

Module 5 : AI Applications

Natural Language Processing

Language is a method of communication with the help of which we can speak, read and write. Natural Language Processing (NLP) is a subfield of Computer Science that deals with Artificial Intelligence (AI), which enables computers to understand and process human language. Natural Language Processing is the technology used to aid computers to understand the human's natural language.

Components of NLP

There are two components of NLP as given –

1 Natural Language Understanding (NLU) Understanding involves the following tasks –

- Mapping the given input in natural language into useful representations.
- Analyzing different aspects of the language.

2 Natural Language Generation (NLG)- It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation. It involves –

- Text planning – It includes retrieving the relevant content from the knowledge base.
- Sentence planning – It includes choosing required words, forming meaningful phrases, setting tone of the sentence.
- Text Realization – It is mapping sentence plan into sentence structure.

Steps in Natural language processing

1. Tokenization's
2. Stemming
3. Lemmatization
4. POS tags
5. Named Entity Recognition

Tokenization

- Cutting the big sentence into small tokens
- Example : Welcome to Last moment tuitions will be divided into tokens [welcome] [to] [last] [moment] [tuitions]

Stemming

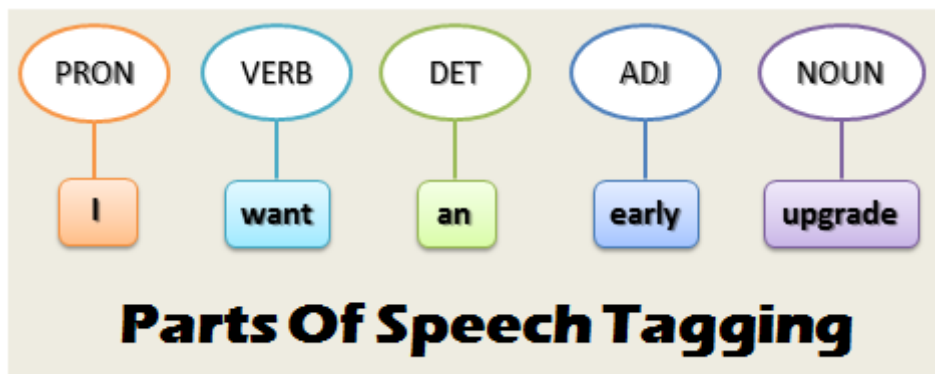
- Normalize words into its base or root forms
- Waits, waited, waiting -> wait (In this example the root meaning or word is wait so instemming we cut remaining part and find the root word)

Lemmatization

- Group together different inflected forms of a word called Lemmatization
- Somehow similar to stemming, as it maps several words into one common root
- Output of lemmatization is a proper word
- Example : Gone, going and went -> Go

POS Tags

- POS stands for Parts of speech tags
- It indicated how a word function in meaning as well as grammatically within a sentence



But problem in POS tag is sometime one word can have different meaning example “ google something on internet” now we know google is a proper noun but here it is use as verb

To solve this problem we use Named entity recognition .

Named Entity Recognition

- It is a process of name entity such as: person , organization , location ,quantities ,etc.

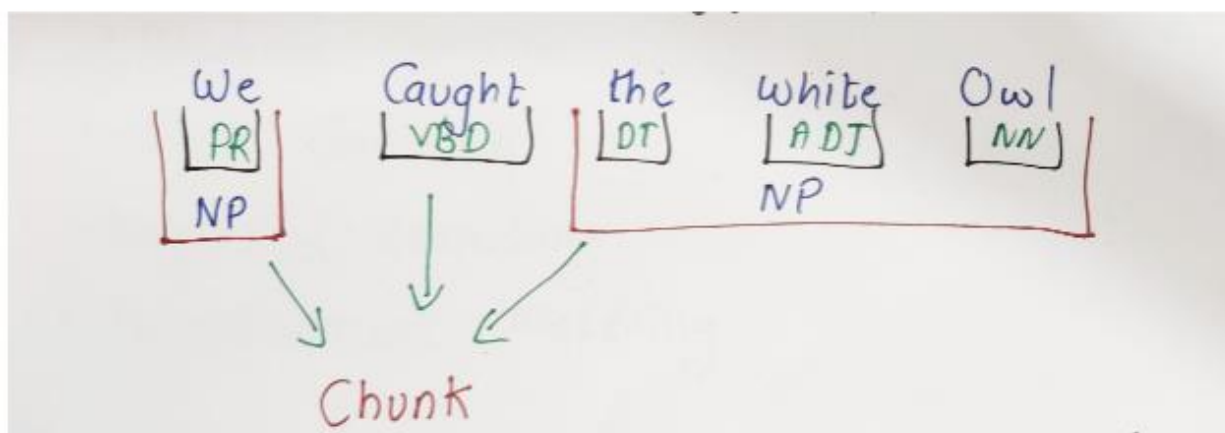
Example



Here Apple is an Organization, Tim cook is a person and cupertino flint center is location. In NER we get exact information like apple is a organization not a fruit.

Chunking

- Picking individual pieces of information and grouping them into bigger pieces.



- This helps get insight and meaningful information from the text

Parsing

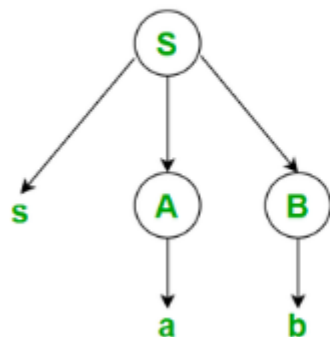
Parsing is a fundamental concept in Natural Language Processing (NLP) that involves analyzing the grammatical structure of a sentence. It refers to the process of breaking down a sentence into its component parts and determining how those parts relate to each other.

In NLP, parsing is typically done using a parser, which is a program that analyzes the syntactic structure of a sentence and produces a parse tree that shows how the sentence can be broken down into its constituent parts. The parse tree represents the structure of the sentence in a hierarchical form, with the root node representing the entire sentence, and the child nodes representing the phrases and words that make up the sentence.

Concept of Parsing

- It is a graphical representation of derivation
- Start symbol is a root of parse tree
- Leaf nodes are terminals
- Interior Nodes are Non-terminals
- If parsed properly it will create input text

Example-1:



Let us take an example of a Grammar (Production Rules).

$S \rightarrow sAB$

$A \rightarrow a$

$B \rightarrow b$

The input string is "sab", then the Parse Tree is :

Types of Parsing

Derivation divides parsing into the followings two types –

- Top-down Parsing
- Bottom-up Parsing

Top-down Parsing

In this kind of parsing, the parser starts constructing the parse tree from the start symbol and then tries to transform the start symbol to the input. The most common form of top down parsing uses a recursive procedure to process the input. The main disadvantage of recursive descent parsing is backtracking.

Working of Top Down Parser :

Let's consider an example where grammar is given and you need to construct a parse tree by using top down parser technique.

Example –

$S \rightarrow aABe$

$A \rightarrow Abc \mid b$

$B \rightarrow d$

Now, let's consider the input to read and to construct a parse tree with top down approach.

Input –

abbcd\$

Now, you will see how the top down approach works. Here, you will see how you can generate an input string from the grammar for a top down approach.

- First, you can start with $S \rightarrow aABe$ and then you will see input string a in the beginning and e in the end.
- Now, you need to generate $abbcd$.
- Expand $A \rightarrow Abc$ and Expand $B \rightarrow d$.
- Now, You have a string like $aAbcde$ and your input string is $abbcd$.
- Expand $A \rightarrow b$.
- Final string, you will get $abbcd$.

Bottom-up Parsing

In this kind of parsing, the parser starts with the input symbol and tries to construct the parser tree up to the start symbol.

Example:

Input string : $a + b * c$

Production rules:

$S \rightarrow E$

$E \rightarrow E + T$

$E \rightarrow E * T$

$E \rightarrow T$

$T \rightarrow id$

Let us start bottom-up parsing

$a + b * c$

Read the input and check if any production matches with the input:

