IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 69-74

Study of Natural Language Processing and Its Factors

Monica N.P Tiwari[1], Prof. Rupali Chikhale[2]
[1]*Department Of Computer Science (MCA), MCA Sem-5, GHRIIT, Nagpur*
[2]*Department Of Computer Science (MCA), GHRIIT, Nagpur*

Abstract: *The Natural language processing is a combination of computer science and artificial intelligence. Basically it is related to artificial intelligence. It is concerned that interaction between the computers and human languages. These include the spoken language systems that integrate speech and natural language. Natural language processing has a role in computer science because many aspects of the field deal with linguistic(related language) features of computation. Natural language processing is an area of an application and research to explore how the computers can be used to manipulates and understands natural language text or speech to do useful things.The paper has been distributed in four phases by discussing different levels of NLP and components ofNatural Language Generation (NLG) the history NLP, state ofthe art presenting the various applications of NLP and current trends andits factors.*

Keywords:*Natural Language Processing; Aspects of Syntactic Analysis; Component of NLP; Matrix*

I. Introduction

Natural Language Processing (NLP) is a sub-field of Artificial Intelligence that is focused on enabling computers to understand and process human languages, to get computers closer to a human-level understanding of language. Natural Language Processing (NLP) is an area of Artificial Intelligence and Linguistics. It is devoted to make computers understand the words or statement written in human languages. It is the process to "understand" natural language command to perform operations like Language Translation and Question Answering.

A. Outline of the Research

This research is commonly depends on the "Applicationsof theNatural Language Processing and its factors". The first section of this study isproviding about an "introduction" to the topic that discussed and defined the background of natural language process. This part will also focus on the aims of the study. The second sections contain levels of Natural language processing. The third one is its metrics and its tables. Natural Language Processing (NLP) assign to AI method of communicating with a brilliant systems using a natural language such as English. Processing of Natural Language is necessary whenever people needed the brilliant system like robots to perform as per their instructions, when people wants to choice from a dialogue based clinical expert system, etc. The area of NLP required making computers to perform useful assignment with the natural languages humans can use. There are two input and output of an NLP system can be defined as follow −

- Speech
- Written Text

The end part of the study is conclusion, which is a comparison of summary of the study and useful execution and guidance for further research.

B. Background of the Study

Natural languageprocessing (NLP) is a main part is associated with the section of computer science, which focuses on the development and improvement in the process. In late 1940s the term wasn't exists, but the work related machine translation (MT) had started. Research in this period was not completely localized. English and Russian were the superior languages for MT, but others, like Chinese were used for MT (Booth, 1967) [10]. MT/NLP research nearly died in year 1966, which wrap up that the +MT is going nowhere. But later on some MT manufacturing systems were providing output to their customers (Hutchins, 1986) [11]. Now this time, the work on the use of computers for classical and linguistic studies had also started.

C. Aims and Objectives

The main aim of NLP is to search a text. This can be through organizedwithdraw, which is removal units of information from unstructured data and converting it into structured data, which rate the subject of a given sentence.

COMPONENT OF NLP

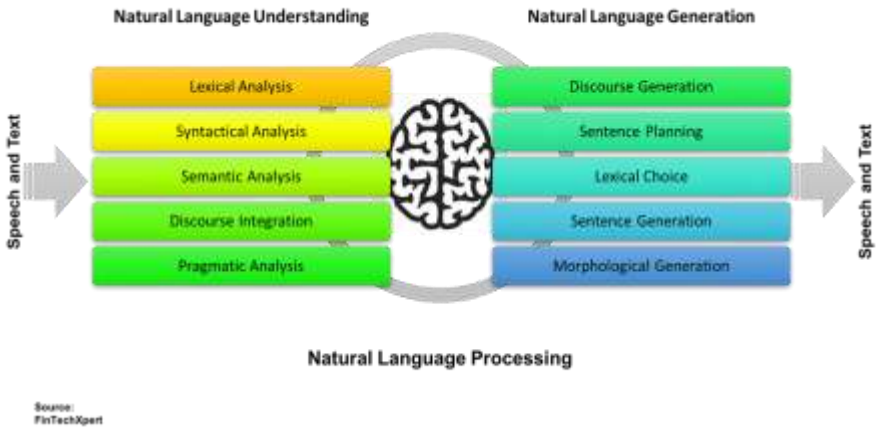

Figure.1- Component of NLP

NLP is divided into two parts as given below −

A. Natural Language Understanding (NLU)

Understanding involves the following tasks −

- Plotting the given input in natural language into useful presentations.
- Analyzing different aspects of the language.
-

