

Q:1

```
import urllib.request  
counts = dict()  
fhand = urllib.request.urlopen('http://data.pr4e.org/romeo.txt')  
for line in fhand:  
    words = line.decode().split()  
    for word in words:  
        counts[word] = counts.get(word, 0) + 1  
print(counts)
```

Problem Solving Using Python and R Lab

Lab11. Retrieving Data From Web and Parsing

Question1. Retrieve data from web page using **URLLIB** and print the frequency of words from that page.

Q:2

=

```
import urllib.request, urllib.parse, urllib.error  
from bs4 import BeautifulSoup  
import ssl  
ctx = ssl.create_default_context()  
ctx.check_hostname = False  
ctx.verify_mode = ssl.CERT_NONE  
url = input('Enter url')  
html = urllib.request.urlopen(url, context=ctx).read()  
soup = BeautifulSoup(html, 'html.parser')  
tags = soup('a')  
for tag in tags:  
    print(tag.get('href', None))
```


Question: 3

```
from IPython.core.display import HTML
que3 = """
<table>
<tr>
<td> ID </td>
<td> Name </td>
<td> mark1 </td>
<td> mark2 </td>
<td> mark3 </td>
</tr>
<tr>
<td> DS01 </td>
<td> Rex </td>
<td> 80 </td>
<td> 55 </td>
<td> 78 </td>
</tr>
<tr>
<td> DS02 </td>
<td> Peter </td>
<td> 68 </td>
<td> 98 </td>
<td> 87 </td>
</tr>
</table>
"""
```

HTML(que3)

Question: 4

```
import JSON
```

```
data = """
[{"id": "ds01",
  "name": "Rex",
  "Semester1": "80,55",
  "Semester2": "78,98"
},
{"id": "ds02",
  "name": "Brent",
  "Semester1": "35,67",
  "Semester2": "65,87"
}
]""
```

Question3. Create a HTML file for the following Student Marks and print the number of students and their names and marks.

ID	name	mark1	mark2	mark3
DS01	rex	87	57	74
DS02	peter	68	98	55

Question4. Create a JSON file for the following Students Marks and print the number of students and their names and marks.

ID: DS01: Name: rex
 Semester1: 80, 55
 Semester2: 50, 70, 82
 ID: DS02: Name: peter
 Semester1: 92, 75

```
info = json.loads(data)
print('user count:', len(info))

for item in info:
    print('Name', item['name'])
    print('Id', item['id'])
    print('Semester1', item['semester1'])
    print('Semester2', item['semester2'])
```

Question: 5

=

```
import requests  
from bs4 import BeautifulSoup  
page = requests.get("https://weather.com/en-IN/weather/tenday/t/d251a574  
fa1fa7a38aef6a630eb7eq1a5c92cc2165acc34a8b2")
```

Page.content

```
bs4 = BeautifulSoup(page.content, 'html.parser')
```

```
bs4.find_all('p')
```

```
(bs4.find_all(class_="DetailsSummary--DetailsSummary--QPFD-"))
```

```
days = [bs4.find_all('h2')[day].get_text() for day in range  
(len(bs4.find_all('h2')))]
```

```
temp = [bs4.find_all(class_="DetailsSummary--temperature-3FMlw")]
```

```
[temp[0].get_text() for temp in range(len(bs4.find_all(class_=  
="DetailsSummary--temperature-3FMlw")))]
```

```
days = days[1:12]
```

```
temp = temp[1:12]
```

```
from IPython.display import display
```

```
import pandas as pd
```

```
b = {"Days": days,  
      "Temperature": temp}
```

```
weather = pd.DataFrame(b)
```

weather

Question5. Crawl Weather of a City and Display 7 Day Forecast

- Find weather data of Tiruchirappalli city from some website such as www.weather.com
 - Exploring page structure with Chrome DevTools
 - Extract information from web page and display the weather forecast for 10 days or 7 days

Reference: <https://www.dataquest.io/blog/web-scraping-tutorial-python/>

Question: 6

```
! pip install yfinance --upgrade --no-cache-dir  
import yfinance as yf  
yf.pdr_override()  
df_info = pdr.get_data_yahoo("INFY", start="2018-01-01").  
reset_index()  
df_info.to_csv('INFY.csv', index=False)  
df_info.head()
```

```
! pip install yfinance --upgrade --no-cache-dir  
import yfinance as yf  
yf.pdr_override()  
df_info = pdr.get_data_yahoo("CTS", start="2018-01-01").  
reset_index()  
df_info.to_csv('CTS.csv', index=False)  
df_info.head()
```

```
import matplotlib.pyplot as plt  
%matplotlib inline  
df_info["Adj Close"].plot(grid=True)
```

```
import matplotlib.pyplot as plt  
%matplotlib inline  
df_cts["Adj Close"].plot(grid=True)
```

Question6. Real Time Stock Prices Crawling and Display of a specified Company

- Crawl prices of a stock such as InfoSys (Stock Code INFY) for a period of month or year.
- Plot a line graph of monthly or yearly price movements
- Crawl prices of one more stock such as CTS
- Update your line graph with the prices movements of two stocks

Reference:

<https://ntguardian.wordpress.com/2018/07/17/stock-data-analysis-python-v2/>

https://github.com/rajkumarbhc/data-science-utils/blob/master/data_science_utils/financial/Web%20Scraping%20and%20Financial%20Data%20Exploration%20in%20Python-Tutorial.ipynb

```
s = "INFY"
infosys, cts = (pdr.get_data_yahoo(s, start = "2018-01-01").reset_index())
for s in ["INFY", "CTS"])
stocks = pd.DataFrame( {"INFY": df_infy["Adj Close"],
                        "CTS": df_cts["Adj Close"]})
stocks.head()
stocks.plot(grid = True)
```