

## Practical – 9 : Web Scraping

**Read a website(with tabular data ) as HTML and perform basic tabular data analysis as practical 7**

### Code:

```
from gettext import install

import bs4 as bs4
import html5lib as html5lib
import pip
import requests as requests

pip install requests
pip install html5lib
pip install bs4

import requests
URL = "https://www.dp.org/data-structures/"
r = requests.get(URL)
print(r.content)

#This will not run on online IDE
import requests
from bs4 import BeautifulSoup
URL = "http://www.values.com/inspirational-quotes"
r = requests.get(URL)
soup = BeautifulSoup(r.content, 'html5lib') # If this line causes an error,
run 'pip install html5lib' or install html5lib
print(soup.prettify())

#Python program to scrape website
#and save quotes from website
import requests
from bs4 import BeautifulSoup
import csv

URL = "http://www.values.com/inspirational-quotes"
r = requests.get(URL)

soup = BeautifulSoup(r.content, 'html5lib')

quotes=[] # a list to store quotes

table = soup.find('div', attrs = {'id':'all_quotes'})

for row in table.findAll('div',attrs = {'class':'col-6 col-lg-3 text-center margin-30px-bottom sm-margin-30px-top'}):
    quote = {}
    quote['theme'] = row.h5.text
    quote['url'] = row.a['href']
    quote['img'] = row.img['src']
    quote['lines'] = row.img['alt'].split(" #")[0]
    quote['author'] = row.img['alt'].split(" #")[1]
    quotes.append(quote)

filename = 'inspirational_quotes.csv'
with open(filename, 'w', newline='') as f:
```

```
w = csv.DictWriter(f, ['theme', 'url', 'img', 'lines', 'author'])
w.writeheader()
for quote in quotes:
    w.writerow(quote)
```

## Output:

```
<div class="row" id="all_quotes">
  <div class="col 6 col lg 3 text center margin 30px bottom sm margin 30px top">
    <a href="/inspirational-quotes/8852-never-mistake-knowledge-for-wisdom-one-helps">
      <img alt="Never mistake knowledge for wisdom. One helps you make a living; the other helps you make a life. #<Author:0x00007f1917948248>"
        class="margin-10px-bottom shadow" src="https://assets.passion.com/quotes/quote_artwork/8852/medium/28208324_tuesday_quote.jpg?1584726475"
        width="318" height="318" data-no-retina>
      </a>
    <h5 class="value_on_red">
      <a href="/inspirational-quotes/8852-never-mistake-knowledge-for-wisdom-one-helps">WISDOM</a>
    </h5>
  </div>
```

```
for row in table.find_all_next('div', attrs = {'class': 'col-6 col-lg-3 text-center margin-30px-bottom sm-margin-30px-top'})
    quote = {}
    quote['theme'] = row.h5.text
    quote['url'] = row.a['href']
    quote['img'] = row.img['src']
    quote['lines'] = row.img['alt'].split(" #")[0]
    quote['author'] = row.img['alt'].split(" #")[1]
    quotes.append(quote)
```

```
quote['theme'] = row.h5.text
```

```
quote['url'] = row.a['href']
```

```
filename = 'inspirational_quotes.csv'
with open(filename, 'w', newline='') as f:
    w = csv.DictWriter(f, ['theme', 'url', 'img', 'lines', 'author'])
    w.writeheader()
    for quote in quotes:
        w.writerow(quote)
```