# Extractive Summary using Enron dataset

**Steps:**

1. Generate email paths
2. Process email thread – For every email check if it is a thread keep it else ignore
3. For each email extract body, to, from, subject – assuming subject of each email in a thread is same. Concatenation of each email body in a thread is also performed in this step.
4. Email Cleaning: All preprocessing steps - Non-alpha numeric characters, stop words removal, lower case
5. Tokenization: Word and Sentence
6. Get word frequency
7. Word Embedding – Extract word vectors from pertained glove model
8. Create Sentence Vectors for each tokenized sentence using word embedding
9. Rank each sentence in an email thread using *TextRank* Algorithm

**Next Steps:**

- Search, implement and compare  other techniques used for extractive summarization
- R&D for abstractive email thread summarization.