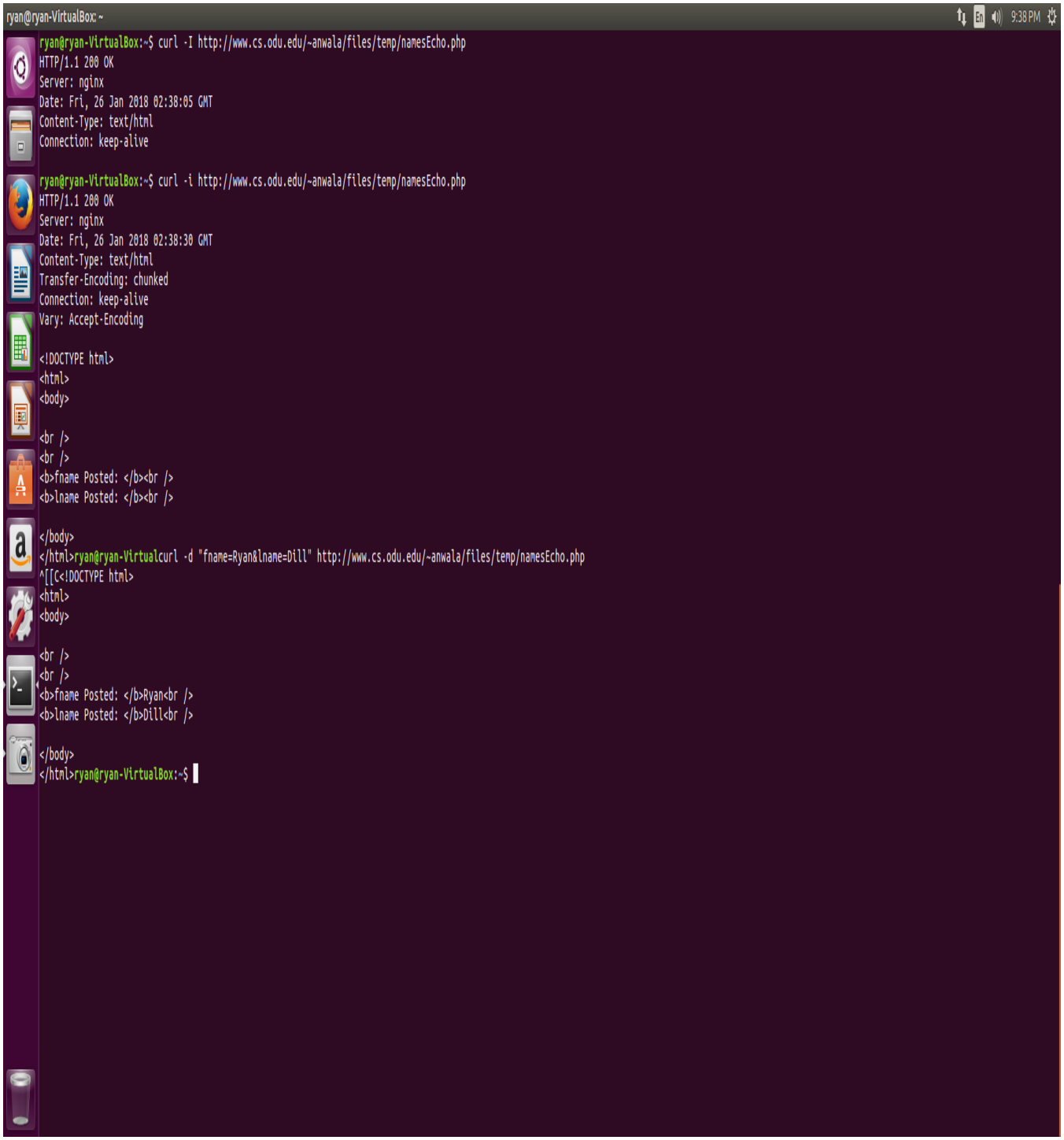
Assignment 1

Ryan Dill
CS 432
January 25, 2018

Question 1: 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.



