Page No.
Date

Experiment NO. : Group 2 - ②

- Title :

Pos (Part - of - speech) tagging.

- Problem statement :

Pos & Taggers for Indian languages.

- objectives :

To assign the correct grammatical tags to each word in a sentence, based on their role and context in the sentence.

- outcome :

co1 : Apply basic principles of elective subjects to problem solving and modelling.

co2 : Use tools and techniques in the area of software development to build mini - projects.

- Theory :

Pos :

- Part of speech tagging is the process of marking each word in a text within the corresponding part of speech, such as the noune verb, adjectives, adverb, etc.

Page No.
Date

- Pos tagging is an important task in the natural language processing and is used in many applications such as machines translation, speech recognition, and text-to-speech system.

- In the Indian languages, (pos) tagging is a challenging task due to the complex morphology and rich inflectional and the derivational system.

- moreover the lack of the standardization and uniformity in the use of scripts, ~~grp~~ orthography, and grammar across different regions and dialects makes it even morechallenging.

- in morechallenging several approaches have been proposed for Pos tagging in Indian language.

- These include rule based, dictionary based, and machine learning, based methods. Rule-based methods rely on set of hand craft rule that define the pattern so for each part of speech.

- Dictionary-based method use a pre-defined dictionary of words and their corresponding parts of speech. machine learning-based methods use annotated training data to learn the statistical patterns and relationships between word

Pos

Page No.
Date

- One popular machine learning - based approach for POS tagging in Indian languages is the conditional Random field (CRF) model.

- In this approach, the sequence of words and their corresponding part - of - speech tags is modeled as a graph. Where the nodes represent the words and the edges represent the transition probabilities between adjacent tags.

- The model is trained on annoted data using the maximum likelihood estimation algorithm to learn the weight of the features that capture the statistical patterns, and the dependencies between the words and their tags.

- Overall POS tagging Indian languages is a challenging task due to the complexity and diversity of the languages.

- However with the availability of larger annoted dataset and advanced machine learning techniques, accurate and efficient POS tagging can be developed for Indian languages, which can help in improving the performance of various NLP operations

- Another approach for POS tagging in Indian languages is to use statistical tagging.

Page No.
Date

Some examples of POS taggers for indian languages

1. NLTK –

The natural language Toolkit (NLTK) provides POS taggers for several Indian language including Hindi, Bengali, Telugu and Urdu.. These are taggers are based on a combination of rule based and statistical techniques.

2. stanford NLP –

The stanfordNLP toolkit provides POS taggers for several Indian languages including Hindi, Bengali, and Telugu. These taggers are based on statistical techniques and have achieved high accuracy on bench mark datasets.

3. ILMT –

The Indian Language machine Translation (ILMT) project provides POS taggers for several Indian language included Hindi, Telugu, Bengali.

Page No.
Date

Conclusion :

In this practical, I have performed the POS Taggers for Indian Languages like Hindi, Bengali and Telugu etc.