ISSN No.: 2454- 2024 (online)

International Journal of Technical Research & Science

NATURAL LANGUAGE PROCESSING: REVIEW

Shridhar B Devamane, Ramesh K.P, Trupthi Rao
E-Mail Id: [1]shri.git@gmail.com, [2]rameshkp1@gmail.com, [3]trupthiraob@gmail.com
ECE Department, APS College of Engineering, Bangalore, Karnataka, India

Abstract- Natural language processing employs computational techniques for the purpose of learning, understanding, and producing human language content. Early computational approaches to language research focused on automating the analysis of the linguistic structure of language and developing basic technologies such as machine translation, speech recognition, and speech synthesis.
Key Words: Natural language processing (NLP), Syntactic, Symantic, Pragmatic, Lexical, Linguistics, Generation, Machine Learning.

1. INTRODUCTION

Natural language processing (NLP) is a field of artificial intelligence concerned with the interactions between computers and human (natural) languages. It can be defined as the automatic (or semi-automatic) processing of human language. The term `NLP' is sometimes used rather more narrowly than that, often excluding information retrieval and sometimes even excluding machine translation. NLP is sometimes contrasted with `computational linguistics'. Nowadays, alternative terms are often preferred, like `Language Technology' or `Language Engineering'. Language is often used in contrast with speech (e.g., Speech and Language Technology).
Computational linguistics, also known as Natural language processing (NLP), is the subfield of computer science concerned with using computational techniques to learn, understand, and produce human language content. Computational linguistic systems can have multiple purposes:
The goal can be aiding human-human communication, such as in machine translation (MT); aiding human-machine communication, such as with conversational agents; or benefiting both humans and machines by analyzing and learning from the enormous quantity of human language contents that are now available online.

2. BLOCK DIAGRAM

A simplified view of Natural Language Processing emphasises four distinct stages shown below. In real systems these stages rarely all occur as separated, sequential processes. In the overview that follows it is assumed that syntactic analysis and semantic analysis will be dealt with by the same mechanism - the parser.

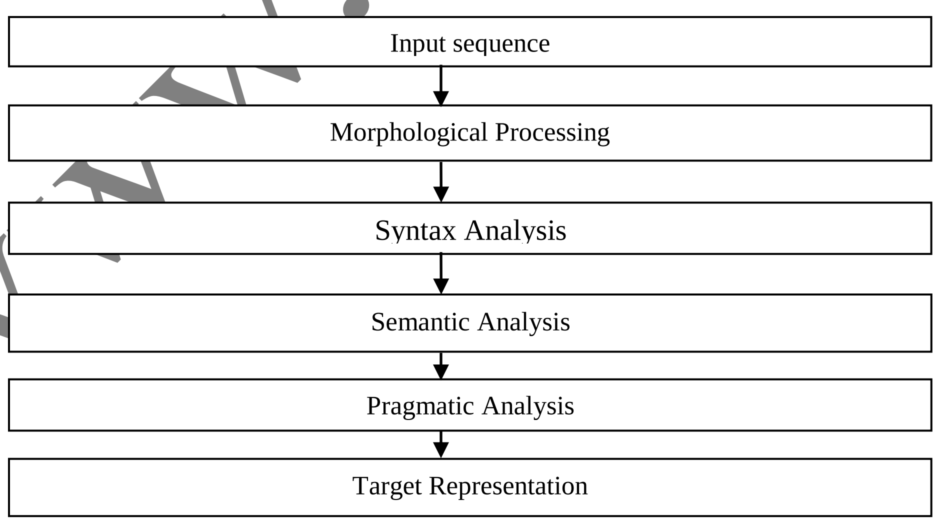

Fig. 2.1 The Logical Steps in Natural Language Processing

3. MORPHOLOGICAL PROCESSING

The preliminary stage which takes place before syntax analysis is morphological processing. The purpose of this stage of language processing is to break strings of language input into sets of tokens corresponding to discrete words, sub-words and punctuation forms. The nature of morphological processing is heavily dependent on the language being analysed. In some languages, single words (used as verbs) contain all the information about the

ISSN No.: 2454- 2024 (online)

International Journal of Technical Research & Science

tense, person and number of a sentence. The output from the morphological processing phase is a string of tokens which can then be used for lexicon lookup. These tokens may contain tense, number, gender and proximity information (depending on the language) and in some cases may also contain additional syntactic information for the parser. The next stage of processing is syntax analysis.

4. SYNTAX ANALYSIS

A language processor must carry out a number of different functions primarily based around syntax analysis and semantic analysis. The purpose of syntax analysis is two-fold: to check that a string of words (a sentence) is well-formed and to break it up into a structure that shows the syntactic relationships between the different words. A syntactic analyzer (or parser) does this using a dictionary of word definitions (the lexicon) and a set of syntax rules (the grammar). A simple lexicon only contains the syntactic category of each word, a simple grammar describes rules which indicate only how syntactic categories can be combined to form phrases of different types.

Lexicon:

Word	Category	Grammar
Cat	Noun	Sentence → Noun Phrase, Verb Phrase
Chased	Verb	Verb Phrase → Verb, Noun Phrase
Large	Adjective	Noun Phrase → Article, Adjective, Noun
The	Article	

In order to carry out semantic analysis the lexicon must be expanded to include semantic definitions for each word it contains and the grammar must be extended to specify how the semantics of any phrase are formed from the semantics of its component parts. For example the grammar rule above Verb Phrase → Verb, Noun Phrase states how the syntactic group called Verb Phrase is formed from other syntactic groups but says nothing about the semantics of any resulting Verb Phrase. Using a simplified form of logic the grammar and lexicon can be expanded to capture some semantic information.

5. GRAMMER FOR LANGUAGES

The single most important consideration for the design of Lkit concerned the types of grammar the toolkit would accept. This section gives a brief overview of the various types of grammar most commonly used to specify natural languages for computer-based analysis.

5.1 Case Grammars

Case Grammars attempt to describe any given sentence in terms of a fixed frame of slots (called cases) which explicitly capture information about any activities described in the sentence, the instigators of those activities, positions, times, etc. For example case analysis of the sentence "Gary repaired the car in the garage on Sunday" could generate the following case frame:

	Repairs
Actor	Gary
Object	Car
Location	Garage
Time	Sunday

5.2 Semantic Grammars

Semantic Grammars (also called Domain Specific Grammars) [Rich & Knight] place no restrictions on the notations used to describe them but use rules built around domain specific phrase types rather than typical abstract categories. For example, in a medical domain the sentence "there is a bleeding ulcer in the stomach towards the lesser curve" the two phrases "in the stomach" and "towards the lesser curve" might be classified by a semantic grammar as a locator-phrase and a location-qualifier respectively. In a more general purpose grammar for English these two phrases would be of the same type: prepositional-phrase.

5.3 Definitive Clause Grammars (DCGs)

Definitive Clause Grammars [Pereira & Warren] use rules which are based on logical implication expressions. These expressions are equivalent to Prolog clauses with the result that a Prolog interpreter can be used as a parser for DCGs. An un-adapted Prolog interpreter would perform a top-down, depth first parse which is not particularly efficient. Example such as Noun Phrase (NP) L Verb Phrase (VP) fi Sentence (append (NP, VP)).

5.4 Lexical Functional Grammars (LFGs)

Lexical grammars present a different approach to grammar/lexicon construction by removing rules from the grammar and embedding them instead in the lexicon. The justification for this is that the valence of words (the number and type of other syntactic groups they associate with) is a feature associated with words rather than their meanings or syntactic classification. The valence of a word defines the legal structure of sentences it may occur in.

