Scraping through Selenium

September 13, 2023

0.1 Importing Libraries

```python
[1]: from selenium import webdriver
     from selenium.webdriver.common.by import By
     import pandas as pd
     import warnings
     warnings.filterwarnings('ignore')
```

```python
[2]: # Set the path to the Chrome driver executable
     path = r'C:\Users\Dipes\Downloads\Resit Smart Data Discovery\FYP␣
       ↪resit\chromedriver-win64\chromedriver-win64\chromedriver.exe'

     # Create a new instance of the Chrome driver
     driver = webdriver.Chrome(path)
```

```python
[3]: # Navigate to the webpage
     web = 'https://www.adamchoi.co.uk/overs/detailed'
     driver.get(web)
```

```python
[4]: # Find an element by XPath
     element = driver.find_element(By.XPATH, '//label[@analytics-event="All␣
       ↪matches"]')
```

```python
[5]: # Perform actions on the element (e.g., click)
     element.click()
```

```python
[6]: # Find all elements with tag name 'tr'
     matches = driver.find_elements(By.TAG_NAME, 'tr')
```

```python
[7]: # Iterate through the 'tr' elements and print their text
     Date=[]
     Hteam=[]
     score=[]
     Ateam=[]
     for match in matches:
         Date.append(match.find_element(By.XPATH,'./td[1]').text)
         Hteam.append(match.find_element(By.XPATH,'./td[2]').text)
         score.append(match.find_element(By.XPATH,'./td[3]').text)
```

```python
        Ateam.append(match.find_element(By.XPATH,'./td[4]').text)
```

[8]:
```python
T=pd.DataFrame({'Date':Date,'Home team':Hteam,'Score':score,'Away Team':Ateam})
```

[9]:
```python
T
```

[9]:
```
          Date       Home team  Score       Away Team
    0   12-08-2023        Arsenal  2 - 1  Nott'm Forest
    1   21-08-2023  Crystal Palace  0 - 1        Arsenal
    2   26-08-2023        Arsenal  2 - 2         Fulham
    3   03-09-2023        Arsenal  3 - 1     Man United
    4   17-09-2023        Everton  ? - ?        Arsenal
    ..         …              …    …              …
    93  14-08-2023     Man United  1 - 0         Wolves
    94  19-08-2023         Wolves  1 - 4       Brighton
    95  26-08-2023        Everton  0 - 1         Wolves
    96  03-09-2023  Crystal Palace  3 - 2         Wolves
    97  16-09-2023         Wolves  ? - ?      Liverpool

    [98 rows x 4 columns]
```

[10]:
```python
# Close the browser when done
driver.quit()
```

[ ]:

Why Selenium?

Selenium offers an array of benefits:

🚀 Versatility: Suitable for a wide range of websites and applications.

🧩 Customization: Tailor your scraping scripts to your specific needs.

🌐 Real-time Interaction: Mimic human browsing for complex tasks.

🔐 Security: Handle authentication and session management securely.

🥷 How Can It Benefit You?

Whether you're a data scientist, market researcher, or business analyst, Selenium can empower you to:

📊 Gather competitor intelligence.

🗹 Monitor prices and product listings.

🔍 Extract job listings and company data.

🗞️ Scrape news articles and social media content.

...and so much more!

🚀 Let's Connect!

I'm passionate about sharing my experiences with Selenium and web scraping, and I'd love to connect with fellow enthusiasts, data professionals, and curious minds. Let's collaborate, exchange insights, and drive innovation together! 🤝

Feel free to reach out to me if you have questions, want to discuss a project, or simply want to learn more about the exciting world of web scraping with Selenium. 💬