**IT412: Natural Language Processing** **(Due Date: xx/xx)**

# Assignment 2:POS tagging using HMM

*Instructor:* Prasenjit Majumder

## 1 Lab Outcome

After performing this assignment you will be able to perform part-of-speech tagging using Hidden Markow Model.

## 2 Problem description

Part-of-Speech tagging is an important part of many natural language processing pipelines where the words in a sentence are marked with their respective parts of speech. In another way we say, It is a process of tagging sentences with part of speech such as nouns, verbs, adjectives, and adverbs, etc. These tags used as features for higher-level tasks such as Named Entity Resolution, Sentiment Analysis, and Question Answering. This assignment mainly focuses on POS tagging using Hidden Markov Model.


Figure 1: Pos tagging

### 2.1 Hidden Markov Model

Hidden Markov Models (HMMs) are a class of probabilistic graphical model that allow us to predict a sequence of unknown (hidden) variables from a set of observed variables. A simple example of an HMM is predicting the weather (hidden variable) based on the type of clothes that someone wears (observed). An HMM can be viewed as a Bayes Net unrolled through time with observations made at a sequence of time steps being used to predict the best sequence of hidden states.

## 3 Implementation

### 3.1 Dataset

The Brown University Standard Corpus of Present-Day American English (or just Brown Corpus) was compiled in the 1960s by Henry Kuera and W. Nelson Francis at Brown University, Providence, Rhode Island as a general corpus (text collection) in the field of corpus linguistics. It contains 500 samples of English-language text, totaling roughly one million words, compiled from works published in the United States in 1961.

### 3.2 Exercise

1. Implement part of speech tagging using a hidden Markov model.
   - Find accuracy
   - Count the number of times particular tag occurs.
2. Implement part of speech tagging using NLTK.
   - Find accuracy
   - Count the number of times particular tag occurs.
3. Compare both accuracy.

# 4 References

- Hidden Markov Model Video Tutorial
  1. https://www.youtube.com/watch?v=kqSzLo9fenk
  2. https://www.youtube.com/watch?v=TPRoLreU9lA
- Hidden Markov Model Article
  1. https://web.stanford.edu/~jurafsky/slp3/A.pdf
  2. https://towardsdatascience.com/introduction-to-hidden-markov-models-cd2c93e6b781
  3. https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab
- Hidden Markov Model code example:
  https://medium.com/@patrickhk/part-of-speech-tagging-with-hidden-markov-models-hmm-4224cb72e9b9

## 4.1 Dataset

- Brown corpus avilable in NLTK