

Natural Language Processing



Natural Language Processing

Content:

1. Introduction
2. Components of NLP
3. NLP Technology
4. Steps in NLP
5. Context - free grammar
6. Top - down parser

It refers to AI process of communicating with an intelligent systems using a natural language such as English.

It involves when you want an intelligent system like robot to perform as per your orders, when you want to listen decision from a dialogue based clinical expert system, etc.

It requires making computers to execute tasks with the natural languages humans use. The input and output of an NLP system are:

- Speech
- Written Text

Components of NLP are:

Natural Language Understanding (NLU)

Understanding involves the following tasks –

- Mapping the given input in natural language into functional representations.
- Analysing different aspects of the language.



Natural Language Generation (NLG)

It is the method of generating meaningful phrases and sentences in the form of natural language from internal representation.

It involves –

- **Text planning** – It includes recover the applicable content from knowledge base.
- **Sentence planning** – It involves selecting required words, forming meaningful phrases.
- **Text Realization** – It is mapping sentence plan into sentence formation.

The NLU is harder than NLG.

Difficulties involved in NLU

It has especially rich form and structure.

It is ambiguous. There can be different levels of ambiguity –

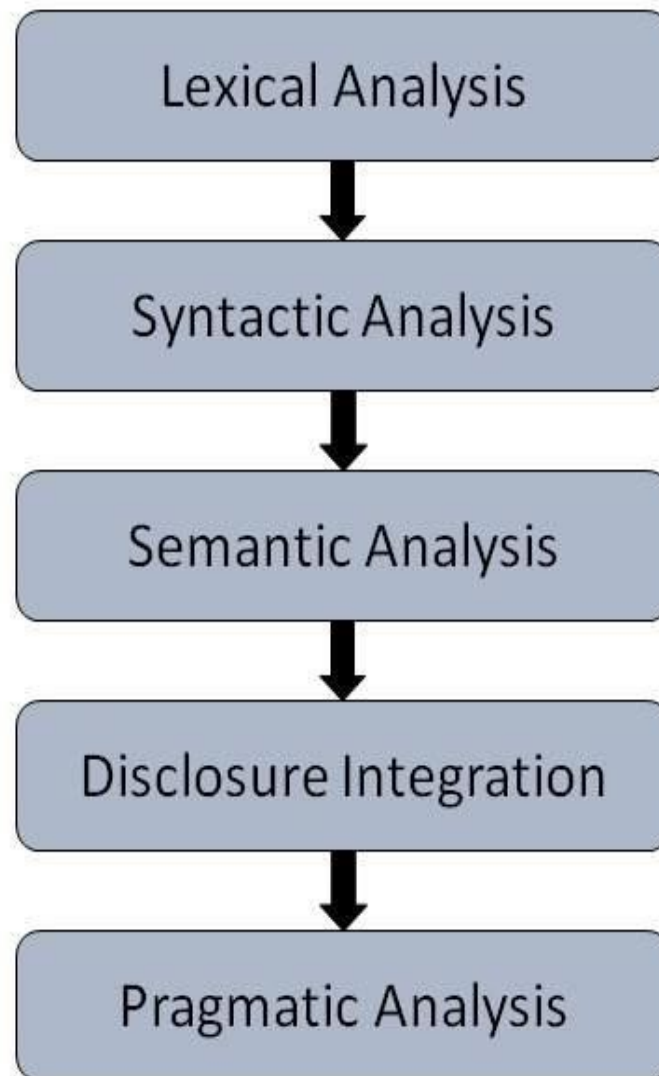
- **Lexical ambiguity** – It is at primitive level such as word-level.
- For example, treating the word “board” as a noun or a verb?
- **Syntax Level ambiguity** – A sentence can defined in different ways.
- For example, “He lifted the beetle with red cap.” – Did he use the cap to lift the beetle or he lifted a beetle that had red cap?
- **Referential ambiguity** – Referring to something using pronouns. For example, Rima went to Gauri. She said, “I am tired.” – Exactly who is tired?
- One input can mean different meanings.
- Many inputs can mean the same thing.

NLP Terminology

- **Phonology** – It is a examination of organizing sound systematically.
- **Morphology** – It is a examination of construction of words from primitive meaningful units.
- **Morpheme** – It is earliest unit of meaning in a language.
- **Syntax** – It refers to order words to make a sentence. It requires to determining the structural role of words in the sentence and in phrases.
- **Semantics** – It is worried with the meaning of words and how to combine words into meaningful phrases and sentences.



- **Pragmatics** – It interacts with using sentences in unlike situations and how the explanation of the sentence is affected.
- **Discourse** – It interacts with how the immediately previous sentence can affect the interpretation of the next sentence.
- **World Knowledge** – It involves the general knowledge about the world.



Steps in NLP

There are five steps in NLP:

- **Lexical Analysis** – It requires to identifying and analysing the structure of words. It means the group of words and phrases in a language. It is dividing the whole chunk of txt into paragraphs, sentences, and words.

