Web scrapping using python

File   Edit   View   Run   Kernel   Settings   Help

Trusted

Code

JupyterLab   Python 3 (ipykernel)

```
[9]:  !pip install selenium pandas webdriver-manager
```

```
Requirement already satisfied: selenium in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (4.30.0)
Requirement already satisfied: pandas in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (2.2.3)
Requirement already satisfied: webdriver-manager in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (4.0.2)
Requirement already satisfied: urllib3[socks]<3,>=1.26 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from seleniu
m) (2.2.3)
Requirement already satisfied: trio~=0.17 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from selenium) (0.29.0)
Requirement already satisfied: trio-websocket~=0.9 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from selenium)
(0.12.2)
Requirement already satisfied: certifi>=2021.10.8 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from selenium) (2
024.8.30)
Requirement already satisfied: typing_extensions~=4.9 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from seleniu
m) (4.12.2)
Requirement already satisfied: websocket-client~=1.8 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from selenium)
(1.8.0)
Requirement already satisfied: numpy>=1.26.0 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2.1.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from pandas)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2024.2)
Requirement already satisfied: requests in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from webdriver-manager) (2.
32.3)
Requirement already satisfied: python-dotenv in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from webdriver-manage
r) (1.0.1)
Requirement already satisfied: packaging in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from webdriver-manager) (2
4.1)
Requirement already satisfied: six>=1.5 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.8.2
->pandas) (1.16.0)
Requirement already satisfied: attrs>=23.2.0 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from trio~=0.17->selen
ium) (24.2.0)
Requirement already satisfied: sortedcontainers in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from trio~=0.17->se
lenium) (2.4.0)
Requirement already satisfied: idna in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from trio~=0.17->selenium) (3.1
0)
Requirement already satisfied: outcome in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from trio~=0.17->selenium)
(1.3.0.post0)
Requirement already satisfied: sniffio>=1.3.0 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from trio~=0.17->sele
nium) (1.3.1)
Requirement already satisfied: cffi>=1.14 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from trio~=0.17->seleniu
m) (1.17.1)
Requirement already satisfied: wsproto>=0.14 in c:\users\ashish sharma\appdata\local\programs\python\python312\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\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from url
lib3[socks]<3,>=1.26->selenium) (1.7.1)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from reques
ts->webdriver-manager) (3.4.0)
Requirement already satisfied: pycparser in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from cffi>=1.14->trio~=0.1
7->selenium) (2.22)
Requirement already satisfied: h11<1,>=0.9.0 in c:\users\ashish sharma\appdata\local\programs\python\python312\lib\site-packages (from wsproto>=0.14->tr
io-websocket~=0.9->selenium) (0.14.0)

[notice] A new release of pip is available: 23.2.1 -> 25.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```
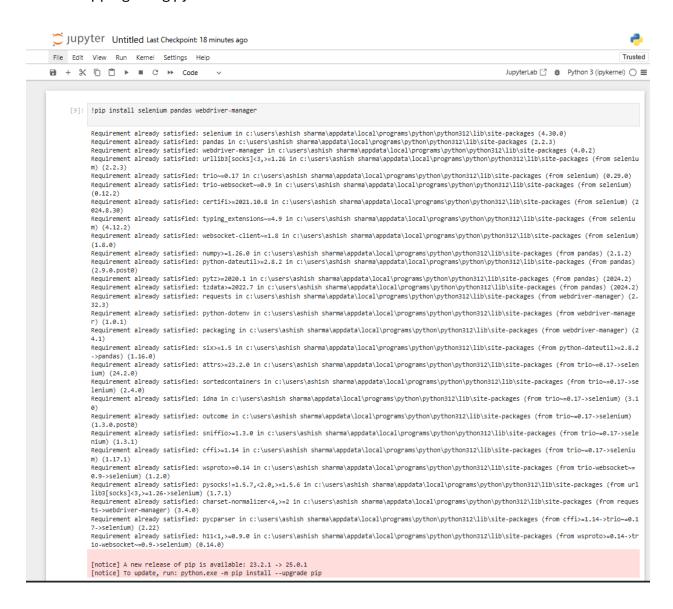
```python
[10]: import time
      import pandas as pd
      from selenium import webdriver
      from selenium.webdriver.common.by import By
      from selenium.webdriver.chrome.service import Service
      from selenium.webdriver.chrome.options import Options
      from webdriver_manager.chrome import ChromeDriverManager
      from selenium.webdriver.common.keys import Keys
```

