

# WEB SCRAPING-ASSIGNMENT3

August 10, 2023

```
[1]: pip install selenium
```

```
Requirement already satisfied: selenium in c:\users\admin\anaconda3\lib\site-  
packages (4.10.0)  
Requirement already satisfied: certifi>=2021.10.8 in  
c:\users\admin\anaconda3\lib\site-packages (from selenium) (2022.9.14)  
Requirement already satisfied: trio~=0.17 in c:\users\admin\anaconda3\lib\site-  
packages (from selenium) (0.22.2)  
Requirement already satisfied: trio-websocket~=0.9 in  
c:\users\admin\anaconda3\lib\site-packages (from selenium) (0.10.3)  
Requirement already satisfied: urllib3[socks]<3,>=1.26 in  
c:\users\admin\anaconda3\lib\site-packages (from selenium) (1.26.11)  
Requirement already satisfied: cffi>=1.14 in c:\users\admin\anaconda3\lib\site-  
packages (from trio~=0.17->selenium) (1.15.1)  
Requirement already satisfied: sortedcontainers in  
c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->selenium) (2.4.0)  
Requirement already satisfied: outcome in c:\users\admin\anaconda3\lib\site-  
packages (from trio~=0.17->selenium) (1.2.0)  
Requirement already satisfied: exceptiongroup>=1.0.0rc9 in  
c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->selenium) (1.1.2)  
Requirement already satisfied: sniffio in c:\users\admin\anaconda3\lib\site-  
packages (from trio~=0.17->selenium) (1.2.0)  
Requirement already satisfied: attrs>=20.1.0 in  
c:\users\admin\anaconda3\lib\site-packages (from trio~=0.17->selenium) (21.4.0)  
Requirement already satisfied: idna in c:\users\admin\anaconda3\lib\site-  
packages (from trio~=0.17->selenium) (3.3)  
Requirement already satisfied: wsproto>=0.14 in  
c:\users\admin\anaconda3\lib\site-packages (from trio-websocket~=0.9->selenium)  
(1.2.0)  
Requirement already satisfied: PySocks!=1.5.7,<2.0,>=1.5.6 in  
c:\users\admin\anaconda3\lib\site-packages (from  
urllib3[socks]<3,>=1.26->selenium) (1.7.1)  
Requirement already satisfied: pycparser in c:\users\admin\anaconda3\lib\site-  
packages (from cffi>=1.14->trio~=0.17->selenium) (2.21)  
Requirement already satisfied: h11<1,>=0.9.0 in  
c:\users\admin\anaconda3\lib\site-packages (from wsproto>=0.14->trio-  
websocket~=0.9->selenium) (0.14.0)  
Note: you may need to restart the kernel to use updated packages.
```

```
[17]: from selenium import webdriver
import time
from selenium.common.exceptions import NoSuchElementException
import warnings
warnings.filterwarnings('ignore')
```

```
[20]: driver = webdriver.chrome("chromedriver.exe")
driver.get('https://www.amazon.in')
time.sleep(5)
```

```
-----
TypeError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_1540\4163656504.py in <module>
----> 1 driver = webdriver.chrome("chromedriver.exe")
      2 driver.get('https://www.amazon.in')
      3 time.sleep(5)

TypeError: 'module' object is not callable
```

```
[21]: import pandas as pd
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.chrome.service import Service as ChromeService
from bs4 import BeautifulSoup
import time
```

```
[25]: def scrape_product_details(search_query, chrome_driver_path):
    service = ChromeService(executable_path=chrome_driver_path)
    driver = webdriver.Chrome(service=service)s

    try:
        data = {
            "Brand Name": [],
            "Name of the Product": [],
            "Price": [],S
            "Return/Exchange": [],
            "Expected Delivery": [],
            "Availability": [],
            "Product URL": [],
        }

        for page in range(1, 4):
            url = f'https://www.amazon.in/s?k={search_query}&page = {page}'
            driver.get(url)
            time.sleep(2)
```

```
File "C:\Users\ADMIN\AppData\Local\Temp\ipykernel_1540\622246034.py", line 19
    time.sleep(2)
```

**SyntaxError:** unexpected EOF while parsing

```
[3]: from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import time
import os
import requests
from bs4 import BeautifulSoup
import urllib.parse
```

```
[10]: def download_images(url, folder_path):
    if not os.path.exists(folder_path):
        os.makedirs(folder_path)

    response = requests.get(url)
    soup = BeautifulSoup(response.content, 'html.parser')
    image_tags = soup.find_all('img', class_='rg_i')

    for i, img in enumerate(image_tags[:10]):
        img_url = img.get('data-src')
        if img_url:
            image_name = os.path.join(folder_path, f"{i + 1}.jpg")
            with open(image_name, 'wb') as f:
                f.write(requests.get(img_url).content)
```

```
[12]: def search_and_scrape(keyword, driver):
    search_bar = driver.find_element_by_name('q')
    search_bar.clear()
    search_bar.send_keys(keyword)
    search_bar.send_keys(Keys.RETURN)
    time.sleep(2)

    for _ in range(3):
        driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
        time.sleep(2)

    soup = BeautifulSoup(driver.page_source, 'html.parser')
    images = soup.find_all('img', class_='rg_i')
    urls = [img.get('data-src') for img in images]

    return urls
```

```
[20]: def main():
    keywords = ['fruits', 'cars', 'Machine Learning', 'Guitar', 'Cakes']
    download_path = os.path.join(os.getcwd(), 'downloads')
    if not os.path.exists(download_path):
        os.makedirs(download_path)

    driver = webdriver.Chrome()

    try:
        for keyword in keywords:
            encoded_keyword = urllib.parse.quote(keyword)
            url = (f"https://www.google.com/search?q={query}")
            driver.get(url)
```

```
File "C:\Users\ADMIN\AppData\Local\Temp\ipykernel_9592\4061739903.py", line 1
```

```
~
```

```
SyntaxError: unexpected EOF while parsing
```

```
[4]: from selenium import webdriver
    from selenium.webdriver.common.keys import Keys
    import time
    import os
    import requests
    from bs4 import BeautifulSoup
    import urllib.parse
```

```
[16]: def scrape_smartphones_details(driver):
    smartphones_data = []

    search_bar = driver.find_element_by_name("q")
    search_term = input("Enter smartphone name: ")
    search_bar.send_keys(search_term)
    search_bar.send_keys(Keys.RETURN)
    time.sleep(2)

    products = driver.find_elements_by_css_selector("div._1AtVbE")

    for product in products:
        details = {}

        for page in range(1, num_pages + 1):
```

```
url = f"https://www.flipkart.com/search?q={urllib.parse.
↳quote(product)}&page={page}"
driver.get(url)
time.sleep(2)

smartphones = driver.find_elements_by_css_selector('div._1AtVbE')
    for smartphone in smartphones:
```

```
File "C:\Users\ADMIN\AppData\Local\Temp\ipykernel_9264\1542251367.py", line 2
    for smartphone in smartphones:
    ^
IndentationError: unexpected indent
```

[ ]: