Natural Language Processing

Tushar B. Kute,
http://tusharkute.com

Natural Language Processing

- Natural Language Processing(NLP) is defined as the branch of Artificial Intelligence that provides computers with the capability of understanding text and spoken words in the same way a human being can.

- It incorporates machine learning models, statistics, and deep learning models into computational linguistics i.e. rule-based modeling of human language to allow computers to understand text, spoken words and understands human language, intent, and sentiment.

Natural Language Processing

- Humans communicate with each other using words and text. The way that humans convey information to each other is called Natural Language. Every day humans share a large quality of information with each other in various languages as speech or text.

- However, computers cannot interpret this data, which is in natural language, as they communicate in 1s and 0s. The data produced is precious and can offer valuable insights. Hence, you need computers to be able to understand, emulate and respond intelligently to human speech.

- Natural Language Processing or NLP refers to the branch of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages.

Natural Language Processing

- NLP combines the field of linguistics and computer science to decipher language structure and guidelines and to make models which can comprehend, break down and separate significant details from text and speech.
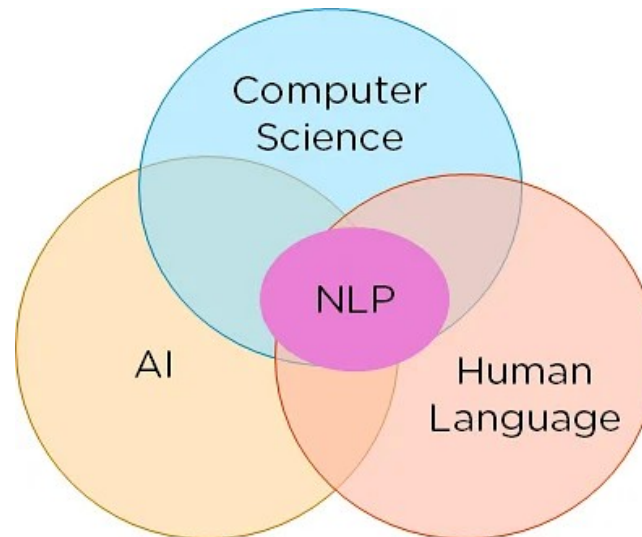
Phases of NLP
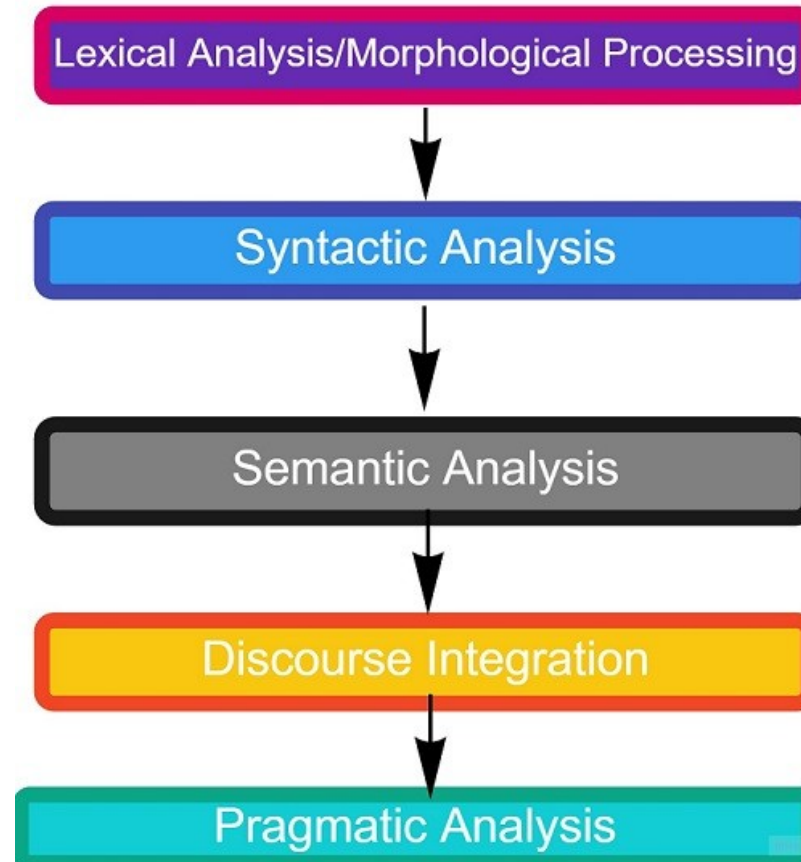
- The first phase is lexical analysis/morphological processing. In this phase, the sentences, paragraphs are broken into tokens.

- These tokens are the smallest unit of text. It scans the entire source text and divides it into meaningful lexemes.

- For example, The sentence "He goes to college." is divided into ['He' , 'goes' , 'to' , 'college', '.'] .

- There are five tokens in the sentence. A paragraph may also be divided into sentences.

Lexical Analysis

TEXT

A paragraph is a distinct section of writing covering one topic. A paragraph will usually contain more than one sentence. A paragraph starts on a new line. Sometimes, paragraphs are indented or numbered. ... It will have detail sentences in the middle and end with a concluding sentence.

SENTENCE DETECTION

['A paragraph is a distinct section of writing covering one topic.',
'A paragraph will usually contain more than one sentence.',
'A paragraph starts on a new line. Sometimes, paragraphs are indented or numbered.',
'It will have detail sentences in the middle and end with a concluding sentence.']

Syntactic Analysis/Parsing

- The second phase is Syntactic analysis. In this phase, the sentence is checked whether it is well-formed or not.

- The word arrangement is studied and a syntactic relationship is found between them. It is checked for word arrangements and grammar.

- For example, the sentence "Delhi goes to him" is rejected by the syntactic parser.

Semantic Analysis

- The third phase is Semantic Analysis. In this phase, the sentence is checked for the literal meaning of each word and their arrangement together.

- For example, The sentence "I ate hot ice cream" will get rejected by the semantic analyzer because it doesn't make sense.

- E.g.. "colorless green idea." This would be rejected by the Symantec analysis as colorless Here; green doesn't make any sense.

Discourse Integration

- The fourth phase is discourse integration. In this phase, the impact of the sentences before a particular sentence and the effect of the current sentence on the upcoming sentences is determined.

- For example, the word "that" in the sentence "He wanted that" depends upon the prior discourse context.

Pragmatic Analysis

- The last phase of natural language processing is Pragmatic analysis. Sometimes the discourse integration phase and pragmatic analysis phase are combined.

- The actual effect of the text is discovered by applying the set of rules that characterize cooperative dialogues.

- E.g., "close the window?" should be interpreted as a request instead of an order.
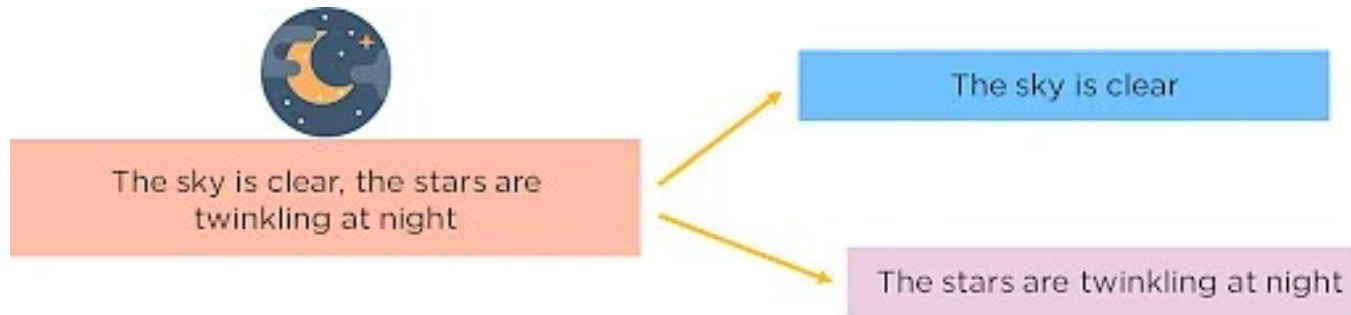
NLP Implementation

- Below, given are popular methods used for Natural Learning Process:

  - Machine learning: The learning nlp procedures used during machine learning. It automatically focuses on the most common cases. So when we write rules by hand, it is often not correct at all concerned about human errors.

    Statistical inference: NLP can make use of statistical inference algorithms. It helps you to produce models that are robust. e.g., containing words or structures which are known to everyone.

tusharkute
.com

NLP Steps

- How to Perform NLP?
  - Segmentation
  - Tokenizing
  - Removing Stop Words:
  - Stemming
  - Lemmatization
  - Part of Speech Tagging
  - Named Entity Tagging

Segmentation

- You first need to break the entire document down into its constituent sentences. You can do this by segmenting the article along with its punctuation like full stops and commas.

The sky is clear, the stars are twinkling at night → The sky is clear / The stars are twinkling at night

Tokenizing

- For the algorithm to understand these sentences, you need to get the words in a sentence and explain them individually to our algorithm.

- So, you break down your sentence into its constituent words and store them. This is called tokenizing, and each world is called a token.

Removing Stop Words

- You can make the learning process faster by getting rid of non-essential words, which add little meaning to our statement and are just there to make our statement sound more cohesive. Words such as was, in, is, and, the, are called stop words and can be removed.

The star are twinkling at night

stars   twinkling   night

Stemming

- It is the process of obtaining the Word Stem of a word. Word Stem gives new words upon adding affixes to them
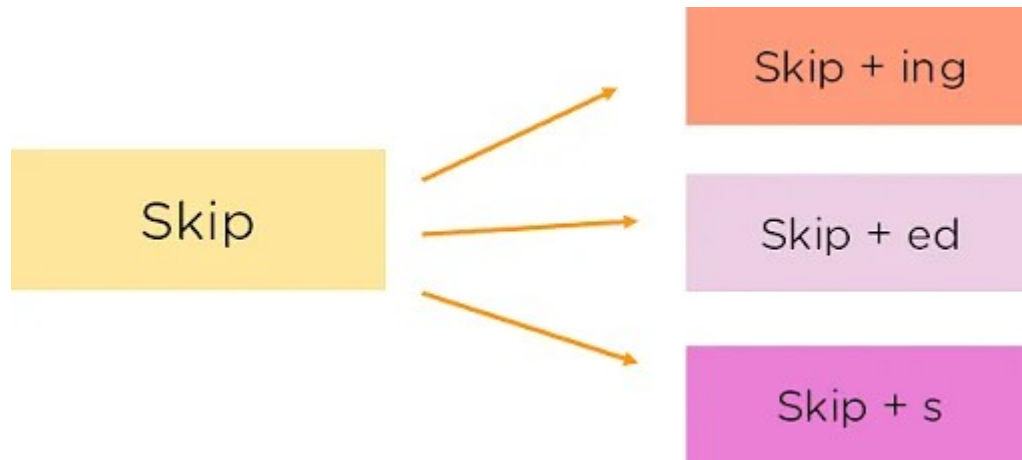
Lemmatization

- The process of obtaining the Root Stem of a word. Root Stem gives the new base form of a word that is present in the dictionary and from which the word is derived. You can also identify the base words for different words based on the tense, mood, gender,etc.
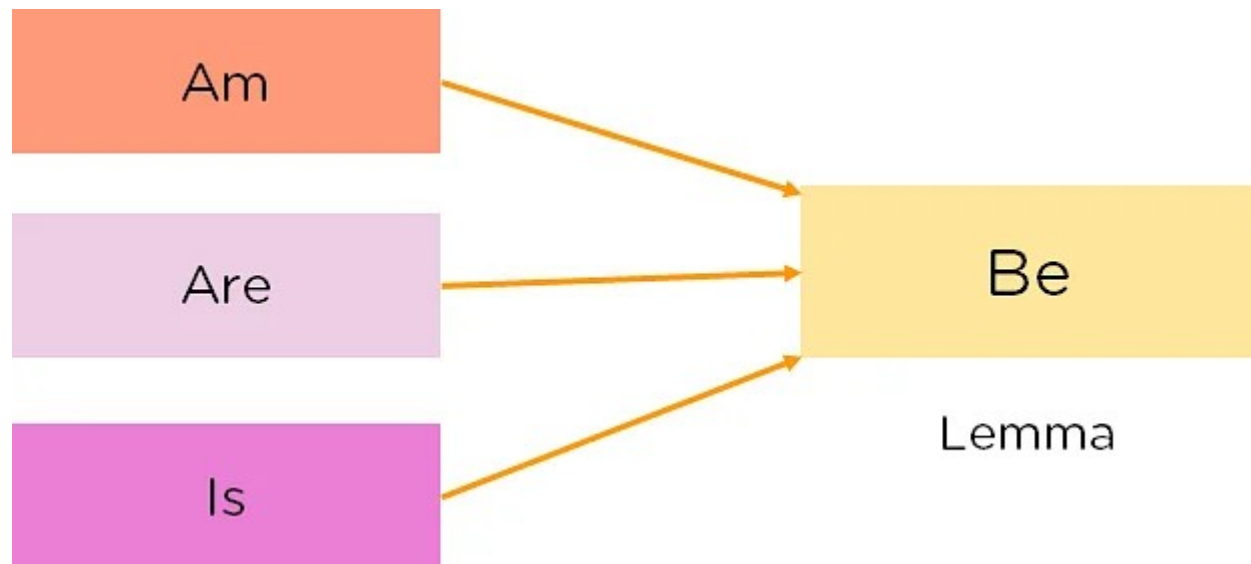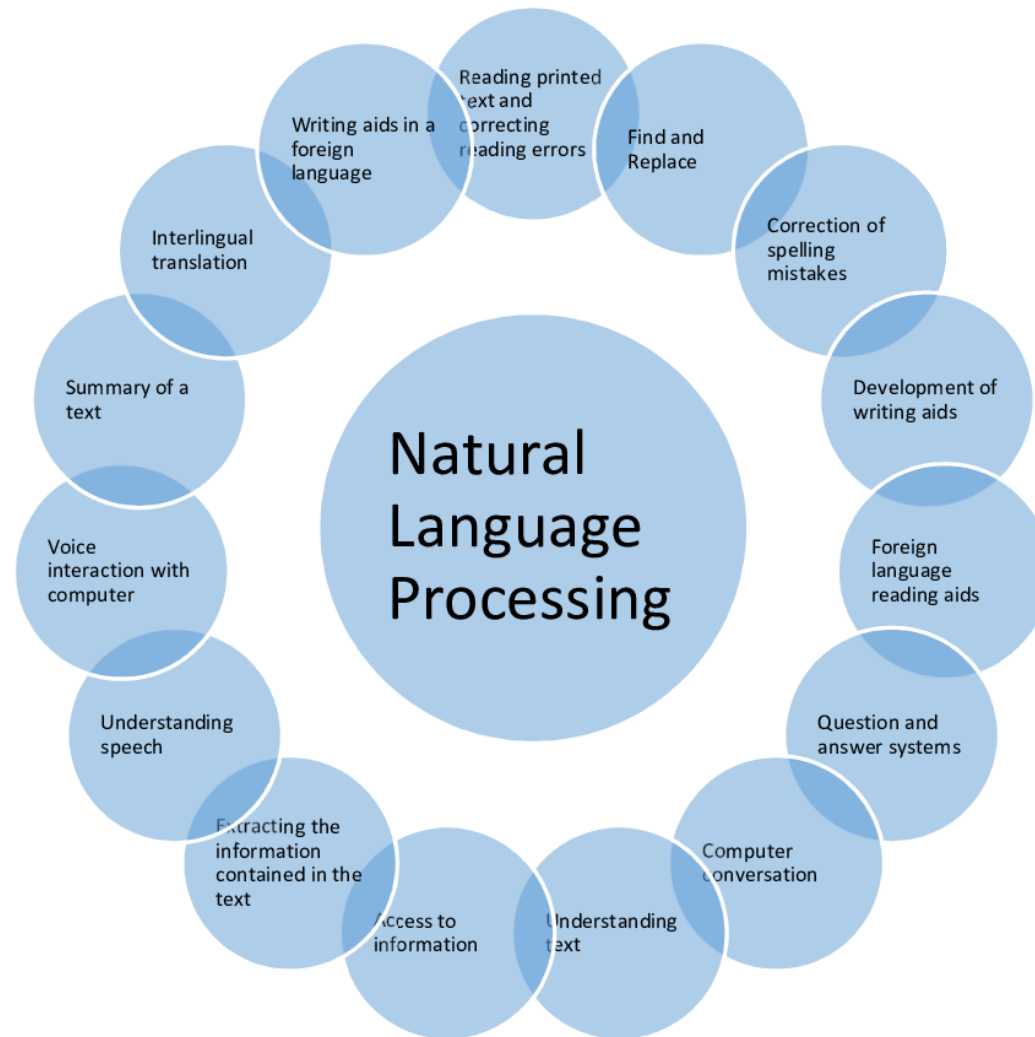
- Now, you must explain the concept of nouns, verbs, articles, and other parts of speech to the machine by adding these tags to our words. This is called 'part of'.

Determiner	Noun	Verb	Adjective	Preposition	Noun
↑	↑	↑	↑	↑	↑
The	stars	are	twinkling	at	night
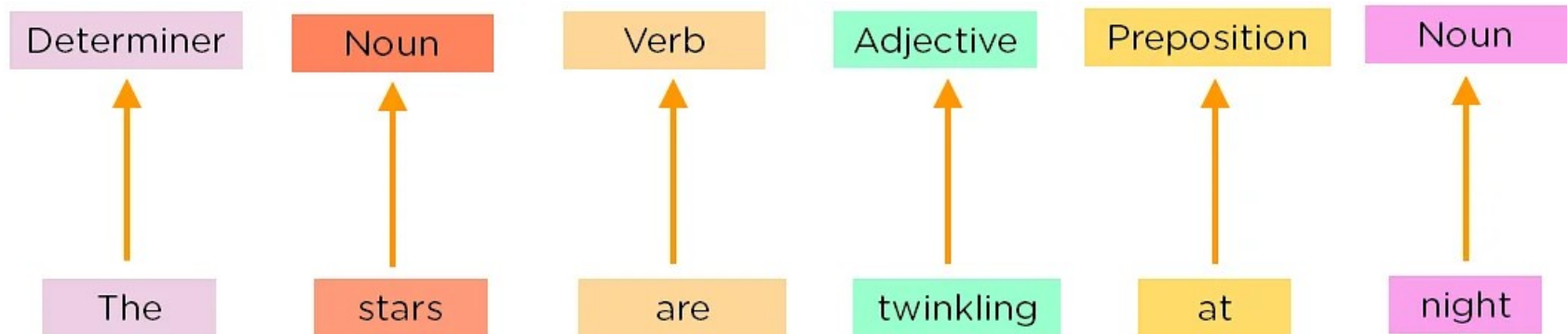
Named Entity Tagging