B. Natural Language Generation (NLG)

It is the process of creating meaningful phrases and sentences in the form of the natural language from some internal depiction. It involves the following −

- **Text planning −** It includes retrieval of the required content from the knowledge base.
- **Sentence planning −** It includes selection of the desired words, forming relevant phrases, setting accent of the sentence.
- **Text Realization −** It is the mapping of sentence plan into the sentence structure.

The NLU is harder than NLG (Natural Language Generation).

II. Levels Of Nlp

The levels of NLP (Natural Language Processing) are described as shown in the following figure: (Figure 2).

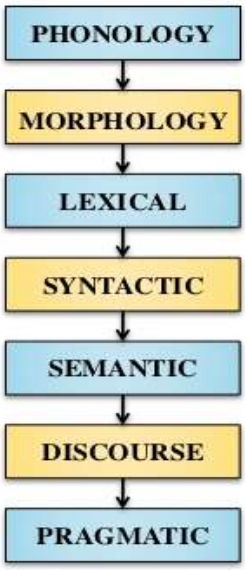

Figure.2-Phases of NLP

Linguistic is the technique which includes the meaning of language, language context and various Factors of the language. The various important terminologies of Natural Language Processing are: -

a. Phonology

Phonology is study of organizing sound systematically. It is the element of Linguistics which mention to the detailed process of sound. The word phonology comes from Ancient Greek and the word phono- its means voice or sound, and the suffix – logy related to speech. In the year of 1993 Nikolai Trubetzkoy indicated that the Phonology is "The study of sound affect to the system of language". Whereas Lass in 1998 wrote that phonology refers widely with the sounds of language, concerned with the to lathe sub discipline of linguistics, although it could be explained as, "phonology properly isto concerned with the function,behavior and organization of thesounds as linguistic items. Phonology includes semantic uses of sound to encode meaning of any Human languages. (Clark et al., 2007) [1].

b. Morphology

Morphology is a study of building of words from ancient meaningful components. The different parts of the word represent the smallest units of meaning said as Morphemes. Morphology which holds of Nature of words, & they are initiated by morphemes. An example of Morpheme could be the word pre-cancellation it can be morphologically analyzed into the three separate morphemes: the prefix pre, the root cancellation, and the suffix-ion. The interpretation of the morpheme stays same throughout all the words, just for understand the meaning the humans can break any of unknown word into morphemes. For example, adding the suffix–ed to a verb, sends that the action of the verb took place in past. The words which cannot be divided and have meaning by themselves are called Lexical morpheme (e.g.: table, chair) .The words like (e.g. -ed, -ing, -est, -ly, -ful) that are combined with the lexical morpheme are said to be Grammatical morphemes (eg. Worked, Consulting, Smallest, Likely, Use). Those grammatical morphemes that can occurs in combination known as the bound morphemes (eg. -ed, -ng).Grammatical morphemes can be divided into two main types:
1) The bound morphemes and
2) Derivational morphemes.

c. Lexical

In Lexical level, humans, as well as NLP systems, clarify the meaning of individual words. Sundry types of processing sprinkled to the word-level understanding – the first of these being a Part-of-speech tag to each of the word. In this processing, the words could act more than one part-Of-speech are appointed as the most probable part-of the speech tag based on the context in which they occur. At the lexical level, the Semantic representations can be replaced by the words that have only single meaning. In NLP system, the nature of the representation differs according to the Semantic theory deployed.

d. Syntactic

It is concerned that with the meanings of words and how to combine words into meaningful phrases and sentences. This level signifiesto analyze the words in a sentence so that it can uncover the grammatical structure of the sentence. Both grammar and the parser are needed in this level. The output of this level of converting is depiction of the sentence that admits the structural dependency relationship between the words. There are various types of grammars that can be blocked and which in rotate, crack option of the parser. Not all the NLP applications needed the overall parse of sentences, therefore for accept challenges in parsing of prepositional phrase attachment and conjunction audit are no longer blocking that appeal for which the exact and casual dependencies are sufficient. Syntax conveys meaning in most languages because the order as well asthe dependency contributes to meaning.