- **Syntactic Analysis (Parsing)** – It requires to analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as “The school goes to boy” is rejected by English syntactic analyser.
- **Semantic Analysis** – It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done with mapping syntactic structures and objects in the task domain. The semantic analyser ignore sentence such as “hot ice-cream”.
- **Discourse Integration** – The meaning of sentence turn on upon the meaning of the sentence just before it. It brings about the meaning of immediately succeeding sentence.
- **Pragmatic Analysis** – During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

Implementation Aspects of Syntactic Analysis

There are many algorithms researchers who have developed for syntactic analysis, but we determine only the following simple methods –

- Context-Free Grammar
- Top-Down Parser

Let us see them in detail –

Context-Free Grammar

It is a grammar that contains rules with a single symbol on the left-hand side of the rewrite rules. Let us create a grammar to define a sentence –

“The bird pecks the grains”

Articles (DET) – a | an | the

Nouns – bird | birds | grain | grains

Noun Phrase (NP) – Article + Noun | Article + Adjective + Noun

= DET N | DET ADJ N

Verbs – pecks | pecking | pecked

Verb Phrase (VP) – NP V | V NP

Adjectives (ADJ) – beautiful | small | chirping



The parse tree smash down the sentence into formation parts so that the computer can easily understand and process it. In order for the parsing algorithm to construct this parse tree, a set of rewrite rules, which describe what tree structures are legal, need to be constructed.

These rules say that a certain symbol may be enlarge in tree by a sequence of other symbols. According to first order logic rule, if there are two strings Noun Phrase (NP) and Verb Phrase (VP), then the string combined by NP followed by VP is a sentence. The rewrite rules for the sentence are as follows –

S → NP VP

NP → DET N | DET ADJ N

VP → V NP

Lexocon –

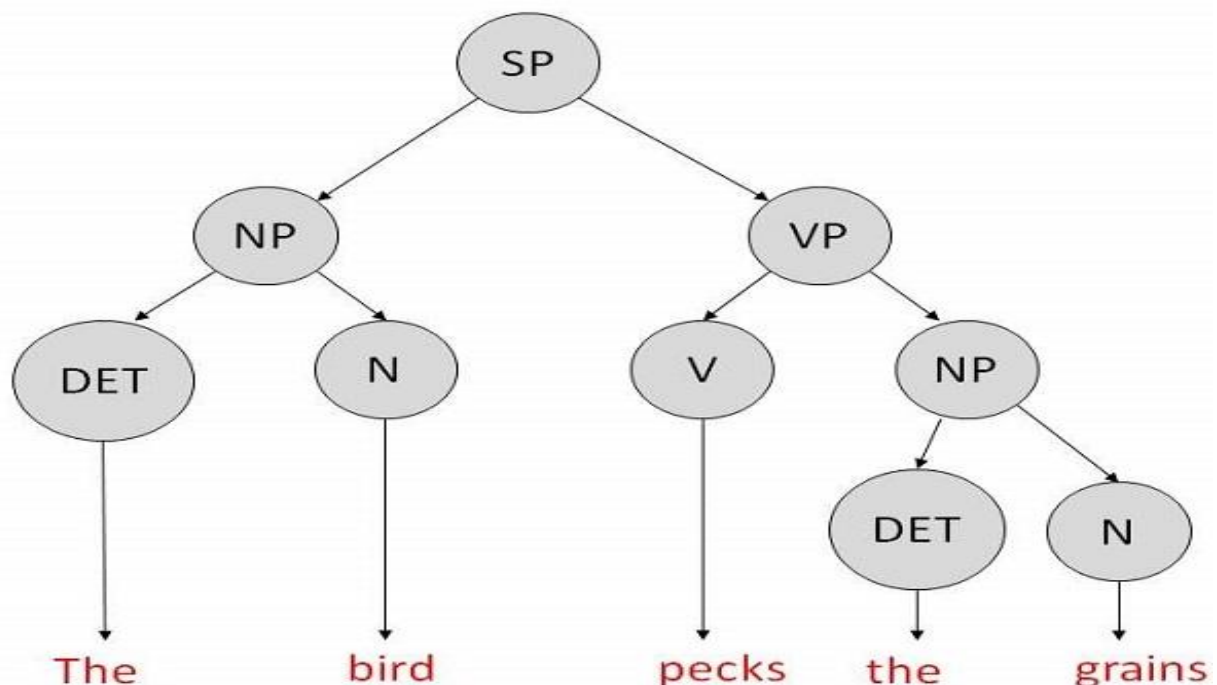
DET → a | the

ADJ → beautiful | perching

N → bird | birds | grain | grains

V → peck | pecks | pecking

The parse tree can be created as shown –



Now consider the above rewrite rules. Since V can be replaced by both, "peck" or "pecks", sentences such as "The bird peck the grains" can be wrongly permitted. i. e. the subject-verb agreement error is approved as correct.

Merit – The simplest style of grammar, therefore extensively used one.

Demerits –

- They are not highly precise. For example, "The grains peck the bird", is a syntactically correct according to parser, but even if it makes no sense, parser takes it as a correct sentence.
- To bring out high precision, multiple sets of grammar need to be prepared. It require a completely different group of rules for parsing singular and plural variations, passive sentences, etc., which can lead to creation of huge set of rules that are unmanageable.

Top-Down Parser

Here, the parser starts with the S symbol and attempts to rewrite it into a sequence of *terminal symbols* that matches the classes of the words in the input sentence until it consists entirely of terminal symbols.

These are then checked with the input sentence to see if it matched. If not, the process is started over again with a different set of rules. This is repeated until a specific rule is found which describes the structure of the sentence.

Merit – It is simple to implement.

Demerits –

- It is inefficient, as the search process has to be repeated if an error occurs.
- Slow speed of working.



Gradeup UGC NET Super Superscription

Features:

1. 7+ Structured Courses for UGC NET Exam
2. 200+ Mock Tests for UGC NET & MHSET Exams
3. Separate Batches in Hindi & English
4. Mock Tests are available in Hindi & English
5. Available on Mobile & Desktop

Gradeup Super Subscription, Enroll Now