

Book Store Web Scraper- Internship Project Report

Submitted by: Prabjot Singh

Domain: Python Programming

Internship Provider: HexSoftwares

Project Title

Automated Book Store Web Scraper using Python

Objective

To extract book details (Title, Price, Rating & Image URLs) from **BooksToScrape.com** automatically and store the results in **CSV and Excel** formats.

Problem Statement

Manual data collection from websites is slow and inefficient.

This scraper **automates data collection** making it quick & structured for analysis.

Tools & Technologies Used

- Python
 - Requests
 - BeautifulSoup
 - CSV
 - Pandas
 - Excel
-

Methodology / Workflow

- 1 Send request to website
- 2 Parse HTML using BeautifulSoup
- 3 Extract book data fields
- 4 Move to next page (Pagination)
- 5 Store data in CSV & Excel

Code Implementation

```
import requests
from bs4 import BeautifulSoup
import csv
import pandas as pd

url = "https://books.toscrape.com/"
base_url = "https://books.toscrape.com/catalogue/"
all_books = []

while True:
    response = requests.get(url)
    soup = BeautifulSoup(response.text, "html.parser")
    books = soup.find_all("article", class_="product_pod")

    for book in books:
        title = book.h3.a["title"]
        price = book.find("p", class_="price_color").text
        rating = book.find("p", class_="star-rating")["class"][1]
        img = book.find("img")["src"].replace("../", "")
        img_link = "https://books.toscrape.com/" + img

        all_books.append([title, price, rating, img_link])

    nxt = soup.find("li", class_="next")
    if nxt:
        next_page = nxt.find("a")["href"]
        url = base_url + next_page
    else:
        break
```

```
file_csv = "books_data.csv"

with open(file_csv, "w", newline="", encoding="utf-8") as f:
    writer = csv.writer(f)
    writer.writerow(["Title", "Price", "Rating", "Image URL"])
    writer.writerows(all_books)

excel_file = "books_data.xlsx"

df = pd.DataFrame(all_books, columns=["Title", "Price", "Rating", "Image URL"])

df.to_excel(excel_file, index=False)
```

Screenshots

- ✓ Output in VS Code
 - ✓ CSV data
 - ✓ Excel data

The screenshot shows a Jupyter Notebook interface with the following details:

- File Explorer:** Shows a project structure with files: basic_1.py, books_data.csv, books_data.xlsx, practice_1.py, and README.md.
- Code Editor:** Displays Python code for scraping data from a website and saving it to CSV and Excel files.
- Output:** Shows the command run in the terminal and the resulting output, which includes printed messages and file paths.

```
PS C:\Users\prabj\OneDrive\Documents\Project 2> & C:\Users\prabj\AppData\Local\Programs\Python\Python313\python.exe "c:/Users/prabj/OneDrive/Documents/Project 2/basic_1.py"
Books With their prices
Books With their prices
DATA SCRAPING COMPLETED SUCCESSFULLY!
CSV file saved as: books_data.csv
Excel file saved as: books_data.xlsx
PS C:\Users\prabj\OneDrive\Documents\Project 2>
```

Conclusion

The project successfully scraped complete book data from the website and converted it to structured formats.

This demonstrates practical usage of **Python automation & data collection**.

References

- <https://books.toscrape.com>
 - <https://beautiful-soup-4.readthedocs.io/>
 - Python documentation

☒ End of Report

