

Welcome Stephanie from Using Python to Access Web Data



Exit

## Following Links in Python

In this assignment you will write a Python program that expands on <http://www.pythonlearn.com/code/urllinks.py> (<http://www.pythonlearn.com/code/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 [https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known\\_by\\_Fikret.html](https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known_by_Fikret.html) ([https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known\\_by\\_Fikret.html](https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/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: [https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known\\_by\\_Zunera.html](https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known_by_Zunera.html) ([https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known\\_by\\_Zunera.html](https://pr4e.dr-chuck.com/tsugi/mod/python-data/data/known_by_Zunera.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: K

## 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:

```
$ python solution.py
Enter URL: http://pr4e.dr-chuck.com/ ... /known_by_Fikret.html
Enter count: 4
Enter position: 3
Retrieving: http://pr4e.dr-chuck.com/ ... /known_by_Fikret.html
Retrieving: http://pr4e.dr-chuck.com/ ... /known_by_Montgomery.html
Retrieving: http://pr4e.dr-chuck.com/ ... /known_by_Mhairade.html
Retrieving: http://pr4e.dr-chuck.com/ ... /known_by_Butchi.html
Last Url: http://pr4e.dr-chuck.com/ ... /known_by_Anayah.html
```

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

**Note:** If you get an error when you run your program that complains about **CERTIFICATE\_VERIFY\_FAILED** when you call **urlopen()**, make the following changes to your program:

```
import urllib
import json
import ssl

...

print 'Retrieving', url
scontext = ssl.SSLContext(ssl.PROTOCOL_TLSv1)
uh = urllib.urlopen(url, context=scontext)
data = uh.read()
```

This will keep your Python code from rejecting the server's certificate.

### Turning in the Assignment

Enter the last name retrieved and your Python code below:

Name:  (name starts with K)

Python code: