Assignment 1

Tarek Fouda

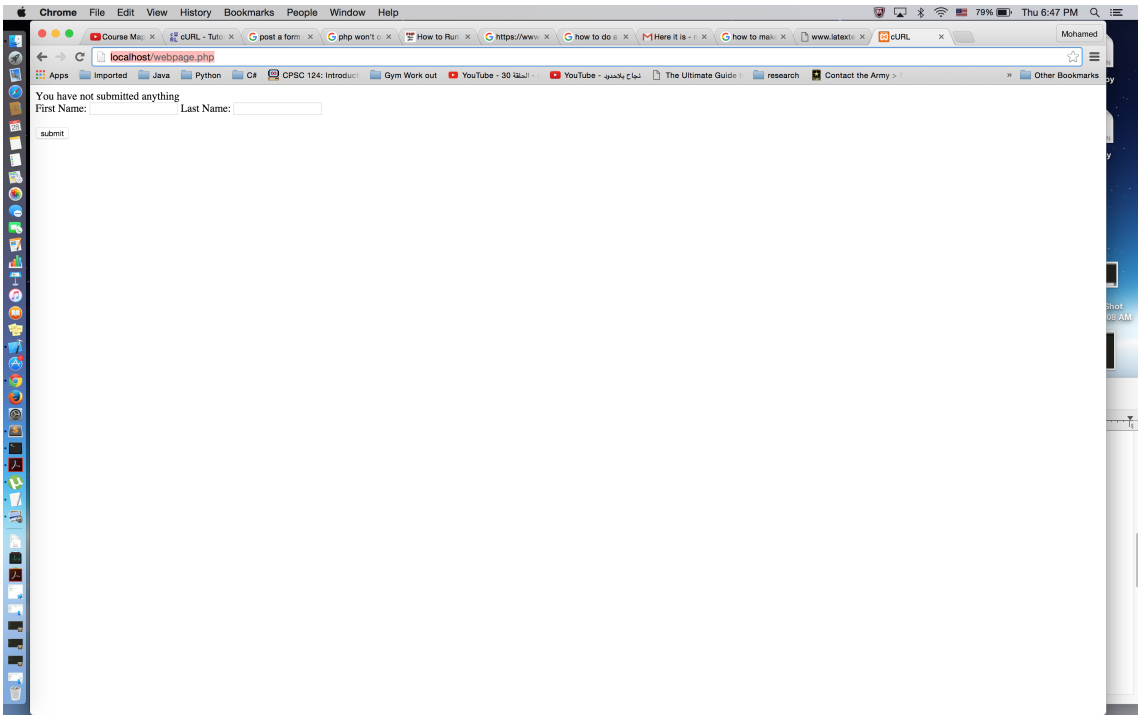
2016-01-27

1 Introduction

This report mainly discusses my approach and how I implemented and solved each of the three problems assigned to me. I will be discussing every problem in a different section. This assignment is due Thursday 01/28/2016.

2 Problem 1- cURL

In this part of the assignment, I was required to demonstrate how I was able to use cURL to POST data to a form. I simply created a .php which has a form whose method = "POST". It basically has two textboxes, The first text box takes your first name as input, and the second textbox takes your last name as input as well. A submit button to send the Data to this form.



ScreenShot 1 shows the webpage

Then we use the curl command to pass the data to this specific form using the localhost/webpage.php (Knowing that webpage.php is the .php file I created.

The cURL command is circled in red in the following screenshot represents the response we get after sending the parameters!

```
Mohameds-MacBook-Pro:desktop mohamedshaaban$ curl -d "firstname=Tarek&lastname=Fouda&submit=submit" http://localhost/webpage.php
<html>
<head>
<meta http-equiv="Content-Type" content = "text/html: charset=utf-8" />
<title>
cURL
</title>
</head>
<body>
<br />Your first name isTarek<br /> Your last name isFouda<form method ="POST" action="webpage.php">
<label> First Name: </label>
<input name="firstname" type="text" />
<label> Last Name: </label>
<input name="lastname" type="text" />
<br/>
<br/>
<input name="submit" type="submit" value="submit"/>
</form>
</body>
</html>Mohameds-MacBook-Pro:desktop mohamedshaaban$
```

```
<?php
if (isset($_POST['submit'])){
        $results = '<br />Your first name is '.$_POST['firstname'];
        $results .= '<br /> Your last name is '.$_POST['lastname'];
} else{
        $results='You have not submitted anything';
}
?>
<html>
<head>
<meta http-equiv="Content-Type" content = "text/html: charset=utf-8" />
<title>
cURL
</title>
</head>
<body>
<?php echo $results; ?>
<form method ="POST" action="webpage.php">
<label> First Name: </label>
<input name="firstname" type="text" />
<label> Last Name: </label>
<input name="lastname" type="text" />
<br/>
<br/>
<input name="submit" type="submit" value="submit"/>
</form>
</body>
</html>
```

3 Problem number 2 - Python

In this problem, it was required to implement a python program that takes a web page such as http://www.cs.odu.edu/ mln/teaching/cs532-s16/test/pdfs.html and print out all the linksand urls in this web page that are PDFs only! Iimplemented a python code that takes this webpage from a user and lists all the links that exist in this specific website, but I misunderstood the requirement of this problem and instead of listing all the links that are PDFs, I listed all the URLs and then imported the FPDF library in python to write all the links in a PDF.

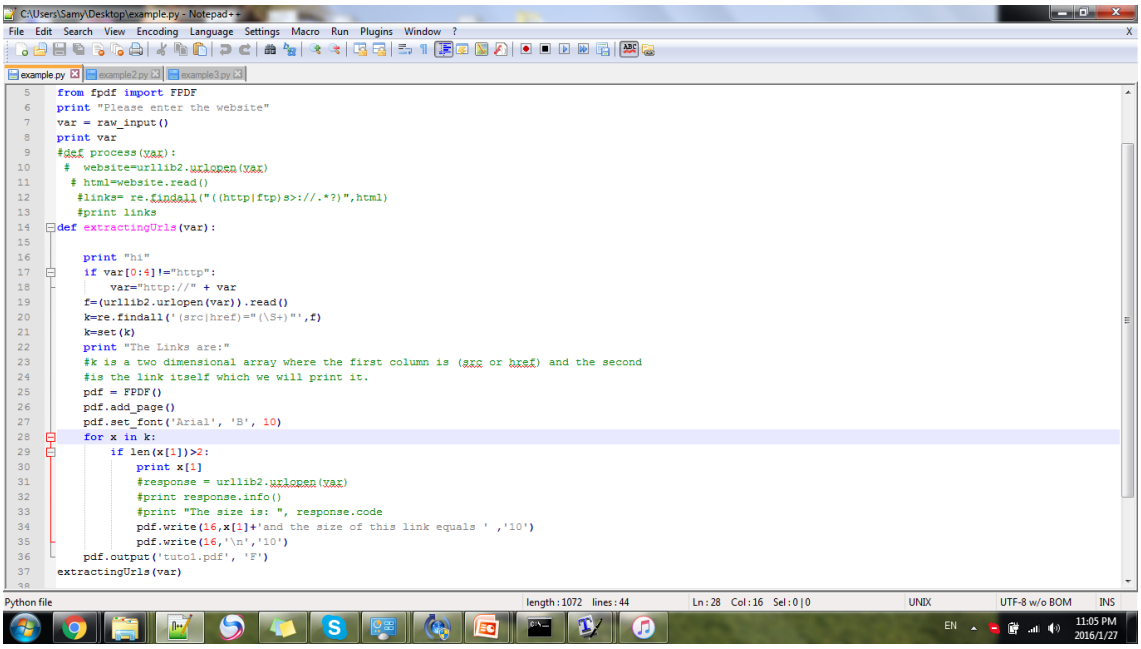


Figure 1 shows the python code to insert links from webpage to a pdf

Keeping in mind that the required is to just print out the links that are PDFs only, so it will not differ much from this code. Basically the above code shown in Figure 1 takes the website from the user and open it using the following lines.

```
print "Please enter the website"
var = raw_input()
```

