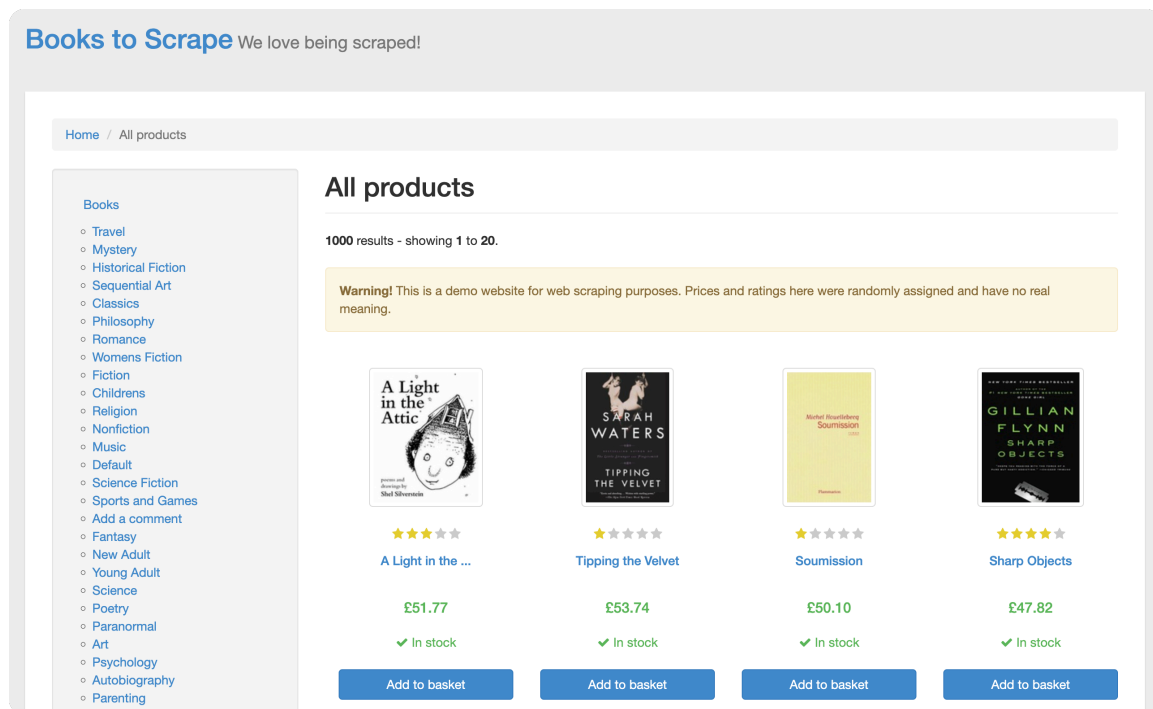


Assignment 1: Web Scraping & Data Analysis

Oct 31, 2021

In this assignment, you should work with data from <http://books.toscrape.com/index.html> (Online Bookstore)



Books to Scrape is a favorite bookstore platform for domestic audiences and provides you with online book purchase services. At the same time, Books to Scrape also provides you with book name, book price, book rankings, categorical information and other information.

Everyone is interested in good books, how can there be so many good books? So, it is very important to scrape high-quality books on the websites. In this project, we will use the requests and regular expressions we have learned before **to scrape all books from the Books to Scrape website.**

Task1. You are required to **scrape 1000 books (all books)** from the website and save result into 'your_name+id.csv'. This file should contain data with the following columns: **(40 marks)**

Title of Book.	5 marks
Number of reviews	5 marks
Rating	5 marks

Description	5 marks
Price	5 marks
Product Type	5 marks
Book category	5 marks
availability	5 marks

You are free to explore data **with more properties** if needed.

Bonus: who is the author for each book? (Hint: Multi-source web scrape via Wikipedia 10 marks)

Task2. You are required to do a data analysis on the data. What do you think is interesting about this data? Tell a story about some interesting thing you have discovered by looking at the data. **(60 marks)**

For example, which one is the book you might read? Does the type of book affect book sales? Which category of books sells the best? What factors affect the inventory of books?

Note: This is an open topic project. You are required to provide a novel topic and demonstrate your hypotheses (view points) with data analysis and figures illustrations.

The reports and running code (web scraping + data analysis) should be submitted using [Jupyter Notebook](#) file.

Submission Checklist:

Yes/No	Items
	Jupyter Notenook code
	your_name+id.csv

Marking Guidelines

Marking Criteria	
Idea (5 marks)	<ul style="list-style-type: none"> • Presents a novel idea •
Knowledge (10 marks)	<ul style="list-style-type: none"> • Clearly demonstrate your viewpoints, with appropriate references.

	<ul style="list-style-type: none"> • Demonstrates good understanding of the topic.
Discussion (30 marks)	<ul style="list-style-type: none"> • Provide convincing arguments to your viewpoints. • Backs up arguments with appropriate data analysis results. • Visualize data analysis results by using more than 5 figures.
Organization (10 marks)	<ul style="list-style-type: none"> • Use of figures to support ideas discussed in the report. • Use of sub-titles and/or clear topic sentences. • A variety of visualization methods
Writing Style (5 marks)	<ul style="list-style-type: none"> • Concise writing style • Strong scientific writing without grammatical errors.