LOWEST PRICE RECOMMENDATION FOR ONLINE SHOPPING

For
(Tools and Technique Lab)

Submitted by

Khushi Kumari (1928034)
Patrali Sarkar (1928043)
Swagata Chanda (1928066)
Abhilash Naskar (1928074)
of
School of Computer Engineering

To
B.Tech Program in Computer Science Engineering

Kalinga Institute of Industrial Technology, Deemed to be University
Bhubaneswar, India

April 2022

INTRODUCTION

The easiest way to get the best shopping deals is to compare prices from different e-commerce sites before making a purchase and that is where the lowest price recommendation system is used. It compares the prices of different products from various e-commerce websites and gives the user a sorted price chart and provides them with the best value for each product.

ABOUT THE PROJECT

In this project we have done web scraping using selenium along with inbuilt python libraries such as pandas and webdriver manager.

Web scraping

Web Scraping is a technique to extract a large amount of data from several websites. The term **"scraping"** refers to obtaining the information from another source (webpages) and saving it into a local file.Web Scraping extracts the data from websites in the unstructured format. It helps to collect these unstructured data and convert it in a structured form.

These are the steps to perform web scraping -

Step -1: Find the URL that you want to scrape

A webpage or website contains a large amount of information. That's why scrap only relevant information. We have provided the URL of three e-commerce websites which are Amazon, Flipkart and Snapdeal.

Step - 2: Inspecting the Page

The data is extracted in raw HTML format, which must be carefully parsed and reduce the noise from the raw data.

Step - 3: Write the code

A code that extracts the information, provides relevant information such as price, and runs the code.

Step - 4: Store the data in the dataframe

We have stored the data in the list and converted it into dataframe.

Selenium

Selenium is an open-source automated testing library. It is used to check browser activities. To install this library, type the following command in your terminal.

Pandas

Pandas library is used for **data manipulation and analysis**. It is used to extract the data and store it in the desired format.

Webdriver manager

WebDriverManager resolves the driver binaries for the browsers Chrome, Firefox, Opera, PhantomJS, Microsoft Edge, and Internet Explorer. For that, it provides several drivers managers for these browsers.

Selenium webdriver

The primary use of Selenium WebDriver is implementing automated tests for web applications.

