

Program Overview:

The program is designed to perform the following tasks:

Scraping Data: Utilizing the requests and BeautifulSoup4 libraries, it extracts book details from the website books.toscrape.com.

Data Processing: It organizes the extracted data into structured CSV files and downloads corresponding book images into categorized folders.

CSV Generation: Each book category is stored in a separate CSV file within the 'Book Scraping Results' directory.

Image Storage: Images are downloaded and stored in respective category folders under 'Book Scraping Results > book images'.

ETL Pipeline Implementation:

This program serves as the initial phase of the envisioned ETL pipeline:

Extraction: It extracts raw book data from the website.

Transformation: The program processes and structures the data into CSV files, ensuring easy analysis and usage.

Loading: Data is organized and stored locally, ready for further processing or integration into the database.

Best,
Sam