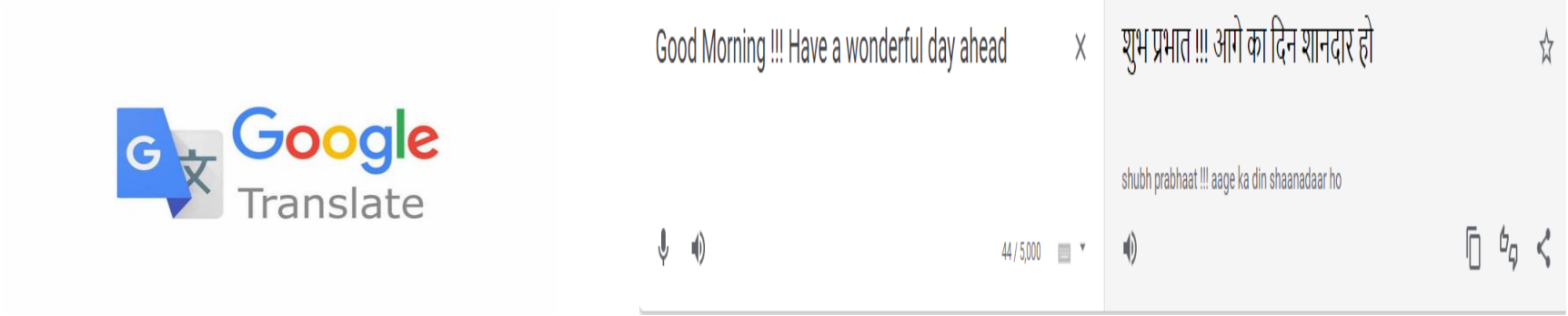
Introduction to Natural Language Processing

Applications of NLP

- Chatbot
- Sentiment Analysis
- Text Classification
- Text Summarization
- Machine Translation
- Text Extraction
- Speech Recognition

