Web Scraping - Web Document, Parser and Object

August 19, 2022

```python
[1]: #import bs4 the beautifulsoup library

     from bs4 import BeautifulSoup


     #create a html document


     html_doc="""<html>
     <body>
     <h1>My First Heading</h1>
     <b><!--This is a comment line--></b>
     <p title="About Me" class="test">My first paragraph.</p>
     <div class="cities">
     <h2>London</h2>
     </div>
     </body>
     </html>"""
```

```python
[2]: soup = BeautifulSoup(html_doc,'html.parser')
     soup
```

```
[2]: <html>
     <body>
     <h1>My First Heading</h1>
     <b><!--This is a comment line--></b>
     <p class="test" title="About Me">My first paragraph.</p>
     <div class="cities">
     <h2>London</h2>
     </div>
     </body>
     </html>
```

```python
[3]: #view the soup type
     type(soup)
```

```
[3]: bs4.BeautifulSoup
```

```
[4]: #view the soup object
     print (soup)
```

```
<html>
<body>
<h1>My First Heading</h1>
<b><!--This is a comment line--></b>
<p class="test" title="About Me">My first paragraph.</p>
<div class="cities">
<h2>London</h2>
</div>
</body>
</html>
```

```
[6]: #create a tag object
     tag1=soup.b    #to access the info related to the command, we can access data␣
      ↪only using tags so here we use b tag to get info
```

```
[7]: #print the tag
     print(tag1)
```

```
<b><!--This is a comment line--></b>
```

```
[10]: #create a comment object type
      tag2=soup.b.string   #i need only the comment
```

```
[11]: #view the comment
      print(tag2)
```

```
This is a comment line
```

```
[ ]:
```