Assignment

Overview

Build a web scraper to automatically search the lowest price of books on e-commerce sites such as flipkart.com, amazon.in and snapdeal.com by using node.js and puppeteer.

Flow:

1. Load the list of ISBN code(Book Id) and book title from excel file.
2. Load the list of ecommerce sites from the ecom.json file.
3. Search the book by its ISBN code on each e-commerce site one by one.
4. If book is found in the search result then match the ISBN code(Exact numeric match) & Book Title(90% Text Match)
5. Scrap the match book data(such as price,Author,Publisher, in stock,page url) and store it in the temporary json.
6. Repeat the steps till you search the ISBN code on the search ecommerce site list.
7. Select the lowest book price site and store the respective scrap value in the excel file.
8. Repeat the step for all the ISBN code in the list.

#! ecom.json Example :

{

“Ecom”:[“flipkart.com”,”amazon.in”,”snapdeal.com”]

}

#2Input Excel Example :

| No | Book Title | ISBN |
| --- | --- | --- |
| 1 | The Subtle Art of Not Giving a F | 9780062641540 |
| 2 | Rich Dad Poor Dad | 9781612680019 |
| 3 | IKIGAI | 9781786330895 |
| 4 | Atomic Habits | 9781847941831 |
| 5 | Zero to One | 9780753555194 |

Output Example

| No | Book Title | ISBN | Site | Found | URL | Price | Author | Publisher | In Stock |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | The Subtle Art of Not Giving a F | 9780062641540 | Snapdeal | Yes | <https://www.snapdeal.com/product/the-subtle-art-of-not/686281189461#bcrumbSearch:9780062641540> | 177 | Mark Manson | Harper Collins | Yes |
| 2 | Rich Dad Poor Dad | 9781612680019 | Amazon | Yes | someurl | 122 | someauthor | somepublisher | Yes |
| 3 | IKIGAI | 9781786330895 | Amazon | Yes | someurl | 122 | someauthor | somepublisher | No |
| 4 | Atomic Habits | 9781847941831 | Flipkart | Yes | someurl | 122 | someauthor | somepublisher | Yes |
| 5 | Zero to One | 9780753555194 | NA | No | NA | NA | NA | NA | NA |

\* This is just an example of output excel to give you an idea about its structure, the actual value within the field will not be the same.

Please Note:

1. One website may give you a multiple search result of the book. You have to select the book

i.which matches the book title text by alteat 90%

1. Ignore the book titles which is inside the (), after : or -

Example

IKIGAI (English, Hardcover, Garcia Hector)

IKIGAI - The Japanese Secret to a Long and Happy Life

IKIGAI : The Japanese Secret to a Long and Happy Life

Ignore the highlighted part and only use IKIGAI while comparing with the input title.

1. Ignore the hyphen in ISBN Code while comparing the input with search result ISBN Code.