Natural Language Processing

Course Code : DLO8012

Dr. Rahul Khokale
Head of the Department
Department of Computer Engineering
St. John College of Engineering and Management, Palghar

Module II : Word Level Analysis

- Morphology Analysis
- Survey of English Morphology
- Inflectional morphology & Derivational morphology
- Lemmatization
- Regular Expressions
- Finite Automata
- Finite State Transducers (FST)
- Morphological parsing with FST
- Lexicon free FST Porter Stemmer
- N-Grams
- N-Gram Language model
- N-gram for spelling correction

Morphology Analysis

Survey of English Morphology

- Morphology is the study of the way words are built up from smaller meaning-bearing units, called **morphemes**.
- A **morpheme** is often defined as the minimal meaning-bearing unit in a language.
- e.g. word ***fox*** consists of single morpheme (*fox*)
- The word ***cats*** consists of two morphemes :

  morpheme *cat* & morpheme *-s*

Morphemes

- Morphemes are classified into two broad classes :
  - -stems
  - -affixes
- Stem is the “main” morpheme of the word supplying the “main” meaning to the word.
- Affixes add “additional” meaning of various kinds to the word.
- Affixes are further divided into :
  - -prefixes
  - -suffixes
  - -infixes
  - -circumfixes

- Prefixes precede the stem
- Suffixes follow the stem
- Circumflexes do both
- Infixes are inserted inside the stem

- **eats--**> eat + (-s)

  stem+suffix

- **unbuckle--**> un- + buckle

  prefix + stem

- In German, the past participle of some verbs is formed by adding *ge-* to the beginning of the stem and *-t* to the end. The participle of the verb *'sagen'* will be *'gesagt'*

- Philippine language - 'Tagalog'

affix='*um*' & stem '*hingi*'

infix form=***humingi***