# CodeAlpha Internship – Task 1: Web Scraping

This project involves extracting structured data from a website using Python.
In this task, we use requests and BeautifulSoup libraries to scrape data from a live webpage and display it in a tabular format using pandas.

- 📌 Internship Domain: Data Analytics
- Tools: Python, Requests, BeautifulSoup, Pandas
- 👤🖥 Performed by: Miju Akshaya P

In [18]:
```python
# Install required libraries (run this only once)
!pip install requests
!pip install beautifulsoup4
!pip install pandas
```

```
Requirement already satisfied: requests in c:\anaconda3_new\lib\site-packages (2.2
2.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\anaconda3_new\lib\site-pack
ages (from requests) (2019.11.28)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\anaconda3_new\lib\site-p
ackages (from requests) (3.0.4)
Requirement already satisfied: idna<2.9,>=2.5 in c:\anaconda3_new\lib\site-packages
(from requests) (2.8)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\anacon
da3_new\lib\site-packages (from requests) (1.25.8)
Requirement already satisfied: beautifulsoup4 in c:\anaconda3_new\lib\site-packages
(4.8.2)
Requirement already satisfied: soupsieve>=1.2 in c:\anaconda3_new\lib\site-packages
(from beautifulsoup4) (1.9.5)
Requirement already satisfied: pandas in c:\anaconda3_new\lib\site-packages (1.0.1)
Requirement already satisfied: numpy>=1.13.3 in c:\anaconda3_new\lib\site-packages
(from pandas) (1.18.1)
Requirement already satisfied: python-dateutil>=2.6.1 in c:\anaconda3_new\lib\site-
packages (from pandas) (2.8.1)
Requirement already satisfied: pytz>=2017.2 in c:\anaconda3_new\lib\site-packages
(from pandas) (2019.3)
Requirement already satisfied: six>=1.5 in c:\anaconda3_new\lib\site-packages (from
python-dateutil>=2.6.1->pandas) (1.14.0)
```

```
In [8]: import requests
        from bs4 import BeautifulSoup
        import pandas as pd
        url = 'http://books.toscrape.com/'
        response = requests.get(url)
        soup = BeautifulSoup(response.text, 'html.parser')

        # Extract book titles and prices
        titles = [book.h3.a['title'] for book in soup.find_all('article', class_='product_po
        d')]
        prices = [book.find('p', class_='price_color').text for book in soup.find_all('artic
        le', class_='product_pod')]

        # Create DataFrame
        df = pd.DataFrame({
            'Book Title': titles,
            'Price': prices
        })

        df.head()  # Display the first 5 results
```

Out[8]:

| | Book Title | Price |
|---|---|---|
| 0 | A Light in the Attic | Â£51.77 |
| 1 | Tipping the Velvet | Â£53.74 |
| 2 | Soumission | Â£50.10 |
| 3 | Sharp Objects | Â£47.82 |
| 4 | Sapiens: A Brief History of Humankind | Â£54.23 |

```
In [21]: df.to_csv("Scrappedbook.csv",index=False)
```

# ✅ Conclusion

This web scraping project successfully extracted book titles and their prices from a practice website using BeautifulSoup and Requests.

The data was stored in a Pandas DataFrame and saved to a CSV file for further analysis or reporting.

Thank you!