International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

An Overview to Natural Language Processing

Pankaj Naruka [1], Rohit Yadav [2], Dr. Himanshu Arora [3], Monika Mehra [4]

[1,2]B.Tech Student, Department of CSE, Arya College of Engineering & Research Centre, Jaipur
[3]Professor, Department of CSE, Arya College of Engineering & Research Centre, Jaipur
[4]Assistant Professor, Department of CSE, Arya College of Engineering & Research Centre, Jaipur

---

Abstract – *Natural Language Processing has recently gained much attention in for understanding and analyzing human language computationally. Its applications include Speech recognition, Text classification and categorization, Named Entity Detection, Paraphrase detection, Character Recognition, Spell Checking etc. The paper gives an introduction to natural language processing followed by its brief history, dependence on linguistics, its usefulness in AI systems and different stages of NLP, namely Natural Language Understanding and Natural Language Generation.*

1. INTRODUCTION

Natural Language processing (NLP) is the technology based on AI that enables the computers to understand human language whereas until some years earlier they were only capable of understanding mathematical language. Human languages are more complex than mathematical language as they do not always follow the specified rules and employs more than just symbols and rules such as phonology, semantics and pragmatics. NLP is based on AI and Linguistics which is the science of language which includes the study of *phonology, morphology, syntax, semantics and pragmatics**[1]***.

With use of NLP developers can structure and organize knowledge to carry out tasks such as speech recognition, translation, relationship extraction, sentiment analysis and automatic summarization.

NLP is divided into two parts, i.e. Natural Language Understanding and Natural Language Generation, to make processing easier.

1.1 Linguistics in NLP

Linguistics is the scientific study of language. It involves analyzing language form, language meaning and language in context**[2]**. It is aimed at forming rules which can be used to inference meaning from a sentence.

NLP algorithms are based upon these rules offered by linguistics to understand the structure of sentences and extract the useful information.

1.2 AI for NLP

NLP plays a vital role in enhancing the AI systems. Without NLP the AI systems can only understand the meaning of language but not the meaning of the words. The NLP allow users to communicate with computers using natural language. Some of the AI systems which use NLP are Google Translate, Google Assistant, Apple's Siri, and Microsoft's Cortana.

2. History

The history of NLP could be said to have started in mid-twentieth century, albeit work being available from earlier periods. In 1950, Alan Turing wrote a paper describing a test for an *intelligent* machine. Shortly after, in 1952, the Hodgkin-Huxley model proposed by Alan Hodgkin and Andrew Huxley showed how neurons interact inside brain to form an electrical network **[8]**.

The series of these events helped inspire the idea of Artificial Intelligence, Natural Language Processing and evolution of computer.

The Georgetown experiment conducted in 1954 involved fully automatic translation of over sixty sentences from Russian to English **[3]**. The authors claimed that within five years the machine translation would be a solved problem. But the real world progress was really slow according an ALPAC report in1966 which resulted in highly reduced funding in further research for machine translation.

Till 1980s most natural language processing systems were based on complex hand written rules. However, starting in the late 1980s there was a revolution in natural language processing with the introduction of machine learning algorithms for natural language processing.

Recent researches has increasingly focused on supervise and semi-supervised algorithms. Such algorithms learn from data that has not been hand annotated for desired answers or from a combination of both annotated and non-annotated data.

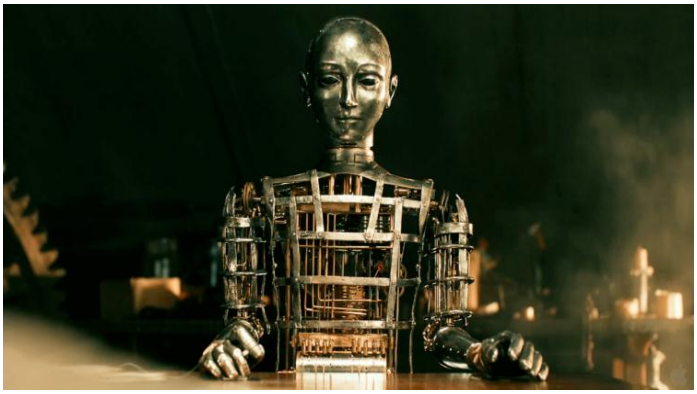

Fig -1: An Automation

Following are some important events in history of NLP:

1950 – NLP started with the publication of an article called "Machine and Intelligence" by Alan Turing.
1950- Attempts for automatic translation between Russian and English.
1960- The work of Chomsky and other linguists on language theory and generative syntax.
1990- Probabilistic and data driven models becomes standard.
2000- A large amount of spoken and textual data become available.

3. STAGES OF NLP

There are two stages of NLP, NLU and NLG, which are further divided into smaller stages. Brief function of these stages is given in the following text.

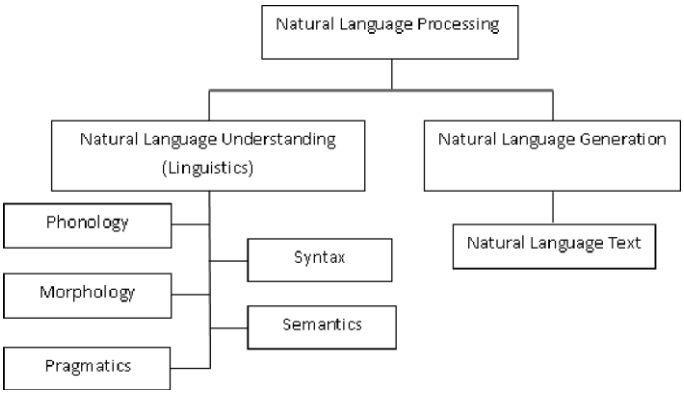

Fig -2: Classification of NLP

3.1 Natural Language Understanding

NLU is the process of conversion of unstructured data (data in human language) to structured data (form compatible for computer operations).

Not every human can understand the intricacies of computer languages. NLU allows humans to communicate with computer using the natural languages so they can focus on the more important work and leave the responsibility of trivial but complex tasks to their computers.

NLU allows computers to understand human languages, perform required operations and make an appropriate decision. It does so in following five steps:

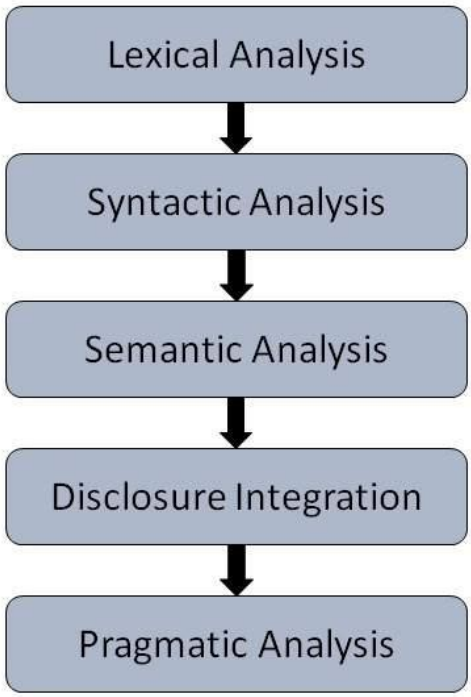

Fig -2: Steps in NLU

Lexical Analysis: Lexical Analysis is the process of dividing whole text into paragraphs, sentences and words. It involves analyzing and identifying structure of words.

Syntactic Analysis: It involves the understanding of the relation between the words using rules defined for a particular language such as English.

Semantic Analysis: The text is checked for meaningfulness. The exact meaning is extracted from the text. It is done by mapping objects with syntactic structures.

Disclosure Integration: The contextual meaning of a sentence also depends on the sentence preceding it and it also brings about the meaning of the next sentence. This step makes sure that the response is related to the current context.

Pragmatic Analysis: In this the original text is reinterpreted to compare with the various inferences that could have been made. The practical meaning instead of just the literal one is drawn out, to select a better response.

3.2 Natural Language Generation

NLG is the technology used by computers to communicate with humans in a way that they can understand. It is the opposite of NLU. In NLU the computer tries to understand the

structure of input data and extract the useful data whereas in NLG it is required to put the data in a structure and present it in a format understandable by humans. The NLG is often used to turn large databases into useful written recommendations which help firm leaders make better decisions.

Fig -2: Natural Language Generation

As proposed by Dale and Reiter **[4]** the typical stages of natural language generation are:

Content Determination: It involves deciding what information to produce in the text after analyzing the input.

Document Structuring: It involves deciding the organization of the information to convey.

Aggregation: It involves the merging of sentences to make naturalness and readability easier.

Lexical Choice: It involves putting words to concepts. E.g., deciding whether to use *medium* or *moderate* to describe the level of fever.

Referring expression generation: It involves creating referring expressions that identify regions and objects like how humans use pronouns and other types of anaphora in natural language.

Realization: It involves creating the actual text which should be correct according to the rules of syntax, morphology and orthography. E.g. using *could be* for past tense of *can be*.

4. CONCLUSIONS

For thousands of years since cognitive revolution humans have used language to communicate and cooperate with each other.

In the mid-twentieth century a new ally emerged which proved to reduce human labor dramatically. This new ally also known as computer could not understand the languages we have developed.

To resolve this problem series of researches and experiments started to make communication between computers and humans easier.

Today with the development of AI, Linguistics, and smart machine learning algorithms the computers are much capable of communicating with the humans and aiding them in their work.

Today most of the AI systems use natural language processing such as Google Assistant, Netflix, Apple's Siri, and Grammarly.

Because of our tendency to become dependent on technology and ever increasing development in technology it seems reasonable to assume that NLP has complex future as the best way humans can communicate is with the help of complex languages we have developed over time.

REFERENCES

[1] Krishna Prakash Kalyanathaya, D. Akila and P. Rajesh, "Advances in Natural Language Processing – A Survey of Current Research Trends, Development Tools and Industry Applications", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7, Issue-5C, February 2019.

[2] A. GelbukhA, "Natural language processing", Fifth International Conference on Hybrid Intelligent Systems (HIS'05), 6-9 Nov. 2005.

[3] Halliday, Michael A.K., Jonathan Webster, "On Language and Linguistics", Continuum International Publishing Group. p. vii. ISBN 978-0-8264-8824-4, June 2006

[4] Hutchins, J., "The history of machine translation in a nutshell", January 2014.

[5] Dale, Robert; Reiter, Ehud, "Building natural language generation systems", Cambridge, U.K.: Cambridge University Press. ISBN 978-0-521-02451-8, March 2000.

[6] "Natural Language Processing." Natural Language Processing RSS. N.p., n.d. Web. 25 Mar. 2017

[7] Alshawi, H. The core language engine. MIT press. 1992

[8] Hodgkin AL, Huxley AF "A quantitative description of membrane current and its application to conduction and excitation in nerve". The Journal of Physiology. 117 (4): 500–44. doi:10.1113/jphysiol.1952.sp004764. PMC 1392413. PMID 12991237 August, 1952.