For example, the two sentences: 'The cat chased the mouse.' and 'The mouse chased the cat.' Have difference only in terms of the syntax, but itprovides quite different meanings.

Implementation Aspects of the Syntactic Analysis

There are a number of algorithms that researchers have developed for syntactic analysis, but we consider only the following simple methods −

- Context-Free Grammar (CFG)
- Top-Down Parser (TDP)

Let us now explore them in detail −

Context-Free Grammar

The Context-Free-Grammar (CFG), it is the grammar whichincludes rules with the single symbol and on the left-hand side of the re-write rules. So let's 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 [1]

The parse tree breaks the sentence into small- structured parts so that the computer can easily and clearly understand and process it. In order for parsing the algorithm to construct this parse tree, a set of rewriting rules, which describe what tree structures are legal, need to be built.

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

The rewritten rules for the sentences are as shown as follows −

S → NP VP

NP → DET N | DET ADJ N

VP → V NP

Lexicon −

DET → a | the

ADJ → beautiful | perching

N → bird | birds | grain | grains

V → peck | pecks | pecking

Figure: The parse tree can be created as shown below−

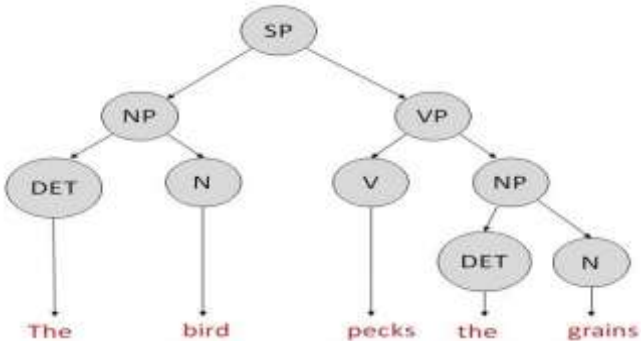

Figure. 3- Parse tree

`Now consider the above rules for rewriting. Since the V can be replaced by both, "peck" or "pecks", sentences such as "The bird pecks the grains" can be mistakenlygranted, i.e. the subject-verb agreement error is approved and it is correct.

The Top-Down Parser

Here, in the Top –Down the parser it starts with the S symbol and tries to rewrite it into a pattern of terminal symbols that matches the classes of the words in the input sentence until it consists complete of the terminal symbols.

These are then checked with the input sentence to see if it matched. If it does notmatch then, the process is started all over again with the different set of rules. This is repeated until a specific rule is found which is used for describing the structure of the sentence.

e. Semantic

In semantic, most people think that the meaning is ruled, however, this is not true, it is all because of the levels that grants to meaning. Semantic processing rules the possible meanings of a sentence by pivoting on the interactions among word-level meanings in the sentence. Thislevel of processing can absorb the semantic rearrange of the words with multiple senses; in a general way to know how syntactic rearranging of the words that can becharged as the multiple parts-of-speech which is skillful at the syntactic level. For example, by

different meanings, 'file' as a noun can mean could be a binder for gathering of the papers, or a tool to form one's fingernails, or a individuals line in a queue (Elizabeth D. Liddy,2001). The semantic level analyzes words for their dictionary clarification, but also for the clarification they derive from the milieu of the sentence. Semantics milieu that most words have more than one clarification but that we can spot the appropriate one by looking at the rest of the sentence.

f. Discourse

Discourse deals with how immediately the preceding sentence can affect the judgment of the next sentence. While the syntax and the semantics work with the sentence-length units, the communication level of NLP works with the no. units of text bigger than a sentence i.e., it does not clarifies the multiple sentence texts as just a sequence sentences, apiece of which can be decoded singly. On the other hand, communication focuses only on the properties of the text as complete that convey meaning by creating connections between the component sentences (Elizabeth D. Lady, 2001).

The two of the most common levels are defined as follows:

A) Anaphora Resolution - Anaphora resolution works by the replacing of words such as pronouns, which are semantically helpless, with the related entity to which they refer.

