Welcome Stephanie from Using Python to Access Web Data  Exit

Scraping Numbers from HTML using BeautifulSoup In this assignment you will write a Python program similar to http://www.pythonlearn.com/code/urllink2.py (http://www.pythonlearn.com/code/urllink2.py). The program will use **urllib** to read the HTML from the data files below, and parse the data, extracting numbers and compute the sum of the numbers in the file.

We provide two files for this assignment. One is a sample file where we give you the sum for your testing and the other is the actual data you need to process for the assignment.

- Sample data: http://python-data.dr-chuck.net/comments_42.html (http://python-data.dr-chuck.net/comments_42.html) (Sum=2482)
- Actual data: http://python-data.dr-chuck.net/comments_211811.html (http://python-data.dr-chuck.net/comments_211811.html) (Sum ends with 29)

You do not need to save these files to your folder since your program will read the data directly from the URL. **Note:** Each student will have a distinct data url for the assignment - so only use your own data url for analysis.

Data Format
The file is a table of names and comment counts. You can ignore most of the data in the file except for lines like the following:

```
<tr><td>Modu</td><td><span class="comments">90</span></td></tr>
<tr><td>Kenzie</td><td><span class="comments">88</span></td></tr>
<tr><td>Hubert</td><td><span class="comments">87</span></td></tr>
```

You are to find all the tags in the file and pull out the numbers from the tag and sum the numbers.
Look at the sample code (http://www.pythonlearn.com/code/urllink2.py) provided. It shows how to find all of a certain kind of tag, loop through the tags and extract the various aspects of the tags.

```
...
# Retrieve all of the anchor tags
tags = soup('a')
for tag in tags:
   # Look at the parts of a tag
   print 'TAG:',tag
   print 'URL:',tag.get('href', None)
   print 'Contents:',tag.contents[0]
   print 'Attrs:',tag.attrs
```

You need to adjust this code to look for **span** tags and pull out the text content of the span tag, convert them to integers and add them up to complete the assignment.

Sample Execution

```
$ python solution.py
Enter - http://python-data.dr-chuck.net/comments_42.html
Count 50
Sum 2482
```

Turning in the Assignment

Enter the sum from the actual data and your Python code below:

Sum: (ends with 29) Submit Assignment

Python code: