

HW1 - Web Science Intro

Logan Bartels
September 20, 2020

Q1

Answer

- IN: E, F, M
- SCC: A, B, C, D, G, P
- OUT: H, L
- Tendrils: N (can reach OUT), O (reachable from IN)
- Tubes: J (serves as the tube connecting IN tendril O to OUT tendril N)
- Disconnected: I, K

Discussion

Figure 1 shows two directed graphs to answer question 1. The top graph can be ignored, as I just drew it as I was reading the links in the list. The second graph is a true directed graph, and represents my answer to the question. The list of nodes on paper was me just brainstorming what the nodes *could* be. E, F, and M are all IN nodes. M only goes to A, which is part of the SCC. E has an out-link to F, which is another part of IN. A and B are part of the SCC because they connect to each other. C, G, and P are part of the SCC because they can all reach each other; you can endlessly go from C to G to P. D is part of the SCC because it connects back to A. H is part of out because it only has an in-link from D, and an out-link to L. L has no out-links, so it is part of OUT. N is a tendril with an out link to L. O is a tendril with an in-link from E. J is the tube that connects N and O. I and K are disconnected because K only links out to I, and K has no in-links at all. I also has no out-links, so both I and K are disconnected.

Q2

Part A

Answer

```
1 lbartels@aspidiske:~/cs432-f20$ curl -i -L -A CS432 http://www.cs.odu.  
  edu/~mweigle/courses/cs532/ua_echo.php  
2 HTTP/1.1 301 Moved Permanently  
3 Server: nginx  
4 Date: Fri, 18 Sep 2020 17:42:18 GMT  
5 Content-Type: text/html  
6 Content-Length: 178  
7 Connection: keep-alive  
8 Location: https://www.cs.odu.edu/~mweigle/courses/cs532/ua_echo.php  
9  
10 HTTP/1.1 200 OK  
11 Server: nginx  
12 Date: Fri, 18 Sep 2020 17:42:18 GMT  
13 Content-Type: text/html; charset=UTF-8  
14 Content-Length: 110  
15 Connection: keep-alive  
16 Vary: Accept-Encoding  
17 X-Powered-By: PHP/5.6.40  
18  
19 <!DOCTYPE html>  
20 <html>  
21 <body>  
22  
23 <br/>USER AGENT ECHO  
24 <br/><br/>  
25 <b>User-Agent:</b> CS432<br/>  
26  
27 </body>  
28 </html>  
29 lbartels@aspidiske:~/cs432-f20$
```

Listing 1: Command and output copy-pasted from the terminal.

Discussion

In order to find the correct command to execute, I had to go to the curl manpage and comb it for the parameters I needed. I knew to include the “-i” to get the HTTP response headers; however, I had to locate the parameters needed to follow redirects and change the User-Agent HTTP request field. Those parameters were “-L” and “-A,” respectively. Since the file changed location, I got a 301 response code. Curl then followed the URI listed next to the “Location.” After curl went to the new URI, a 200 response code, meaning that it was successful. Once the file was found, the file’s content was output to the terminal. The User Agent was also changed to “CS432” because that parameter was specified at the command line with the “-A” parameter.

Part B

Answer

```

1 lbartels@aspidiske:~/secure_html$ curl -L -A CS432 http://www.cs.odu.
  edu/~mweigle/courses/cs532/ua_echo.php -o output.html
2  % Total    % Received % Xferd  Average Speed   Time    Time     Time
   Current
3                                  Dload  Upload  Total  Spent  Left
   Speed
4  0      0    0     0    0      0     0    0  --:--:-- --:--100    178
   100    178    0     0  35600    0  --:--:-- --:--:-- 35600
5 100    110  100    110    0     0   3793    0  --:--:-- --:--:--
   --:--:-- 3793

```

Listing 2: Terminal output after executing command for part b.

Discussion

I'll admit, I don't really know what to make of the terminal output, but obtaining the result here was simple. For the command, take out the "-i" to not include the response headers. Then add an "-o" at the end and a filename to specify the output file. The User Agent is "CS432" in the output file because that is what the command in part A specified. The command changed the User Agent. Figure 2 shows the screenshot of the page I opened in Firefox near the end of the document.

Part C

Answer

The answer is Figure 3, located near the end of the document.

Discussion

Figure 3 shows a screenshot the webpage I obtained after clicking the link posted in GitHub. The User Agent is different this time because I did not change it at the command line. I just clicked on the URI listed on GitHub, and the User Agent was taken by my Firefox browser.

Q3

Answer

```
1 from bs4 import BeautifulSoup
2 import requests
3 import sys
4 import re
5
6 # A test to confirm the command line argument was correctly passed in
7 print("URI requested: ", sys.argv[1])
8
9 response = requests.get(sys.argv[1])
10
11 soup = BeautifulSoup(response.text)
12 for links in soup.find_all('a'):
13     resp = requests.get((links.get('href')))
14     if resp.headers['Content-Type'] != "application/pdf":
15         continue
16     else:
17         print('\n')
18         print('URI: ', links.get('href'))
19         print('Final URI: ', resp.url)
20         print('Content Length: ', resp.headers['Content-Length'], '
    bytes')
21         print('\n')
```

Listing 3: My Python program to find PDFs on a webpage.

URI 1 Output

```
1 lbartels@scorpii:~/cs432-f20$ python3 findPDFs.py https://www.cs.odu.
  edu/~mweigle/courses/cs532/pdfs.html
2 URI requested:  https://www.cs.odu.edu/~mweigle/courses/cs532/pdfs.html
3 findPDFs.py:11: GessedAtParserWarning: No parser was explicitly
  specified, so I'm using the best available HTML parser for this
  system ("html.parser"). This usually isn't a problem, but if you run
  this code on another system, or in a different virtual environment,
  it may use a different parser and behave differently.
4
5 The code that caused this warning is on line 11 of the file findPDFs.py
  . To get rid of this warning, pass the additional argument 'features
  ="html.parser"' to the BeautifulSoup constructor.
6
7 soup = BeautifulSoup(response.text)
8
9
```

```
10 URI: http://www.cs.odu.edu/~mln/pubs/ht-2015/hypertext-2015-temporal-
    violations.pdf
11 Final URI: https://www.cs.odu.edu/~mln/pubs/ht-2015/hypertext-2015-
    temporal-violations.pdf
12 Content Length: 2184076 bytes
13
14
15
16
17 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-annotations.
    pdf
18 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-
    annotations.pdf
19 Content Length: 622981 bytes
20
21
22
23
24 URI: http://arxiv.org/pdf/1512.06195
25 Final URI: https://arxiv.org/pdf/1512.06195.pdf
26 Content Length: 1748959 bytes
27
28
29
30
31 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-off-topic.pdf
32 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-off-
    topic.pdf
33 Content Length: 4308768 bytes
34
35
36
37
38 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-stories.pdf
39 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-
    stories.pdf
40 Content Length: 1274604 bytes
41
42
43
44
45 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-profiling.pdf
46 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-
    profiling.pdf
47 Content Length: 639001 bytes
48
49
```

```
50
51
52 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-brunelle-
      damage.pdf
53 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-
      brunelle-damage.pdf
54 Content Length:  2205546 bytes
55
56
57
58
59 URI:  http://bit.ly/1ZDatNK
60 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-
      temporal-intention.pdf
61 Content Length:  720476 bytes
62
63
64
65
66 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.pdf
67 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.
      pdf
68 Content Length:  1254605 bytes
69
70
71
72
73 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic-sites.
      pdf
74 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic
      -sites.pdf
75 Content Length:  709420 bytes
76
77
78
79
80 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-dictionary.
      pdf
81 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-
      dictionary.pdf
82 Content Length:  2350603 bytes
```

Listing 4: Output after running the findPDFs.py program at the command line with `https://www.cs.odu.edu/~mweigle/courses/cs532/pdfs.html` as the command line parameter.

URI 2 Output

```
1 lbartels@betelgeuse:~/cs432-f20$ python3 findPDFs.py https://www.cs.odu
```

```
.edu/~mweigle/
2 URI requested: https://www.cs.odu.edu/~mweigle/
3 findPDFs.py:11: GessedAtParserWarning: No parser was explicitly
  specified, so I'm using the best available HTML parser for this
  system ("html.parser"). This usually isn't a problem, but if you run
  this code on another system, or in a different virtual environment,
  it may use a different parser and behave differently.
4
5 The code that caused this warning is on line 11 of the file findPDFs.py
  . To get rid of this warning, pass the additional argument 'features
  ="html.parser"' to the BeautifulSoup constructor.
6
7 soup = BeautifulSoup(response.text)
8
9
10 URI: http://www.cs.odu.edu/~mweigle/files/CV.pdf
11 Final URI: https://www.cs.odu.edu/~mweigle/files/CV.pdf
12 Content Length: 115063 bytes
13
14
15
16
17 URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-aturban-
  archivenow.pdf
18 Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-
  aturban-archivenow.pdf
19 Content Length: 3998654 bytes
20
21
22
23
24 URI: http://www.cs.odu.edu/~mkelly/papers/2017_jcdl_countingMementos.
  pdf
25 Final URI: https://www.cs.odu.edu/~mkelly/papers/2017
  _jcdl_countingMementos.pdf
26 Content Length: 274265 bytes
27
28
29
30
31 URI: http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.pdf
32 Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.
  pdf
33 Content Length: 1254605 bytes
34
35
36
```

```
37
38 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic-sites.
      pdf
39 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic
      -sites.pdf
40 Content Length:  709420 bytes
41
42
43
44
45 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-misenet14.pdf
46 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-misenet14
      .pdf
47 Content Length:  1231147 bytes
48
49
50
51
52 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-brunelle-
      damage.pdf
53 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-
      brunelle-damage.pdf
54 Content Length:  2205546 bytes
55
56
57
58
59 URI:  http://www.cs.odu.edu/~mweigle/papers/olariu-misenet13.pdf
60 Final URI:  https://www.cs.odu.edu/~mweigle/papers/olariu-misenet13.pdf
61 Content Length:  405542 bytes
62
63
64
65
66 URI:  http://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_149.pdf
67 Final URI:  https://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_149.pdf
68 Content Length:  813692 bytes
69
70
71
72
73 URI:  http://www.cs.odu.edu/~mweigle/papers/yan-soli09.pdf
74 Final URI:  https://www.cs.odu.edu/~mweigle/papers/yan-soli09.pdf
75 Content Length:  278184 bytes
76
77
78
```



```
79
80 URI:  https://www.cs.odu.edu/~mweigle/files/CV.pdf
81 Final URI:  https://www.cs.odu.edu/~mweigle/files/CV.pdf
82 Content Length:  115063 bytes
```

Listing 5: Output after running the findPDFs.py program at the command line with `https://www.cs.odu.edu/~mweigle/` as the command line parameter.

