

Assignment One

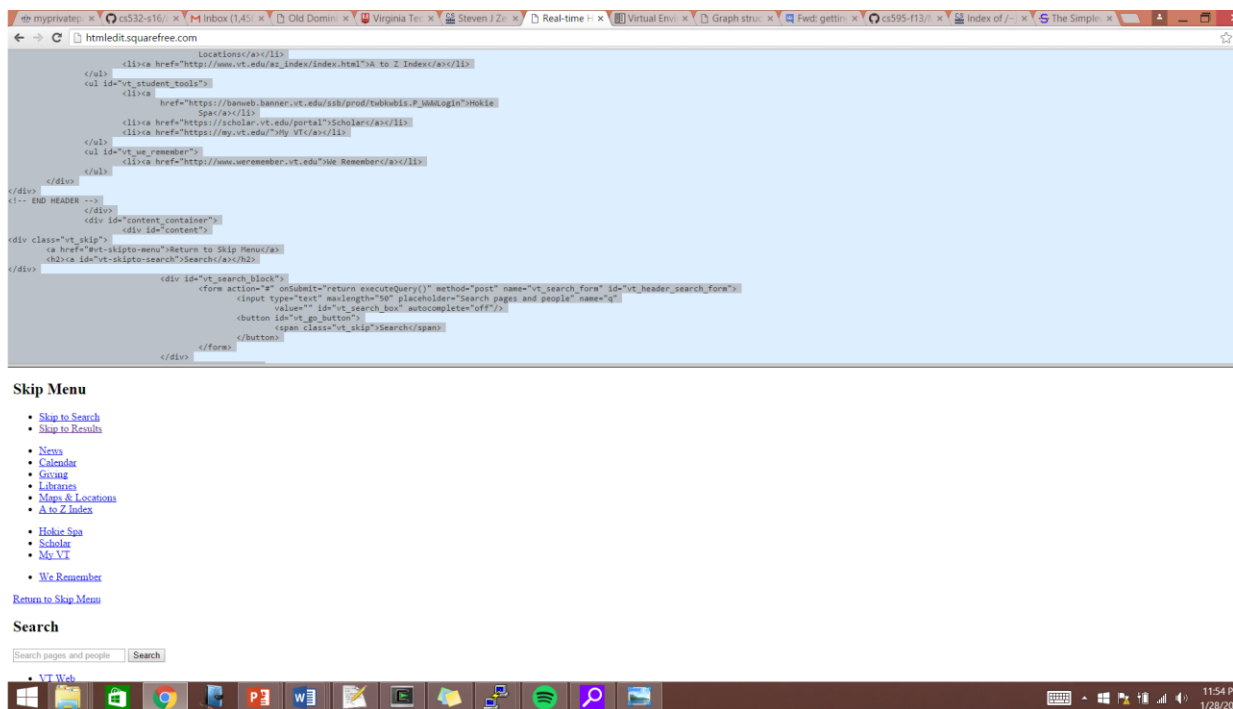
Jacob Berlin

CS432 – Spring 2016

Question One:

Demonstrate that you know how to use "curl" well enough to correctly POST data to a form.

Show that the HTML response that is returned is "correct". That is, the server should take the arguments you POSTed and build a response accordingly. Save the HTML response to a file and then view that file in a browser and take a screen shot.



Question Two:

Write a Python program that

1. takes as a command line argument a web page
2. extracts all the links from the page
3. lists all the links that result in PDF files, and prints out the bytes for each of the links. (note: be sure to follow all the redirects until the link terminates with a "200 OK".)
4. show that the program works on 3 different URIs, one of which needs to be:

<http://www.cs.odu.edu/~mln/teaching/cs532-s16/test/pdfs.html>

```
$ python ./assignment_one.py http://www.cs.odu.edu/~mln/teaching/cs532-s16/test/pdfs.html
This is the name of the script: ./assignment_one.py
This is the name of the URI to check: http://www.cs.odu.edu/~mln/teaching/cs532-s16/test/pdfs.html
200
Server: nginx
Date: Fri, 29 Jan 2016 03:26:50 GMT
Content-Type: text/html
Transfer-Encoding: chunked
Connection: close
Vary: Accept-Encoding
*****
Now I am going to post all of the linked pdf files!
*****
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/ht-2015/hypertext-2015-temporal-violations.pdf
Content-Type: application/pdf
Number of bytes in file: 218406
Response code: 200
*****
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/tpd1-2015/tpd1-2015-annotations.pdf
Content-Type: application/pdf
Number of bytes in file: 622981
Response code: 200
*****
*****PDF FILE*****
http://arxiv.org/pdf/1512.06195
Content-Type: application/pdf
Number of bytes in file: 1748961
Response code: 200
*****
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/tpd1-2015/tpd1-2015-off-topic.pdf
Content-Type: application/pdf
Number of bytes in file: 4308768
Response code: 200
*****
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/tpd1-2015/tpd1-2015-stories.pdf
Content-Type: application/pdf
Number of bytes in file: 1274604
Response code: 200
*****
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/tpd1-2015/tpd1-2015-profiling.pdf
Content-Type: application/pdf
Number of bytes in file: 639001
```

