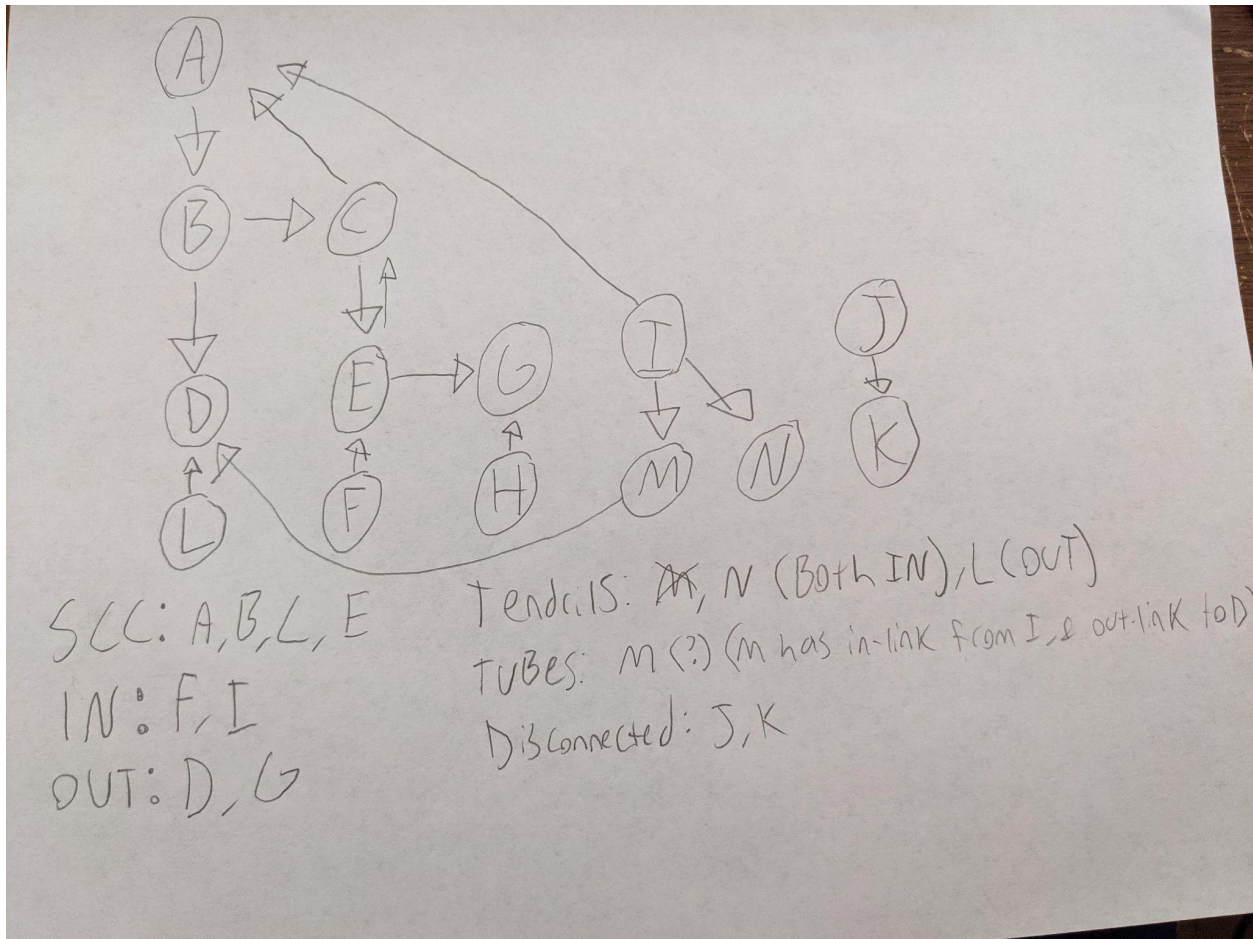


Andrew Juers  
Professor M. Weigle  
CS 432  
9/19/21

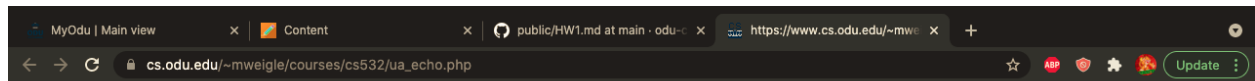
## Homework 1 Report

Q1



Q2

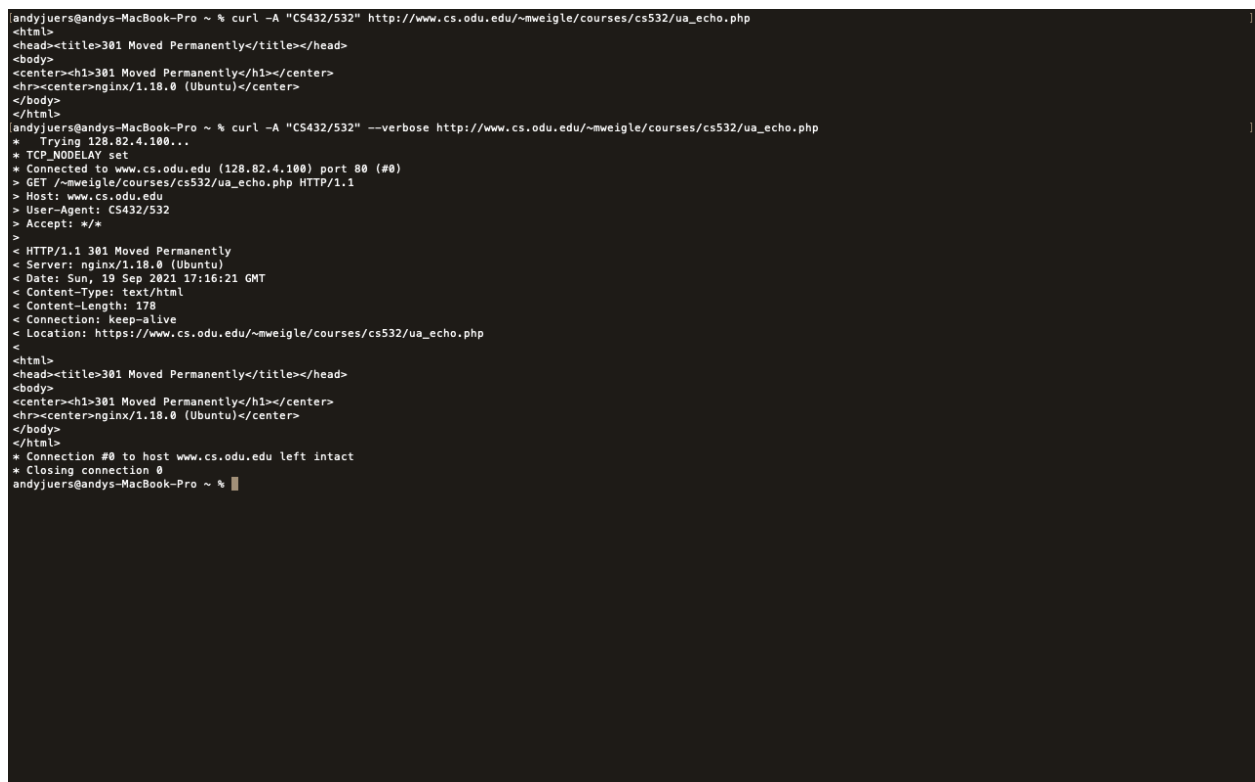
a



USER AGENT ECHO

**User-Agent:** Mozilla/5.0 (Macintosh; Intel Mac OS X 10\_15\_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0.4577.63 Safari/537.36

b



C

```
~/desktop/cs432/hw1 — zsh                                ~/desktop/cs432/hw1 — ssh cs_ajuer001@linux.cs.odu.edu +
cs_ajuer001@sol:~/cs432/hw1$ python3 main.py https://www.cs.odu.edu/~mweigle/courses/cs532/pdfs.html
main.py:7: GuessedAtParserWarning: 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.

The code that caused this warning is on line 7 of the file main.py. To get rid of this warning, pass the additional argument 'features="html.parser"' to the BeautifulSoup constructo
r.

soup = BeautifulSoup(response.text)
URI: https://www.cs.odu.edu/~mln/pubs/ht-2018/hypertext-2018-nwala-bootstrapping.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/ht-2018/hypertext-2018-nwala-bootstrapping.pdf
Content Length: 994153

URI: https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-atkins-news-similarity.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-atkins-news-similarity.pdf
Content Length: 18995885

URI: https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-jones-off-topic.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-jones-off-topic.pdf
Content Length: 3119205

URI: https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-jones-archiveit.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/ipres-2018/ipres-2018-jones-archiveit.pdf
Content Length: 2639215

URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-nwala-scraping-serps-seeds.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-nwala-scraping-serps-seeds.pdf
Content Length: 2172494

URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-kelly-private-public-web-archives.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-kelly-private-public-web-archives.pdf
Content Length: 2553579

URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-aturban-archivenow.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-aturban-archivenow.pdf
Content Length: 3998654

URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-alam-archive-banner.pdf
Final URI: https://www.cs.odu.edu/~mln/pubs/jcdl-2018/jcdl-2018-alam-archive-banner.pdf
Content Length: 596000

cs_ajuer001@sol:~/cs432/hw1$
```

Q3

<https://github.com/fudgenuggets1/cs432hw1>

I am aware that my program currently only works on very specific html websites, but I felt it was sufficient for this question. I tried hard but I had a hard time learning how to use bs4 with requests. I hope I can learn it so I can master this material.