Q4-Amazon Web Scraper Report

#### S Vignesh

#### 17BCE1216

#### 7358638481

#### vighneshss@gmail.com

#### N S R Yokesh

#### 17BCE1043

#### 8056017548

#### yokesh.nsr2010@gmail.com

#### Sainath G

#### 17BCE1107

#### 9884026065

#### sainathganesh1@gmail.com

**Introduction :**

The Amazon Web Scraper is used for thread-safe, asynchronous scraping of the Amazon website for extracting details (product name, price, etc.) about Warren Buffet books using Python.

Python3 dependencies/libraries used:

* LXML
* Beautiful Soup
* Selenium
* Matplotlib
* And other built-in utility functions that comes along with python

**Process:**

1. Using Selenium package, we first create a browser instance.
2. Then using JavaScript, we get the page source.
3. This page source is then parsed partially using Beautiful Soup and the rest is parsed using LXML.
4. After parsing, the data is stored in a JSON file.
5. The data in the JSON file is then arranged in ascending order with respect to the item price.
6. Data is then retrieved from the JSON file and is used to plot the bar chart using Matplotlib package.

**User manual:**

(Optional) Download and place a new chromedriver from in the same dir <https://sites.google.com/a/chromium.org/chromedriver/>

* python3 new\_scrapper.py gen\_dat

The above command is used to start the asynchronous scraper. Abruptly stopping it wont cause data loss.

python3 new\_scraper.py plot\_dat

The above command is to start the plotting of the scraped data stored in the Json file.

