Demo _Web Scraping - Navigating the Tree (1)

August 19, 2022

[1]:
```
#import the required library
from bs4 import BeautifulSoup
```

[2]:
```
#create html document
book_html_doc = """<catalog>
    <head><title>The web book catalog </title></head>
    <p class="title"><b>The Book Catalog</b></p>
    <books>

    <book id="bk001">
    <author>Hightower, Kim</author>
    <title>The First Book</title>
      <genre>Fiction</genre>
      <price>44.95</price>
      <pub_date>2000-10-01</pub_date>
      <review>An amazing story of nothing.</review>
   </book>


   <book id="bk002">
      <author>Nagata, Suanne</author>
      <title>Becoming Somebody</title>
      <genre>Biography</genre>
      <review>A masterpiece of the fine art of gossiping.</review>
   </book>



   <book id="bk003">
      <author>Oberg, Bruce</author>
      <title>The Poet's First Poem</title>
      <genre>Poem</genre>
      <price>24.95</price>
      <review>The least poetic poems of the decade.</review>
   </book>

   </books>
```

```
   </catalog>"""
```

[3]:
```
booksoup=BeautifulSoup(book_html_doc,"html.parser")
```

[4]:
```
print(booksoup)
```

```
<catalog>
<head><title>The web book catalog </title></head>
<p class="title"><b>The Book Catalog</b></p>
<books>
<book id="bk001">
<author>Hightower, Kim</author>
<title>The First Book</title>
<genre>Fiction</genre>
<price>44.95</price>
<pub_date>2000-10-01</pub_date>
<review>An amazing story of nothing.</review>
</book>
<book id="bk002">
<author>Nagata, Suanne</author>
<title>Becoming Somebody</title>
<genre>Biography</genre>
<review>A masterpiece of the fine art of gossiping.</review>
</book>
<book id="bk003">
<author>Oberg, Bruce</author>
<title>The Poet's First Poem</title>
<genre>Poem</genre>
<price>24.95</price>
<review>The least poetic poems of the decade.</review>
</book>
</books>
</catalog>
```

[5]:
```
print(booksoup.title.string)
```

```
The web book catalog
```

[6]:
```
print(booksoup.get_text())
```

```


The web book catalog
The Book Catalog


Hightower, Kim
The First Book
Fiction
```

```
44.95
2000-10-01
An amazing story of nothing.


Nagata, Suanne
Becoming Somebody
Biography
A masterpiece of the fine art of gossiping.


Oberg, Bruce
The Poet's First Poem
Poem
24.95
The least poetic poems of the decade.
```

[7]:
```
# Remove the empty space
for i in booksoup.stripped_strings:  #striioed string removes the space which␣
 ↪is unwanted
    print(i)
```

```
The web book catalog
The Book Catalog
Hightower, Kim
The First Book
Fiction
44.95
2000-10-01
An amazing story of nothing.
Nagata, Suanne
Becoming Somebody
Biography
A masterpiece of the fine art of gossiping.
Oberg, Bruce
The Poet's First Poem
Poem
24.95
The least poetic poems of the decade.
```

[8]:
```
#craete a sibling object
sibiling1=booksoup.catalog.books.book
sibiling1
```

[8]:
```
<book id="bk001">
<author>Hightower, Kim</author>
<title>The First Book</title>
<genre>Fiction</genre>
<price>44.95</price>
<pub_date>2000-10-01</pub_date>
<review>An amazing story of nothing.</review>
</book>
```

[9]:
```
sibling2=sibiling1.next_sibling.next_sibling
sibling2
```

[9]:
```
<book id="bk002">
<author>Nagata, Suanne</author>
<title>Becoming Somebody</title>
<genre>Biography</genre>
<review>A masterpiece of the fine art of gossiping.</review>
</book>
```

[10]:
```
sibling3=sibling2.next_sibling.next_sibling
sibling3
```

[10]:
```
<book id="bk003">
<author>Oberg, Bruce</author>
<title>The Poet's First Poem</title>
<genre>Poem</genre>
<price>24.95</price>
<review>The least poetic poems of the decade.</review>
</book>
```

[11]:
```
# previous Sibling

previou=sibling3.previous_sibling.previous_sibling
previou
```

[11]:
```
<book id="bk002">
<author>Nagata, Suanne</author>
<title>Becoming Somebody</title>
<genre>Biography</genre>
<review>A masterpiece of the fine art of gossiping.</review>
</book>
```

[12]:
```
first=previou.previous_sibling.previous_sibling
first
```

[12]:
```
<book id="bk001">
<author>Hightower, Kim</author>
```

```
<title>The First Book</title>
<genre>Fiction</genre>
<price>44.95</price>
<pub_date>2000-10-01</pub_date>
<review>An amazing story of nothing.</review>
</book>
```

[ ]: