iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38015

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Natural Language Processing: Components, Advances, Tools and Industrial Applications

Hima Yeldo[1], Chandni Singh[2], Shambhavi Pandey[3]
[1, 2, 3]*B.Tech, Department Of Cs & It, Sam Higginbottom University Of Agriculture, Technology & Sciences, Prayagraj, Uttar Pradesh, India*

Abstract: Natural Language Processing is the study that focuses the interplay between computer and the human languages NLP has spread its applications in various fields such as an email
Spam detection, machine translation, summation, information extraction, and question answering etc. Natural Language Processing classifies two parts i.e. Natural Language Generation and Natural Language understanding which evolves the task to generate and understand the text.

I. INTRODUCTION

Natural Language Processing (NLP) is a program which help in making computer understands the human language. NLP is the sub field of artificial intelligence and computer science which is concerned with the interaction between human language and computer. NLP came into existence to ease the work of users and to satisfy them to communicate in human language.
Linguistics analyze with the nature of language and communication. It is the science of language which includes Phonology which refers to the study of sound, Morphology refers to the study of construction of words, Pragmatics deals with understanding sentences in different situations, Syntax refers to arranging of words and Semantics determines the possible meanings of a sentence. [6]

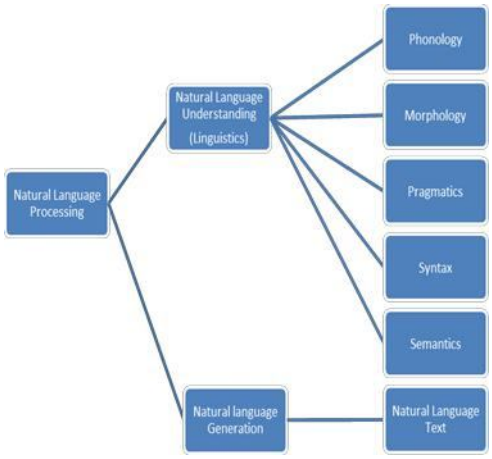


Fig.1. Broad Classification of NLP

The goal of NLP is to read, understand, decipher and make human language valuable. NLP is used to analyze text, and allow the machines to understand how humans speak. This human-computer interaction allows real-world applications like sentiment analysis, automatic text summation, named entity recognition, stemming, relationship extraction, and more. NLP is often used for machine translation, text mining and automatic question answering. [2][7]

II. BRIEF HISTORY

From around 1950, Natural Language Processing started its brief development like what is a “Turing Machine” and in 1957; it brings a lot of set structures. Until 1990, computational sets and other systems become slow or limited due to some rules but later on with the introduction of machine learning, python and other deep learning, the growing stage of sets and systems increasing.
While, NLP is facing still lot of challenges and difficulties especially because of the human interface or interaction but in other way it also opened a way to many techniques, and digital transformation. [2][7]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

III. LEVELS OF NLP

Natural Language Processing works on various levels and these levels are one of the most explanatory methods which help us to generate NLP text by realizing Sentence Planning, Content Planning and Surface Realization phase. [6]

Fig.2. Phases of NLP architecture

A. Phonology

Phonology refers to the systematic arrangement of sound. The term phono-which means sound or voice and logy-which means speech or word has come from Ancient Greek. Phonology is concerned with the behavior function, and organization of sound. Phonology includes semantic use of sound to encode human language. [6]

B. Morphology

Morphology refers to the study of construction of words from primitive meaningful units. These primitive meaningful units are termed as morphemes like prefix, suffix and base. Morphology
Which comprise of mature of words, are initiated by morphemes? Morphemes are minimal units that cannot be divided further and have their own meanings. For example the word “dog” has one morpheme but the word “dogs” have two morphemes as –s denotes plurality. [6]

C. Lexical

Lexical analysis involves analyzing and identifying the structure of words and deals with the study of level of words with their respective lexical meanings and part of speech. This level utilizes the language’s lexicon *(collection of words and phrases)*, which is a collection of individual lexemes. Lexical analysis is used to divide chunks into words, paragraph and sentences. [6]

D. Syntactic

Syntactic analysis also known as *parsing* involves arranging of words in a manner that shows relationship among the words. It allows extraction of phrases which convey more meaning than the individual word. It is a process of analyzing a string of symbol in nlp, data structures and computer language by following the rules of formal grammar.

E. Semantic

The Semantic analysis deals with determination of what a sentence really means by relating syntactic features with multiple definitions. It draws the exact meaning from the text. Semantic is known as the study of meaning which is of two types’ associative meaning and conceptual meaning. For example conceptual meaning of sea is something filled with salt water, very large and so on but associative Meaning of sea is storm, shipwreck and so on. [5][6]

F. Discourse

The discourse level deals with the analysis of meaning of a text beyond a single sentence and focus on the properties of text that convey meaning by making connections between words and sentences.

©IJRASET: All Rights are Reserved

G. Pragmatics

Pragmatics deals with the use of real world knowledge and understanding the sentences in different situations and how it effects the interpretation of the sentence.

IV. NATURAL LANGUAGE GENERATION

Natural Language Generation (NLG) is the process of producing meaningful phrases, sentence and paragraphs from an internal representation. It generates narrative that summarizes, describes and explains structured data. It is a part of NLP that identifies goals, plans and evaluate the situation to achieve the goal. [6]

Fig.3.Components of Natural Language Generation

A. Components OF NLG

1) *Speaker And Generator:* We need to have a speaker or an application to generate a text that renders the application's intentions into a fluent phrase relevant to the situation.

B. Components And Level Of Representation

The process of language generation involves four tasks as follows:

1) *Textual Organization:* The information must be textually organized according to the grammar. It must be ordered both in terms of linguistic relations and sequentially.
2) *Realization:* The organized and selected resources must be realized as an actual voice or text output.
3) *Content Selection:* Information should be included and selected in the set. It depends on how the information can be parsed into the representational units. Some parts are added by default while other parts of the units may have to be removed. [6][7]
4) *Linguistic Resources:* To support the realization of the information, we must be chose linguistic resources. The choices of particular words, syntactic constructs, idioms are the resources which will come down.

C. Application Or Speaker

It is used to maintain the model of the situation. Here the application or the speaker just initiates the process but does not take place in the generation of language. It stores the structure and history of the contents and deploys a presentation of what it actually knows. The only Requirement here is the speaker has to make sense.

V. ADVANCES OF NATURAL LANGUAGE PROCESSING

A. Machine Translation

Machine translation was the first non-numeric application of computers and it was started in the late 1990's .However, it achieved very limited success. A change occurs in the field during early 1990's when researchers at IBM obtain a large number of each other, known as parallel text. These data allowed them to collect some of word translations and word sequences of parallel text and word sequence to build a sample model of MT. [9]

©IJRASET: All Rights are Reserved

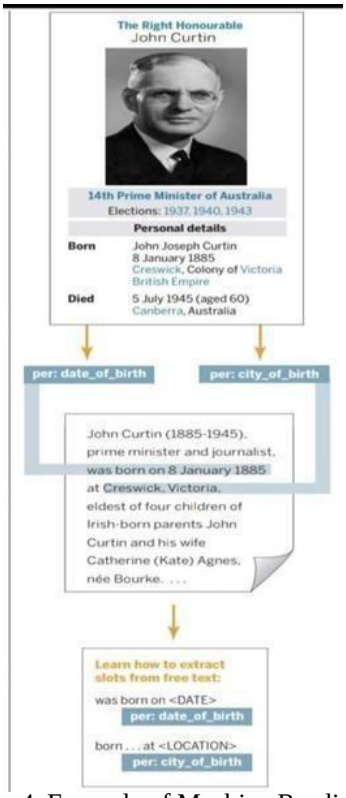


Fig.4. Example of Machine Reading

In the late 1990's the golden age brought the combination of sample online text, including some quantities of parallel text And inexpensive computing and a new idea for making statistical phrase -based MT system. Rather than translating word by word, the key notices that small group of words have some distinctive translation.

After all these additions and change, today we have services which allow free and instant translation between two different languages , But it still produces that type of translations that are only for determining core of passage. Finally, we have seen the development of an approach of MT with the use of deep learning based sequence models. [1]

B. Spoken Dialogue Systems And Convertsions Agents

Dialogue becomes a popular topic in NLP research since 1980's. Now a day's word on text based dialogue are more improve by including spoken dialogue with mobile devices like Apple's Siri, Google Now and Microsoft Crotona for information access and text based apps. Spoken dialogue system also used in robots to help people with simple task. [1]

The creation of SDSs, whether between humans and artificial agents, it requires some tools such as:-

1) Automatic Speech Recognition (ASR) to identify what a user or human says.

2) Dialogue Management (DM) to determine what actually user want .It obtains humans requested information.

3) Text-to-speech (TTS) to send the information back in human spoken form.

SDSs need to be ready to interact with user or human when there is an error in speech recognition what user actually said, either Automatically or the repeating the dialogue of the user.

In speech to speech translation system, MT components are also needed to help dialogue between speakers of different languages and system to identify the mistranslation. [1]

There many challenges in building SDSs the primary problem is improving the accuracy basic ASR, DM, TTS blocks. These includes basic problems of recognizing and producing normal human conversational behavior such as coordination and clear spoken words. Sometimes human gives cues while speaking such as filled pause e.g. "um" or "uh". Non-humanlike delay makes SDS difficult to carry response, as user should wait for seconds to receive the system response. To address this incremental processing of ASR, MT and TTS Modules are done so that systems can respond more quickly. [1]

©IJRASET: All Rights are Reserved

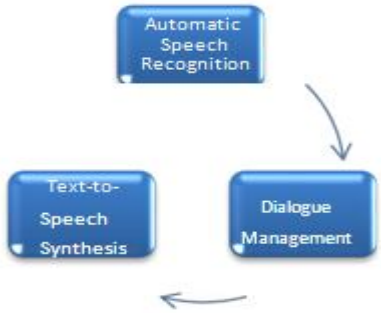


Fig.5. A Spoken Dialogue System

C. Machine Reading

Machine reading is the idea in which machines can become intelligent, a can usefully integrate and summarize information for human by reading and Understanding the vast quantities of text which are available.

In the early days of artificial intelligence , many researchers were focused on the approach of trying to enable intelligent machines by manually giving large structured knowledge bases in a formal logical languages and developing methods for deriving further facts from this knowledge. [1]

However, in the modern online world, we mainly have huge repositories of online information which is coded in human languages .One place where the above statement is true is in the scientific literature, where the findings are reported entirely in human languages text. The quantity of scientific literature is growing day by day. Thus, there is an increase need of machine reading for the purpose of comprehending and summarizing the literature and also for extracting facts and hypotheses from this material. [2]

The initial goal of machine reading is to extract basic facts as most commonly a relation between two entities for example “child of” (for instances, Bill Clinton, Chelsea Clinton). This is referred as relation extraction.

For a particular domain-specific relations, there are many systems have been successfully built. One technique used is handwritten pattern that is match of linguistic expression of relations e.g. <PERSON>’s daughter, <PERSON>. More better results can be obtain by the use of ML.A structured prediction classifier give instances of such relation which are based on extracted features from the sequence of word and grammatical Structure of sentences. Such systems are the fact-extraction tools in fields such as biomedicine. [9][1]

VI. DEVELOPMENT FRAMEWORKS AND TOOLS FOR NLP

The development frameworks and tools help us to build industrial application. These frameworks and tools give in build libraries and also fulfill the specific needs of industry.

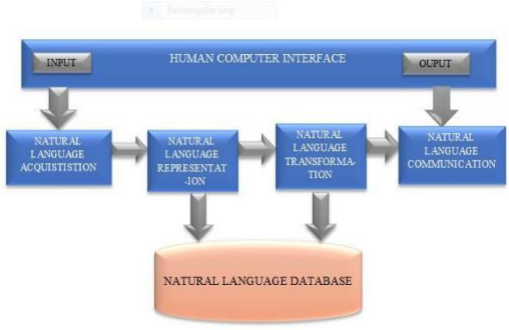


Fig.6. Block representation of stages in the development of NLP tools

As figure show the block diagram of various stages of NLP application development. The Natural language acquisition block build with help of the speech processing, any acquisition tool which place the natural language text into the system. [8]

Natural language representation block use graph model or structured tree to present the Natural language understanding.

Natural language database is storage of natural language data such as MNIST etc., and then these are used by machine algorithm to perform other tasks of NLP.

©IJRASET: All Rights are Reserved

This database can be accessed by representation transformation blocks to do their tasks. Natural language
Transformation will consist of various learning and extraction algorithm to extract the meaningful actions from the NLP tasks.
Natural language communication is presentation of the action that user want i.e., an outcome of NLP tasks. The generally outcome may be either Natural language or in the form of computer action.

Table 1 Natural language processing tools

Tool	Developer/ Supported by	Native Language Support	Usage	Application areas
CoreNLP	Stanford Group	Java	Open source GPL[3]	Common NLP tasks and Analytics
NLTK[2]	Steven Bird and Edward Loper	Python	Apache License	Common NLP tasks' and Analytics Supported by Apache Software Foundation
TextBlob	Steven Loria	Python	Open source	Specialised library for Text Analytics on top of NLTK
Gensim	RaRe Technologies	Python	GNU lesser GPL[3]	Specialized library of Topic modelling
Spacy	Explosion.ai	Python	MIT License	Common NLP tasks[1] and seamlessly integrates with Python 's AI echo system
Open NLP	Apache Software foundation	Python	Apache License	Supports common NLP tasks[1] and useful to build text analytics applications
UIMA[2]	IBM / Apache Software foundation	Java and C++	Apache License	Supports text analytics and multimodal analytics
GATE[4]	GATE Research Team University of Sheffield UK	Java	GNU lesser GPL[3]	GUI based development tool for text mining and language processing
Chainer	Preferred Networks Inc	Python	Open source	Supports various types of network architectures including feed-forward nets, convnets, recurrent nets and recursive nets
Deeplearning4j	Open Community	Java and Scala	Apache License	Supports topic modelling, parallel GPUs and RNN, word2vec, doc2vec, GloVe algorithms
Deepnl	GiuseppeAttardi	Python	Open source GPL[3]	Supports common NLP tasks[1]
Dynet[5]	Carnegie Mellon University	C++ with rappers for Python	Open source	Supports common NLP tasks[1]
Nlpnet	MIT	Python	MIT License	Supports common NLP tasks[1]
OpenNMT	MIT	Python	MIT License	Specialized library for machine translation, summarization, image to text, speech recognition, language modelling and sequence tagging
PyTorch	Facebook and Uber AI team	C++, Python	Open source	Supports development of Computer Vision and NLP tasks
TensorFlow	Google Brain Team	C++ and Python	Apache license	Supports many deep learning algorithms and interface API for Python, C, C++, Java, Scala
TFLearn	TFLearn contributors	C++ and Python	MIT License	Built on top of TensorFlow, Easy to use Deep Neural Network API, helper functions
Theano	Montreal Institute for Learning Algorithms	Python	Open source (BSD License)	Supports well for mathematical and matrix operations
Keras	FrancoisChollet	Python	MIT License	Supports easy prototyping, various deep neural networks
Lingpipe	Alias-Inc	Java	Alias-i license	Specialised in text processing. It has Java APIs
MALLET[7]	Andrew McCallum	Java	Common Public License	Supports common NLP tasks[1]
LUIS[8]	Microsoft	Microsoft SDK	Pay per use	Cloud based APIs that supports rapid development of bots, conversation systems using pre-built domain specific models

VII. ANALYSIS AND GENERTATION OF SPEAKER STATE

Much of the work in NLP has focused on identification of positive or negative orientation of textual language and identification of belief states on the basis of lexical and syntactic information. Both
Sentiment and beliefs attitude towards events and proposition, where sentiment can also concern attitude towards organization, people and abstract concepts. Detection of sentiment and emotion in text can be determine by lexical and sentence level information. [8]
Sentiment can be determine by words conveying positive or negative orientation
For example , "sad","worried" ,"difficult", and "weak" are the words with negative orientation and "comfortable" , "important", "successful" and "interesting" are word with positive orientation. There are online sentiment dictionaries such as
Whistle's dictionary of affect and systems created from subject ranked terms, such as
Tausizik and Penne baker's LIWC (Linguistic Inquiry and Word Count) can be used to assess positive and negative sentiment in the content .More approaches to sentiment analysis also see to identify the source as well as the object of the sentiment. [8]
There are some researches using features that have proven important in recognizing classic emotions to identify speaker's state like medical condition, speaker characteristics such as gender, age, pathology and personality and speaker conditions drunkenness, sleepiness etc. Corpora collected such studies that been used in the Interspaces Paralinguistic Challenges. Emotion generation has been proven more difficult challenges for TTS synthesis. There are some systems like MARY that attempt to generate emotions such as depression, aggression or cheerfulness.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