**Features:**

Command line utility - quick access and outputs

Headless selenium instance - used for scraping on massive servers

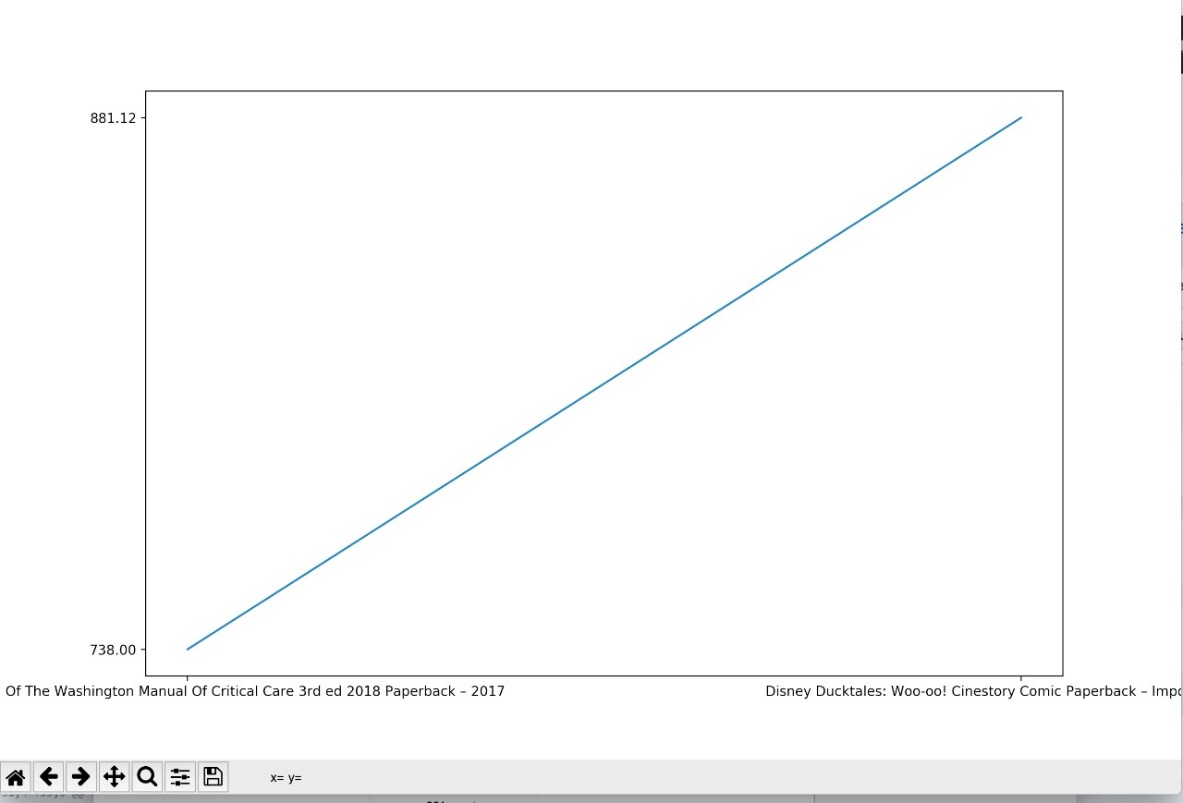
Threaded scraping - faster performance

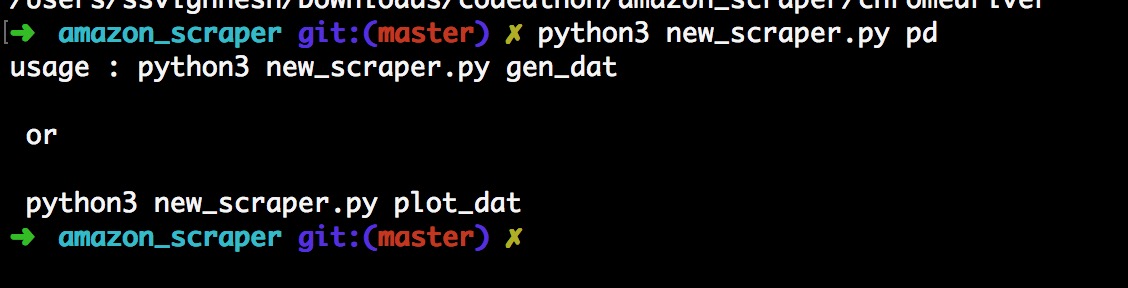
Thread-safe file processing - prevents data corruption

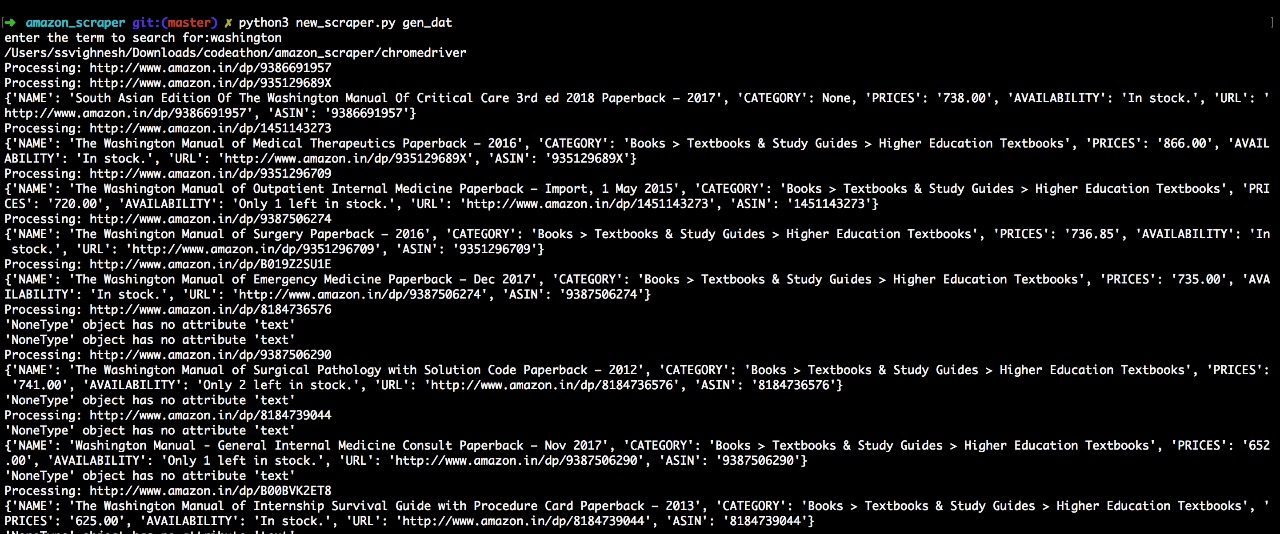
Matplotlib - interactive plotting based on data scraped

Json file - universally usable using slight modifications

**Outputs:**







**Conclusion:**

The Amazon Web scraper can be used to obtain details of various products among various pages under one search term using ASIN of each product.

Thank you for the opportunity! It was fun!