

Documentation: Amazon Best Sellers Web Scraper

Overview

This Python script uses Selenium to scrape product details from Amazon's Best Sellers section. It logs into Amazon using user credentials, navigates through selected categories, and extracts information about products with discounts greater than 50%. The data is saved in a structured CSV format.

Features

1. **Authentication**: Logs into Amazon using user-provided credentials.
2. **Scraping Criteria**:
 - Extracts up to 1500 best-selling products from 10 predefined categories.
 - Focuses on products with discounts greater than 50%.
3. **Extracted Data**:
 - Product Name
 - Product Price
 - Sale Discount
 - Best Seller Rating
 - Ship From
 - Sold By
 - Rating
 - Product Description
 - Number Bought in the Past Month (if available)
 - Category Name
 - All Available Images
4. **Data Storage**: Saves extracted data in CSV or JSON formats.
5. **Error Handling**: Manages exceptions during scraping and ensures the program continues running.

Prerequisites

1. Python 3.x installed on your machine.
2. Google Chrome installed.
3. ChromeDriver corresponding to your Chrome version.
4. Selenium Python library installed.
5. A valid Amazon account with credentials.

Setup Instructions

Step 1: Install Dependencies

Install the required Python libraries:

```
```bash
```

```
pip install selenium
'''
```

### Step 2: Download ChromeDriver

1. Visit the [ChromeDriver download page](https://chromedriver.chromium.org/downloads).
2. Download the driver version matching your Chrome browser.
3. Place the driver executable in a directory accessible via your PATH or specify its location in the script.

### Step 3: Configure Credentials

Update the `username` and `password` variables in the script with your Amazon login credentials:

```
'''python
username = "your_email@example.com"
password = "your_password"
'''
```

## Usage Instructions

### Step 1: Run the Script

Run the Python script:

```
'''bash
python amazon_scraper.py
'''
```

### Step 2: Monitor Execution

1. The script will open a Chrome browser window.
2. It will log in to Amazon and navigate to the Best Sellers section.
3. It will scrape product data from the specified categories and save the results to a CSV file.

### Step 3: Access the Data

The scraped data will be saved in a file named `amazon\_best\_sellers.csv` in the current working directory. You can modify the filename or format (e.g., JSON) in the `save\_data` function.

## Code Structure

### Functions

1. `init_driver()`: Initializes the Selenium WebDriver with Chrome options.
2. `login_amazon(driver, username, password)`: Logs into Amazon using provided credentials.
3. `scrape_category(driver, category_url)`: Scrapes product details from a given category URL.
4. `save_data(data, filename, file_format)`: Saves the scraped data to a file in CSV or JSON

format.

5. `**`main()`**`: Coordinates the execution of the script.

### Disclaimer

This script is for educational purposes only. Ensure compliance with Amazon's terms of service when using it. Excessive scraping may result in account restrictions or bans.