## HW 7: Beautiful Soup

In this assignment you will complete two Python functions in **soupHW.py** that are similar to the code in  http://www.py4e.com/code3/urllink2.py. The first function (*getSumSpans*) will use **urllib** to read the HTML from a URL and extract the numbers in the *text* of the *span* tag and then return the sum of the numbers (as an integer).  The second function (*followLinks*) will also use **urllib** to read the HTML from a URL and extract the *a (anchor)* tags and then repeatedly follow the link at a particular position relative to the first name in the list. It will return the *text* from the last a (anchor) tag that it processes which will be a person's name.

We provided unit tests that use two files to test the first function.

- Sample data: http://py4e-data.dr-chuck.net/comments_42.html (Sum=2553)
- Actual data: http://py4e-data.dr-chuck.net/comments_132199.html (Sum= 2714)

We provided unit tests that use two files to test the second function.

- Start at http://py4e-data.dr-chuck.net/known_by_Fikret.html
  Find the link at position **3** (the first name is at position 1). Follow that link. Repeat this process **4** times. The answer is the last name that you retrieve.
  Sequence of names: Fikret Montgomery Mhairade Butchi Anayah
  Last name in the sequence: Anayah
- Start at: http://py4e-data.dr-chuck.net/known_by_Charlie.html
  Find the link at position **18** (the first name is 1). Follow that link. Repeat this process **7** times.
  Last name in the sequence: Shannah

**Note:**
The web pages used in testing the second function tweak the height between the links and hide the page after a few seconds to make it difficult for you to do the assignment without writing a Python program. But frankly with a little effort and patience you can overcome these attempts to make it a little harder to complete the assignment without writing a Python program. But that is not the point. The point is to write a clever Python program to solve the program.

### Turning in the Assignment

In canvas turn in the link to your github repository.

### Grading

- 15 points for passing test_sumSpan1
- 15 points for passing test_sumSpan2
- 15 points for passing test_followLinks1
- 15 points for passing test_followLinks2

### Total 60

You can earn 1 point of extra credit for a possible total of 3 points for each non-trivial commit that you make before Friday Oct 26 at 10pm. Each commit must be at least 3 hours apart.

You can earn 3 more points for correctly completing the *getGradeHistogram* function.   This function must create a dictionary where the key is the grade range (90, 80, 70, etc) and the value is the number of grades in that range in the text of the span tags at the given URL.  It will return a sorted list of tuples in descending order.  There is a unittest to test this function already in the file. Remove the comment to test it.