Project-1: **100** Points
Due date: 02/13

Project description: This will be a semester-long project. Project-1 will be on collecting/scrapping certain news articles. Choose any news agency and go to their website and select a category (politics, sports, entertainment ,etc) and click on any 5 links within the category. For example 5 sports related news. Copy those URLs and paste it in a text file.

Now create a python program that is going to visit each and every URLs present in the text file and download/copy/scrap the content of that URL. You can use any python package to accomplish this with no restriction. You want to write your own downloader/scraper. Go for it. Remember we are interested in the news articles not the advertisement and all other junk/information present in the page.

Example: python3 my_news_downloader.py . The output will be 5 files containing the articles mentioned in the 5 URLs.

Instructions:

1) Create a github repo and make it public, name the repo appropriately since you will be able to put this in your resume (please do not copy code from other students, learn/refer to the internet, but please do not copy from another student(s). You are here to learn not to collect letters (A,B,C,D). Project-5 will be a personal interview where I am going to ask you about the code, so please understand what you are doing. I trust you are not going to indulge in plagiarism activities).
2) In this repo push your python files.
3) Export your conda environment so that the TAs may be able to run your code. The export environment file name should be **requirements.yml**.
4) Write an elaborate README.md file. Mentioning what the software does and any other information that you want to mention. Then write instructions on how to use your software. This is important. Think of it as all the steps required to run your software. Like how to initialize the conda environment with your requirements.yml file and how to run your python code and what is the output. Give as much information as possible. I was talking to a hiring person of a company near St. Louis. The hiring person was saying that a student got the job because the readme was so elaborate that he was able to reproduce the results during the phone interview. Documentation is very important in software engineering and also for you to remember what you have done.
5) Push the output files (5 of then) and the input file (containing 5 URLs) to the repo.

Rubric: 100 points

- Able to download 5 news articles and store them in five different files: 60 points
- requirements.yml file is present: 20 points
- README.md is written well: 20 points (this is subjective so try to impress the TAs with an elaborate readme)

You have to send the public github link to your repo to the TAs, send it to any one of them. **Please be sure that the subject of the email that you are going use should be "Project-1 CS325" and in the body do not forget to mention your name with your 800 number.** They will grade randomly so there is no fixed TA to a corresponding student.
TAs email IDs: ntavlee@siue.edu and nimorga@siue.edu

ZIP ALL THE FILES AND SUBMIT IT TO MOODLE

IF YOU HAVE ANY QUESTION ASK ME AFTER THE CLASS OR DURING THE CLASS, I AM HAPPY TO EXPLAIN.