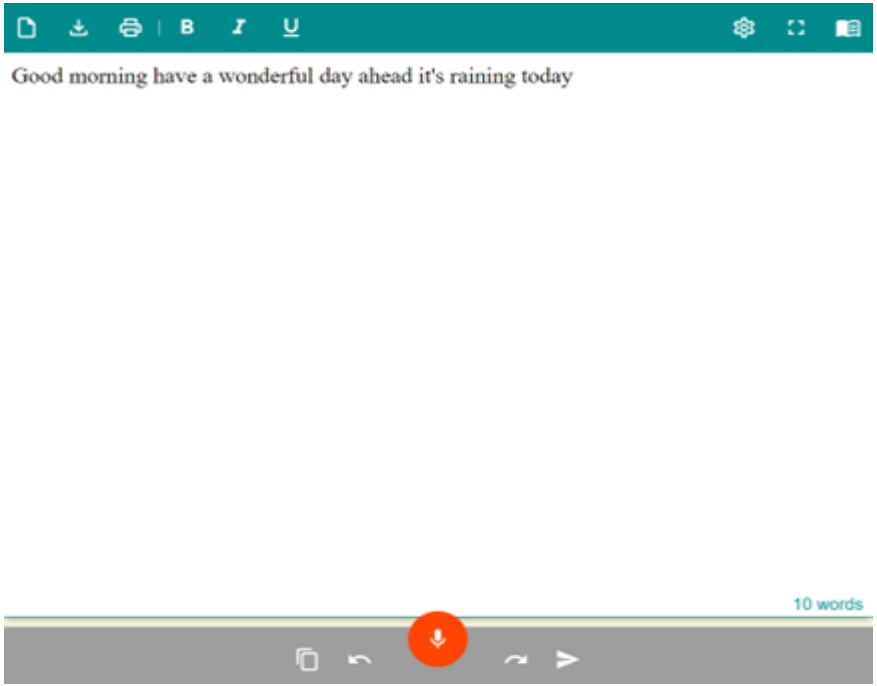
Applications of NLP

Machine Translation



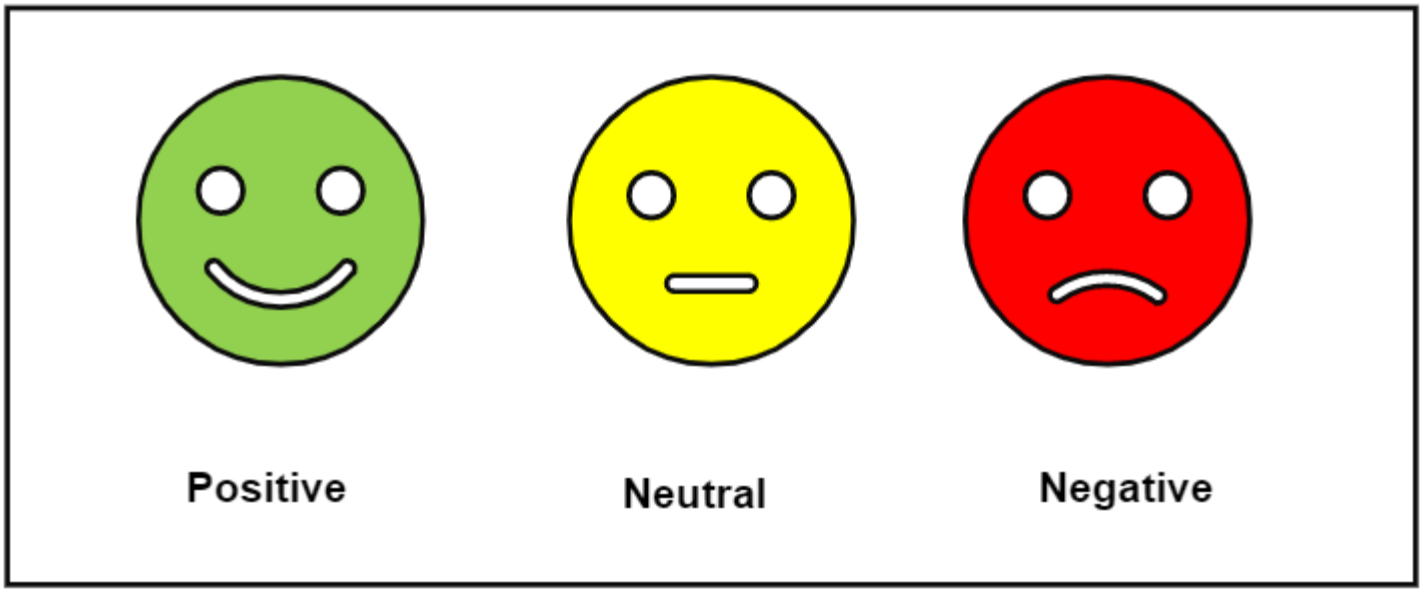
Applications of NLP

Speech to text



Applications of NLP

Sentiment Analysis



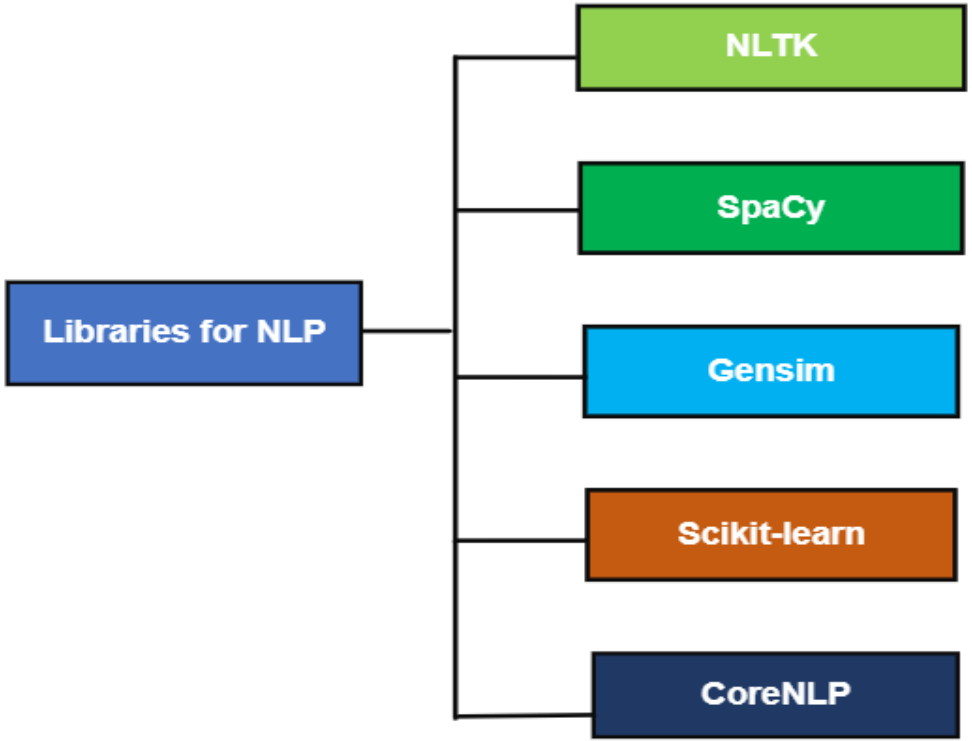
