Documentation: Amazon Best Sellers Scraper Using Selenium

Yash Dhage
yashdhage20@gmail.com

Features

- **Automated Browsing**: Uses Selenium to navigate through Amazon's best-seller pages.
- **Data Extraction**: Collects:
  - Product Name
  - Price
  - Discount Percentage
  - Rating
  - Shipping Information
  - Seller Details
  - Product Description
  - Number of Purchases in the Past Month
  - Available Product Images
- **Delays**: Introduces randomized delays to mimic human behavior and reduce the risk of bot detection.
- **Data Storage**: Saves the scraped data into a structured CSV file.

Setup Instructions

1. **Prerequisites**
   - **Python**: Ensure Python 3.7+ is installed on your system.
   - **Google Chrome and ChromeDriver**:
     - Download and install Google Chrome.
     - Download the appropriate version of ChromeDriver to match your browser version.
   - **Brave Browser (Optional)**:
     - If using Brave, ensure it is installed and provide its binary path in the script.
2. **Install Required Libraries**

   Install Selenium using pip

3. **Set Up the Environment**
   - Place the `chromedriver` binary in an accessible directory.
   - Update the script with the correct `driverPath` and Brave binary path, if applicable.
4. **Enable Headless Browsing**
   - To run without opening a visible browser window, the script uses `--headless` mode. Remove the argument in the `options` configuration if you want to see the browser interactively.

Usage Instructions

Change the path in the code to your chromedriver and if you are not using Brave comment line 21
```
options.binary_location = "/Applications/BraveBrowser.app/Contents/MacOS/Brave Browser"
```

1. **Run the Script**

   Run the script using the python3 command in terminal.

   ```
   python3 main.py
   ```

2. **Input Data**
   - The script is pre-configured to scrape 10 categories from Amazon's Best Sellers section. Modify the `categories` list in the script to include other categories or change URLs as needed.
3. **Scraping Workflow**
   - The script will:
     1. Open each category URL in the browser.
     2. Scroll through the page to load all elements.
     3. Scrape the details of products with discounts greater than 50%.
     4. Navigate through up to 30 pages of each category (can be changed if needed in the loop).
4. **Output**
   - Scraped data is saved in a CSV file named `amazon_best_sellers.csv` in the current directory. The file contains all the extracted fields for each product.