**Using the Plunkett Research Database and Python to Parse and Format Director Names**

The Plunkett Research Database contains a wide range of reliable information on corporations, including lists of their directors. The following is information on how to use this database to extract the directors, as well as how to use a Python script to parse and format the information into Excel.

Step 1:

Use the [Plunkett Database](http://www.plunkettresearch.com.silk.library.umass.edu/) to conduct a search on the companies you are interested in. Once you have the search results, select “Export Search Results.” In General Information, only select “Stock Ticker.” Export the data in text format.

Step 2:

Once a text file of the data is obtained, open the Director Parsing Python Script in a Python Editor (located in Dropbox->EURA Spring 2022->Fortune Board Member Project-> Director Parsing Code.py).

NOTE: Make sure that your text file is in the same directory as your Python Script.

Replace ‘F500DATA.txt’ with the name of your text file.

Rename ‘F500PARSED.txt’ with the name you would like your parsed file to be.

Run the script.

Step 3:

Return to the folder/ directory that your Python Script and original text file is located in. There now should be another text file with the parsed data.

Import this text file into Excel by starting in Excel and selecting File->Open->insertnamehere.txt

Select ‘Comma’ as the only delimiter when importing the data.

Highlight Column A and go to data, select ‘Stocks.’

The stock tickers should be converted to standardized company names, and you have your final edgelist!

NOTES/ QUESTIONS:

* It is unclear what year these data are from. A cross reference between 10-K forms and the exported data shows that the information for most companies is consistent with 2020 boards. The only differences are additions to board of directors in 2021. There was no case that I found in which a board member from 2020 was not included in the list, even if they were removed from the board in 2021.
* Archives are available, but only for a ‘Pro’ membership. Would it be worth checking with UMass about upgrading?
* Is there a way to use R or excel to filter the Fortune500 companies on the list? Otherwise I can do it manually.
* Look over 10-K parsing script as potential long-term solution.
  + Issues with anti-scraping, issues with RegEx