

## Extracting Data from XML

In this assignment you will write a Python program somewhat similar to <http://www.py4e.com/code3/geoxml.py>. The program will prompt for a URL, read the XML data from that URL using `urllib` and then parse and extract the comment counts from the XML data, 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.xml](http://py4e-data.dr-chuck.net/comments_42.xml) (Sum=2553)
- Actual data: [http://py4e-data.dr-chuck.net/comments\\_1419498.xml](http://py4e-data.dr-chuck.net/comments_1419498.xml) (Sum ends with 56)

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 and Approach

The data consists of a number of names and comment counts in XML as follows:

```
Command Prompt
Microsoft Windows [Version 10.0.22000.856]
(c) Microsoft Corporation. All rights reserved.

C:\Users\miroy>cd "C:\Users\miroy\OneDrive\Desktop"

C:\Users\miroy\OneDrive\Desktop>"C:\Users\miroy\OneDrive\Desktop\extractinghtml.py"
Retrieving http://py4e-data.dr-chuck.net/comments_1419498.xml
Retrieved 4224 characters
50
2556.0

C:\Users\miroy\OneDrive\Desktop>_
```