6. AUGMENTED TRANSITION NETWORKS (ATNs)

ATNs are usually described as a highly developed and extended form of State Transition Network (STN). A basic STN for natural language would have arcs labelled as some terminal syntactic category. Recursive Transition Networks (RTNs) are a development of STNs. They consist of collections of small STNs with arcs which can be labelled with names of other networks (non-terminals).

7. SEMANTICS

Semantic representations need to capture details of objects and their relationships, events and the chains of causality that tie them together. A detailed discussion of semantic representations is beyond the scope of this document but the following section intends to highlight the issues of particular importance to semantic representations used specifically for NLP. The target representation is basically:

- The conceptual level: What exactly is represented and in which manner. A simple representation may only capture details of objects and object-object relationships.
- The abstract form: Most representations can be described in abstract forms, using diagrams for example. These diagrams show what the representation makes explicit and what must be inferred.
- The physical realization of the representation: The implementation form, data structures and inference mechanisms.

8. TARGET REPRESENTATION

The target basically indicates different steps applied to the given sentence and the final representation of the sentence.

9. APPLICATIONS OF NLP

The basic applications of natural language processing are listed below:

9.1 Text Classification

The goal of text categorization is to classify the topic or theme of a document. A popular classification example is sentiment analysis where class labels represent the emotional tone of the source text such as “positive” or “negative”.

9.2 Language Modelling

Language modelling is really a subtask of more interesting natural language problems, specifically those that condition the language model on some other input. The problem is to predict the next word given the previous words. The task is fundamental to speech or optical character recognition, and is also used for spelling correction, handwriting recognition, and statistical machine translation.

9.3 Speech Recognition

Speech recognition is the problem of understanding what was said. The task of speech recognition is to map an acoustic signal containing a spoken natural language utterance into the corresponding sequence of words intended by the speaker. The system analyses the person’s specific voice and uses it to fine-tune the recognition of that person’s speech, resulting in increased or better accuracy.

9.4 Caption Generation

Caption generation is the problem of describing the contents of an image. Given a digital image, such as a photo, generate a textual description of the contents of the image. A language model is used to generate the caption that is conditioned on the image.

9.5 Machine Translation

Machine translation is the problem of converting a source text in one language to another language. The automatic translation of text or speech from one language to another is one of the most important applications of NLP.

9.6 Document Summarization

Document summarization is the task where a short description of a text document is created. As above, a language model is used to output the summary conditioned on the full document. Some examples of document summarization include: Creating a heading for a document, creating an abstract of a document.

CONCLUSION

The natural language Processing uses computer to process speech and texts and has the ability to abstract the data from audio, video, text data and process it and makes use of artificial intelligence and machine learning algorithms. The focus is mainly on the interaction among computers and human.

REFERENCES

[1] P. Langley, "Intelligent Behavior in Humans and Machines, Advances in Cognitive Systems", Vol. 3, No. 12, December 2012.
[2] Russel, Norvig, "Artificial Intelligence-A Modern Approach", 1995, Prentice Hall.
[3] Artificial Intelligence by Elaine Rich and Kevin Knight, 1991, McGraw-Hill.
[4] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", pages 86-105, Upper Saddle River: Pearson Education, 2009.
[5] Manning C, Schutze H, "Foundations of Statistical Natural Language Processing", MIT Press, Cambridge, MA, 1999.
[6] Jurafsky D, Martin J, "Speech and Language Processing", Prentice Hall, Upper Sale River, NJ 2000.
[7] P. Langley, "Intelligent Behavior in Humans and Machines, Advances in Cognitive Systems", Vol. 3, No. 12, December 2012.
[8] Dr. Mariana Navis, "Natural Language Processing", Hasso Plattner Institute, SoSe 2016.
[9] Sebastian Shaumyan, "Applicative grammar as a semantic theory of natural language", University of Chicago, Press, 1977.
[10] The Artificial Neural Network, http://en.wikipedia.org/wiki/Artificial_neural_network.