- Next, introduce your machine to pop culture references and everyday names by flagging names of movies, important personalities or locations, etc that may occur in the document.

- You do this by classifying the words into subcategories. This helps you find any keywords in a sentence. The subcategories are person, location, monetary value, quantity, organization, movie.

- After performing the preprocessing steps, you then give your resultant data to a machine learning algorithm like Naive Bayes, etc., to create your NLP application.

Natural Language Processing

- Reading printed text and correcting reading errors
- Find and Replace
- Correction of spelling mistakes
- Development of writing aids
- Foreign language reading aids
- Question and answer systems
- Computer conversation
- Understanding text
- Access to information
- Extracting the information contained in the text
- Understanding speech
- Voice interaction with computer
- Summary of a text
- Interlingual translation
- Writing aids in a foreign language

- NLP is one of the ways that people have humanized machines and reduced the need for labor. It has led to the automation of speech-related tasks and human interaction. Some applications of NLP include :

  - Translation Tools: Tools such as Google Translate, Amazon Translate, etc. translate sentences from one language to another using NLP.

  - Chatbots: Chatbots can be found on most websites and are a way for companies to deal with common queries quickly.

Applications of NLP

- Virtual Assistants: Virtual Assistants like Siri, Cortana, Google Home, Alexa, etc can not only talk to you but understand commands given to them.

- Targeted Advertising: Have you ever talked about a product or service or just googled something and then started seeing ads for it? This is called targeted advertising, and it helps generate tons of revenue for sellers as they can reach niche audiences at the right time.

- Autocorrect: Autocorrect will automatically correct any spelling mistakes you make, apart from this grammar checkers also come into the picture which helps you write flawlessly.

- Information retrieval & Web Search: Google, Yahoo, Bing, and other search engines base their machine translation technology on NLP deep learning models. It allows algorithms to read text on a webpage, interpret its meaning and translate it to another language.

- Grammar Correction:

- NLP technique is widely used by word processor software like MS-word for spelling correction & grammar check.

Applications of NLP

- Question Answering
  - Type in keywords to ask Questions in Natural Language.
- Text Summarization
  - The process of summarising important information from a source to produce a shortened version
- Machine Translation
  - Use of computer applications to translate text or speech from one natural language to another.

tusharkute
.com

Future of NLP

- Human readable natural language processing is the biggest Al- problem. It is all most same as solving the central artificial intelligence problem and making computers as intelligent as people.

- Future computers or machines with the help of NLP will able to learn from the information online and apply that in the real world, however, lots of work need to on this regard.

- Natural language toolkit or nltk become more effective

- Combined with natural language generation, computers will become more capable of receiving and giving useful and resourceful information or data.

Natural Language vs. Computer Language

Parameter	Natural Language	Computer Language
Ambiguous	They are ambiguous in nature.	They are designed to unambiguous.
Redundancy	Natural languages employ lots of redundancy.	Formal languages are less redundant.
Literalness	Natural languages are made of idiom & metaphor	Formal languages mean exactly what they want to say

Advantages of NLP

- Users can ask questions about any subject and get a direct response within seconds.

- NLP system provides answers to the questions in natural language

- NLP system offers exact answers to the questions, no unnecessary or unwanted information

- The accuracy of the answers increases with the amount of relevant information provided in the question.

- NLP process helps computers communicate with humans in their language and scales other language-related tasks

- Allows you to perform more language-based data compares to a human being without fatigue and in an unbiased and consistent way.

- Structuring a highly unstructured data source

Disadvantages of NLP

- Complex Query Language- the system may not be able to provide the correct answer it the question that is poorly worded or ambiguous.

- The system is built for a single and specific task only; it is unable to adapt to new domains and problems because of limited functions.

- NLP system doesn't have a user interface which lacks features that allow users to further interact with the system

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
http://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com