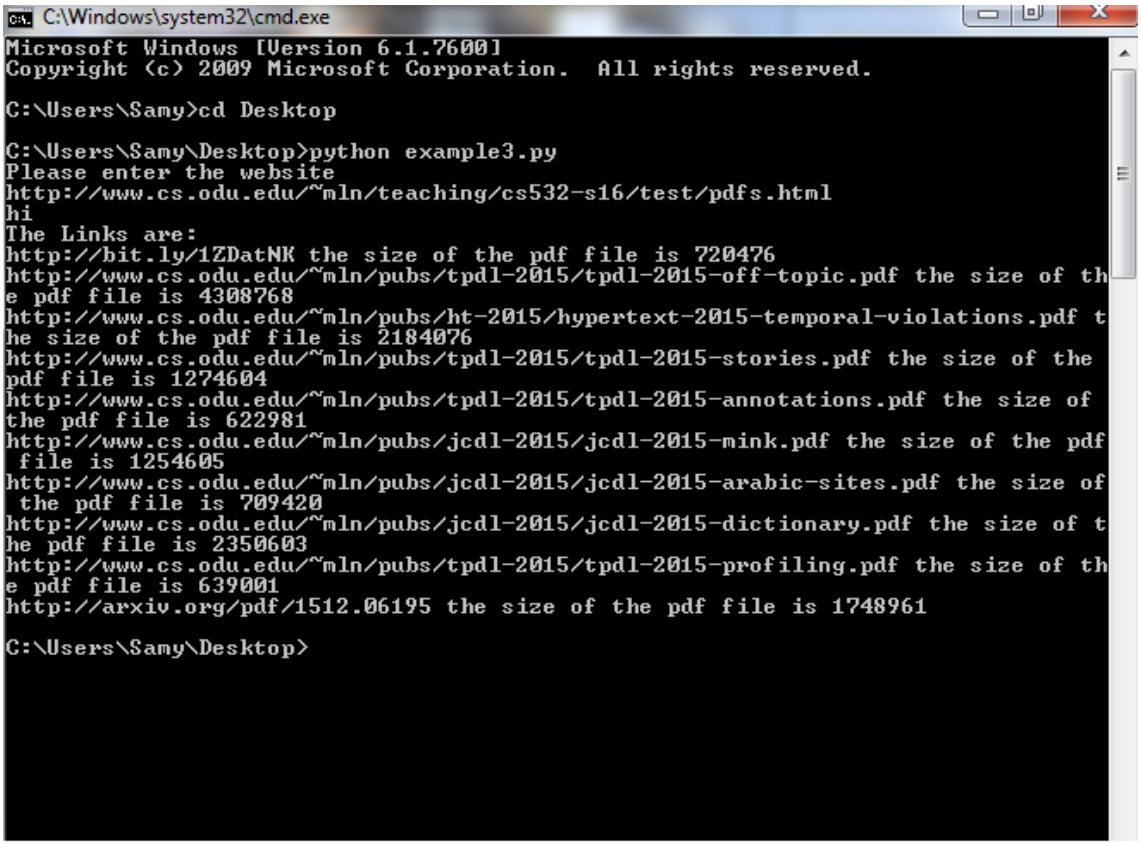
where var now contains the website. By implementing a function called extractingUrls which uses the var as a parameter, we make sure that the first 4 characters of var is http otherwise we will have to append http:// to the url itself. why? because the urlopen(url) function requires the url to be written starting with http://. and this is described in the next 3 lines.

```
    if var[0:4]!="http":
        var="http://" + var
    f=(urllib2.urlopen(var)).read()
```

after that we start looking for all urls that exist in this webpage using findall() function. This way we will have the URLs in a two dimensional array where the link is in the array[1]. so we need to loop on this list to find all the URLs and check a condition, whether opening this specific link will result in a HTTP Content-Type= application/pdf or not. Ofcourse we are only interested in the PDFs.

```
k=re.findall('(src|href)="(\S+)"',f)
    k=set(k)
    print "The Links are:"
    #k is a two dimensional array where the first column
    #is (src or href) and the second
    #is the link itself which we will print it.
    pdf = FPDF()
    pdf.add_page()
    pdf.set_font('Arial', 'B', 10)
    for x in k:
        if len(x[1])>2:
            response = urllib2.urlopen(x[1])
            if response.info()["Content-Type"] ==
            'application/pdf':
                print x[1]+" the size of the pdf file is "
            + response.info()["Content-Length"]
```

As shown in this piece of code, we are only interested in the PDF links so we will get their size from response.info()["Content-Length"] command.

