Web Scraping Project: Kitapsepeti Book Data

Development by Alperan Çakar

Project Objective

- This project involves web scraping to collect book information from the "Roman" category on Kitapsepeti.com, including titles, star ratings, and prices. The data is then stored in JSON and CSV formats, with proper categorization and formatting for ease of access and analysis.

Project Requirements

- ❖ **Python**: Programming language used.
- ❖ **Libraries:**
  - ➢ **Requests**: Library for fetching data from web pages.
  - ➢ **BeautifulSoup**: Library for parsing and analyzing HTML data.
  - ➢ **Pandas**: Library for organizing data and saving it in CSV format.
  - ➢ **Numpy**: Library for handling NaN values.
  - ➢ **JSON**: Format used for storing data in JSON files.

Code Explanation

Data Collection and Processing

1. **Data Fetching**: Book information is collected from pages 1 to 5 of Kitapsepeti.com.
2. **Data Parsing**: For each book, title, star rating, and price information are parsed using BeautifulSoup.
   - Title: Extracted from the a tag with class product-title text-center.
   - Star Rating: Calculated from the width of the span tag with class stars-fill.
   - Price: Extracted from the span tag with class product-price.
3. **Data Storage**: Collected data is added to a list (Books) in dictionary format.

JSON Operations

4. **Creating JSON File**: Collected data is saved in JSON format in the file books.json.
5. **Adding Category**: A category field ("Roman") is added to each book object in the JSON file, and the updated data is saved in books_updated.json.

CSV Conversion

6. **Creating CSV File**: Updated JSON data is converted to CSV format using pandas DataFrame and saved in Books.csv.

Outputs

- **books.json**: The initial JSON file containing book data.
- **books_updated.json**: The updated JSON file with added category information.
- **Books.csv**: The CSV file containing book information with category as the header.

Conclusion

- This project serves as a foundational example in web scraping and data analysis. The collected data can be utilized for various analyses, such as examining book prices, calculating average ratings, and more.