Finding bigrams in a sentence

Possible parse trees for a sentence

Topic modelling

* LDA
* Parameters of LDA and their significance

Clustering , Classification

* Similarity and difference
* Naive Bayes algorithm
* Accuracy improvment

Regular expression

Predictive Analysis

Probabilistic context free grammar : Create parse trees and grammar rules for sentence, based on probabilistic rule find the best parse tree

POS Tagging

Markov model

* Bayes rule

Features of text corpus

Parsing strategies : top down , bottom up

Keyword normalization technique : lemmetization stemming

Conditional Random Fields , Hidden Markov Model