$a + b * c$

$T + b * c$

$E + b * c$

$E + T * c$

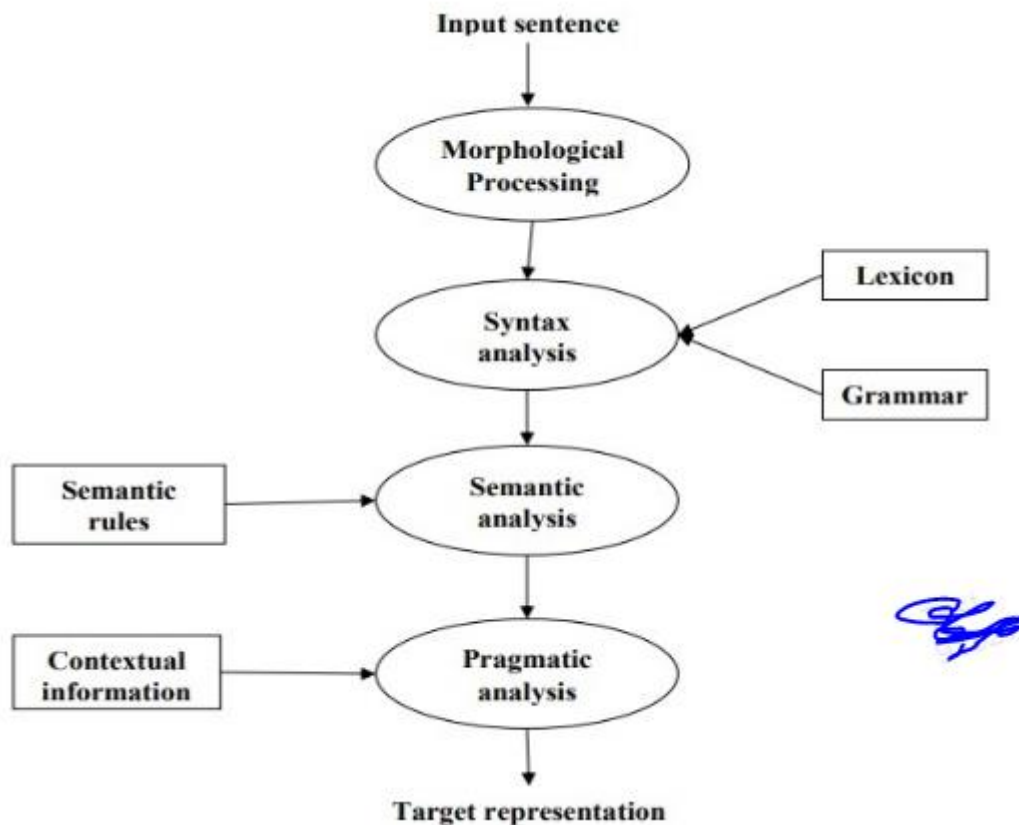
$E * c$

$E * T$

E

S

Phases of NLP (Natural Language Processing)



Morphological Processing

- It is the first phase of NLP.
- The purpose of this phase is to break chunks of language input into sets of tokens corresponding to paragraphs, sentences and words.
- For example, a word like “uneasy” can be broken into two sub-word tokens as “un-easy”

Syntax Analysis

It is the second phase of NLP.

The purpose of this phase is two folds:

1. To check that a sentence is well formed or not
 - Eg: “The school goes to the boy” would be rejected by syntax analyzer or parser.
2. Taking input data and giving structural representation of input data.
 - Example : “ The chef cooks the soup”

Semantic Analysis

- It is the third phase of NLP.
- The purpose of this phase is to draw exact meaning, or you can say dictionary meaning from the text.
- The text is checked for meaningfulness.
- For example, semantic analyzer would reject a sentence like “Hot ice-cream”.

Pragmatic Analysis

- It is the fourth phase of NLP.
- Pragmatic analysis simply fits the actual objects/events, which exist in a given context with object references obtained during the last phase (semantic analysis).
- For example, the sentence “The car hit the pole while it was moving” can have two semantic interpretations and pragmatic analyzer will choose between these two possibilities

Application of NLP

Application of Natural Language Processing

- Machine Translation
- Sentimental Analysis
- Automatic Summarization
- Question answering
- Speech recognition

Machine Translation

- Machine translation (MT), the process of translating one source language or text into another language, is one of the most important applications of NLP.
- There are different types of machine translation systems. Let us see what the different types are.
- Bilingual MT systems produce translations between two particular languages.
- Multilingual MT systems produce translations between any pair of languages. They may be either unidirectional or bi-directional in nature.
- Example : Google translator

