searching_by_class_and_regexes_solution

March 19, 2023

1 TODO: Find All Tags With Attribute `class='section'`

In the cell below, use the `.find_all()` method to find all the tags in the `sample.html` file that have the attribute `class="section"`. 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 `class="section"` from the `page_content` object. Loop through the list and print each tag in the list. 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')

# Print the tags that have the attribute class_ = 'h2style'
for tag in page_content.find_all(class_ = 'section'):
    print(tag.prettify())
```

```
<div class="section">
 <h2 class="h2style" id="hub">
  Student Hub
 </h2>
 <p>
  Student Hub is our real time collaboration platform where you can work with peers and mentors.
 </p>
</div>

<div class="section">
 <h2 class="h2style" id="know">
  Knowledge
 </h2>
 <p>
  Search or ask questions in
  <a href="https://knowledge.udacity.com/">
   Knowledge
```

```
  </a>
 </p>
</div>
```

2 TODO: Find All Tags The Start With The Letter `h`

In the cell below, pass a regular expression to the `.find_all()` method to find all the tags whose names start with the letter `h`. 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 whose names start with the letter `h` by passing a regular expression to the `.find_all()` method. Loop through the list and print each tag in the list.

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

        # Import the re module
        import re

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

        # Print only the tag names of all the tags whose names start with the letter h
        for tag in page_content.find_all(re.compile(r'^h')):
            print(tag.name)
```

```
html
head
h1
h2
hr
h2
h3
```