For the first file included I used the base test case. As shown, only the pdf files were chosen along with the formatting required for each specific value.

4 | Assignment One

```
$ python ./assignment_one.py http://www.cs.odu.edu/~mln/teaching/
This is the name of the script: ./assignment_one.py
This is the name of the URI to check: http://www.cs.odu.edu/~mln/teaching/
200
Server: nginx
Date: Fri, 29 Jan 2016 04:43:34 GMT
Content-Type: text/html
Transfer-Encoding: chunked
Connection: close
Vary: Accept-Encoding
*****
Now I am going to post all of the linked pdf files!
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/phd/salaheldeen-phd-dissertation.pdf
Content-Type:
application/pdf
Number of bytes in file:
10875917
Response code:
200
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/phd/alsum-phd-dissertation.pdf
Content-Type:
application/pdf
Number of bytes in file:
23475427
Response code:
200
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/phd/cartledge-phd-dissertation.pdf
Content-Type:
application/pdf
Number of bytes in file:
20240922
Response code:
200
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/phd/klein-phd-dissertation.pdf
Content-Type:
application/pdf
Number of bytes in file:
7873533
Response code:
200
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/phd/smith-phd-dissertation.pdf
Content-Type:
application/pdf
Number of bytes in file:
4018713
Response code:
200
*****PDF FILE*****
http://www.harding.edu/facconn/pubs/lazy-preservation-dissertation.pdf
Content-Type:
application/pdf
Number of bytes in file:
4301157
Response code:
200
*****PDF FILE*****
http://www.cs.odu.edu/~mln/pubs/ms/jones-ms-2015.pdf
```

For my second test case, I used a more robust page. The regular Michael Nelson ODU page sufficed for this.

```
$ python ./assignment_one.py http://www.cs.odu.edu/~tkennedy
This is the name of the script: ./assignment_one.py
This is the name of the URI to check: http://www.cs.odu.edu/~tkennedy
200
Server: nginx
Date: Fri, 29 Jan 2016 04:45:45 GMT
Content-Type: text/html
Transfer-Encoding: chunked
Connection: close
Vary: Accept-Encoding
*****
Now I am going to post all of the linked pdf files!
*****PDF FILE*****
(venv)
```

For my last case, I used Thomas Kennedy's CS page on the website. I chose this case because there were no PDF links on that page, allowing the user to see that there will not always be PDF's on the page.

Question Three:

Consider the "bow-tie" graph in the Broder et al. paper (fig 9):

<http://www9.org/w9cdrom/160/160.html>

Now consider the following graph:

A --> B

B --> C

C --> D

C --> A

C --> G

E --> F

G --> C

G --> H

I --> H

I --> J

I --> K

J --> D

L --> D

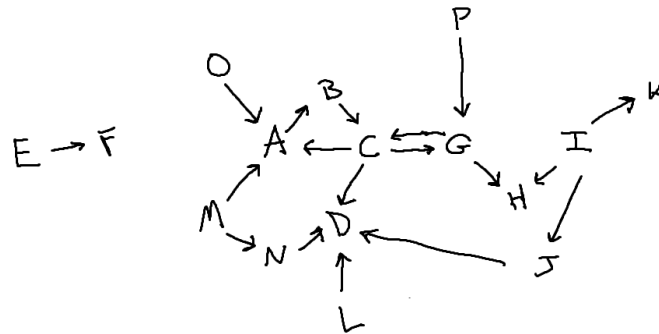
M --> A

M --> N

N --> D

O --> A

P --> G



IN: I, M

SCC: G, C, A, B

Out: D, H

Tendrils: K, P, O, L

Tubes: N, J

Disconnected: E, F