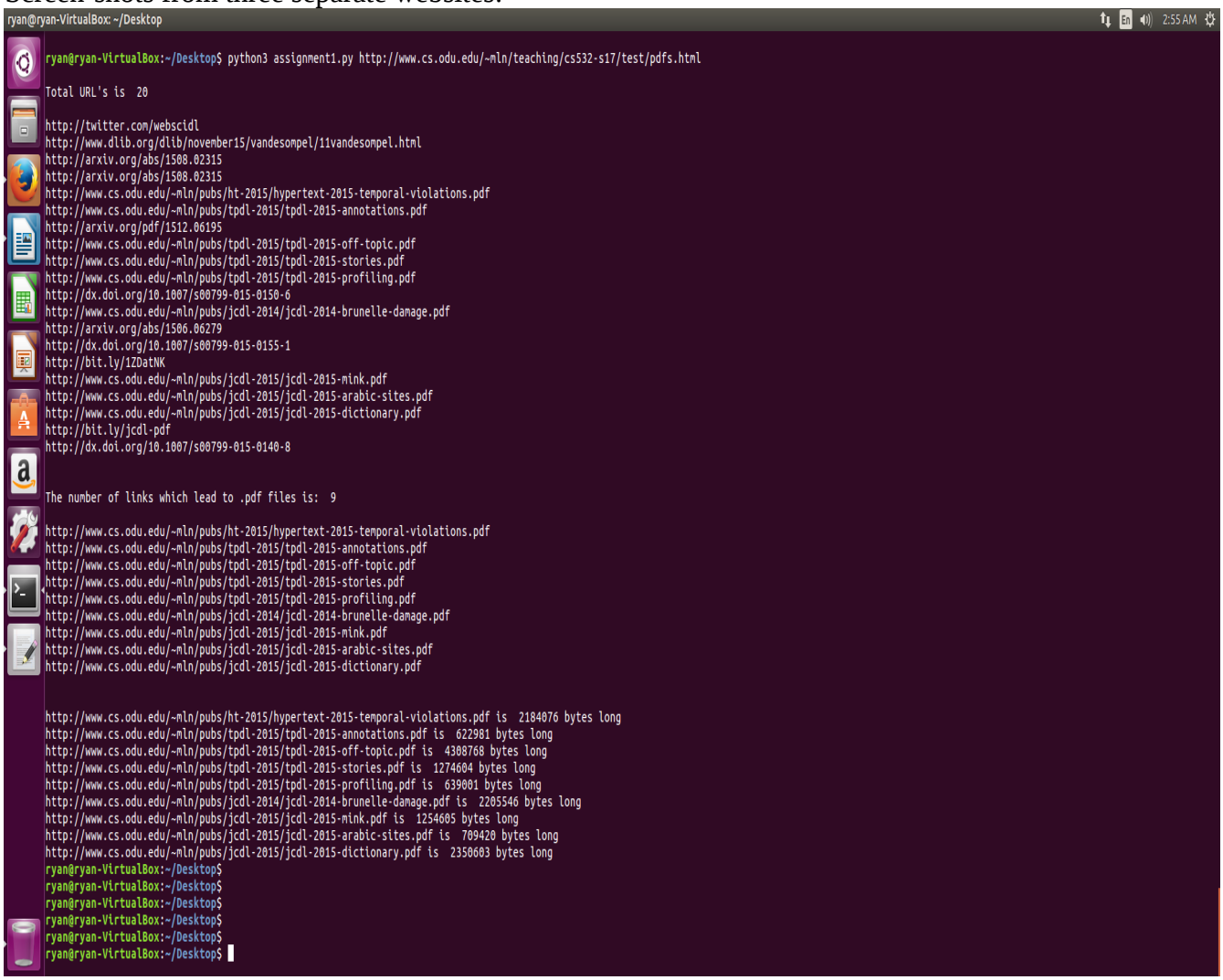
Question 2: Write a Python program that:

