Python for Informatics

Assignment 6

"URL Reader"

Description:

1. Rename the ***socket1.py*** program from our textbook to ***URL_reader.py***.
2. Modify the ***URL_reader.py*** program to use ***urllib*** instead of a ***socket***. This code should still process the incoming data in ***chunks***, and the size of each chunk should be 512 characters. The idea is that to allow for the processing of very large files, which may be too large to fit into working memory. You cannot use a chunk (buffer) that grows beyond the 512 character limit!
3. Add code that prompts the user for the URL so it can read any web page.
4. Add error checking using ***try*** and ***except*** to handle the condition where the user enters an improperly formatted or non-existent URL.
5. Count the number of characters received, and stop displaying any text after it has shown ***exactly*** 3000 characters. Since your chunk size will not divide evenly into 3000, you will need to add some logic to ensure that exactly 3000 characters (no more, and no less) are displayed.
6. Continue to retrieve the entire document, count the total number of characters, and display (i.e., *print*) the total number of characters.

Note: Because you are printing characters one chunk at a time, and you need to stop printing when you reach 3000 characters, there's a point where ***you will need to print only the portion of the chunk*** that enables you to reach the 3000 character print limit. **This printing of calculated portion of the last printable chunk is arguably the most challenging aspect of this assignment.**

Deliverable:

Two files as attachments at our course shell assignment page. The first file should be a Python ***URL_reader.py*** file with the specified functionality. The second file should be a screenshot image file (.png or .jpg) demonstrating the correct execution of your program with an URL of your choice. Please ensure that your full name is specified as a Python comment at the top of the ***URL_reader.py*** file.

Submission Deadline:

Please see the course schedule in our syllabus for all assignment submission deadlines.