorzystaniem biblioteki BeautifulSoup, służącej do pobierania odpowiednich danych ze znaczników z wczytanej strony. Aplikacja działa tylko i wyłącznie dla aktów prawnych.

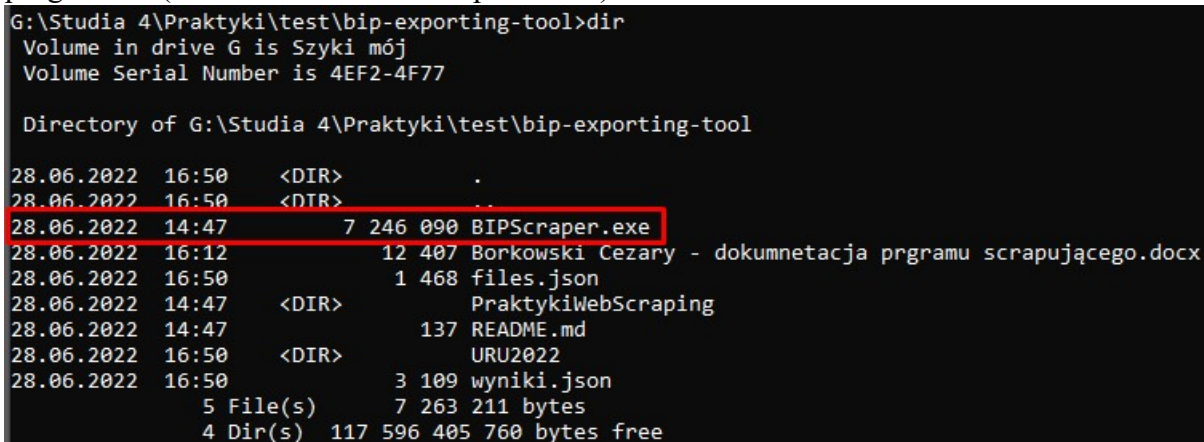
Możliwości uruchamiania

Program można uruchomić na kilka sposobów:

- poprzez plik wykonywalny .exe, klikając na niego
- poprzez uruchomienie pliku .py lub .exe z poziomu wiersza poleceń i podanie argumentu po nazwie, czyli linku do odpowiedniej zakładki aktów prawnych (brak argumentu oznacza przeszukiwanie od początku)

Możliwości uruchamiania – z poziomu CMD

Krok 1: Przejdźcie do odpowiedniego folderu na komputerze gdzie znajdują się plik z programem (zalecane uruchomienie pliku .exe)



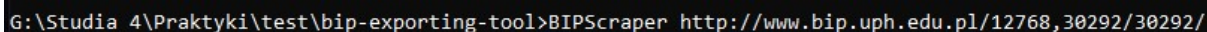
```
G:\Studia 4\Praktyki\test\bip-exporting-tool>dir
Volume in drive G is Szyki mój
Volume Serial Number is 4EF2-4F77

Directory of G:\Studia 4\Praktyki\test\bip-exporting-tool

28.06.2022  16:50    <DIR>          .
28.06.2022  16:50    <DIR>          ..
28.06.2022  14:47         7 246 090 BIPScrapper.exe
28.06.2022  16:12         12 407 Borkowski Cezary - dokumnetacja prgramu scrapującego.docx
28.06.2022  16:50           1 468 files.json
28.06.2022  14:47    <DIR>          PraktykiWebScraping
28.06.2022  14:47          137 README.md
28.06.2022  16:50    <DIR>          URU2022
28.06.2022  16:50           3 109 wyniki.json
                5 File(s)          7 263 211 bytes
                4 Dir(s)  117 596 405 760 bytes free
```

Rysunek 1 Folder zawierający program

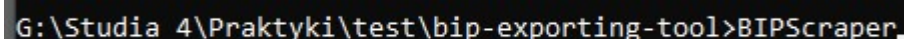
Krok 2a: Uruchomienie programu z argumentem



```
G:\Studia 4\Praktyki\test\bip-exporting-tool>BIPScrapper http://www.bip.uph.edu.pl/12768,30292/30292/
```

Rysunek 2 Uruchomienie programu z argumentem

Krok 2b: Uruchomienie programu bez argumentu



```
G:\Studia 4\Praktyki\test\bip-exporting-tool>BIPScrapper
```