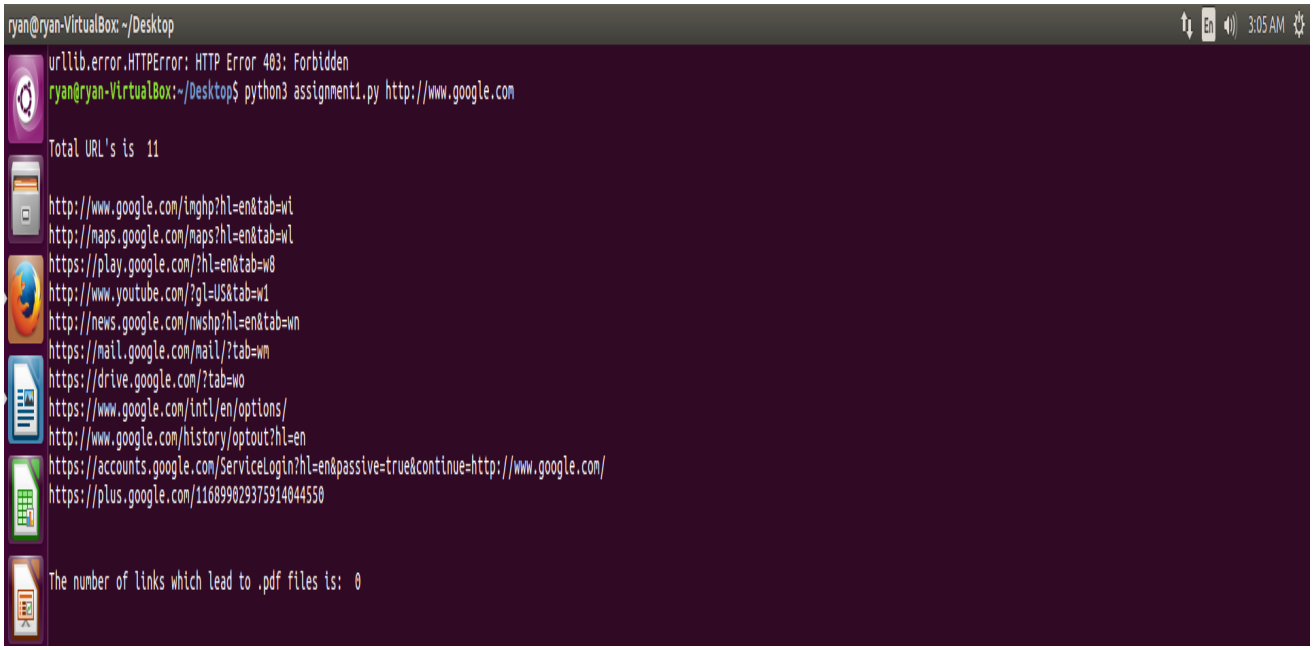
1. takes as a command line argument a web page
2. extracts all the links from the page
3. lists all the links that result in PDF files, and prints out the bytes for each of the links. (note: be sure to follow all the redirects until the link terminates with a "200 OK".)
4. show that the program works on 3 different URIs, one of which needs to be: http://www.cs.odu.edu/~mln/teaching/cs532-s17/test/pdfs.html