Applications of NLP

Text Classification

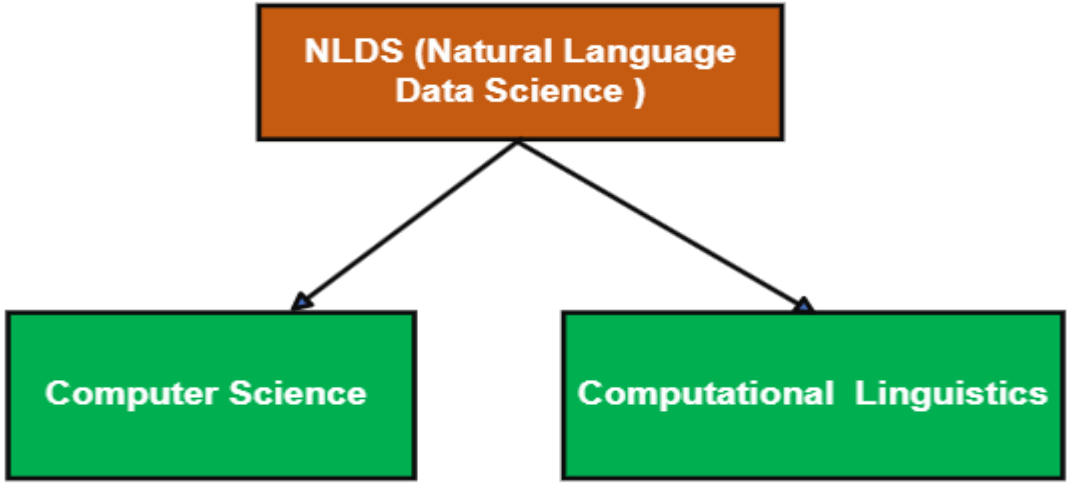
	label	message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

```
Enter SMS Content: Welcome to NLP Module!!!!
Given SMS is  ham
```

