## Using MD&A of company's SEC filing to predict future stock price

CS410 Project Progress Report by Dmitry Villevald (<a href="mailto:dmitryv2@illinois.edu">dmitryv2@illinois.edu</a>)

Project Progress Report:

Q: Which tasks have been completed?

A: The following tasks have been completed:

- I completed the code which pulls the index files from the SEC EDGAR database for each quarter and extracts Q-10 and K-10 filings for the selected companies.
- From every filing I built a dictionary including all words references in the document with their frequencies (i.e. bag of words representation.)
- Finally, these term frequencies were compared with Loughran and McDonald Sentiment Word Lists and four sentiment scores were calculated for each filing -Positive, Negative, Uncertain and Litigious.

Q: Which tasks are pending?

A: Pending tasks:

 Upload the stock returns data for the selected companies (forward from the date of filing) and estimate performance of several sentiment-based investment strategies.

Q: Are you facing any challenges?

A: Not really. It turned out that extracting only the MD&A section from the filing is difficult because in many filings this section is a placeholder containing a list of links to the other sections of the document where the related topics are discussed. Because of this I chose to build the bag of words for the entire filing. This, in my opinion, would be a better option (compared to selecting only the filings where MD&A is a well-structured separate section) because it will ensure that the sentiment words are captured regardless where in the document they appear.