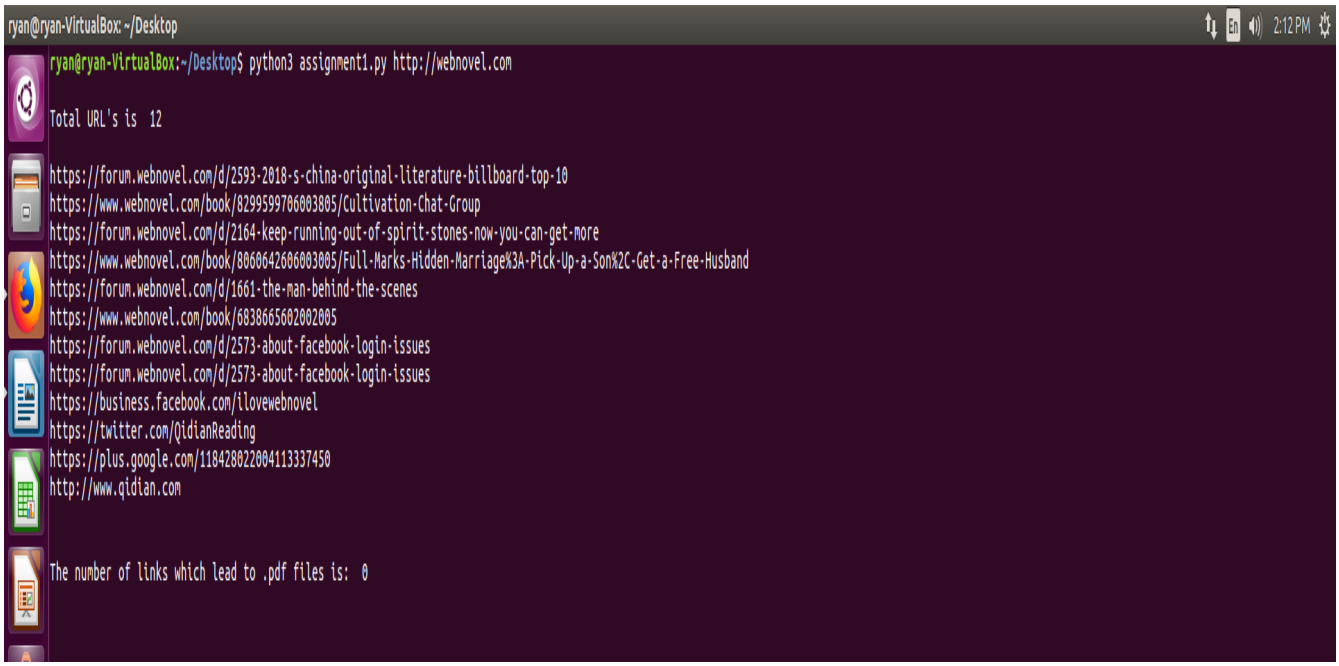
The code used to extract links and find .pdf files

```
import sys
from bs4 import BeautifulSoup
import urllib.request
import requests
import re
if len(sys.argv) == 2:
        url =sys.argv[1]

else:
        print("Not enough arguments. Please retry.")
        sys.exit()

website = urllib.request.urlopen(url)

beauty = BeautifulSoup(website, "html.parser")

link_array = []

for link in beauty.findAll('a', attrs={'href': re.compile("^http")}):
        link_array.append(link.get('href'))

print("\nTotal URL's is ", len(link_array),"\n")
print(*link_array,sep='\n')

print("\n")

pdf_array = [s for s in link_array if ".pdf" in s]
print("The number of links which lead to .pdf files is: ",len(pdf_array),"\n")

print(*pdf_array,sep='\n')

print("\n")

for index, newurl in enumerate(pdf_array):
        response = requests.get(newurl)
        print(newurl, "is ", response.headers['content-length'], "bytes long")
```

Screen-shots from three separate websites:







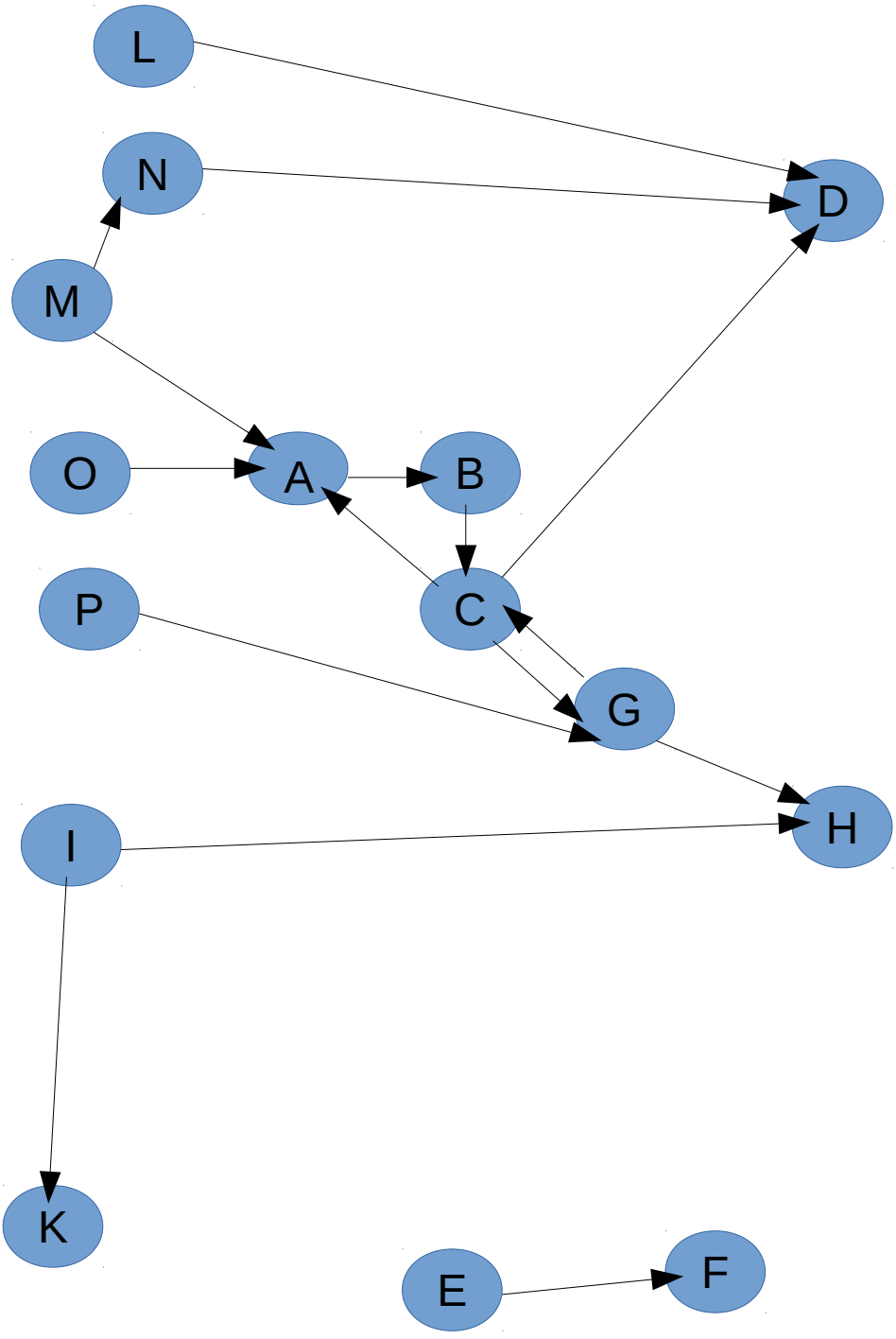
3. Consider the "bow-tie" graph in the Broder et al. paper (fig 9):
   `http://www9.org/w9cdrom/160/160.html`

   Now consider the following graph:

   A --> B
   B --> C
   C --> D
   C --> A
   C --> G
   E --> F

```
G --> C
G --> H
I --> H
I --> K
L --> D
M --> A
M --> N
N --> D
O --> A
P --> G

For the above graph, give the values for:

IN:
SCC:
OUT:
Tendrils:
Tubes:
Disconnected:
```

IN: I, P, O, M, N, L
SCC: A, B, C, G
OUT: H, D
TENDRILS: I → K
TUBES: L → D
DISCONNECTED: E → F


L
N
D
M
O
A
B
P
C
G
H
I
K
E
F