To analyze the financial statements for public companies, first we have to find a way to massively, and automatically download the data from SEC official websites

We are focusing on 10-Qs. It turns out that every year a company only produces 3 10-Qs, for the first 3 quarters. The financials for the fourth quarter is always in 10-K. so if we want to ever track company performance in quarters. 10-Q is not enough.

First we need to obtain a detailed list of all the companies listed on NYSE. A text file containing the ticker symbols and company names is obtained from EODData. The list also includes some tickers like XXX-X or XXX.X, which are unavailable on SEC’s EDGAR database. After a simple parsing through python, we could obtain a clean, searchable list. So far all the tickers are stored in a txt file separated by an empty space, in a single line. The line can be read and parsed later. The searchable companies are estimated to be around 2658.

Maybe csv would be more helpful?

NASDAQ provided a comprehensive csv file for all the listed companies, so its ticker list is easy to obtain. For NASDAQ, its tickers are subtler because the additional tickers for the same company may not follow a certain format, like that of NYSE companies. So to save time, I’ll leave those companies there. If a search is failed on EDGAR, nothing will be returned except a <p> statement “No matching Ticker Symbol.” I’ll let the future program capture the statement instead of spending tons of time figuring out which NASDAQ ticker is unsearchable.

Next step, write a program to navigate into the filing page for each ticker, find the list of 10-Q, and go into each 10-Q’s excel page.