Practical 6.1

Write following python programs on Beautiful Soup Perform following operations on a HTML document.

- Find the title tag from a given html document.
- Count and retrieve all the paragraph tags and extract the text in the first paragraph tag.
- Find the text of the first "a" tag and length of the text of the first "h2" tag.
- Find the href of the first "a" tag.
- Find the first tag with a given attribute value in an html document.

In [3]:
```
from bs4 import BeautifulSoup
import requests
```

In [22]:
```
html_doc = """
<html>
<head>
<title>Enter a title, displayed at the top of the window.</title>
</head>
<body>
<h2>Enter the main heading, usually the same as the title.</h2>
<p>Be <b>bold</b> in stating your key points. Put them in a list: </p>
<ul>
<li>The first item in your list</li>
<li>The second item; <i>italicize</i> key words</li>
</ul>
<p>Improve your image by including an image. </p>
<p><img src="http://www.mygifs.com/CoverImage.gif" alt="A Great HTML Resour
ce"></p>
<p>Add a link to your favorite <a href="https://www.dummies.com/">Web site
</a>.
Break up your page with a horizontal rule or two. </p>
<hr>
<div id="special" class="s" >Special Div
<p>Finally, link to <a href="page2.html">another page</a> in your own Web s
ite.</p>
<p>© Wiley Publishing, 2011</p></div>
</body>
</html>
"""
```

In [23]:
```
soup = BeautifulSoup(html_doc,'html.parser')
```

In [24]:
```
print(soup.prettify())
```

```
<html>
 <head>
  <title>
   Enter a title, displayed at the top of the window.
```

```
  </title>
 </head>
 <body>
  <h2>
   Enter the main heading, usually the same as the title.
  </h2>
  <p>Be
   <b>
    bold
   </b>
   in stating your key points. Put them in a list:
  </p>
  <ul>
   <li>
    The first item in your list
   </li>
   <li>
    The second item;
    <i>
     italicize
    </i>
    key words
   </li>
  </ul>
  <p>
   Improve your image by including an image.
  </p>
  <p>
   <img alt="A Great HTML Resource" src="http://www.mygifs.com/CoverImage.
gif"/>
  </p>
  <p>
   Add a link to your favorite
   <a href="https://www.dummies.com/">
    Web site
   </a>
Break up your page with a horizontal rule or two.
  </p>
  <hr/>
  <div class="s" id="special">
   Special Div
   <p>
    Finally, link to
    <a href="page2.html">
     another page
    </a>
    in your own Web site.
   </p>
   <p>
    © Wiley Publishing, 2011
   </p>
  </div>
 </body>
</html>
```

1. Find the title tag from a given html document.

In [25]:
```
soup.title.string
```

Out[25]:
```
'Enter a title, displayed at the top of the window.'
```

2. Count and retrieve all the paragraph tags and extract the text in the first paragraph tag.

In [26]:
```
count = 0
for link in soup.find_all('p'):
    print(link.text)
    count += 1

print("\nTotal number of paragraph tags : ",count)
```

```
Be bold in stating your key points. Put them in a list:
Improve your image by including an image.

Add a link to your favorite Web site.
Break up your page with a horizontal rule or two.
Finally, link to another page in your own Web site.
© Wiley Publishing, 2011

Total number of paragraph tags :  6
```

In [27]:
```
soup.find('p').text
```

Out[27]:
```
'Be bold in stating your key points. Put them in a list: '
```

3. Find the text of the first "a" tag and length of the text of the first "h2" tag.

In [35]:
```
soup.find('a').text
```

Out[35]:
```
'Web site'
```

In [29]:
```
len(soup.find('h2').text)
```

Out[29]:
```
54
```

4. Find the href of the first "a" tag.

In [30]:
```
soup.find('a').get('href')
```

Out[30]:
```
'https://www.dummies.com/'
```

5. Find the first tag with a given attribute value in an html document.

In [36]:
```
id = "special"
soup.find(id=id)
```

Out[36]:
```
<div class="s" id="special">Special Div
<p>Finally, link to <a href="page2.html">another page</a> in your own Web
site.</p>
<p>© Wiley Publishing, 2011</p></div>
```