Main Libraries : NLP



Natural Language Data Science



NLP for Feature Generation

Natural Language Generation

Natural Language Understanding

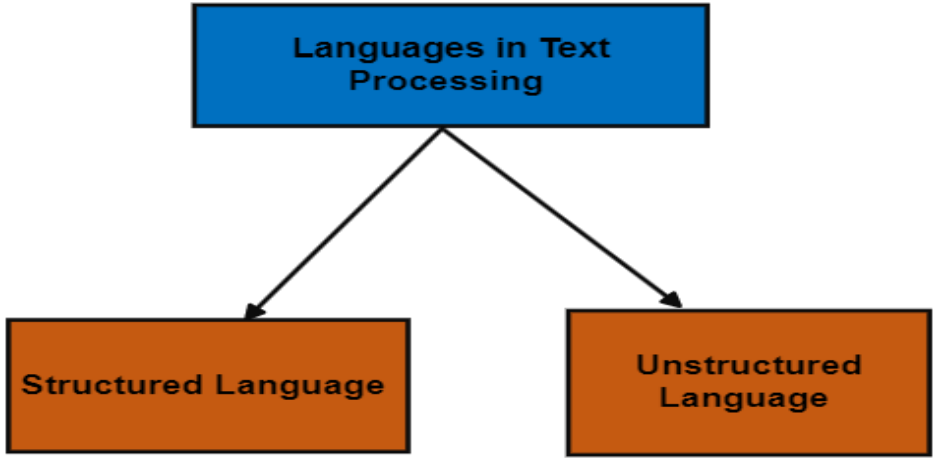
NLP : Definition

Natural Language : Human language that human speaks for communication purpose.

Natural Language Processing is an area of AI research which processes natural languages such as english, hindi etc .

It requires converting natural language to numerical data that can be used by machines for learning.

Languages in Text Processing



(grammar is well defined and adhered strictly)

(grammar is well defined and may be adhered)

Problem

- Ambiguity
- Context Interpretation

For Ex. Unstructured Language
The sofa didn't fit through the door because it was too narrow

Text Processing Stages

Data Acquisition

Getting and acquiring the raw data from diverse sources

-- file

-- csv

-- online format (emitted from web services : XML, JSON, HTML)

Cleaning Stage

Applicable for data extracted through online format

In cleaning stage, the tags are removed from online format and text is extracted.

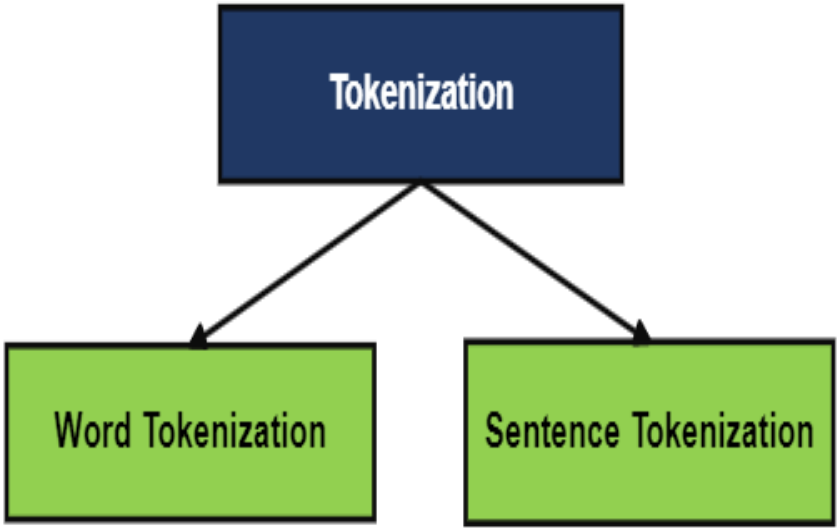
Normalization

Convert the text into one case i.e. either to lower or upper case(text normalization)

removal of punctuations

Tokenization

Conversion of data to sequence of tokens.



Word Tokenization : It will convert the text data into sequence of tokens.

Sentence Tokenization : It will convert the text data into sequence of sentences

Stop words removal

The words which are ignored during processing of natural languages are called as stop words.

Stop words are the commonly used words like pronoun, article, preposition etc. and do not add value to the text.

