Ch-12 : Natural Language Processing

Q1: Describe Natural Language Processing?

Natural Language Processing (NLP) refers to AI method of communicating with intelligent systems using a natural language such as English.

Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.

The field of NLP involves making computers to perform useful tasks with the natural languages humans use. The input and output of an NLP system can be −

- Speech
- Written Text

There are two components of NLP as given −

Natural Language Understanding (NLU)

Understanding involves the following tasks − Mapping the given input in natural language into useful representations and Analyzing different aspects of the language.

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 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.

The NLU is harder than NLG.

Difficulties in NLU

NL has an extremely rich form and structure.

It is very ambiguous. There can be different levels of ambiguity −

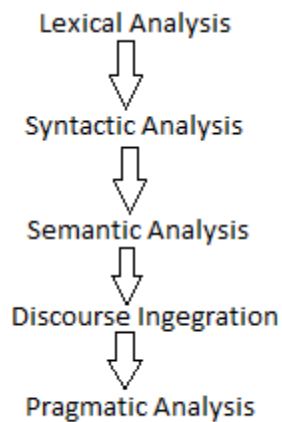- **Lexical ambiguity** − It is at very primitive level such as word-level.

- For example, treating the word "board" as noun or verb?

- **Syntax Level ambiguity** − A sentence can be parsed in different ways.

- For example, "He lifted the beetle with red cap." − Did he use 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 study of organizing sound systematically.

- **Morphology** − It is a study of construction of words from primitive meaningful units.

- **Morpheme** − It is primitive unit of meaning in a language.

- **Syntax** − It refers to arranging words to make a sentence. It also involves determining the structural role of words in the sentence and in phrases.

- **Semantics** − It is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.

- **Pragmatics** − It deals with using and understanding sentences in different situations and how the interpretation of the sentence is affected.

- **Discourse** − It deals with how the immediately preceding sentence can affect the interpretation of the next sentence.

- **World Knowledge** − It includes the general knowledge about the world.

Q2: Describe various steps of Natural Language Processing.

Steps in NLP:

Lexical Analysis
⬇
Syntactic Analysis
⬇
Semantic Analysis
⬇
Discourse Ingegration
⬇
Pragmatic Analysis

Lexical Analysis − It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

Syntactic Analysis (Parsing) − It involves 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 analyzer.

It is the grammar that consists rules with a single symbol on the left-hand side of the rewrite rules. Let us create grammar to parse 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 breaks down the sentence into structured 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 expanded in the 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

DET → a | the

ADJ → beautiful | perching

N → bird | birds | grain | grains

V → peck | pecks | pecking

The parse tree can be created as shown –

[The bird pecks the grains]
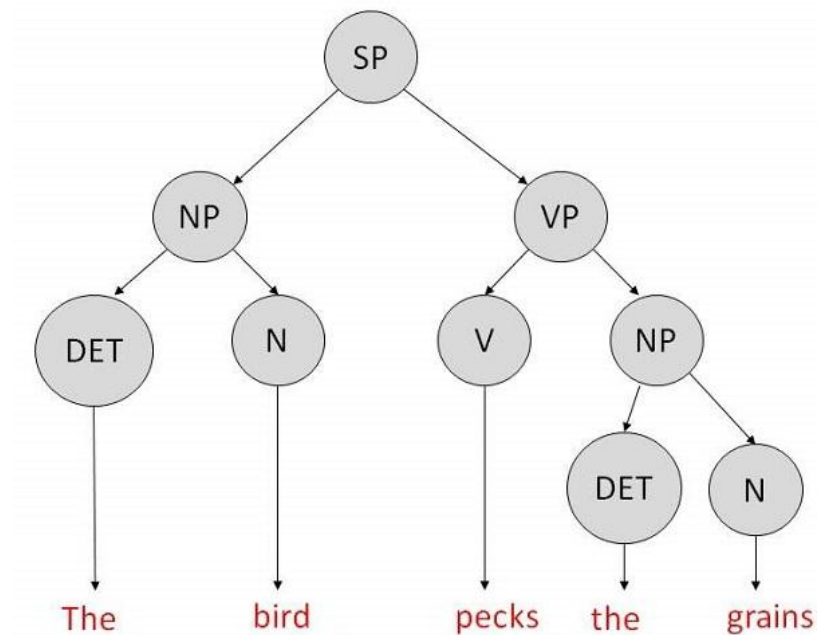
Semantic Analysis − It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as "hot ice-cream".

Discourse Integration − The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence. To pin down these references requires an appeal to a model of the current discourse context, from which we can learn that the current user is USER068 and that the only person named "Bill" about whom we could be talking is USER073.

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. The final step toward effective understanding is to decide what to do as a result. One possible thing to do is to record what was said as a fact and be done with it. For some sentences, whose intended effect is clearly declarative, that is precisely correct thing to do. But for other sentences, including this one, the intended effect is different. We can discover this intended effect by applying a set of rules that characterize cooperative dialogues. The final step in pragmatic processing is to translate, from the knowledge based representation to a command to be executed by the system.

Q3: What features of Natural Language makes it difficult to process using computing systems?

➔ Problem 1: English sentences are incomplete descriptions of the information that they are intended to convey.
➔ Problem 2: The same expression means different things in different context.
➔ Problem 3: No natural language program can be complete because new words, expressions, and meanings can be generated quite freely.