Rysunek 3 Uruchomienie programu z argumentem

Krok 3: Wyniki programu

W CMD wyświetla się aktualnie przerabiany akt:

```
Przedmiot aktu: w sprawie wyrażenia opinii dotyczącej zmian w statucie Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach
Status aktu: 60d8b61ccce322714ed30c40
Numer rocznik aktu: 3/2022
Data wydania: 2022-03-24T23:00:00Z
RodzajAktu: 60dade91cce322714ed30c54
2022-04-07T13:50:00Z
2022-06-07T14:48:00Z

UPH

Przedmiot aktu: w sprawie zaopiniowania sprawozdania z realizacji strategii uczelni w roku 2021
Status aktu: 60d8b61ccce322714ed30c40
Numer rocznik aktu: 2/2022
Data wydania: 2022-02-10T23:00:00Z
RodzajAktu: 60dade91cce322714ed30c54
2022-02-28T14:22:00Z
2022-06-07T14:48:00Z

UPH
```

Rysunek 4 Wynik w CMD

Dane o aktach zapisywane są do pliku wyniki.json:

```
studia 4 > Praktyki > test > bip-exporting-tool > {!} wyniki.json > ...
[{"_id":{"$_id":"62bb1548c5b24717712d404c"},"przedmiotAktu":"w sprawie wyrażenia opinii dotyczącej użyczenia nieruchomości DS Nr 3 przy ul. Bema"},
{"_id":{"$_id":"62bb1549c5b24717712d404e"},"przedmiotAktu":"w sprawie zaopiniowania zmian w Strategii Uniwersytetu Przyrodniczo-Humanistycznego"},
{"_id":{"$_id":"62bb1549c5b24717712d4050"},"przedmiotAktu":"w sprawie wyrażenia opinii dotyczącej zmian w statucie Uniwersytetu Przyrodniczo-Humanistycznego"},
{"_id":{"$_id":"62bb154ac5b24717712d4052"},"przedmiotAktu":"w sprawie zaopiniowania sprawozdania z realizacji strategii uczelni w roku 2021", "nu
```

Rysunek 5 Zawartość wyniki.json

Dane o plikach zapisywane są w pliku files.json:

```
studia 4 > Praktyki > test > bip-exporting-tool > {!} files.json > ...
[{"_id":{"$_id":"62bb1548c5b24717712d404d"},"sciezka":"URU2022/URU-005-22.pdf","nrZalacznika":{"isPresent":0},"dataDodania":{"$date":"2022-04-01T23:00:00Z"}},
{"_id":{"$_id":"62bb1549c5b24717712d404f"},"sciezka":"URU2022/URU-004-22.pdf","nrZalacznika":{"isPresent":0},"dataDodania":{"$date":"2022-03-24T23:00:00Z"}},
{"_id":{"$_id":"62bb154ac5b24717712d4051"},"sciezka":"URU2022/URU-003-22.pdf","nrZalacznika":{"isPresent":0},"dataDodania":{"$date":"2022-03-24T23:00:00Z"}},
{"_id":{"$_id":"62bb154bc5b24717712d4053"},"sciezka":"URU2022/URU-002-22.pdf","nrZalacznika":{"isPresent":0},"dataDodania":{"$date":"2022-02-10T23:00:00Z"}},
```

Rysunek 6 Zawartość files.json

Pliki zapisywane są w odpowiednich folderach z odpowiednią nazwą:



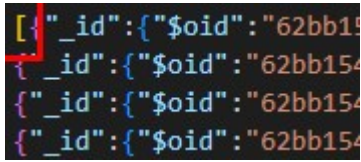
Rysunek 7 Tworzone foldery

- ① URU-011-21.pdf
- ① URU-011-21-dc.pdf
- ① URU-012-21.pdf
- ① URU-012-21-dc.pdf
- ① URU-013-21.pdf
- ① URU-013-21-dc.pdf

Rysunek 8 Zawartość folderu URU2021

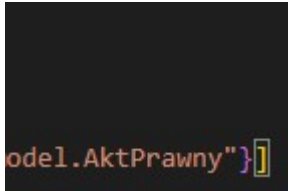
Ważne

Aby poprawnie importować dane do MongoDB należy uprzednio na początku i na końcu plików JSON dodać nawiasy kwadratowe. Jak na poniższych obrazkach



```
[{"_id":{"$_oid":"62bb154"}, {"_id":{"$_oid":"62bb154"}, {"_id":{"$_oid":"62bb154"}, {"_id":{"$_oid":"62bb154"}
```

Rysunek 9 Początek pliku



```
odel.AktPrawny"}]
```

Rysunek 10 Koniec pliku