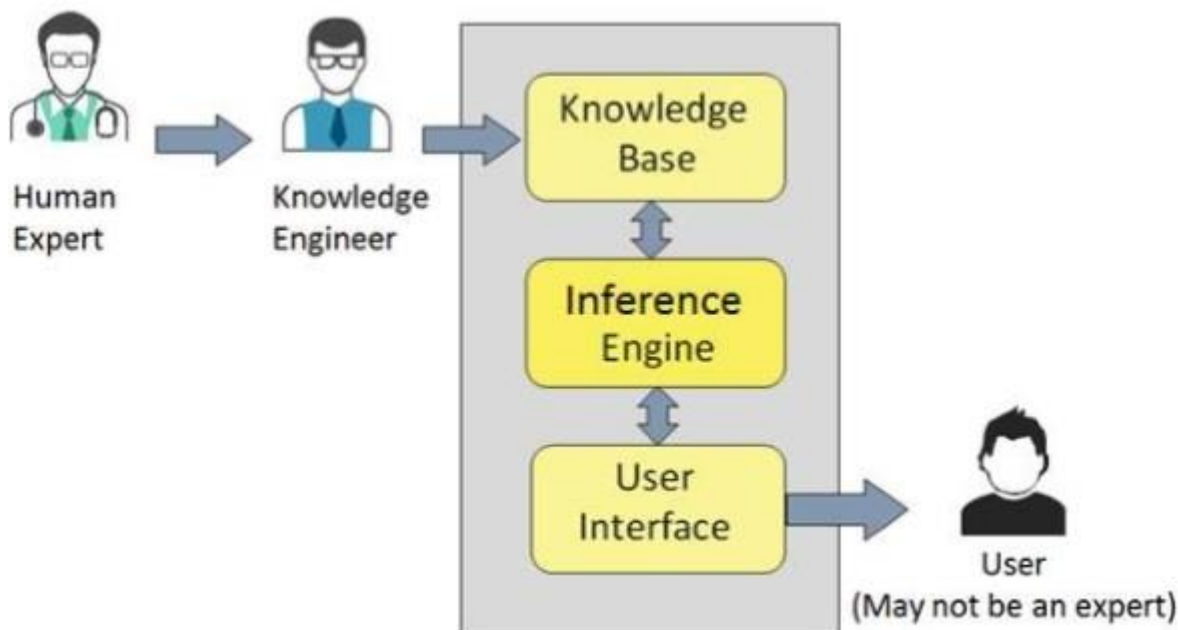
Sentiment Analysis

- Another important application of natural language processing (NLP) is sentiment analysis.
- As the name suggests, sentiment analysis is used to identify the sentiments among several posts.
- It is also used to identify the sentiment where the emotions are not expressed explicitly.
- Companies are using sentiment analysis, an application of natural language processing (NLP) to identify the opinion and sentiment of their customers online.
- It will help companies to understand what their customers think about the products and services.
- Companies can judge their overall reputation from customer posts with the help of sentiment analysis.
- In this way, we can say that beyond determining simple polarity, sentiment analysis understands sentiments in context to help us better understand what is behind the expressed opinion.

Automatic Summarization

- In this digital era, the most valuable thing is data, or you can say information.
 - However, do we really get useful as well as the required amount of information?
- The answer is 'NO' because the information is overloaded and our access to knowledge and information far exceeds our capacity to understand it.
- We are in a serious need of automatic text summarization and information because the flood of information over the internet is not going to stop.
 - Text summarization may be defined as the technique to create short, accurate summary of longer text documents.
 - Automatic text summarization will help us with relevant information in less time. Natural language processing (NLP) plays an important role in developing an automatic text summarization. Question-answering
 - Another main application of natural language processing (NLP) is question-answering. Search engines put the information of the world at our fingertips, but they are still lacking when it comes to answering the questions posted by human beings in their natural language.
 - We have big tech companies like Google are also working in this direction.
 - Question-answering is a Computer Science discipline within the fields of AI and NLP.
 - It focuses on building systems that automatically answer questions posted by human beings in their natural language.
 - A computer system that understands the natural language has the capability of a program system to translate the sentences written by humans into an internal
 - The exact answers can be generated by doing syntax and semantic analysis of the questions. Lexical gap, ambiguity and multilingualism are some of the challenges for NLP in building a good question answering system

Architecture of Expert System and what are the Limitations of Expert System



An expert system is an artificial intelligence system that uses a knowledge base and inference engine to solve complex problems that would typically require human expertise. The architecture of an expert system consists of several components, each of which plays a crucial role in the functioning of the system. Here's a brief description of each component:

1. **Knowledge Base:** This component is the repository of knowledge that an expert system uses to make decisions. It consists of a collection of facts, rules, and heuristics that have been acquired through the expertise of domain experts. The knowledge base is typically organized into a hierarchical structure that allows the system to access and use the relevant knowledge quickly.
2. **Inference Engine:** This component is responsible for reasoning and making decisions based on the knowledge contained in the knowledge base. It uses various techniques, such as forward and backward chaining, to infer new information from the knowledge base and apply it to the problem at hand.
3. **User Interface:** This component allows users to interact with the expert system and provide input or receive output. The user interface can be text-based or graphical, depending on the complexity of the problem and the user's expertise.
4. **Explanation Facility:** This component provides the user with explanations of the system's decision-making process. It helps to build trust in the system by providing transparency and allowing the user to understand how the system arrived at a particular decision.

Limitations of Expert Systems:

1. **Limited domain expertise:** Expert systems are designed to solve problems within a specific domain, and their knowledge base is limited to that domain. They may not be able to handle problems outside of their domain.
2. **Lack of common sense reasoning:** Expert systems lack common sense reasoning and may not be able to deal with situations that humans would consider obvious.
3. **Dependence on domain experts:** The development and maintenance of expert systems require a high level of expertise from domain experts. As such, they can be expensive and time-consuming to develop.
4. **Inflexibility:** Expert systems are often designed to solve specific problems and may not be able to adapt to changing circumstances or new information.