

Scraping Numbers from HTML using BeautifulSoup In this assignment you will write a Python program similar to <http://www.py4e.com/code3/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://py4e-data.dr-chuck.net/comments_42.html (Sum=2553)
- Actual data: http://py4e-data.dr-chuck.net/comments_1419496.html (Sum ends with 64)

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.

```
C:\Users\mirey\OneDrive\Desktop>BSHTML.py
Enter - http://py4e-data.dr-chuck.net/comments_1419496.html
2764
```

```
C:\Users\mirey\OneDrive\Desktop>BSHTML.py
Enter - http://py4e-data.dr-chuck.net/comments_42.html
2553
```