# Lab 5 – Poszukiwanie bibliotek o określonej funkcjonalności (Web Scraping)

## 👤 Autor

### Tomasz Królikowski, numer albumu: 153790

### Zadanie znajduje się w repozytorium GIT pod adresem:

### https://github.com/krolikowski80/studia\_WSB/tree/main/Python/intro/zad\_5

## 1. Cel zadania

Celem niniejszego laboratorium było zapoznanie się z bibliotekami do automatycznego pobierania danych z internetu (**web scraping**) i ich wykorzystanie w praktycznych przykładach.  
Zastosowano dwie różne biblioteki: - **BeautifulSoup** – do prostego parsowania HTML, - **Scrapy** – do budowy wydajnych crawlerów.

Przedstawiono działanie obu rozwiązań na przykładzie rzeczywistych stron WWW, porównano ich możliwości oraz przygotowano działające skrypty demonstracyjne.

## 

## 🧩 Wykorzystane biblioteki

### 🔹 BeautifulSoup

Biblioteka do prostego parsowania kodu HTML/XML.  
Używana z requests do pobierania i analizowania zawartości stron internetowych.

### 🔹 Scrapy

Pełnoprawny framework do tworzenia crawlerów, obsługujący: - wiele adresów URL, - automatyczne przechodzenie po podstronach, - eksport danych do JSON/CSV/XML.

## 

## 📁 Struktura katalogów

ZAD\_5/  
├── examples/  
│ ├── bs4\_example\_1.py # Nagłówki H2 z Wikipedii  
│ ├── bs4\_example\_2.py # Linki ze strony  
│ ├── scrapy\_example/  
│ │ ├── scrapy\_example/  
│ │ │ ├── spiders/  
│ │ │ │ └── quotes\_spider.py # Spider Scrapy – cytaty  
│ │ │ └── ...  
│ │ └── scrapy.cfg  
│ └── scrapy\_example\_advanced.py # Uruchamianie spidera z poziomu Pythona  
├── raport.md  
└── README.md

## ▶️ Jak uruchomić przykłady

### 🔹 BeautifulSoup

#### 📘 bs4\_example\_1.py

Pobiera nagłówki <h2> z Wikipedii:

cd examples  
python bs4\_example\_1.py

#### 📘 bs4\_example\_2.py

Pobiera pierwsze 10 linków ze strony:

python bs4\_example\_2.py

### 🔹 Scrapy

#### 🕷 quotes\_spider.py

Zbiera cytaty i autorów z wielu stron i zapisuje je do pliku quotes.json:

cd examples/scrapy\_example  
scrapy crawl quotes -O quotes.json

#### 🧪 scrapy\_example\_advanced.py

Uruchamia spidera z poziomu Pythona:

cd examples  
python scrapy\_example\_advanced.py

## 💻 Wymagania

Zainstaluj wymagane biblioteki (najlepiej w środowisku wirtualnym):

pip install beautifulsoup4 lxml requests scrapy

## 🔗 Linki do dokumentacji

* [BeautifulSoup – dokumentacja](https://www.crummy.com/software/BeautifulSoup/bs4/doc/)
* [Scrapy – dokumentacja](https://docs.scrapy.org/en/latest/)