

Forward School

Program Code: J620-002-4:2020

Program Name: FRONT-END SOFTWARE DEVELOPMENT

Title : Case Study - Webscrapping on Trip Advisor

Name: Ooi Caaron 

IC Number: 990701-07-5837

Date : 4/7/23

Introduction : Exercise using BeautifulSoup and Selenium is a practical exercise that involves utilizing two powerful tools, BeautifulSoup and Selenium, for web scraping and web automation tasks.

Conclusion : Still need to practice more and do revision

Instruction: Use Selenium and Beautiful Soup to extract the first 5 comments and their titles from the trip advisor link.

Url: https://www.tripadvisor.com/Airline_Review-d8729220-Reviews-AirAsia-AirAsia-Berhad-Malaysia
(https://www.tripadvisor.com/Airline_Review-d8729220-Reviews-AirAsia-AirAsia-Berhad-Malaysia)

In [1]:

```
from selenium import webdriver
from bs4 import BeautifulSoup
import time

driver = webdriver.Chrome('C:\GoogleDriver\chromedriver')
driver.get("https://www.tripadvisor.com/Airline_Review-d8729220-Reviews-AirAsia-AirAsia-
```

In [2]:

```

# driver.find_element_by_xpath('/html/body/div[2]/div/div/div[5]/div/div/div[1]/div[2]/d
# soup = BeautifulSoup(driver.page_source, 'html.parser')
# title = soup.findAll("a", attrs={"class": "Qwuub"})
# comment = soup.findAll("div", attrs={"class": "fIrGe _T"})
# elements = soup.find_all("a", class_="Qwuub")

# Iterate over the first five elements and print their text
# for i in range(5):
#     if i < len(title):
#         print(f'Title: {title[i].text}')
#         print(f'Comment: {comment[i].text}')
#         print('')
#     else:
#         break
all_comments = []
for page in range(1, 6):
    soup = BeautifulSoup(driver.page_source, 'html.parser')
    titles = soup.findAll("a", attrs={"class": "Qwuub"})
    comments = soup.findAll("div", attrs={"class": "fIrGe _T"})

    for i in range(5):
        title = titles[i].text
        comment = comments[i+1].text
        all_comments.append(f'{title}:\n{comment}')
    if page == 1:
        next_button = driver.find_element_by_xpath('//*[@id="component_1"]/div/div[5]/di
        next_button.click()
    else:
        next_button = driver.find_element_by_xpath('//*[@id="component_1"]/div/div[5]/di
        next_button.click()
    time.sleep(2)

for i, comment in enumerate(all_comments):
    print(f"Comment {i+1}: {comment}")
    print('')

```

[illegible]

```
nk"><span><span>Not good</span></span></a>, <a class="Qwuub" href="/ShowUserReviews-g1-d8729220-r898536056-AirAsia_AirAsia_Berhad_Malaysia-World.html" rel="noopener" target="_blank"><span><span>Don't use this airline</span></span></a>, <a class="Qwuub" href="/ShowUserReviews-g1-d8729220-r896952984-AirAsia_AirAsia_Berhad_Malaysia-World.html" rel="noopener" target="_blank"><span><span>The staffs are very very rude</span></span></a>, <a class="Qwuub" href="/ShowUserReviews-g1-d8729220-r896094812-AirAsia_AirAsia_Berhad_Malaysia-World.html" rel="noopener" target="_blank"><span><span>No in-flight entertainment</span></span></a>]
```

Reference: <https://medium.com/ymedialabs-innovation/web-scraping-using-beautiful-soup-and-selenium-for-dynamic-page-2f8ad15efe25> (<https://medium.com/ymedialabs-innovation/web-scraping-using-beautiful-soup-and-selenium-for-dynamic-page-2f8ad15efe25>)