open.michigan

Unless otherwise noted, the content of this course material is licensed under a Creative Commons Attribution 3.0 License.
http://creativecommons.org/licenses/by/3.0/.

Copyright © 2009, Charles Severance.

You assume all responsibility for use and potential liability associated with any use of the material. Material contains copyrighted content, used in accordance with U.S. law. Copyright holders of content included in this material should contact open.michigan@umich.edu with any questions, corrections, or clarifications regarding the use of content. The Regents of the University of Michigan do not license the use of third party content posted to this site unless such a license is specifically granted in connection with particular content. Users of content are responsible for their compliance with applicable law. Mention of specific products in this material solely represents the opinion of the speaker and does not represent an endorsement by the University of Michigan. For more information about how to cite these materials visit http://michigan.educommons.net/about/terms-of-use.

Any medical information in this material is intended to inform and educate and is not a tool for self-diagnosis or a replacement for medical evaluation, advice, diagnosis or treatment by a healthcare professional. You should speak to your physician or make an appointment to be seen if you have questions or concerns about this information or your medical condition. Viewer discretion is advised: Material may contain medical images that may be disturbing to some viewers.

UNIVERSITY OF MICHIGAN

CC BY

UNIVERSITY OF MICHIGAN SCHOOL OF INFORMATION

SI 502: Networked Computing: Storage, Communication, and Processing

Assignment 8 – Programming Python

Due Date: Friday March 27, 2009 at 11:50PM

Overview

You are to produce a program which will retrieve web pages and (a) display the contents of the page, (b) count the number of lines in the page, and (c) count the number of links in the page (anchor tags) as shown below:

python netprog.py

```
Enter Command: dump http://www.dr-chuck.com/page1.htm
Retrieving: http://www.dr-chuck.com/page1.htm
<h1>The First Page</h1>
<p>
If you like, you can switch to the
<a href="http://www.dr-chuck.com/page2.htm">
Second Page</a>.
</p>

Enter Command: links http://www.dr-chuck.com/page1.htm
Retrieving: http://www.dr-chuck.com/page1.htm
Found 1 lines with links
Enter Command: lines http://www.dr-chuck.com/page1.htm
Retrieving: http://www.dr-chuck.com/page1.htm
Found 7 lines
Enter Command: links http://www.umich.edu/
Retrieving: http://www.umich.edu/
Found 190 lines with links
Enter Command: lines http://www.umich.edu/
Retrieving: http://www.umich.edu/
Found 355 lines
Enter Command: links
Please enter 'links' or 'lines' followed by a URL
Enter a Command:
```

You will be given a number of working code samples from the lectures. These can be used as sample code to get you started. You should no do any socket programming for this assignment (i.e. don't start with browser.py, browser2.py, http.py sock1.py, or sock2.py).

You should start by making a copy of **links2.py** and then make changes to make to links2.py to get to this assignment.

Instead of reading in a URL and opening it, read in the command line and then split the command line into words. The first work (i.e. [0]) is the command and the second word is the url. All of the commands need to retrieve the data for the URL using returl and splitting it into lines.

Once the data is retrieved you can make a series of checks to see if the command is "dump", "lines", or links. Depending on the command, you can do a different loop through the retrieved lines to either print the lines out, count the lines, or search for lines that have href's in them.

You should recover from several error conditions:

$ **python netprog.py**
Enter Command: **zap**
Please enter 'links' or 'lines' followed by a URL
Enter Command: **zap http://www.dr-chuck.com/**
Retrieving: http://www.dr-chuck.com/
Please enter 'dump', 'links' or 'lines' followed by a URL
Enter Command: **links http://www.98sd09ds09.com/**
Retrieving: http://www.98sd09ds09.com/
Could not open document: http://www.98sd09ds09.com/
Enter Command:

Turn in

You should turn in a screen shot of your program demonstrating the basic features working and the source code to your program in CTools.