FLOWCHART

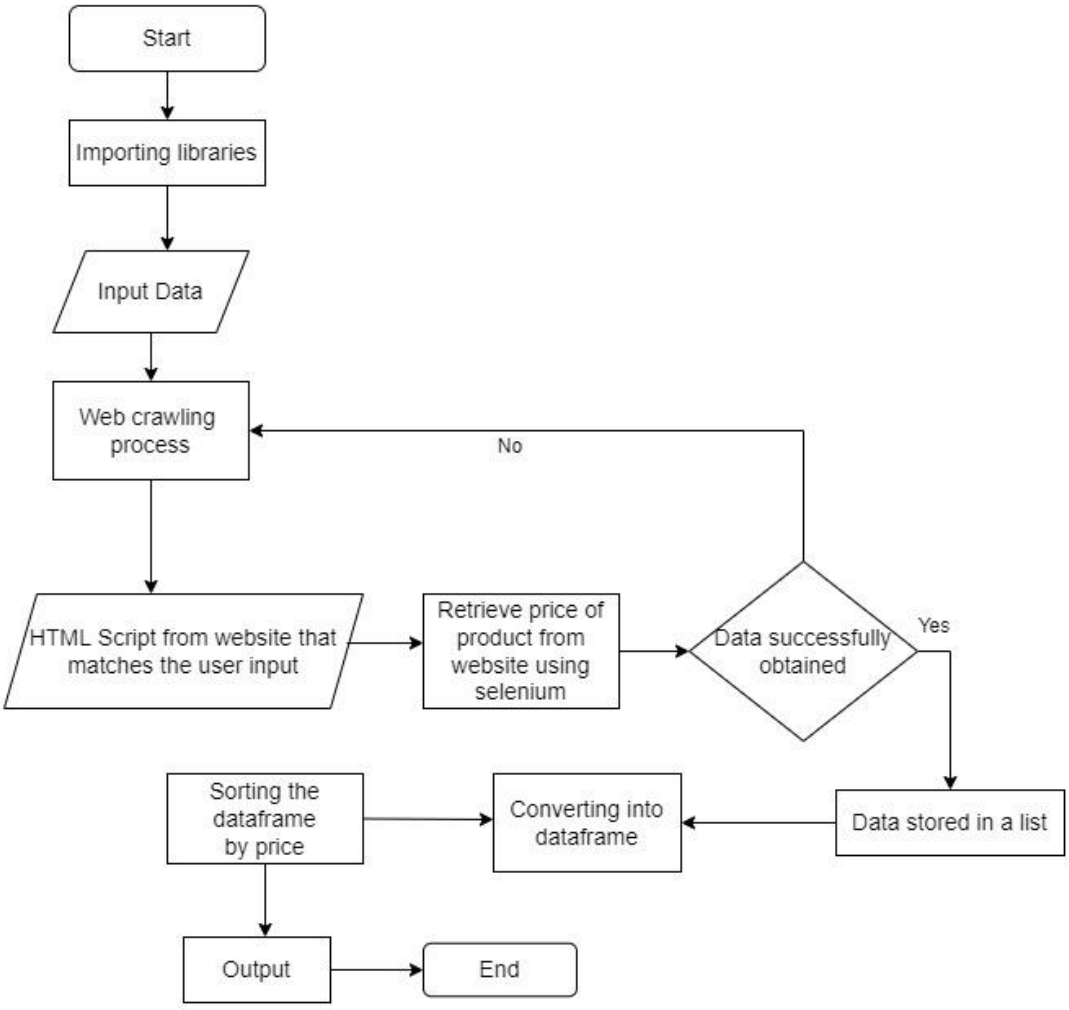

Fig. 1.1

RESULT

Below are the results of our project where we have extracted the prices and sorted them in ascending order.

1.Webscraping from amazon

```
    driver.get(f'https://www.amazon.in/s?k={find_query}')
    all_product = driver.find_elements(By.XPATH, '//span[@class="a-size-medium a-color-base a-text-normal"]')
    all_product_price = driver.find_elements(By.XPATH, '//span[@class="a-price-whole"]')
    for idx in range(0, len(all_product)):
        eachProduct = all_product[idx]
        list_str = []
        for substring in subset:
            list_str.append(f'([^.]*{substring}[^.]*)')
        product_match = re.search("".join(list_str),eachProduct.text , re.IGNORECASE)
        if(product_match):
            print(eachProduct.text, " ", all_product_price[idx].text, "\n")
            temp_dict={"site": "amazon" , "name" : eachProduct.text, "price" : all_product_price[idx].text.replace(",","")}
            product_list.append(temp_dict)
9]                                                                                                                    Python

Dell New Inspiron 5410 2in1 Laptop Intel i3-1125G4, 8GB, 256GB SSD, Win 11 + MSO'21, 14" (35.56 cms) Touch FHD 60Hz Display,
Platinum Silver Color, FPR + Backlit KB (D560725WIN9SE), 1.5Kgs   53,990

Dell New Inspiron 5418 Laptop Intel i5-11320H, 16GB DDR4, 512Gb SSD, Win 11 + MS Office'21, 14" (35.56 cms) FHD 250 Nits
Display, Platinum Silver, FPR + Backlit KB (D560633WIN9S), 1.46Kgs   65,990
```

Fig. 1.2

2.Webscraping from flipkart

```
    driver.get(f'https://www.flipkart.com/search?q={find_query}')
    if(driver.find_elements(By.XPATH, '//div[@class="_4rR01T"]')==[]):
        all_product = driver.find_elements(By.XPATH, '//a[@class="s1Q9rs"]')
    else:
        all_product = driver.find_elements(By.XPATH, '//div[@class="_4rR01T"]')
    if(driver.find_elements(By.XPATH, '//div[@class="_4rR01T"]')==[]):
        all_product_price = driver.find_elements(By.XPATH, '//div[@class="_30jeq3"]')
    else:
        all_product_price = driver.find_elements(By.XPATH, '//div[@class="_30jeq3 _1_WHN1"]')
    for idx in range(0, len(all_product)):
        eachProduct = all_product[idx]
        list_str = []
        for substring in subset:
            list_str.append(f'([^.]*{substring}[^.]*)')
        product_match = re.search("".join(list_str),eachProduct.text , re.IGNORECASE)
        if(product_match):
            print(eachProduct.text, " ", all_product_price[idx].text[1:], "\n")
            temp_dict={ "site" : "flipkart","name" : eachProduct.text, "price" : all_product_price[idx].text[1:].replace(",","")}
            product_list.append(temp_dict)

)]                                                                                                                    Python

DELL Vostro Core i3 11th Gen - (8 GB/256 GB SSD/Windows 10) VOSTRO 3400 Thin and Light Laptop   38,990

DELL Vostro Core i3 10th Gen - (8 GB/256 GB SSD/Windows 10 Home) Vostro 3401 Thin and Light Laptop   37,990
```

