Books Online

Requirements for Price Monitoring System

This document contains the requirements and instructions for developing a beta version of a price monitoring system for Books Online.

Notes:

- While working, store your code in a GitHub repository and make frequent commits.
- Remember to commit a requirements.txt file, but don't store your virtual environment in your repository.
- You also shouldn't commit your CSV files.
- Write a README.md file and add it to the repository, giving instructions on how to run your code successfully and output some data.

Phase 1

Pick any single product page (i.e., a single book) on Books to Scrape, and write a Python script that visits this page and extracts the following information:

- product_page_url
- universal_ product_code (upc)
- book_title
- price_including_tax
- price_excluding_tax
- quantity_available
- product_description
- category
- review_rating
- image_url

Write the data to a CSV file using the above fields as column headings.

Phase 2

Now that you have obtained the information for one book, you should get all of the necessary information for one category.

Pick any book category on Books to Scrape, and write a Python script that visits this category page and extracts the product page URL for each book in the category.

Then combine this script with the work you have completed in Phase 1 to extract the product data for each book in your category and write the data to a single CSV file.

Note: some category pages have more than 20 books listed, spread across different pages. This is referred to as pagination. Your application should be able to handle this scenario automatically.

Phase 3

Write a script that visits Books to Scrape, extracts all the book categories available, and then extracts product information for all books across all categories. You should write the data for each book category in a separate CSV file.

Phase 4

Finally, extend your existing work to download and save the image file for each product page you visit.