National Conference on Trending Technology for Achieving Sustainable Development Goals
NCTTASDG 2023 Organized by Shri Shankarprasad Agnihotri College of Engineering, Wardha

THE ANALYTICAL STUDY NATURAL LANGUAGE PROCESSING

Sahil Shende*1, Sahil Warudkar*2

*1,2B.Tech. First Year Student, Faculty of Engineering and Technology, Datta Meghe Institute of Medical Sciences (Deemed to be University), Sawangi (Meghe)

Wardha, Maharashtra, India.

ABSTRACT

In today's generation, Natural Language Processing is the most preferred way of communication between machines and human beings. Natural Language Processing is a branch of artificial intelligence and a subfield of machine learning. Before attempting to make sense of the presented information, Natural Language Processing first reads it. Once the data has been correctly processed, the machine will either finish the task or respond with a response. In this paper, I tried to explain Natural Language Processing in simpler terms with my understanding of it. This paper contains the definition of NLP, itsworks, and how it is used in the python programming languagewith the help of NLTK library, and its applications.

Keywords: Natural Language Processing (NLP), Sentiment analysis, Natural Language ToolKit (NLTK), Artificial Intelligence, Machine learning

I. INTRODUCTION

Today's world is dependent on technology in various sectors, it can be a small thing in our lives that we use on daily basis to the most advanced techniques as well as things. In the past three years, the world has become more reliant on and accepting towards internet services. AI researchers have made the life of the common people much easier and more convenient with their continuous efforts in their research. The objective of Natural Language Processing (NLP) has always been to shorten the gap in communication between machines and humans[1]. There are many techniques for Sentiment Analysis is one example of the Natural Language Processing techniques that researchers have developed to help computers comprehend and analyse human language. Known as Entity Recognition, Summarization, etc.. Natural language processing is the ability of computer software to comprehend human language as it is spoken and written (NLP). It is a part of machine intelligence (AI). The field of natural language processing goes back to the 1950s. Although it was not recognised at the time as a different issue from artificial intelligence, Alan Turing presented the now-famous Turing test as a measure of intelligence in a 1950 article titled "Computing Machinery and Intelligence." The suggested test consists of a task that calls for the automatic synthesis and interpretation of natural language [2].This paper aims to give information about what NLP is,Why it is important, how it works, and its applications in simple words. The first section of this paper explains the fundamentals of NLP. In the second section, I explained its working, its implementation in python using the NLTK library and its applications. The last section contains the conclusion.

II. PROCESSING OF NATURAL LANGUAGES

Natural Language Processing Categories

The objective of natural language processing (NLP), a subfield of artificial intelligence in computer science, is to assist computers in comprehending human writing and spoken language. Because there is a lot of unstructured data available, this is a complicated task. How people speak and write also known as their "tone of voice" is particular to every person and is constantly changing to match current use. Natural language understanding and natural language generation, which advance the tasks of comprehending and creating text, are the two primary divisions of natural language processing. [3].

There are several crucial terms used in natural language processing, including:

The area of linguistics known as phonology is concerned with the systematic ordering of sound. The word is derived from the Greek words "phono" and "logy," which both mean "word or speech" and "voice or sound," respectively.

III.   MORPHOLOGY

The foundation of morphology, the character of words, is morphemes. The several parts that make up a word are an example of a morpheme, which is the smallest unit of meaning. Because morphemes have the same meaning in all words, humans can break down any unfamiliar word into morphemes to understand its meaning.

Syntax- Most languages use syntax to convey meaning even though order and dependency are factors in connotation. This stage of processing corresponds to the sentence that illustrates the connections between the structural dependencies of the word. Numerous grammars can hinder parsers, which reduces their chances of working. These parsing challenges are o longer subject to prepositional phrase attachment or conjunction audit preventing the argument that phrasal and clausal dependencies are sufficient for NLP applications that do not demand a complete parse of sentences [4].

Semantic- The semantic level examines words for both the elucidation provided by their dictionary definitions and the elucidation provided by the context of the sentence. Most words According to semantics, most words have several meanings identify the correct one by examining the rest of the sentence[5].

Pragmatic- It tries to understand the purpose and explain how new meaning is read into texts without really being written in them. Pragmatics is concerned with the firm use of language in situations. This necessitates thorough familiarity with the outside world, particularly familiarity with intentions, plans, and ambitions. [4].
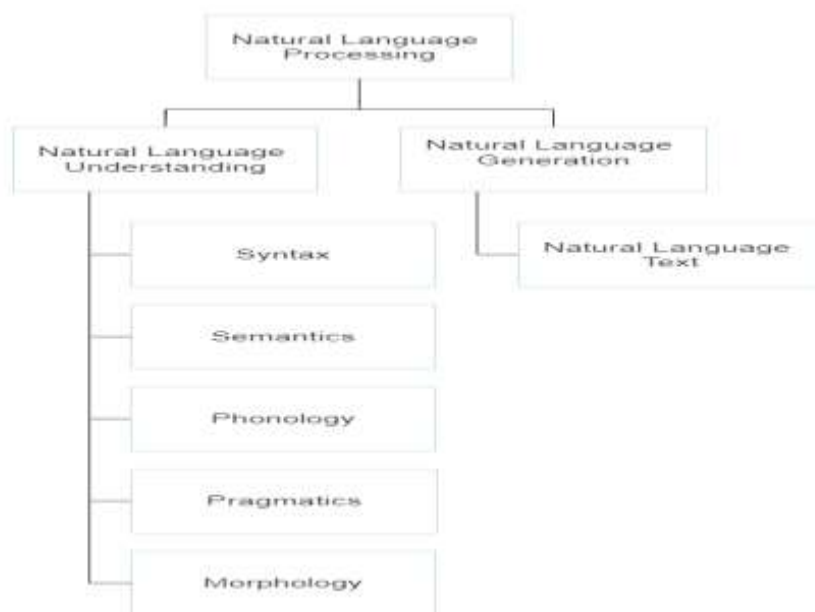

Figure 1: Natural Language Processing classification

Natural Language Processing in Action- Computers can now comprehend natural language just like people do thanks to NLP. Natural language processing uses artificial intelligence to take real-world input, analyse it, and make sense of it in a way that a computer can understand, regardless of whether the language is spoken or written. Computers have reading programmes and microphones for audio collection, just as how humans have diverse sensors like ears for hearing and eyes for seeing. Computers have programmes to digest information similarly to how people utilise their brains to do so. During processing, the input is ultimately converted into computer-readable code. Natural language processing involves two primary steps: algorithm development and data preparation. Thanks to NLP, computers can now understand natural language exactly as people do Natural

language processing uses artificial intelligence to take real-world input, analyse it, and make sense of it in a way that a computer can understand, regardless of whether the language is spoken or written. Computers have reading programmes and microphones for audio collection, just as how humans have diverse sensors like ears for hearing and eyes for seeing. Similar to how people use their brains to process information, computers have programmes to do the same. The input is finally transformed into computer-readable code during processing.The two core processes in natural language processing are algorithm development and data preprocessing. Text data must be prepared and "cleaned" in the data preparation process before computers can use it. analyse it. Preprocessing makes data usable and draws attention to text features that an algorithm can make use of. There are various ways to do this, including:

Tokenization: The text has now been cut up into manageable chunks.

Stop word deletion: Common words are removed in this case, leaving only the unusual words that provide the greatest context for the text. Words are broken down into their simplest components at this step of processing via lemmatization and stemming.

Speech-act tagging: Words are then classified as nouns, verbs, or adjectives, depending on the part of speech they belong to. An algorithm is created to process the data after the data preparation stage. The two core categories of natural language processing algorithms are the ones that are most frequently employed, despite the fact that there are many distinct kinds of them.:

Rule-based system: This system's linguistic rules were thoughtfully created. This method has been used ever since natural language processing was first developed.Utilizing machine learning as a system Statistical techniques are used by machine learning algorithms. Training data is given to them in order to instruct them on how to carry out jobs. As more training data is handled, their techniques evolve. Natural language processing techniques use machine learning, deep learning, and neural networks in order to modify their own rules through iterative processing and learning.

Python-based Natural Language Processing- The Natural Language Toolkit (NLTK) is a set of program modules, datasets, tutorials and exercises used in python for performing Natural Language Processing. Three different types of documentation are present. The toolkit's usage is explained in the tutorials. Every toolkit module, interface, class, and data component is described in the API documentation. Technical reports describe the toolkit's implementation and design.After installing NLTK, import this library then pass a string to process the further step. Now, for tokenizing the words we have to import regex and then pass the split function of the string, the words in the string are separated. Now, use this'nltk.word_tokenize(text)' code to tokenize words without being negated. Here only words are tokenized and not the whitespace and any other thing. After that stemming process is done by using the porter stemmer algorithm. It applies some rules like you're moving some additional characters and using the porter stemmer we are going to import the porter stemmer from analytic a dot stem and then the object should be initialized and next the string is stemming and f is used in front of the string to get the formatted string we are going to define variables inside the string and these are going to call variables themselves then pass the function that stem the string then, pass the words so it displays the stem and its corresponding stem. The is also another stemming method called Snowball Stemmer it has an advantage over porter stemmer that it works on different languages [6].

IV.   APPLICATIONS OF NLP

1.   Chatbots

Chatbots are programmed to interact with users in a human-like manner. Depending on their complexity, chatbots can either simply respond to certain phrases or even carry o conversations that make it hard to distinguish them t from real people. They understand the complexities of the English language because Machine Learning is used in its creation.As they have more conversations they get more knowledge of Encantoto to understand the precise meaning of the sentence and continue to improve with time.

2.   Autocomplete and Autocorrect

The suggestion provided while typingin the search engine is provided by using autocomplete and autocorrect to guess what the user wants to search and uses Natural Language Processing for it. When users enter certain words, search engines use their huge data sets to determine what their users are likely typing and recommend the most realistic options They use natural language processing to understand these words and how they relate to constructing different phrases

3.   Voice Assistants

The voice assistants use a complex combination of speech recognition and Natural language Processing to interpret what the user is saying. Voice assistants' long-term objective is to connect people to the internet and offer a wide range of services solely through voice interaction. Given that voice assistants occasionally still cannot grasp what you are saying, they are still some distance from achieving this goal.

4.   Translation

Sequence-to-sequence to sequence modelling, a method used in natural language processing, is used by translators. It enables the algorithm to translate a string of words from one language to another. Earlier, language translators employed statistical machine translation (SMT), which involved analysing millions of texts that had already been translated from one language into another and searching for the language's fundamental patterns and terminology. However, this approach fell short of sequence-to-sequence to sequence modelling in terms of accuracy.

5.   Grammar Checkers

Millions of sentences are used to train the NLP algorithm so that it can recognise the proper structure. Because of this, it may offer a better synonym, the proper verb tense, or a clearer sentence structure than what you have typed. Grammarly and WhiteSmoke are two of the most well-known NLP-based grammar checkers.

V.   CONCLUSION

Today's world cannot be recognized without NLP, it is used by almost every system that is based on AI and Machine Learning. The mostfamous systems are Google Assistant, Apple's Siri and Grammarly. From this article, we conclude that The capacity given to computers by software to make them understand human language, how it is written and spoken, and the precise meaning of the sentences is known as natural language processing (NLP). NLP sequentially process this information by going through each process syntax, semantics, pragmatics, phonology etc. NLP is executedusing python using Natural Languagtoolkitit (NLTK) library, an open-source library. Some of the systems where NLP is used are voice assistants, translators, chatbots, grammar checkers etc.

VI.   REFERENCES

[1]   Rajarshi SinhaRoy, "A Study on the journey of Natural Language Processing models: from Symbolic Natural Language Processing to Bidirectional Encoder Representations from Transformers", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 6, pp. 331-345, November-December 2021

[2]   Natural language processing. (2022, November 20). In *Wikipedia*. https://en.wikipedia.org/wiki/Natural_language_processing

[3]   Diksha Khurana, Aditya Koli, Kiran Khatter and Sukhdev Singh, "Natural Language Processing: State of The Art, Current Trends and Challenges", Accendere Knowledge Management Services Pvt. Ltd., India

[4]   Liddy, E. D. (2001). Natural language processing.

[5]   Feldman, S. (1999). NLP Meets the Jabberwocky: Natural Language Processing in Information Retrieval. ONLINE-WESTON THEN WILTON-, 23, 62-73.

[6]   https://www.youtube.com/watch?v=WYge0KZBhe0

[7]   Hutchins, W.J. Machine translation, Chichester, England: Ellis Horwood, 1986.

[8]   Hutchins, W.J. and Somers, H.L. An introduction to machine translation, London: Academic Press, 1992.

[9]   Jacobs, PS. (ed) Text-based intelligent systems, Hillsdale, NJ: Lawrence Erlbau Associates, 1992.

[10] Joshi, A.K., Webber, B.L. and Sag, I. A. (eds.) Elements of discourse understanding, Cambridge: Cambridge University Press, 1981.