

November 3, 2023

Build the web crawler to pull product information and links from an e-commerce website

```
[1]: import requests
from bs4 import BeautifulSoup

# Define the URL of the webpage you want to scrape
url = "https://webscraper.io/test-sites"

# Send a GET request to the URL
response = requests.get(url)

# Check if the request was successful (status code 200)
if response.status_code == 200:
    # Parse the HTML content of the page using BeautifulSoup
    soup = BeautifulSoup(response.text, 'html.parser')

    # Find all anchor (a) tags with class 'site-heading' within a div with
    ↪class 'test-site'
    test_site_links = soup.find_all('div', class_='test-site')

    # Extract and print the links
    for site in test_site_links:
        site_link = site.find('a')['href']
        print("Test Site Link:", site_link)
else:
    print("Failed to retrieve the page. Status code:", response.status_code)
```

```
Test Site Link: /test-sites/e-commerce/allinone
Test Site Link: /test-sites/e-commerce/static
Test Site Link: /test-sites/e-commerce/ajax
Test Site Link: /test-sites/e-commerce/more
Test Site Link: /test-sites/e-commerce/scroll
Test Site Link: /test-sites/tables
```

```
[3]: import requests
from bs4 import BeautifulSoup

# Define the URL of the specific test site (E-commerce site)
```

```

url = "https://webscraper.io/test-sites"

# Send a GET request to the URL
response = requests.get(url)

# Check if the request was successful (status code 200)
if response.status_code == 200:
    # Parse the HTML content of the page using BeautifulSoup
    soup = BeautifulSoup(response.text, 'html.parser')

    # Find all product divs
    product_divs = soup.find_all('div', class_='col-sm-4 col-lg-4 col-md-4')

    # Loop through each product div and extract product information
    for product_div in product_divs:
        # Extract product name
        product_name = product_div.find('a', class_='title').text.strip()

        # Extract product price
        product_price = product_div.find('h4', class_='pull-right').text.strip()

        # Print or save the extracted data
        print("Product Name:", product_name)
        print("Product Price:", product_price)
        print("\n")
    else:
        print("Failed to retrieve the page. Status code:", response.status_code)

```