Assignment 1 - Part 1

Student name: Anh Nguyen.

Project description:

In the class Major Studio 1, I am starting to find my interest. My pursuit starts with an interest in communication and journalism. After digging deeper into the fields of communication and journalism, I nail down the topic to "how does conspiracy become a new propaganda weapon to alter information censorship for power?"

The hypothesis question might change, but the general theme of conspiracy and propaganda remains primary. I am curious about the current news, for example, around the QAnon cabal. Assignment 1 in Python Fall 2020 is a great opportunity to create an assisting tool for my research.

Both conspiracy and propaganda are new topics to me, so I anticipate many reading and case studies. To work more efficiently around tight deadlines of the semester, a machine that gathers news articles would help me a great deal in collecting information. My aim for assignment 1 is to make a tool that feeds me articles with keywords from my chosen sources.

Project aim:

- First of all, the tool can gather articles from my chosen sources within a specific time window based on my input keywords.
  - Besides the main function that is collecting articles, I am also curious to see if I can track:
    - Comments.
    - Shares.
    - Views.

- Below is my list of chosen sources. The list might change due to the timeline and scope of the assignment.

    - The New York Times.

    - The New Yorker.

    - Vox media.

    - The Atlantic.

    - The Time.

    - The Wall Street Journal.

    - ProPublica.

- After collecting the news, the software will export the results into a CVS file for a record. Within the CVS file, there are links to the article, headlines, authors, published dates.

Existing project, examples:

1. [Web scraping with Beautiful Soup.](#)

2. [Web scraping the President's lies in 16 lines of Python.](#)

3. [Python script to search for News based on keywords.](#)

4. [Top stories from Africa every morning.](#)

Tools/libraries:

1. API from sources.

    - Either a specific API of a specific source such as [the New York Times API,](#) or the [news API](#) as Ajinkya Sonawane suggested in his project.

2. Beautiful Soup 4.

    - Beautiful Soup might not be the best for my project because websites can block me if there are many requests. The library is better for small to

medium websites. I want to include it in my description as another approach for further exploration.

3. Request library.

4. Pandas library.

- Pandas can help me parsing the results into an output in a CVS file.