URI 3 Output

```
1 lbartels@betelgeuse:~/cs432-f20$ python3 findPDFs.py https://www.cs.odu
  .edu/~mweigle/Main/PubsByYear
2 URI requested:  https://www.cs.odu.edu/~mweigle/Main/PubsByYear
3 findPDFs.py:11: GessedAtParserWarning: No parser was explicitly
  specified, so I'm using the best available HTML parser for this
  system ("html.parser"). This usually isn't a problem, but if you run
  this code on another system, or in a different virtual environment,
  it may use a different parser and behave differently.
4
5 The code that caused this warning is on line 11 of the file findPDFs.py
  . To get rid of this warning, pass the additional argument 'features
  ="html.parser"' to the BeautifulSoup constructor.
6
7 soup = BeautifulSoup(response.text)
8
9
10 URI:  http://www.cs.odu.edu/~mweigle/files/CV.pdf
11 Final URI:  https://www.cs.odu.edu/~mweigle/files/CV.pdf
12 Content Length:  115063 bytes
13
14
15
16
17 URI:  https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-jones-
  archiveit.pdf
18 Final URI:  https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-
  jones-archiveit.pdf
19 Content Length:  2639215 bytes
20
21
22
23
24 URI:  https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-jones-off-
  topic.pdf
25 Final URI:  https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-
  jones-off-topic.pdf
26 Content Length:  3119205 bytes
27
```

```
28
29
30
31 URI:  https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-atkins-
      news-similarity.pdf
32 Final URI:  https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-
      atkins-news-similarity.pdf
33 Content Length:  18995885 bytes
34
35
36
37
38 URI:  https://www.cs.odu.edu/~mln/pubs/ht-2018/hypertext-2018-nwala-
      bootstrapping.pdf
39 Final URI:  https://www.cs.odu.edu/~mln/pubs/ht-2018/hypertext-2018-
      nwala-bootstrapping.pdf
40 Content Length:  994153 bytes
41
42
43
44
45 URI:  https://www.cs.odu.edu/~mweigle/papers/mccoy-kddl18.pdf
46 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mccoy-kddl18.pdf
47 Content Length:  1002468 bytes
48
49
50
51
52 URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-aturban-
      archivenow.pdf
53 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-
      aturban-archivenow.pdf
54 Content Length:  3998654 bytes
55
56
57
58
59 URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-alam-archive-
      -banner.pdf
60 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-alam-
      archive-banner.pdf
61 Content Length:  596000 bytes
62
63
64
65
66 URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-nwala-
```

```
scraping-serps-seeds.pdf
67 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-nwala-
    scraping-serps-seeds.pdf
68 Content Length:  2172494 bytes
69
70
71
72
73 URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-kelly-
    private-public-web-archives.pdf
74 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-kelly-
    private-public-web-archives.pdf
75 Content Length:  2553579 bytes
76
77
78
79
80 URI:  http://www.cs.odu.edu/~mweigle/papers/alkwai-tois17-preprint.pdf
81 Final URI:  https://www.cs.odu.edu/~mweigle/papers/alkwai-tois17-
    preprint.pdf
82 Content Length:  1430568 bytes
83
84
85
86
87 URI:  http://www.cs.odu.edu/~mkelly/papers/2017_jcdl_wail.pdf
88 Final URI:  https://www.cs.odu.edu/~mkelly/papers/2017_jcdl_wail.pdf
89 Content Length:  412476 bytes
90
91
92
93
94 URI:  http://www.cs.odu.edu/~mweigle/papers/alnoamany-websci17.pdf
95 Final URI:  https://www.cs.odu.edu/~mweigle/papers/alnoamany-websci17.
    pdf
96 Content Length:  6962016 bytes
97
98
99
100
101 URI:  http://www.cs.odu.edu/~mkelly/papers/2017_jcdl_countingMementos.
    pdf
102 Final URI:  https://www.cs.odu.edu/~mkelly/papers/2017
    _jcdl_countingMementos.pdf
103 Content Length:  274265 bytes
104
105
```

```
106
107
108 URI:  http://www.cs.odu.edu/~mweigle/papers/alam-jcdl17.pdf
109 Final URI:  https://www.cs.odu.edu/~mweigle/papers/alam-jcdl17.pdf
110 Content Length:  1600140 bytes
111
112
113
114
115 URI:  http://www.cs.odu.edu/~anwala/files/publications/NwalaJCDL_LMP.
      pdf
116 Final URI:  https://www.cs.odu.edu/~anwala/files/publications/
      NwalaJCDL_LMP.pdf
117 Content Length:  17623699 bytes
118
119
120
121
122 URI:  http://www.cs.odu.edu/~mweigle/papers/brunelle-jcdl17.pdf
123 Final URI:  https://www.cs.odu.edu/~mweigle/papers/brunelle-jcdl17.pdf
124 Content Length:  1276346 bytes
125
126
127
128
129 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-
      nanonetbookchapter-preprint17.pdf
130 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-
      nanonetbookchapter-preprint17.pdf
131 Content Length:  714649 bytes
132
133
134
135
136 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-computer17.pdf
137 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-
      computer17.pdf
138 Content Length:  498301 bytes
139
140
141
142
143 URI:  http://www.cs.odu.edu/~mweigle/papers/pardue-vis16-2pg-poster.pdf
