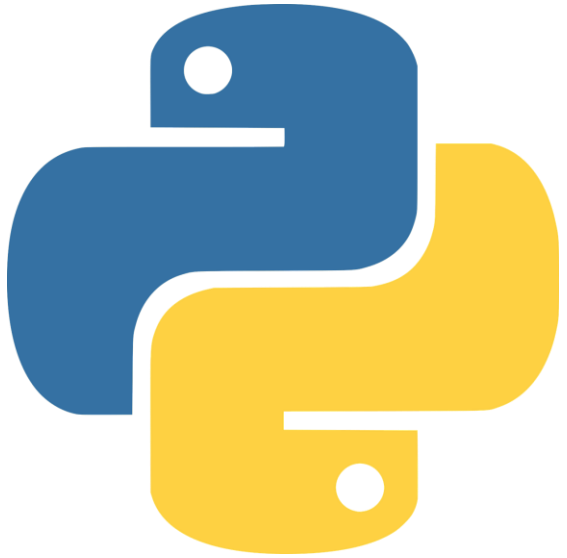
Web Scraping with Python Using Beautiful Soup

I will show how to modify, view, and search for HTML tags and much more.



Searching For Tags

The **find** method return a single BeautifulSoup object

web_scraping2.py

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")

result = doc.find("option")
print(result)
```

The **find_all** method returns a list of BeautifulSoup objects

web_scraping2.py

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")

result = doc.find_all("option")
print(result)
```

Tag Attributes

To see all of the attributes:

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tag = doc.find_all("option")
print(tag.attrs)
```

web_scraping2.py

To modify attributes:

```
from bs4 import BeautifulSoup

with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")

tag = doc.find_all("option")
tag['selected'] = 'false'
tag['color'] = "blue"
print(tag)
```

web_scraping2.py

Find Multiple Tags

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tags = doc.find_all(["p", "div", "li"])
print(tags)
```

web_scraping2.py

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")

doc.find_all(["option"], text="Undergraduate")
print(tags)
```

web_scraping2.py

Find Attributes

web_scraping2.py

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tags = doc.find_all(["option"], text="Undergraduate", value="undergraduate")
print(tags)
```

Find Class Names

```
from bs4 import BeautifulSoup


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tags = doc.find_all(class_="btn-item")
print(tags)
```

web_scraping2.py

Find Regular Expressions

```
from bs4 import BeautifulSoup

import re


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tags = doc.find_all(text=re.compile("\$.*"))
for tag in tags:
    print(tag.strip())
```

web_scraping2.py

Find Limit

web_scraping2.py

```
from bs4 import BeautifulSoup
import re


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tags = doc.find_all(text=re.compile("\$.*"), limit=1)
for tag in tags:
    print(tag.strip())
```

Count of results

Save Modified HTML

```
from bs4 import BeautifulSoup
import re


with open("index2.html", "r") as f:
    doc = BeautifulSoup(f, "html.parser")


tags = doc.find_all("input", type="text")
for tag in tags:
    tag['placeholder'] = "I changed you!"

with open("changed.html", "w") as file:
    file.write(str(doc))
```

web_scraping2.py