|  |
| --- |
|  |
| CS432 Spring 2018 |
| Assignment 1 |
|  |
| **Jonathan-Kruszewski** |
| **1/28/2018** |

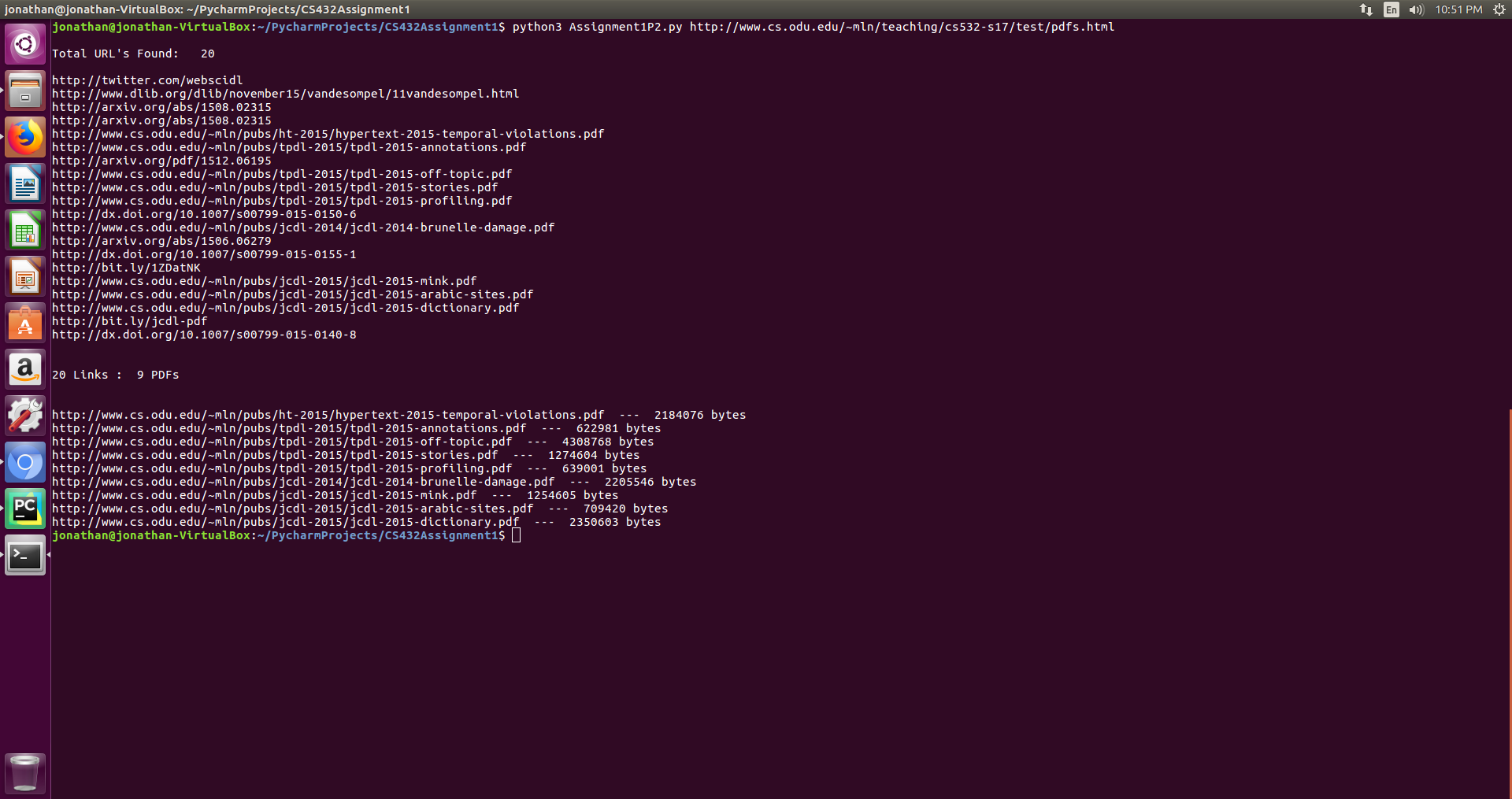
|  |
| --- |
|  |

**Part 1:**



**Part 2:**

|  |
| --- |
|  |
|  | from bs4 import BeautifulSoup  import sys |
|  | import requests |
|  | import urllib.request |
|  | import re |
|  |  |
|  | if len(sys.argv) <= 1: |
|  | print("ERROR: INSUFFICIENT ARGUMENTS!") |
|  | else: |
|  | url = sys.argv[1] |
|  | html\_page = urllib.request.urlopen(url) |
|  | soup = BeautifulSoup(html\_page, "html.parser") |
|  | linksFound = [] |
|  | for link in soup.findAll('a', attrs={'href': re.compile("^http")}): |
|  | linksFound.append(link.get('href')) |
|  | print("\nTotal URL's Found: ", len(linksFound), "\n") |
|  | print(\*linksFound,sep='\n') |
|  | print("\n") |
|  |  |
|  | pdfFound = [] |
|  | for s in linksFound: |
|  | if".pdf" in s: |
|  | pdfFound.append(s) |
|  | print(len(linksFound), "Links : ", len(pdfFound), "PDFs") |
|  | print("\n") |
|  | for i, urls in enumerate(pdfFound): |
|  | response = requests.get(urls) |
|  | print(urls, " --- ", response.headers['content-length'], "bytes") |



**Part 3:**

IN: O, M  
SCC: A, B, C, G  
OUT: H, D  
Tendrils: L, I, K  
Tubes: N, D   
Disconnected: E, F  