B) Discourse/Text Structure Recognition - Discourse/text structure recognition sway the functions of the sentences in the text, which, in turn will, adds to the meaningful representation oftext.

g. Pragmatics

Pragmatics deals with the using and understanding of the sentences in various situations and how the analysis of the sentence is disturbed. Pragmatic is concerned with the tough use of the language in situations and handle basic over and above basic of the text for understanding the main goal and to explain it on how meaning is read into texts without actually being cryptograph in them. This requisite much world knowledge, including the understanding of the intentions, plans, and goals. The following example of the two sentences required objective of an anaphoric terms 'they', but this intention to require logical or world knowledge- (Elizabeth D. Lady, 2001).

III. Mathematical Calculation (Matrices)

Here the matrices is given as below-

Key Metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

Here,
TP => True Positive
TN => True Negative
FP => False Positive
FN => False Negative

IV. Results And Discussions

A. Applications of NLP

Machine Translation:

As the name makes its usage obviously, & natural language processing can be busy in the translation on the web. Data is broadly available on the internet. People from different regions are able to access it from any part of the world. With this the dissimilar in the quantity of users speaking and understanding different languages, NLP helps with translation of the particular languages into their natural languages. Google is the carrier of this along with many of other companies that are using the NLP for machine translation. This may come with its own set of challenges. The important challenge is to obtain the original meaning of the translated text.

Summarization:

NLP can be used in the narration of the texts for avoiding the overflow of the information on the internet. A viral trend is slowly taking over with the ability to narrate the large block of text or document into a meaningful summary. This is an uphill task if you look at it properly. Extraction of correct meaning from the large string of the text contains many hurdles. This is very useful technique that can help the companies for understanding the emotions behind the huge summary of data which they collect from their products and the services. NLP for sure change the pattern in which the market works.

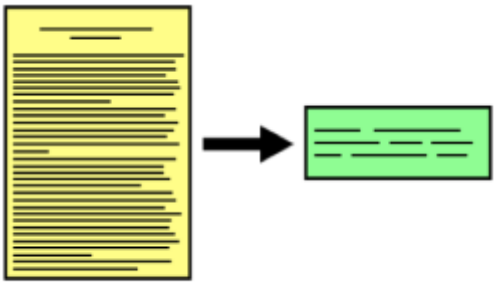

Figure. 4

Question Answering:

If you may have ever used Siri on iOS or maybe any other digital assistant, then you will definitely know what purpose NLP (Natural Language Processing) solves in the 'answering questions'. A QA (Question Answering) application is a system that is capable of answeringthe valid human request logically. The NLP has the capability to understand the human languages can be done in two ways: A) either in text-only format or B) spoken dialogue. This technology is still in the phase of development as of now, but can still be seen generally used everywhere in for form of Siri, Ok Google, Amazon Echo Dot and the chat boxes.

Well, these are pretty much of the major uses of Natural Language Processing (NLP) that we can see and also we can use. It is still a very interesting technology that is rapidly extensively growing.

B. Relationship between Language and Text

Results are dependent on the secondary sources and also theoretical grounds reveal that the text and technology have convenient relationships. Students are unable to understand the content without appropriate understanding of the language. With the lack of understanding the language, it is not possible for the trainee to understand and absorb the information. The measure of natural language acquisition is one of the most essentialplan that can guide in the action of language acquisition. The natural language acquisition measureonward with the encouragement from the teachers can deliver as a main source for leading improvement in the academic process of the students. Teachers as well as the learners both the two of us can concentrate on the strategies, which are powerful enough for utilization of the language in educational setting. For example, the research studies show that context analysis cannot be accomplished without producing the understanding of text.

V. Conclusion

It is concluded that, the Natural language processing afield of the computer science and AI that mainly focuses only on the interaction between the computers and the humans. NLP actually makes the job easier but it still requires a human interference. Peoples and the industry are afraid of NLP that it would begin a trend of jobs grabbing which is true and it makes sense but it mainly cannot work the way it does without the help of the human inputs. The passion to work and serve to the loopholes or bugs in a machine is the task of a human who is handling it.

References

[1] https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.html.
[2] https://www.tutorialspoint.com/artificial_intelligence
[3] https://arxiv.org/pdf/1708.05148