searching_the_parse_tree_solution

March 18, 2023

1 TODO: Find All The <p> Tags

In the cell below, use the `.find_all()` method to find all the `<p>` tags in the `sample.html` file. Start by opening the `sample.html` file and passing the open filehandle to the BeautifulSoup constructor using the `lxml` parser. Save the BeautifulSoup object returned by the constructor in a variable called `page_content`. Then use the `.find_all()` method to find all the `<p>` tags from the `page_content` object. Save the list returned by the `.find_all()` method in a variable called `p_list`. Finally, loop through the list and print each tag in the list. Since the `<p>` tags contain subtags, use the `.prettify()` method to improve readability.

In [1]:
```
# Import BeautifulSoup
from bs4 import BeautifulSoup

# Open the HTML file and create a BeautifulSoup Object
with open('./sample.html') as f:
    page_content = BeautifulSoup(f, 'lxml')

# Find all the p tags
p_list = page_content.find_all('p')

# Print each tag in the p_list
for tag in p_list:
    print(tag.prettify())
```

```
<p>
 Student Hub is our real time collaboration platform where you can work with peers and mentors.
</p>

<p>
 Search or ask questions in
 <a href="https://knowledge.udacity.com/">
  Knowledge
 </a>
</p>

<p>
 Good luck and we hope you enjoy the course
```

```
</p>
```

2 TODO: Find All The <a> and <link> Tags

In the cell below, use the `.find_all()` method to find all the `<a>` and `<link>` tags in the `sample.html` file. Start by opening the `sample.html` file and passing the open filehandle to the BeautifulSoup constructor using the `lxml` parser. Save the BeautifulSoup object returned by the constructor in a variable called `page_content`. Then find all the `<a>` and`<link>` tags from the `page_content` object by passing a list to the `.find_all()` method. Loop through the list and print each tag in the list. Use the `.prettify()` method to improve readability.

```
In [2]: # Import BeautifulSoup
        from bs4 import BeautifulSoup

        # Open the HTML file and create a BeautifulSoup Object
        with open('./sample.html') as f:
            page_content = BeautifulSoup(f, 'lxml')

        # Print all the a and link tags
        for tag in page_content.find_all(['a', 'link']):
            print(tag.prettify())
```

```
<link href="./teststyle.css" rel="stylesheet"/>

<a href="https://knowledge.udacity.com/">
 Knowledge
</a>
```

3 TODO: Find All The <h1> Tags With The Attribute id='intro'

In the cell below, use the `.find_all()` method to find all the `<h1>` tags in the `sample.html` file that have the attribute `id="intro"`. Start by opening the `sample.html` file and passing the open filehandle to the BeautifulSoup constructor using the `lxml` parser. Save the BeautifulSoup object returned by the constructor in a variable called `page_content`. Then find all the `<h1>` tags that have the attribute `id="intro"` from the `page_content` object. Loop through the list and print each tag in the list.

```
In [3]: # Import BeautifulSoup
        from bs4 import BeautifulSoup

        # Open the HTML file and create a BeautifulSoup Object
        with open('./sample.html') as f:
            page_content = BeautifulSoup(f, 'lxml')
```

```
        # Print all the h1 tags with id = intro
        for tag in page_content.find_all('h1', id = 'intro'):
            print(tag)
```

```
<h1 id="intro">Get Help From Peers and Mentors</h1>
```

4 TODO: Find All Tags With Attribute `id='hub'`

In the cell below, use the `.find_all()` method to find all the tags in the `sample.html` file that have the attribute `id="hub"`. Start by opening the `sample.html` file and passing the open filehandle to the BeautifulSoup constructor using the `lxml` parser. Save the BeautifulSoup object returned by the constructor in a variable called `page_content`. Then find all the tags that have the attribute `id="hub"` from the `page_content` object. Loop through the list and print each tag in the list.

```
In [4]: # Import BeautifulSoup
        from bs4 import BeautifulSoup

        # Open the HTML file and create a BeautifulSoup Object
        with open('./sample.html') as f:
            page_content = BeautifulSoup(f, 'lxml')

        # Print all the tags with id = hub
        for tag in page_content.find_all(id = 'hub'):
            print(tag)
```

```
<h2 class="h2style" id="hub">Student Hub</h2>
```