144 Final URI:  https://www.cs.odu.edu/~mweigle/papers/pardue-vis16-2pg-
      poster.pdf
145 Content Length:  583877 bytes
146
```

```
147
148
149
150 URI:  http://www.cs.odu.edu/~mweigle/papers/pardue-vis16-poster.pdf
151 Final URI:  https://www.cs.odu.edu/~mweigle/papers/pardue-vis16-poster.
    pdf
152 Content Length:  614446 bytes
153
154
155
156
157 URI:  http://www.cs.odu.edu/~mln/pubs/tpdl-2016/tpdl-2016-kelly.pdf
158 Final URI:  https://www.cs.odu.edu/~mln/pubs/tpdl-2016/tpdl-2016-kelly.
    pdf
159 Content Length:  953454 bytes
160
161
162
163
164 URI:  http://www.cs.odu.edu/~mweigle/papers/alnoamany-ijdl16a.pdf
165 Final URI:  https://www.cs.odu.edu/~mweigle/papers/alnoamany-ijdl16a.
    pdf
166 Content Length:  3741559 bytes
167
168
169
170
171 URI:  http://www.cs.odu.edu/~mweigle/papers/alnoamany-ijdl16b.pdf
172 Final URI:  https://www.cs.odu.edu/~mweigle/papers/alnoamany-ijdl16b.
    pdf
173 Content Length:  4697304 bytes
174
175
176
177
178 URI:  http://www.cs.odu.edu/~mweigle/papers/brunelle-ijdl16.pdf
179 Final URI:  https://www.cs.odu.edu/~mweigle/papers/brunelle-ijdl16.pdf
180 Content Length:  7847111 bytes
181
182
183
184
185 URI:  http://www.cs.odu.edu/~mweigle/papers/aturban-tpdl15.pdf
186 Final URI:  https://www.cs.odu.edu/~mweigle/papers/aturban-tpdl15.pdf
187 Content Length:  622981 bytes
188
189
```

```
190
191
192 URI:  http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-stories.pdf
193 Final URI:  https://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-
        stories.pdf
194 Content Length:  1274604 bytes
195
196
197
198
199 URI:  http://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-off-topic.pdf
200 Final URI:  https://www.cs.odu.edu/~mln/pubs/tpdl-2015/tpdl-2015-off-
        topic.pdf
201 Content Length:  4308768 bytes
202
203
204
205
206 URI:  http://www.cs.odu.edu/~mweigle/papers/brunelle-ijdl15b.pdf
207 Final URI:  https://www.cs.odu.edu/~mweigle/papers/brunelle-ijdl15b.pdf
208 Content Length:  5700649 bytes
209
210
211
212
213 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-tmbmc15.pdf
214 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-tmbmc15.
        pdf
215 Content Length:  3387091 bytes
216
217
218
219
220 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic-sites.
        pdf
221 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-arabic
        -sites.pdf
222 Content Length:  709420 bytes
223
224
225
226
227 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.pdf
228 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2015/jcdl-2015-mink.
        pdf
229 Content Length:  1254605 bytes
230
```

```
231
232
233
234 URI:  http://www.cs.odu.edu/~mweigle/papers/rawat-ahsw15.pdf
235 Final URI:  https://www.cs.odu.edu/~mweigle/papers/rawat-ahsw15.pdf
236 Content Length:  1039622 bytes
237
238
239
240
241 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-jsac14.pdf
242 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-jsac14.
    pdf
243 Content Length:  1992270 bytes
244
245
246
247
248 URI:  http://www.cs.odu.edu/~mweigle/papers/wang-milcom14.pdf
249 Final URI:  https://www.cs.odu.edu/~mweigle/papers/wang-milcom14.pdf
250 Content Length:  402145 bytes
251
252
253
254
255 URI:  http://www.cs.odu.edu/~mweigle/papers/ruffing-milcom14.pdf
256 Final URI:  https://www.cs.odu.edu/~mweigle/papers/ruffing-milcom14.pdf
257 Content Length:  498835 bytes
258
259
260
261
262 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-misenet14.pdf
263 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-misenet14
    .pdf
264 Content Length:  1231147 bytes
265
266
267
268
269 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-kelly-mink.
    pdf
270 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-kelly-
    mink.pdf
271 Content Length:  556863 bytes
272
273
```

```
274
275
276 URI:  http://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-brunelle-
      damage.pdf
277 Final URI:  https://www.cs.odu.edu/~mln/pubs/jcdl-2014/jcdl-2014-
      brunelle-damage.pdf
278 Content Length:  2205546 bytes
279
280
281
282
283 URI:  http://www.cs.odu.edu/~mkelly/papers/2014_dl_acid.pdf
284 Final URI:  https://www.cs.odu.edu/~mkelly/papers/2014_dl_acid.pdf
285 Content Length:  541843 bytes
286
287
288
289
290 URI:  http://www.cs.odu.edu/~mweigle/papers/alnoamany-ijdl14.pdf
291 Final URI:  https://www.cs.odu.edu/~mweigle/papers/alnoamany-ijdl14.pdf
292 Content Length:  1725946 bytes
293
294
295
296
297 URI:  http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-nanocom14.pdf
298 Final URI:  https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-nanocom14
      .pdf
299 Content Length:  1773662 bytes
300
301
302
303
304 URI:  http://www.cs.odu.edu/~mweigle/papers/olariu-mccr14.pdf
305 Final URI:  https://www.cs.odu.edu/~mweigle/papers/olariu-mccr14.pdf
306 Content Length:  876816 bytes
307
308
309
310
