FLIP ROBO

ASSIGNMENT-1

WEB SCRAPING

In all the following questions, you have to use BeautifulSoup to scrape different websites and collect data as per the requirement of the question.

Every answer to the question should be in form of a python function which should take URL as the parameter. Use Jupyter Notebooks to program, upload it on your GitHub and send the link of the Jupyter notebook to your SME.

1) Write a python program to display all the header tags from **wikipedia.org**.

ANS- http://localhost:8889/notebooks/Untitled7.ipynb?kernel_name=python3#

2) Write a python program to display **IMDB's Top** rated **100 movies'** data (i.e. **name, rating, year of release**) and make **data frame**.

ANS- http://localhost:8889/notebooks/Untitled13.ipynb?kernel_name=python3

3) Write a python program to display **IMDB's Top** rated **100 Indian movies'** data (i.e. **name, rating, year of release**) and make **data frame**.

ANS- http://localhost:8889/notebooks/Untitled16.ipynb?kernel_name=python3#

4) Write s python program to display list of respected former presidents of India(i.e. Name , Term of office) from https://presidentofindia.nic.in/former-presidents.htm

ANS- http://localhost:8889/notebooks/Untitled17.ipynb?kernel_name=python3#

5) Write a python program to scrape cricket rankings from **icc-cricket.com**. You have to scrape:

a) Top **10 ODI teams** in men's cricket along with the records for **matches, points and rating**.

ANS - http://localhost:8889/notebooks/Untitled20.ipynb?kernel_name=python3

b) Top **10 ODI Batsmen** along with the records of their **team and rating.**

ANS -

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
html = urlopen ('https://www.icc-cricket.com/rankings/mens/player-rankings/odi')
bs = BeautifulSoup(html, "html.parser")
first_data = soup.find ("div",attrs={"data-cricket-scope":"odi"}).find("div",class_="rankings-block__top-player").get_text(strip=True,separator=" ").split(" ")
other_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find_all("tr",class_="table-body")
final_lst=[]
final_lst.append(first_data)
for i in data:
    split_lst=i.get_text(strip=True,separator=" ").split(" ")
    final_lst.append(split_lst)
```

c) Top **10 ODI bowlers** along with the records of their **team and rating.**

ANS-

```
first_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find("div",class_="rankings-block__top-bowlers").get_text(strip=True,separator=" ").split(" ")

other_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find_all("tr",class_="table-body")

final_lst=[]
final_lst.append(first_data)
```

```
for i in data:
split_lst=i.get_text(strip=True,separator=" ").split(" ")
final_lst.append(split_lst)
```

6) Write a python program to scrape cricket rankings from **icc-cricket.com**. You have to scrape:

a) Top **10 ODI teams** in women's cricket along with the records for **matches, points and rating**.

ANS- http://localhost:8889/notebooks/Untitled22.ipynb?kernel_name=python3

b) Top **10 women's ODI Batting** players along with the records of their **team and rating**.

ANS- first_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find("div",class_="rankings-block__top-player").get_text(strip=True,separator=" ").split(" ")

```
other_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find_all("tr",class_="table-body")


final_lst=[]
final_lst.append(first_data)


for i in data:
   split_lst=i.get_text(strip=True,separator=" ").split(" ")
        final_lst.append(split_lst)
```

c) Top **10 women's ODI all-rounder** along with the records of their **team and rating**.

ANS- first_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find("div",class_="rankings-block_top-womens all rounders").get_text(strip=True,separator=" ").split(" ")

```
other_data=soup.find("div",attrs={"data-cricket-scope":"odi"}).find_all("tr",class_="table-body")

final_lst=[]
final_lst.append(first_data)

for i in data:
   split_lst=i.get_text(strip=True,separator=" ").split(" ")
   final_lst.append(split_lst)
```

7) Write a python program to scrape mentioned news details from https://www.cnbc.com/world/?region=world :

i) Headline

ii) Time

iii) News Link

ANS- http://localhost:8889/notebooks/Untitled24.ipynb?kernel_name=python3

8) Write a python program to scrape the details of most downloaded articles from AI in last 90 days.

https://www.journals.elsevier.com/artificial-intelligence/most-downloaded-articles

Scrape below mentioned details :

i) Paper Title

ii) Authors

iii) Published Date

iv) Paper URL

ANS- http://localhost:8889/notebooks/Untitled25.ipynb?kernel_name=python3

9) Write a python program to scrape mentioned details from **dineout.co.in** :

i) Restaurant name

ii) Cuisine

iii) Location

iv) Ratings

v) Image URL

ANS- http://localhost:8889/notebooks/Untitled26.ipynb?kernel_name=python3

10) Write a python program to scrape the details of top publications from Google Scholar from https://scholar.google.com/citations?view_op=top_venues&hl=en

i) Rank

ii) Publication

iii) h5-index

iv) h5-median

ANS- http://localhost:8889/notebooks/Untitled27.ipynb?kernel_name=python3