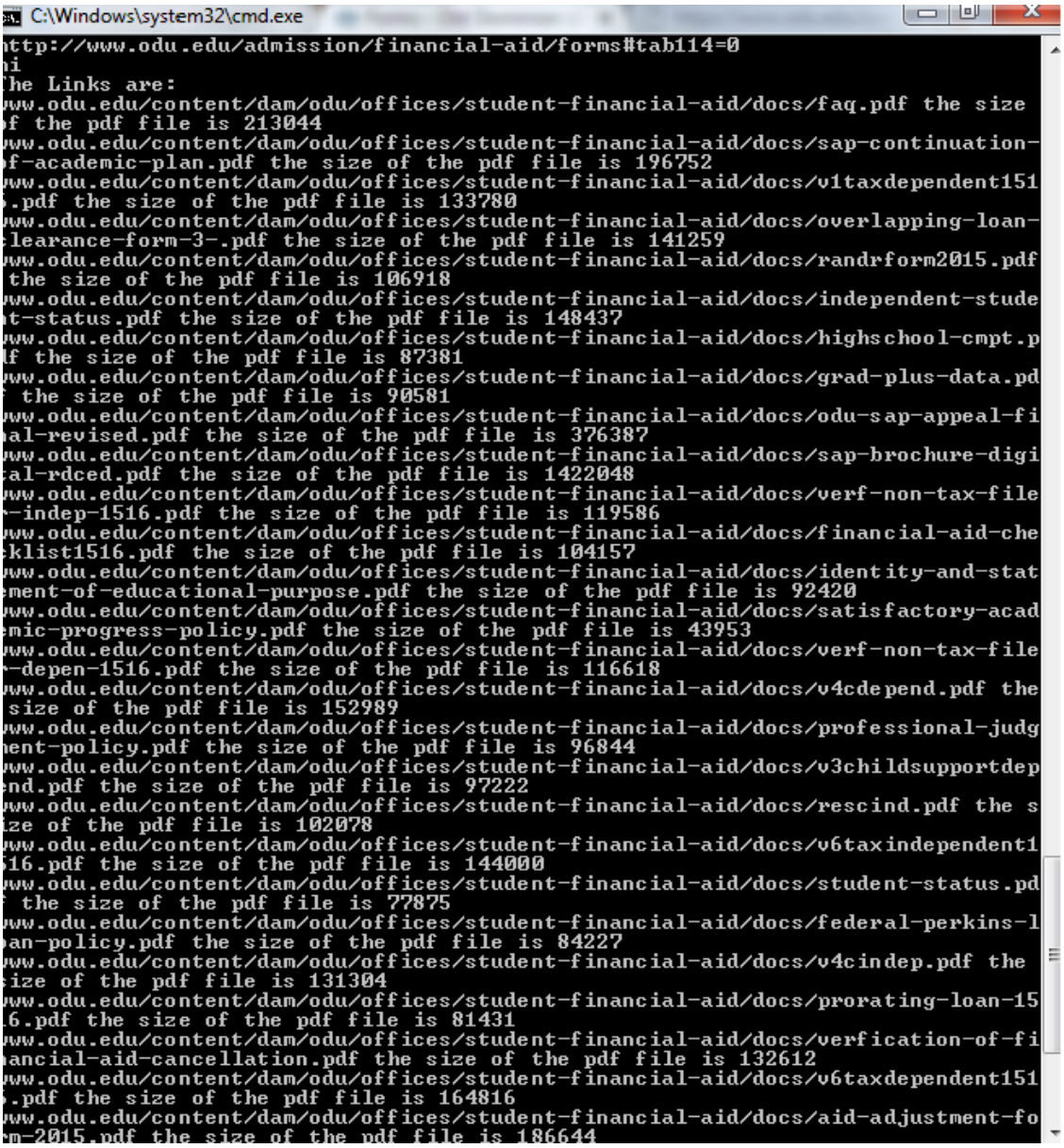


```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7600]
Copyright (c) 2009 Microsoft Corporation.  All rights reserved.

C:\Users\Samy>cd Desktop

C:\Users\Samy\Desktop>python example3.py
Please enter the website
http://www.cs.odu.edu/~mln/teaching/cs532-s16/test/pdfs.html
hi
The Links are:
http://bit.ly/1ZDatNK the size of the pdf file is 720476
http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-off-topic.pdf the size of th
e pdf file is 4308768
http://www.cs.odu.edu/~mln/pubs/ht-2015/hypertext-2015-temporal-violations.pdf t
he size of the pdf file is 2184076
http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-stories.pdf the size of the
pdf file is 1274604
http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-annotations.pdf the size of
the pdf file is 622981
http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.pdf the size of the pdf
 file is 1254605
http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic-sites.pdf the size of
 the pdf file is 709420
http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-dictionary.pdf the size of t
he pdf file is 2350603
http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-profiling.pdf the size of th
e pdf file is 639001
http://arxiv.org/pdf/1512.06195 the size of the pdf file is 1748961

C:\Users\Samy\Desktop>
```

Figure 2 shows the result on entering http://www.cs.odu.edu/ mln/teaching/cs532-s16/test/pdfs.html as a website

```
C:\Users\Samy\Desktop>python example3.py
Please enter the website
www.cs.odu.edu
hi
The Links are:
http://www.cs.odu.edu/StrategicPlan0515_2010.pdf the size of the pdf file is 909
323
http://www.cs.odu.edu/files/csdeptresearch.pdf the size of the pdf file is 26339
84
http://www.cs.odu.edu/files/cs_systems_services.pdf the size of the pdf file is
412031
http://www.cs.odu.edu/files/cs_systems_it_infrastructure_2012.pdf the size of th
e pdf file is 2049957
http://www.cs.odu.edu/studentappointmentinfo.pdf the size of the pdf file is 636
560

C:\Users\Samy\Desktop>
```

Figure 3 shows the result on entering http://www.cs.odu.edu, still searching for the pdfs is a way to handle exception due to errors in opening some urls!



```
C:\Windows\system32\cmd.exe
ttp://www.odu.edu/admission/financial-aid/forms#tab114=0
i
he Links are:
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/faq.pdf the size
f the pdf file is 213044
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/sap-continuation-
f-academic-plan.pdf the size of the pdf file is 196752
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/v1taxdependent151
.pdf the size of the pdf file is 133780
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/overlapping-loan-
learance-form-3-.pdf the size of the pdf file is 141259
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/randrform2015.pdf
the size of the pdf file is 106918
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/independent-stude
t-status.pdf the size of the pdf file is 148437
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/highschool-cmpt.p
f the size of the pdf file is 87381
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/grad-plus-data.pd
the size of the pdf file is 90581
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/odu-sap-appeal-fi
al-revised.pdf the size of the pdf file is 376387
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/sap-brochure-digi
al-rdced.pdf the size of the pdf file is 1422048
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/verf-non-tax-file
-indep-1516.pdf the size of the pdf file is 119586
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/financial-aid-che
klist1516.pdf the size of the pdf file is 104157
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/identity-and-stat
ment-of-educational-purpose.pdf the size of the pdf file is 92420
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/satisfactory-acad
mic-progress-policy.pdf the size of the pdf file is 43953
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/verf-non-tax-file
-depen-1516.pdf the size of the pdf file is 116618
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/v4cdepend.pdf the
size of the pdf file is 152989
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/professional-judg
ent-policy.pdf the size of the pdf file is 96844
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/v3childsupportdep
nd.pdf the size of the pdf file is 97222
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/rescind.pdf the s
ize of the pdf file is 102078
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/v6taxindependent1
16.pdf the size of the pdf file is 144000
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/student-status.pd
the size of the pdf file is 77875
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/federal-perkins-l
an-policy.pdf the size of the pdf file is 84227
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/v4cindep.pdf the
ize of the pdf file is 131304
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/prorating-loan-15
6.pdf the size of the pdf file is 81431
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/verfication-of-fi
ancial-aid-cancellation.pdf the size of the pdf file is 132612
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/v6taxdependent151
.pdf the size of the pdf file is 164816
ww.odu.edu/content/dam/odu/offices/student-financial-aid/docs/aid-adjustment-fo
m-2015.pdf the size of the pdf file is 186644
```

Figure 4 shows the result on entering http://www.odu.edu/admission/financial-aid/formstab114=0