311 URI:  http://www.cs.odu.edu/~mweigle/papers/almalag-iccvel13.pdf
312 Final URI:  https://www.cs.odu.edu/~mweigle/papers/almalag-iccvel13.pdf
313 Content Length:  321704 bytes
314
315
316
317
```



```
318 URI: http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-globecom13.pdf
319 Final URI: https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-
    globecom13.pdf
320 Content Length: 347912 bytes
321
322
323
324
325 URI: http://www.cs.odu.edu/~mkelly/papers/2013
    _ieevis_boxofficeprediction.pdf
326 Final URI: https://www.cs.odu.edu/~mkelly/papers/2013
    _ieevis_boxofficeprediction.pdf
327 Content Length: 122738 bytes
328
329
330
331
332 URI: http://www.cs.odu.edu/~mkelly/posters/2013_vis_boxoffice.pdf
333 Final URI: https://www.cs.odu.edu/~mkelly/posters/2013_vis_boxoffice.
    pdf
334 Content Length: 373897 bytes
335
336
337
338
339 URI: http://www.cs.odu.edu/~mweigle/papers/olariu-misenet13.pdf
340 Final URI: https://www.cs.odu.edu/~mweigle/papers/olariu-misenet13.pdf
341 Content Length: 405542 bytes
342
343
344
345
346 URI: http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-misenet13.pdf
347 Final URI: https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-misenet13
    .pdf
348 Content Length: 1093385 bytes
349
350
351
352
353 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_134.pdf
354 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_134.pdf
355 Content Length: 296096 bytes
356
357
358
359
```

```
360 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_149.pdf
361 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_149.pdf
362 Content Length: 813692 bytes
363
364
365
366
367 URI: http://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_126.pdf
368 Final URI: https://www.cs.odu.edu/~mln/pubs/tpdl-2013/paper_126.pdf
369 Content Length: 9850923 bytes
370
371
372
373
374 URI: http://www.cs.odu.edu/~mln/pubs/jcdl-2013/fp105-AlNoamany.pdf
375 Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2013/fp105-AlNoamany.
pdf
376 Content Length: 290745 bytes
377
378
379
380
381 URI: http://www.cs.odu.edu/~mweigle/papers/almalag-vanet-ch17.pdf
382 Final URI: https://www.cs.odu.edu/~mweigle/papers/almalag-vanet-ch17.
pdf
383 Content Length: 563385 bytes
384
385
386
387
388 URI: http://www.cs.odu.edu/~mweigle/papers/almalag-mp2p13.pdf
389 Final URI: https://www.cs.odu.edu/~mweigle/papers/almalag-mp2p13.pdf
390 Content Length: 398562 bytes
391
392
393
394
395 URI: http://www.cs.odu.edu/~mweigle/papers/yan-its13.pdf
396 Final URI: https://www.cs.odu.edu/~mweigle/papers/yan-its13.pdf
397 Content Length: 837182 bytes
398
399
400
401
402 URI: http://www.cs.odu.edu/~mweigle/papers/mohrehkesh-casemans12.pdf
403 Final URI: https://www.cs.odu.edu/~mweigle/papers/mohrehkesh-
casemans12.pdf
```

```
404 Content Length: 390810 bytes
405
406
407 Traceback (most recent call last):
408   File "/usr/local/lib/python3.6/dist-packages/urllib3/connection.py",
      line 160, in _new_conn
409     (self._dns_host, self.port), self.timeout, **extra_kw
410   File "/usr/local/lib/python3.6/dist-packages/urllib3/util/connection.
      py", line 61, in create_connection
411     for res in socket.getaddrinfo(host, port, family, socket.
      SOCK_STREAM):
412   File "/usr/lib/python3.6/socket.py", line 745, in getaddrinfo
413     for res in _socket.getaddrinfo(host, port, family, type, proto,
      flags):
414 socket.gaierror: [Errno -2] Name or service not known
415
416 During handling of the above exception, another exception occurred:
417
418 Traceback (most recent call last):
419   File "/usr/local/lib/python3.6/dist-packages/urllib3/connectionpool.
      py", line 677, in urlopen
420     chunked=chunked,
421   File "/usr/local/lib/python3.6/dist-packages/urllib3/connectionpool.
      py", line 392, in _make_request
422     conn.request(method, url, **httplib_request_kw)
423   File "/usr/lib/python3.6/http/client.py", line 1264, in request
424     self._send_request(method, url, body, headers, encode_chunked)
425   File "/usr/lib/python3.6/http/client.py", line 1310, in _send_request
426     self.endheaders(body, encode_chunked=encode_chunked)
427   File "/usr/lib/python3.6/http/client.py", line 1259, in endheaders
428     self._send_output(message_body, encode_chunked=encode_chunked)
429   File "/usr/lib/python3.6/http/client.py", line 1038, in _send_output
430     self.send(msg)
431   File "/usr/lib/python3.6/http/client.py", line 976, in send
432     self.connect()
433   File "/usr/local/lib/python3.6/dist-packages/urllib3/connection.py",
      line 187, in connect
434     conn = self._new_conn()
435   File "/usr/local/lib/python3.6/dist-packages/urllib3/connection.py",
      line 172, in _new_conn
436     self, "Failed to establish a new connection: %s" % e
437 urllib3.exceptions.NewConnectionError: <urllib3.connection.
      HTTPConnection object at 0x7f5bba3d2240>: Failed to establish a new
      connection: [Errno -2] Name or service not known
```

Listing 6: Output after running the findPDFs.py program at the command line with `https://www.cs.odu.edu/~mweigle/Main/PubsByYear` as the command line parameter.

Discussion

The first thing I had to figure out was how to access the command line parameter within the code. That's where *Command Line Arguments in Python* came in. Their example code involved printing the name of the script, which is stored in `sys.argv[0]`. So all I had to do to access the URI passed in was reference `sys.argv[1]`. From week 2 materials, I got the idea of passing in `links.get('href')` into `requests.get()`, and storing it in an additional variable called "resp." From there, it's just checking if the content type is a pdf. If it isn't, skip that iteration. If it is, print out the URI, final URI, and content length. Admittedly, I could not figure out how to put commas in the numbers for the content length. The program ran perfectly for the first two URIs, but on the third, I got a few exceptions. I think they were mainly the result of one or more PDFs no longer existing. The one I believe the exception was highlighting was `http://research.nokia.com/files/public/mdc-final450-mohrekesh-new.pdf`. Upon trying to click the link myself, I was met with Figure 4.

References

Every report must list the references that you consulted while completing the assignment. If you consulted a webpage, you must include the URL.

- Curl manpage, <https://curl.haxx.se/docs/manpage.html>
- Command Line Arguments in Python, <https://www.geeksforgeeks.org/command-line-arguments-in-python/>
- Beautiful Soup Documentation, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/#extract>
- Week-02 Python Lecture Slides, https://docs.google.com/presentation/d/1_BtQ5b7aG07xSw-B1cnCJT78qXxL2BqYEzug5Mz6cVU/edit?usp=sharing
- Week-02 Python Collab Notebook, https://colab.research.google.com/github/cs432-websci-fall20/assignments/blob/master/432_Week_02_Python.ipynb
- Week-01 Lecture Slides, https://docs.google.com/presentation/d/17ltjnT3AY8vm04e-rV3-_vus_i18HIhxmXNupvOyOQs/edit?usp=sharing

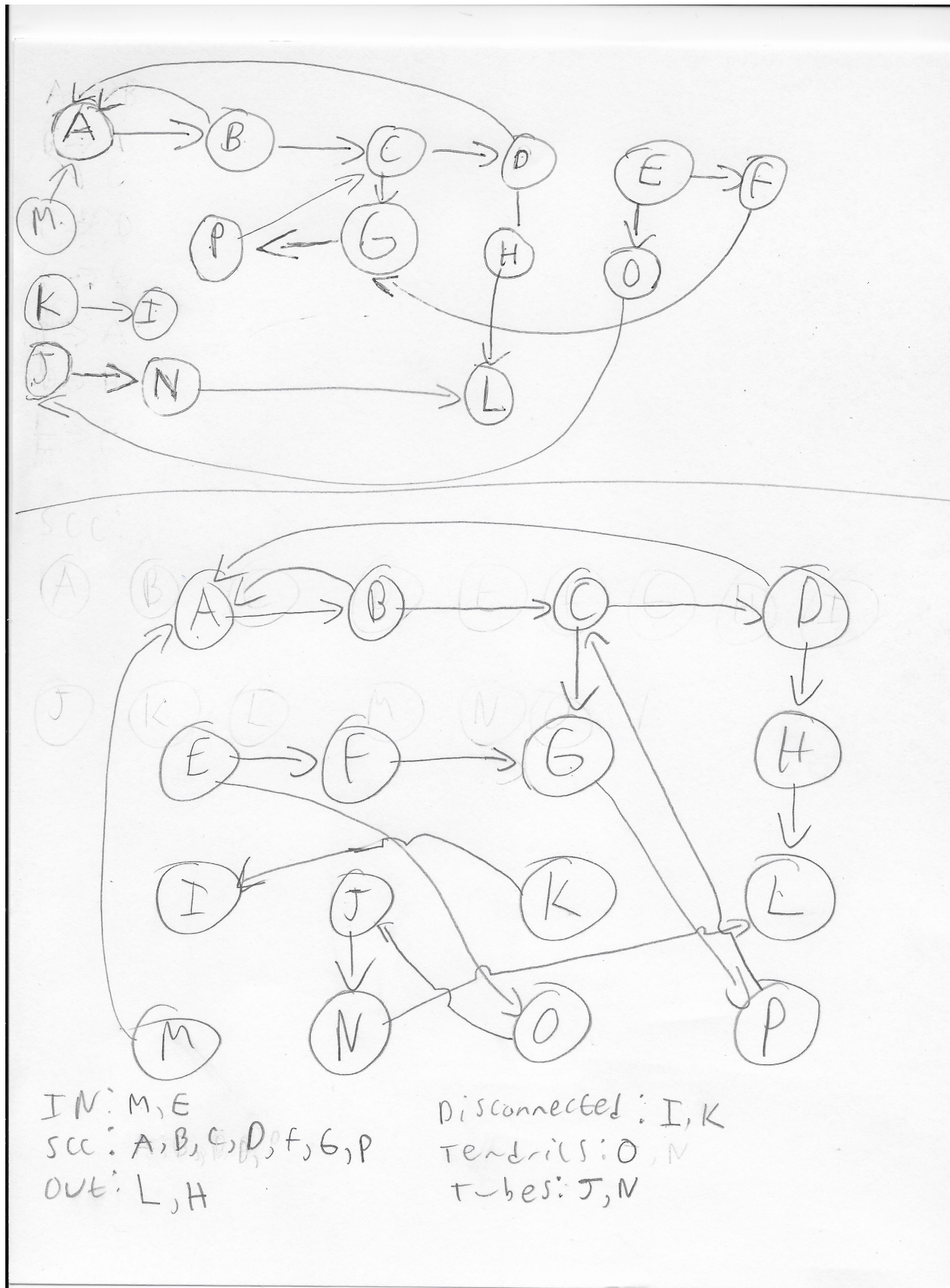


Figure 1: My directed graphs for question1. The lower graph is my final graph. Ignore the list of nodes. The list of nodes on my report is my final answer.

USER AGENT ECHO

User-Agent: CS432

Figure 2: Screenshot of the output file obtained from Firefox in question 2 part B.

USER AGENT ECHO

User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:80.0) Gecko/20100101 Firefox/80.0

Figure 3: Screenshot of the URI obtained from Firefox in question 2 part C.

Enhanced Error Page

COX

research.nokia.com cannot be found

Displaying search results for research.nokia.com

SPONSORED RESULTS

Online Research - Online Research

www.fastquickssearch.com/Online Research/Here

Search for Online Research on fastquickssearch.com! Find Online Research. Relevant Results on fastquickssearch

Online Research - Online Research

find.zoo.com/Here/Online Research

Search for Online Research at Zoo.com Now! Search Online Research. Compare Results on Zoo.com

Online Research - Online Research

www.directhit.com/Online Research/Here

Search for Online Research at Directhit.com. Compare Results. Find Online research at Directhit.com

Online Research - Information

www.kensaq.com/Information/Online Research

Welcome to Kensaq.com. Find Online Research Today! Find Online Research Now at Kensaq.com!

Online Research - Online Research

www.top10answers.com/Online Research/Save_Time

Search For Online Research Here. Search For Online Research Now!

SPONSORED RESULTS

Online Research - Online Research

www.alhea.net/Online Research/Here

Find Online Research Here. Multiple Results at Once! Online Research Results For You on Alhea Today!

Online Research - Online Research

www.productopia.com/Online Research/With us

Search For Online Research at Productopia.com. Search For Online Research that are Great for You!

Online Research - Online Research

www.simpli.com/Now/Online Research

Search For Online Research that are Right for You! Search For Online Research at Simpli.com

Nokia and ABI Research identify key trends in manufacturing ...

www.nokia.com/about-us/news/releases/2020/05/06/...

Nokia and ABI Research identify key trends in manufacturing investment to enable Industry 4.0. 74% of manufacturing decision-makers surveyed plan to upgrade comms and control networks in next two years to advance digital transformation and Industry 4.0 ; More than 90% are investigating use of 4G/5G for their operations

Nokia - Wikipedia

en.wikipedia.org/wiki/Nokia_Research_Center

Nokia Bell Labs is a research and scientific development firm that was once the R&D arm of the American Bell System. It became a subsidiary of Nokia Corporation after the takeover of Alcatel-Lucent in 2016. NGP Capital

5G consumer market research | Nokia

www.nokia.com/networks/research/5g-consumer...

Nokia conducted in-depth 5G research with more than 3,000 consumers in markets where 5G is already deployed: the US, the UK and South Korea. This major new 5G research report study reveals what consumers expect from 5G market trends, which 5G services they most want to receive, and how they'd like to receive them.

Marketing Research: Nokia - 2763 Words | Report Example

ivypanda.com/essays/marketing-research-nokia

The bottom line as suggested in this research is the fact that Nokia as a company needs to change its technology in the manufacture of phones and come up with strategies that can effectively compete with the rivals (Buckley 2011).

Nokia Bell Labs: Research

www.bell-labs.com/our-research

Nokia Bell Labs is building the network of the future – one network connecting all humans, all senses, all things, all systems, all infrastructure, all processes. Current Research Applications Augmented Human

SPONSORED RESULTS

Online Research - Online Research

www.alhea.net/Online Research/Here

Find Online Research Here. Multiple Results at Once! Online Research Results For You on Alhea Today!

Online Research - Online Research

www.productopia.com/Online Research/With us

Search For Online Research at Productopia.com. Search For Online Research that are Great for You!

Online Research - Online Research

www.simpli.com/Now/Online Research

Search For Online Research that are Right for You! Search For Online Research at Simpli.com

research.nokia.com

Q

About This Page

Our Policies

Privacy Policy

Make a Request

Figure 4: Screenshot of the error page I was taken to by clicking a link to a PDF that no longer exists.