```python
[11]: options = Options()
      options.add_argument("--start-maximized")  # Open Chrome in full-screen
      driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()), options=options)
```

```python
[12]: job_title = "Data Analyst"
      location = "India"

      url = f"https://www.linkedin.com/jobs/search/?keywords={job_title}&location={location}"
      driver.get(url)
      time.sleep(5)  # Allow the page to load
```

```python
[13]: scroll_pause_time = 3
      total_scrolls = 50  # Adjust this to get more results

      for _ in range(total_scrolls):
          driver.find_element(By.TAG_NAME, "body").send_keys(Keys.END)  # Scroll down
          time.sleep(scroll_pause_time)

          # Click 'See More Jobs' button if available
          try:
              see_more_button = driver.find_element(By.CLASS_NAME, "infinite-scroller__show-more-button")
              driver.execute_script("arguments[0].click();", see_more_button)
              time.sleep(3)
          except:
              pass  # If button isn't found, continue scrolling
```

```python
[14]: job_list = []
      jobs = driver.find_elements(By.CLASS_NAME, "base-card")

      for job in jobs:
          try:
              title = job.find_element(By.CLASS_NAME, "base-search-card__title").text
              company = job.find_element(By.CLASS_NAME, "hidden-nested-link").text
              location = job.find_element(By.CLASS_NAME, "job-search-card__location").text
              job_url = job.find_element(By.TAG_NAME, "a").get_attribute("href")

              job_list.append({"Title": title, "Company": company, "Location": location, "Job URL": job_url})

          except Exception as e:
              print(f"Error: {e}")

      driver.quit()  # Close the browser
```

Error: Message: no such element: Unable to locate element: {"method":"css selector","selector":".hidden-nested-link"}
  (Session info: chrome=134.0.6998.166); For documentation on this error, please visit: https://www.selenium.dev/documentation/webdriver/troubleshootin
g/errors#no-such-element-exception
Stacktrace:
        GetHandleVerifier [0x00ADC7F3+24435]
        (No symbol) [0x00A62074]
        (No symbol) [0x009306E3]
        (No symbol) [0x00978B39]
        (No symbol) [0x00978E8B]
        (No symbol) [0x0096E1F1]
        (No symbol) [0x0099D804]
        (No symbol) [0x0096E114]
        (No symbol) [0x0099DA34]
        (No symbol) [0x009BF20A]
        (No symbol) [0x0099D5B6]
        (No symbol) [0x0096C54F]
        (No symbol) [0x0096D894]
        GetHandleVerifier [0x00DE70A3+3213347]
        GetHandleVerifier [0x00DFB0C9+3295305]
        GetHandleVerifier [0x00DF558C+3271948]
        GetHandleVerifier [0x00B77360+658144]
        (No symbol) [0x00A6B27D]
        (No symbol) [0x00A68208]
        (No symbol) [0x00A683A9]
        (No symbol) [0x00A5AAC0]
        BaseThreadInitThunk [0x76D95D49+25]
        RtlInitializeExceptionChain [0x7740CE3B+107]
        RtlGetAppContainerNamedObjectPath [0x7740CDC1+561]

Error: Message: no such element: Unable to locate element: {"method":"css selector","selector":".hidden-nested-link"}
  (Session info: chrome=134.0.6998.166); For documentation on this error, please visit: https://www.selenium.dev/documentation/webdriver/troubleshootin
g/errors#no-such-element-exception
Stacktrace:
        GetHandleVerifier [0x00ADC7F3+24435]

```python
[15]: df = pd.DataFrame(job_list)
      df.to_csv("job_listings.csv", index=False)
      print(f"✅ Scraped {len(df)} job listings!")
      df.head()
```

✅ Scraped 448 job listings!

[15]:

	Title	Company	Location	Job URL
0	Data Analyst II (MS Excel, Power BI, SQL)	Walmart Global Tech India	Greater Chennai Area	https://in.linkedin.com/jobs/view/data-analyst...
1	Data Analyst	KPMG India	Delhi, India	https://in.linkedin.com/jobs/view/data-analyst...
2	Data Analyst	slice	Bengaluru, Karnataka, India	https://in.linkedin.com/jobs/view/data-analyst...
3	Data Analyst	EaseMyTrip.com	Mumbai, Maharashtra, India	https://in.linkedin.com/jobs/view/data-analyst...
4	Advance Analytics	Pidilite Industries Limited	Mumbai, Maharashtra, India	https://in.linkedin.com/jobs/view/advance-anal...

[ ]: