

**POS Tagging of Hindi Language**

**ABSTRACT**

Part of Speech Tagger is an important tool that is used to develop language translator and information extraction. The main goal of analysis of NLP (natural language processing) is to understand natural languages by parsing them. In the practice of analyzing natural languages there exist various sub-tasks. These sub-tasks depend on inbuilt structure of language and do not require complete knowledge and understanding of language. Part-of-speech tagging is one of them. Part-of-speech tagging is basically a practice of assigning language-specific grammatical tags to each word of language-specific input text, according to word’s appearance in the text.

These tags can be like noun, adverb, number, negative, etc. There exist a variety of taggers for most popular language in the world, i.e., English. But such taggers cannot be used for morphologically rich Hindi language as difference exists between structures of both languages.

**Keywords:** *Parts of speech,**Natural Language,**NLTK*

**Introduction:**

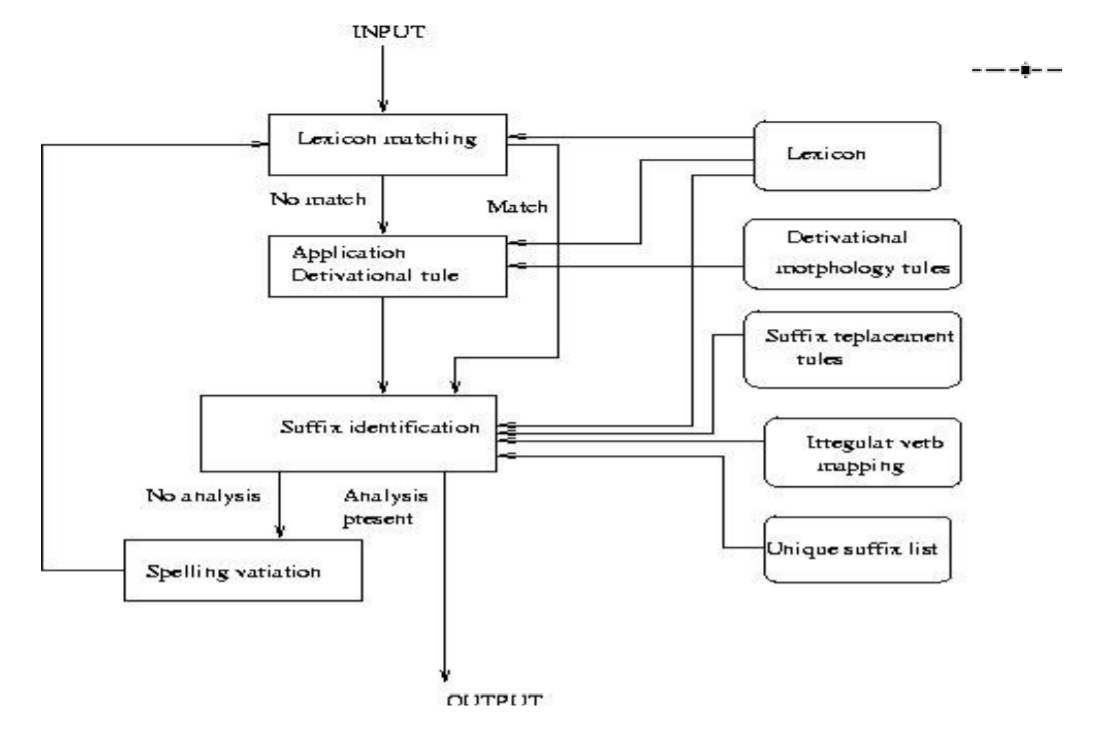
The main goal of analysis of NLP (natural language processing) is to understand natural languages by parsing them. In the practice of analyzing natural languages there exist various sub-tasks. Part of Speech Tagger is an important tool that is used to develop language translator and information extraction.

These sub-tasks depend on inbuilt structure of language and do not require complete knowledge and understanding of language. Part-of-speech tagging is one of them. Part-of-speech tagging is basically a practice of assigning language-specific grammatical tags to each word of language-specific input text, according to word’s appearance in the text.

In order to perform POS tagging we’ve using libraries which are provided by NLTK Python library. The tagging is performed in following stages:

* Preprocessing dataset
* Stemming and Lemmatization
* Training POS Tagger
* Testing new data

**Block Diagram:**

****

**Problem Statement:**

To write a program to identify and tag the different parts of speech in the given text in the Hindi language.

**Goals and Objectives:**

1. To divide the given input text into parts
2. To identify and tag the parts according to their respective part of speech

**Relevant mathematics associated with the Project:**

**System Description**:

* **Input:**

Text file containing the text in Hindi language.

* **Output:**

The sentence is divided and tagged as per the required parts of speech.

**Project Group ID:** 27

**Project Members: Project Guide:**

Devashish Katoriya Prof. P. P. Vaidya

Ashutosh Bhawsar

Bhushan Shilawat

Ninad Kapadnis