Fig. 1.3

3.Webscraping from Snapdeal

```
    driver.get(f'https://www.snapdeal.com/search?keyword={find_query}')
    all_product = driver.find_elements(By.XPATH, '//p[@class="product-title"]')
    all_product_price = driver.find_elements(By.XPATH, '//span[@class="lfloat product-price"]')
    for idx in range(0, len(all_product)):
        eachProduct = all_product[idx]
        list_str = []
        for substring in subset:
            list_str.append(f'([^.]*{substring}[^.]*)')
        product_match = re.search("".join(list_str),eachProduct.text , re.IGNORECASE)
        if(product_match):
            print(eachProduct.text, " ", all_product_price[idx].text, "\n")
            temp_dict={"site": "snapdeal" , "name" : eachProduct.text, "price" : all_product_price[idx].text[4::].replace(",","")}
            product_list.append(temp_dict)
```

Python

```
KALARKARI Laptop Skin DellBrand Premium Matte vinyl HD printed Easy to Install Laptop Skin/Sticker/Decal/Vinyl/Cover for all
size laptops upto 15.6   Rs. 169

Laptop Skin Dell_symbol Premium Matte vinyl HD printed Easy to Install Laptop Skin/Sticker/Decal/Vinyl/Cover for all size
laptops upto 15.6   Rs. 169

Laptop Skin Dell logo_symbol Premium Matte vinyl HD printed Easy to Install Laptop Skin/Sticker/Decal/Vinyl/Cover for all size
```

Fig 1.4

4.Final result

	site	name	price
29	snapdeal	Laptop Skin Dell logo_symbol Premium Matte vin...	169.0
28	snapdeal	Laptop Skin Dell_symbol Premium Matte vinyl HD...	169.0
27	snapdeal	KALARKARI Laptop Skin DellBrand Premium Matte ...	169.0
31	snapdeal	RADANYA Laptop Skins Stickers | Fits Dell, Hp,...	189.0
30	snapdeal	MeSleep Dell Vs Apple Laptop Decal	189.0
15	flipkart	DELL Vostro Core i3 10th Gen - (8 GB/256 GB SS...	37990.0
16	flipkart	DELL Vostro Core i3 10th Gen - (4 GB/256 GB SS...	38190.0
6	amazon	Dell New Inspiron 5410 2in1 Laptop Intel i3-11...	38490.0
3	amazon	Dell New Vostro 3400 Laptop Intel i5-1135G7, 1...	38890.0
19	flipkart	DELL Vostro Core i3 11th Gen - (8 GB/256 GB SS...	38990.0
14	flipkart	DELL Vostro Core i3 11th Gen - (8 GB/256 GB SS...	38990.0
22	flipkart	DELL Ryzen 3 Dual Core 3250U - (8 GB/256 GB SS...	39390.0
21	flipkart	DELL Inspiron Core i3 10th Gen - (4 GB/256 GB ...	39490.0
8	amazon	Dell Vostro 3510 Laptop |15.6-inch FHD | Core ...	41199.0
17	flipkart	DELL Inspiron Core i3 11th Gen - (8 GB/1 TB HD...	42590.0
25	flipkart	DELL Inspiron Core i3 10th Gen - (8 GB/512 GB ...	47990.0
23	flipkart	DELL Inspiron Core i3 11th Gen - (8 GB/256 GB ...	49990.0
11	amazon	Dell New 14 AMD Ryzen 3-3250U 14 inches FHD Di...	50590.0
9	amazon	Dell Inspiron 14 5410 Laptop |14" inch FHD |I...	53990.0
24	flipkart	DELL Vostro Ryzen 5 Hexa Core 5500U - (8 GB/51...	53990.0

Fig. 1.5

DISCUSSION OF RESULT

Price comparison websites extract essential details such as product prices, reviews, features, and descriptions from multiple sites. These details are then compiled on the price comparison website and tailored accordingly for easy access. So, when a buyer searches for a product on the website, the site quickly compares and lists similar products from a number of retailers. This process simplifies the buying decision of the buyer since they can compare factors such as price deals, shipping costs, and other features. Our project will help users to compare prices from three different websites which will help them decide which product to buy in an easier and effective way rather than wasting their effort and time in searching every website separately.

