# Mid-term Review

| | |
|---|---|
| Utkarsh Kumar | 130050022 |
| Sivaprasad S | 130050085 |
| Shantanu Thakoor | 13D100003 |

# Problem Statement

We aim to capture the grammar and writing style of the author of several sample text. Based on this we intend to create a random text which may not already exist in the corpus, but will make sense at least locally. Coherence within a sentence is difficult, within a paragraph is almost impossible (with this method).

# Deliverables

## Input

- A large enough text for us build an accurate model, preferably by the same author on the same topic
- Currently we are handling only reported text, as handling quotations etc. requires global knowledge

## Output

- Sample text that resembles the writing styles of the author

# Approach

- We have taken an HMM based approach
- For now each represents a word and a transition indicates a word followed by another
- These transitions will have probabilities which we have learned from our corpus

# Improvements

- Every node represents a bigram rather than a word, improves the locality of coherence
- Incorporate smoothing techniques to introduce bigrams and trigrams which are not present in the corpus
- Treat quotations, commas and other punctuations in a more sophisticated manner so that our corpus can be varied more widely and we can build a model accordingly.

# Current Status

## generateModel.py

- Read from the corpus and populate a hash table with the number of occurrences of each bigram and word; handle punctuations smartly
- Write the generated model in json format on disk to be read by executeModel.py

## executeModel.py

- Load the Markov model from disk and unpack it
- $P(W_i|W_{i-1}) = c(W_{i-1},W_i)/c(W_{i-1})$
- Generate the document sentence by sentence until a given word limit is reached

**From one came to declare unto him the officers in thy days and the land of Moab, and the temple of Simon, I bring thee the seas, and the strength in the office.**

**Of the silver, I spare them that were the chariots, as much abroad into the king.**

*Casing has been edited

She had a leprous men of Gershon, therefore hearken unto her people out against thy dwelling without the soul, had wholly followed hard by a time.

And to the angel of the church in Philadelphia write; These things saith he that is holy, he that is true, he that hath the key of David, he that openeth, and no man shutteth, and shutteth, and no man openeth.

*Casing has been edited

# Demo