The following is the python code for such a program

```
-- This program takes a website as a command line argument
-- and extract all the links that are PDFs and show each size.
import urllib2
import BeautifulSoup
import sys
import re
from fpdf import FPDF
print "Please enter the website"
var = raw_input()
def extractingUrls(var):

    print "hi"
    if var[0:4]!="http":
        var="http://" + var
    f=(urllib2.urlopen(var)).read()
    k=re.findall('(src|href)="(\S+)"',f)
    k=set(k)
    print "The Links are:"
    #k is a two dimensional array where the first column
   # is (src or href) and the second
    #is the link itself which we will print it.
    pdf = FPDF()
    pdf.add_page()
    pdf.set_font('Arial', 'B', 10)
    for x in k:
        if len(x[1])>2:
            try:
              ah = var[7:]
              --print ah
              ba = ah.partition("/")[0]
              --print ba +"ajaj"
              print x[1]
              if x[1][0:4]=="http":
                response = urllib2.urlopen(x[1])
                if response.info()["Content-Type"] ==
               'application/pdf':
                  print x[1]+" the size of the pdf file is "
               + response.info()["Content-Length"]
              elif x[1][-4:] ==".pdf" and x[1][0:4]!="http":
                if x[1][0:1] == "/":
                  --print ba + x[1]
                  response = urllib2.urlopen("http://" +
                  ba+x[1])
```

```
                    if response.info()["Content-Type"]
                  == 'application/pdf':
                      print ba + x[1]+
                  " the size of the pdf file is "+
                    response.info()["Content-Length"]
                  else:
                    --print ba + x[1]
                    response = urllib2.urlopen("http://" +
                    ba+'/'+x[1])
                    if response.info()["Content-Type"] ==
                  'application/pdf':
                      print ba +'/'+ x[1]+
                    " the size of the pdf file is "+
                    response.info()["Content-Length"]
            except:
              pass
                        --print response.info()
            --print "The size is: ", response.code
extractingUrls(var)
```

4 Problem 3- Bow-Tie

In this problem, It is requiredtodraw the edges and specify the SCC which is the strong component cycle that you can reach any node in the SCC from another node in the SCC as well. The IN nodes are the nodes which you can go to the SCC but none of the nodes in the SCC can reach the IN nodes. The OUT nodes are the nodes which are reachable from the SCC but none of the OUT nodes can reach any of the SCC. Tendrils are the nodes which go to the IN or OUT, also tubes are the nodes that connect IN and OUT without reaching the SCC. Based on the Graph I drew from the givens, I figured out the Solution as shown in the next image.

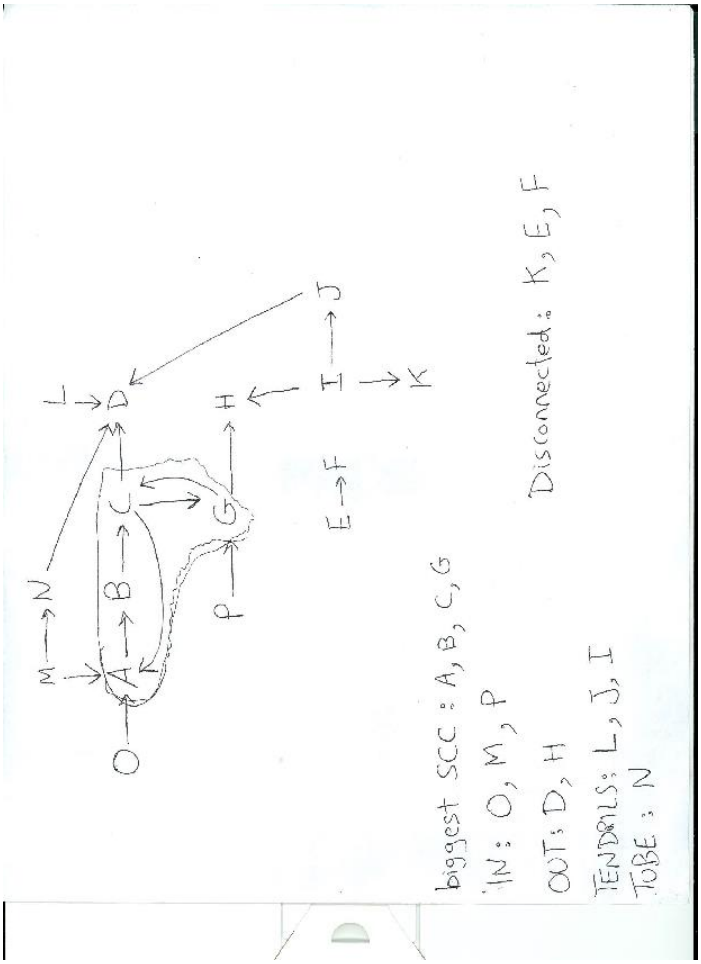


Figure 5 shows the solution