SOURCE CODE

```
from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.by import By
import re
import pandas as pd
driver = webdriver.Chrome(ChromeDriverManager().install())
driver.maximize_window()
find_query = input("Search for any product...")
print(find_query)

product_list = []
subset = find_query.split()

driver.get(f'https://www.amazon.in/s?k={find_query}')
all_product = driver.find_elements(By.XPATH, '//span[@class="a-size-medium
a-color-base a-text-normal"]')
all_product_price = driver.find_elements(By.XPATH, '//span[@class="a-price-whole"]')
for idx in range(0, len(all_product)):
    eachProduct = all_product[idx]
    list_str = []
 for substring in subset:
        list_str.append(f'([^.]*{substring}[^.]*)')
```

```
    product_match = re.search("".join(list_str),eachProduct.text , re.IGNORECASE)
    if(product_match):
        print(eachProduct.text, " ", all_product_price[idx].text, "\n")
        temp_dict={"site": "amazon" , "name" : eachProduct.text, "price" :
all_product_price[idx].text.replace(",","")}
        product_list.append(temp_dict)\

driver.get(f'https://www.flipkart.com/search?q={find_query}')
if(driver.find_elements(By.XPATH, '//div[@class="_4rR01T"]')==[]):
    all_product = driver.find_elements(By.XPATH, '//a[@class="s1Q9rs"]')
else:
    all_product = driver.find_elements(By.XPATH, '//div[@class="_4rR01T"]')
if(driver.find_elements(By.XPATH, '//div[@class="_4rR01T"]')==[]):
    all_product_price = driver.find_elements(By.XPATH, '//div[@class="_30jeq3"]')
else:
    all_product_price = driver.find_elements(By.XPATH, '//div[@class="_30jeq3
_1_WHN1"]')
for idx in range(0, len(all_product)):
    eachProduct = all_product[idx]
    list_str = []
    for substring in subset:
        list_str.append(f'([^.]*{substring}[^.]*)')
    product_match = re.search("".join(list_str),eachProduct.text , re.IGNORECASE)
    if(product_match):
        print(eachProduct.text, " ", all_product_price[idx].text[1:], "\n")
        temp_dict={ "site" : "flipkart","name" : eachProduct.text, "price" :
all_product_price[idx].text[1:].replace(",","")}
        product_list.append(temp_dict)


driver.get(f'https://www.snapdeal.com/search?keyword={find_query}')
all_product = driver.find_elements(By.XPATH, '//p[@class="product-title"]')
all_product_price = driver.find_elements(By.XPATH, '//span[@class="lfloat
product-price"]')
for idx in range(0, len(all_product)):
    eachProduct = all_product[idx]
    list_str = []
    for substring in subset:
        list_str.append(f'([^.]*{substring}[^.]*)')
    product_match = re.search("".join(list_str),eachProduct.text , re.IGNORECASE)
```

```
    if(product_match):
        print(eachProduct.text, " ", all_product_price[idx].text, "\n")
        temp_dict={"site": "snapdeal" , "name" : eachProduct.text, "price" :
all_product_price[idx].text[4::].replace(",","")}
        product_list.append(temp_dict)
print(product_list)

df = pd.DataFrame(product_list)
df.head(20)
df['price'] = pd.to_numeric(df['price'])
df.info()
df.sort_values(by=['price'],inplace=True)
df.head(20)
```

OUTPUT

	site	name	price
29	snapdeal	Laptop Skin Dell logo_symbol Premium Matte vin...	169.0
28	snapdeal	Laptop Skin Dell_symbol Premium Matte vinyl HD...	169.0
27	snapdeal	KALARKARI Laptop Skin DellBrand Premium Matte ...	169.0
31	snapdeal	RADANYA Laptop Skins Stickers | Fits Dell, Hp,...	189.0
30	snapdeal	MeSleep Dell Vs Apple Laptop Decal	189.0
15	flipkart	DELL Vostro Core i3 10th Gen - (8 GB/256 GB SS...	37990.0
16	flipkart	DELL Vostro Core i3 10th Gen - (4 GB/256 GB SS...	38190.0
6	amazon	Dell New Inspiron 5410 2in1 Laptop Intel i3-11...	38490.0
3	amazon	Dell New Vostro 3400 Laptop Intel i5-1135G7, 1...	38890.0
19	flipkart	DELL Vostro Core i3 11th Gen - (8 GB/256 GB SS...	38990.0
14	flipkart	DELL Vostro Core i3 11th Gen - (8 GB/256 GB SS...	38990.0
22	flipkart	DELL Ryzen 3 Dual Core 3250U - (8 GB/256 GB SS...	39390.0
21	flipkart	DELL Inspiron Core i3 10th Gen - (4 GB/256 GB ...	39490.0
8	amazon	Dell Vostro 3510 Laptop |15.6-inch FHD | Core ...	41199.0
17	flipkart	DELL Inspiron Core i3 11th Gen - (8 GB/1 TB HD...	42590.0
25	flipkart	DELL Inspiron Core i3 10th Gen - (8 GB/512 GB ...	47990.0
23	flipkart	DELL Inspiron Core i3 11th Gen - (8 GB/256 GB ...	49990.0
11	amazon	Dell New 14 AMD Ryzen 3-3250U 14 inches FHD Di...	50590.0
9	amazon	Dell Inspiron 14 5410 Laptop |14" inch FHD |I...	53990.0
24	flipkart	DELL Vostro Ryzen 5 Hexa Core 5500U - (8 GB/51...	53990.0

Below is the link to our project:

https://github.com/swagatachanda/lowest-price-recommendation-website.git