Assignment: Following Links in HTML Using BeautifulSoup

In this assignment you will write a Python program that expands on https://www.py4e.com/code3/urllinks.py. The program will use **urllib** to read the HTML from the data files below, extract the href= values from the anchor tags, scan for a tag that is in a particular position from the top and follow that link, repeat the process a number of times, and report the last name you find.

This course uses a third-party tool, Assignment: Following Links in HTML Using BeautifulSoup, to enhance your learning experience. The tool will reference basic information like your name, email, and Coursera ID.

I, **Ishaan Narula**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Open Tool

Done

Welcome Ishaan Narula from Using Python to Access Web Data

Following Links in Python

In this assignment you will write a Python program that expands on http://www.py4e.com/code3/urllinks.py. The program will use **urllib** to read the HTML from the data files below, extract the href= vaues from the anchor tags, scan for a tag that is in a particular position relative to the first name in the list, follow that link and repeat the process a number of times and report the last name you find.

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

- Sample problem: Start at http://py4e-data.dr-chuck.net/known_by_Fikret.html
  Find the link at position **3** (the first name is 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 sequence: Anayah
- Actual problem: Start at: http://py4e-data.dr-chuck.net/known_by_Ryhanna.html
  Find the link at position **18** (the first name is 1). Follow that link. Repeat this process **7** times. The answer is the last name that you retrieve.
  Hint: The first character of the name of the last page that you will load is: I

Strategy

The web pages 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.

Sample execution

Here is a sample execution of a solution:

```
$ python3 solution.py
Enter URL: http://py4e-data.dr-chuck.net/known_by_Fikret.html
Enter count: 4
Enter position: 3
Retrieving: http://py4e-data.dr-chuck.net/known_by_Fikret.html
Retrieving: http://py4e-data.dr-chuck.net/known_by_Montgomery.html
Retrieving: http://py4e-data.dr-chuck.net/known_by_Mhairade.html
Retrieving: http://py4e-data.dr-chuck.net/known_by_Butchi.html
Retrieving: http://py4e-data.dr-chuck.net/known_by_Anayah.html
```

The answer to the assignment for this execution is "Anayah".

Turning in the Assignment

Done

Enter the last name retrieved and your Python code below:

Name: (name starts with l) Submit Assignment

Python code: