CS 410 Final Project Progress report: Sentiment Analysis on Financial Statements(Free Topics Track)

Team Member:
  Hui Huang – huih3@illinois.edu

1) Which tasks have been completed?
   - Managed to download 'Risk Factors' section of 10-Ks for several tickers from SEC using sec_api Python API.
   - Preprocess/Clean up the dataset including removing stop words, word lemmatization
   - Compute Jaccard Similarity/Cosine Similarity of the 10-ks and across time using sentiment bag of words model
2) Which tasks are pending?
     Other neural network related models
     Model performance comparison by evaluating the alpha factors
3) Are you facing any challenges?
   One reviewer suggests me to expand either on the depth or breadth of the topic, I am trying to read any state-of-the-art technology or related papers for this topic.