Installing PyCharm: [Make sure python v3.9.x or above is installed]

https://www.jetbrains.com/pycharm/download/#section=windows

- Download the community version.
- Run the installer and in the steps which come
- Choose the following options when prompted:
    - Create Desktop Shortcut and Add “Open Folder as Project”
- Once it is installed create a new project by the name of the scraper(any name) and open it.

Installing Selenium, Beautiful Soup, WebDriver:

Step-1: Open your scraper project in pycharm and go to its terminal.

Step-2: Run the following commands to install the modules:

Selenium: pip install selenium
Beautiful Soup: pip install beautifulsoup4
WebDriver Manager: pip install webdriver_manager

Step-3: Make sure that the chrome browser is installed in your system

Step-4: Run the following code (in main.py) to ensure that everything is working correctly.

```
import time
from selenium import webdriver
from bs4 import BeautifulSoup
from webdriver_manager.chrome import ChromeDriverManager

driver = webdriver.Chrome(ChromeDriverManager().install())

driver.get("https://google.com")

time.wait(5)
```

Tutorial of BeautifulSoup:
https://youtu.be/XQgXKtPSzUI [This will give you a basic idea of how to use Beautiful Soup]

Code of Today's session:

```
# import time
# import os
import time

from selenium import webdriver
from bs4 import BeautifulSoup
from webdriver_manager.chrome import ChromeDriverManager


driver = webdriver.Chrome(ChromeDriverManager().install())




urls = ["https://www.codechef.com/problems/XYSTR",
"https://www.codechef.com/problems/SUBINC"]
cnt=0
for url in urls:
   driver.get(url)
   cnt+=1
   time.sleep(5)

   html = driver.page_source
   soup = BeautifulSoup(html, 'html.parser')

   problem_text = soup.find('div', {"class": "problem-statement"}).get_text()
   # print(problem_text)
   problem_text = problem_text.encode("utf-8")
   problem_text = str(problem_text)

   with open("problem"+str(cnt)+".txt", "w+") as f:
       f.write(problem_text)

# driver.get("https://www.codechef.com/tags/problems/dynamic-programming")
#
# time.sleep(5)
#
# html = driver.page_source
# soup = BeautifulSoup(html, 'html.parser')

# all_ques_div = soup.findAll("div", {"class": "problem-tagbox-inner"})
#
# all_ques = []
```

```
#
# for ques in all_ques_div:
#     all_ques.append(ques.findAll("div")[0].find("a"))
#
# urls = []
# titles = []
#
# for ques in all_ques:
#     urls.append("https://www.codechef.com"+ques['href'])
#     titles.append(ques.text)

# with open("problem_urls.txt", "w+") as f:
#     f.write('\n'.join(urls))

# with open("problem_titles.txt", "w+") as f:
#     f.write('\n'.join(titles))
```

Note:

Provide appropriate delays between your get requests as it gives your browser time to visit the page and also make your automatic scrapper behave like a human so that the website doesn’t blocks you.