

## BASIC WEB SCRAPING

```
import pandas
import requests
from bs4 import BeautifulSoup
url = 'https://quotes.toscrape.com/'
html = requests.get(url)
soup = BeautifulSoup(html.text, 'html.parser')
soup
print(soup.prettify())
print(soup.get_text())
soup.title
soup.title.name
soup.title.string
soup.a
soup.span.text
soup.small.text
soup.a.text
quotes = soup.find_all("div", {"class": "quote"})
quotes
tag = soup.find_all("div", {"class": "tags"})
tag
for i in quotes:
    print((i.find('span', {'class': 'text'})).text)
for i in soup.findAll("div", {"class": "tags"}):
    print((i.find("meta"))['content'])
# store root url without page number
root = 'http://quotes.toscrape.com/page/'
# create empty lists
quotes = []
authors = []
tags = []
# loop over page 1 to 10
for pages in range(1,10):
    html = requests.get(root+ str(pages))
    soup = BeautifulSoup(html.text)
    for i in soup.findAll("div", {"class": "quote"}):
        quotes.append((i.find("span", {"class": "text"})).text)
    for j in soup.findAll("div", {"class": "quote"}):
        authors.append((j.find("small", {"class": "author"})).text)
    for k in soup.findAll("div", {"class": "tags"}):
        tags.append((k.find("meta"))['content'])
quotes
authors
tags
df=pandas.DataFrame({"Quotes" : quotes, "Authors": authors, "Tags" : tags})
print(df)
```