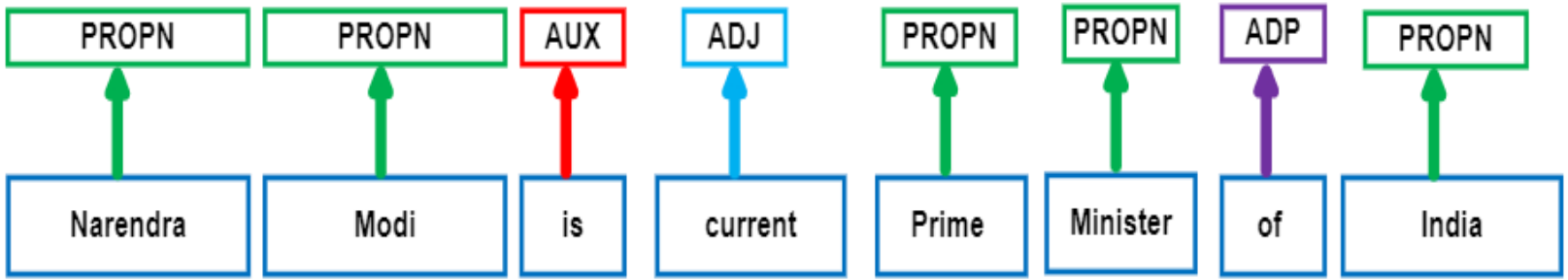
When the stop words are removed, low level information is removed from the text so that we can focus on important information.

Parts of speech Tagging

Parts of speech is a part of English grammar which includes entities like noun, pronoun, adverb, adjective etc.

Disadvantage :- It works only with structured language dataset.

For Ex. Narendra Modi is Current PM of India

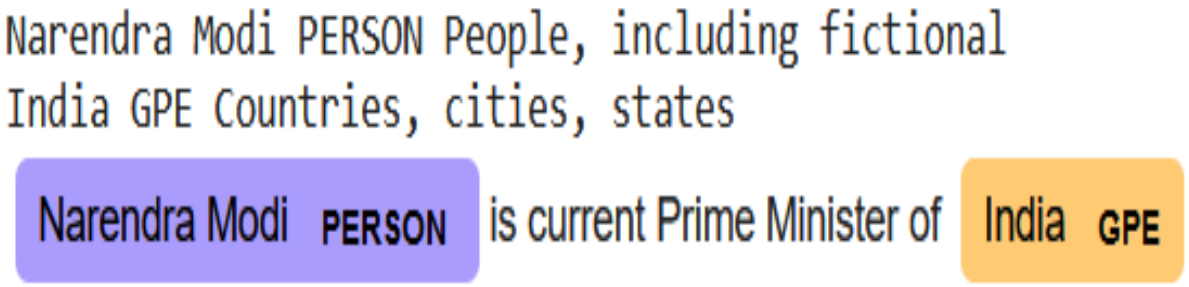


Name Entity Recognition

Recognizing an entity based on (Name ,thing, animal, city, state, org, country etc.)

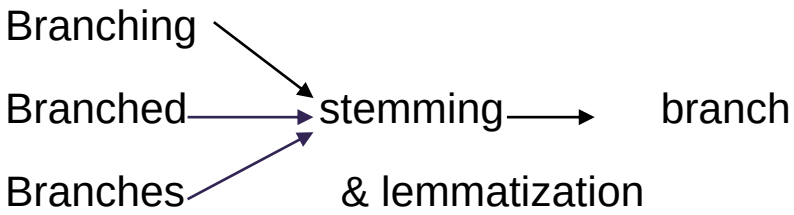
NER helps to understand the context of sentence.

For Ex. Narendra Modi is Current PM of India



Stemming and Lemmatization(Advanced form of normalization)

Stemming and Lemmatization are techniques for deriving the root word.



Caching
Cached ---> stemming --> cach
Caches

Caching
Cached --> lemmatization --> Cache
Caches

Compare

Parameters	Stemming	Lemma
Memory (RAM)	Do not use RAM	It uses RAM to check the words in look up dictionary.
meaningfulness	It may be maintained. There is no guarantee,	Guaranteed Yes
Turn Around Time for response	It is Fast	It is Slow

Thank You !!!