Subject: Re: Price Extraction Program - Phases 1 to 4

Hi Sam,

I hope this message finds you well. I've successfully completed Phases 1 to 4 of the price extraction program for Books Online. Below is a summary of the progress:

Phase 1:

I've developed a Python script that extracts detailed information from a single product page. The script sends a request to the specified product URL, parses the HTML content, and extracts essential details such as Product Page URL, UPC, Title, Prices (including and excluding tax), Quantity Available, Product Description, Category, Review Rating, and Image URL. The extracted data is neatly organized and saved into a CSV file.

Phase 2:

The script has been extended to scrape all book URLs within a single category from Books to Scrape. It dynamically handles pagination to ensure all books across multiple pages are scraped. The data for each book within the chosen category is then combined and stored in a single CSV file.

Phase 3:

A new script has been created to extract all book categories available on Books to Scrape. The script then extracts product information for all books across these categories, storing the data for each category in separate CSV files.

Phase 4:

Building upon the existing scripts, I've incorporated functionality to download and save the image file from each product page visited. The images are stored in the /book\_images directory.

Project Structure:

I've organized the project structure to include the main script (scrape\_books.py), a requirements file (requirements.txt), a directory for category-specific CSV files (/book\_categories), and a directory for downloaded images (/book\_images).

Prerequisites and Installation:

I've provided clear instructions in the README.md file regarding the prerequisites for running the script and the installation process. Users can refer to this documentation for a seamless setup.

How to Run:

To run the script, users can follow the instructions in the README.md file. After execution, the script will create category-specific CSV files in the /book\_categories directory and download corresponding images to the /book\_images directory.

I've ensured the documentation is comprehensive, providing a step-by-step guide for anyone looking to understand or contribute to the project.

Please find the updated project link here (https://github.com/Yemaiiii01/BooksDataExtraction/commits/master).

Feel free to reach out if you have any questions or if there are additional features you'd like to incorporate into the next phases.

Best regards,

Adeyemi Aiyegbusi

Senior Developer