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

-Tushar Bhagat Roll_No. 07

In [13]:
```
pip install requests beautifulsoup4
```

Requirement already satisfied: requests in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (2.31.0)
Requirement already satisfied: beautifulsoup4 in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (4.12.3)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (from requests) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (from requests) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (from requests) (2.2.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (from requests) (2024.2.2)
Requirement already satisfied: soupsieve>1.2 in c:\users\tmbha\appdata\local\programs\python\python310\lib\site-packages (from beautifulsoup4) (2.5)
Note: you may need to restart the kernel to use updated packages.

In [24]:
```python
import requests
from bs4 import BeautifulSoup

# Function to fetch product information from the webpage
def fetch_books(url):
    try:
        # Send a GET request to the URL
        response = requests.get(url)
        response.raise_for_status()  # Raise an error for bad responses

        # Parse the webpage content
        soup = BeautifulSoup(response.text, 'html.parser')

        # Find product listings
        products = soup.find_all('article', class_='product_pod')

        book_info = []

        # Loop through the products and extract information
        for product in products:
            title = product.h3.a['title']  # Book title
            price = product.find('p', class_='price_color').text  # Book price
            link = 'http://books.toscrape.com/catalogue/' + product.h3.a['href'].replace('../../../', '')  # Full L

            book_info.append({
                'title': title,
                'price': price,
                'link': link
            })

        return book_info

    except requests.exceptions.RequestException as e:
        print(f"An error occurred: {e}")
        return []

# Example usage
url = 'http://books.toscrape.com/catalogue/page-1.html'
books = fetch_books(url)

# Display the fetched book information
if books:
    for book in books:
        print(f"Title: {book['title']}, Price: {book['price']}, Link: {book['link']}")
else:
    print("No books found.")
```

Title: A Light in the Attic, Price: £51.77, Link: http://books.toscrape.com/catalogue/a-light-in-the-attic_1000/index.html

Title: Tipping the Velvet, Price: £53.74, Link: http://books.toscrape.com/catalogue/tipping-the-velvet_999/index.html

Title: Soumission, Price: £50.10, Link: http://books.toscrape.com/catalogue/soumission_998/index.html

Title: Sharp Objects, Price: £47.82, Link: http://books.toscrape.com/catalogue/sharp-objects_997/index.html

Title: Sapiens: A Brief History of Humankind, Price: £54.23, Link: http://books.toscrape.com/catalogue/sapiens-a-brief-history-of-humankind_996/index.html

Title: The Requiem Red, Price: £22.65, Link: http://books.toscrape.com/catalogue/the-requiem-red_995/index.html

Title: The Dirty Little Secrets of Getting Your Dream Job, Price: £33.34, Link: http://books.toscrape.com/catalogue/the-dirty-little-secrets-of-getting-your-dream-job_994/index.html

Title: The Coming Woman: A Novel Based on the Life of the Infamous Feminist, Victoria Woodhull, Price: £17.93, Link: http://books.toscrape.com/catalogue/the-coming-woman-a-novel-based-on-the-life-of-the-infamous-feminist-victoria-woodhull_993/index.html

Title: The Boys in the Boat: Nine Americans and Their Epic Quest for Gold at the 1936 Berlin Olympics, Price: £22.60, Link: http://books.toscrape.com/catalogue/the-boys-in-the-boat-nine-americans-and-their-epic-quest-for-gold-at-the-1936-berlin-olympics_992/index.html

Title: The Black Maria, Price: £52.15, Link: http://books.toscrape.com/catalogue/the-black-maria_991/index.html

Title: Starving Hearts (Triangular Trade Trilogy, #1), Price: £13.99, Link: http://books.toscrape.com/catalogue/starving-hearts-triangular-trade-trilogy-1_990/index.html

Title: Shakespeare's Sonnets, Price: £20.66, Link: http://books.toscrape.com/catalogue/shakespeares-sonnets_989/index.html

Title: Set Me Free, Price: £17.46, Link: http://books.toscrape.com/catalogue/set-me-free_988/index.html

Title: Scott Pilgrim's Precious Little Life (Scott Pilgrim #1), Price: £52.29, Link: http://books.toscrape.com/catalogue/scott-pilgrims-precious-little-life-scott-pilgrim-1_987/index.html

Title: Rip it Up and Start Again, Price: £35.02, Link: http://books.toscrape.com/catalogue/rip-it-up-and-start-again_986/index.html

Title: Our Band Could Be Your Life: Scenes from the American Indie Underground, 1981-1991, Price: £57.25, Link: http://books.toscrape.com/catalogue/our-band-could-be-your-life-scenes-from-the-american-indie-underground-1981-1991_985/index.html

Title: Olio, Price: £23.88, Link: http://books.toscrape.com/catalogue/olio_984/index.html

Title: Mesaerion: The Best Science Fiction Stories 1800-1849, Price: £37.59, Link: http://books.toscrape.com/catalogue/mesaerion-the-best-science-fiction-stories-1800-1849_983/index.html

Title: Libertarianism for Beginners, Price: £51.33, Link: http://books.toscrape.com/catalogue/libertarianism-for-beginners_982/index.html

Title: It's Only the Himalayas, Price: £45.17, Link: http://books.toscrape.com/catalogue/its-only-the-himalayas_981/index.html