VIII. INDUSTRIAL APPLICATIONS OF NLP

NLP aims where machine interaction to human can be easily possible. NLP still continues to influence the unstructured data's and make it meaningful to a machine. Robotics, health care and financial services are some of the sectors that can continue to be advanced by NLP.

One of the initial applications of NLP in early years of 2000 was on machine translation which converts the one human language into another. But it rapidly found its acceptance in customer service industry. Most popular application of NLP in customer service is called as "Chabot" or Virtual assistant.

Industrial applications of NLP are classified into major 3 categories: Conversational systems, Text Analytics, Machine translation.

A. Conversational Systems

Conversational system allows making the conversation with the automated systems in a natural language via voice or text. They help to automate the complex workflows in an organization with 24x7 supports to its users.

Most common type of conversation devices are Chabot and Virtual Assistants. Today, these two devices are used by banks, e-commerce, social media and other self-service point of sales systems that provide various services to their customers. [8]

B. Text Analytics

Text Analytics which also known as Text mining aims to extract meaningful contents, words or some parts from text, documents, either in emails or communications such as tweets and SMS.

C. Machine Translation

It is the task of translating one natural language into another. Most popular application of machine translation is Google translator. Other software's are also used for speech translation and teaching.

Now, we will look at some industrial applications for the following domain areas. [7][8]

1) *Healthcare:* Hospitals are implementing Virtual Assistants which are developed with the combination of Natural Language Processing, Computer Vision and Machine learning which will automatically create and recollect patient's history by interacting with the patients. Virtual Assistant handles the routine tasks such as scheduling appointments and registration of the patients. [8]
2) *Automotive:* Self-driving cars are one of the most significant innovations in this field. NLP enables human computer interface in what is now called as "in-car assistants" which is gaining popularity in the industries. [8]
3) *Finance:* NLP are developed in those applications such as credit score, sentiment analysis and document search. The credit scoring application helps in assessing the credibility of Banks and financial Institutions and providing credit scores using NLP and machine learning. In the sentiment analysis applications, news site and social media automate text mining functions for real-time information on the market and then document classification And unit name is given to filter out the most relevant information for investor's needs. In document search applications, the banks or financial institutions uses Chabot interface that enables to its customers to search for the information and get answers for transactional queries. [8]
4) *Manufacturing:* Robotics and Process Automation are the two promising applications areas that employ Natural Language Processing capabilities. Using NLP, Robot on a manufacturing floor can interact with human operator at a remote place and process instructions for assembly and movement of machines and products.
5) *Retail:* Using Natural Language Processing, a Virtual Assistant is placed in front of the Retail shop which can identify and know what customer wants and provide them quick information about the products. The specialized virtual assistant also can understand the shopper's comments and complaints regarding the products and provide them a quick direction for resolution. Many platforms are currently using conversational agents for customer support. [8]
6) *Education:* A combination of Natural Language Processing and Computer vision creates a platform to deliver virtual classroom for students. We have already seen the digital assistants which helps the students solving the problems from expert knowledge available in digital libraries.

The top vendors delivering NLP solutions are IBM, Google, and Amazon. [8]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

IX. CONCLUSION

So far, In Natural Language Processing it helps computer to understand the human language, the goals of Natural Language Processing and how human-computer interaction enables real world applications. After that in NLP, discussed about the classification of NLP through which, understudied the different terminologies of Natural Language Understanding and Components of Natural Language Generation.

The situation is changing rapidly over five years. Huge improvements have been done in speech recognition, talking to your phone a common place activity, especially for young people. Web search engines are increasing day by day and successfully in understanding complex queries, and MT can at least yield the gist of material in different languages, even if it cannot yet produce human quality translation.

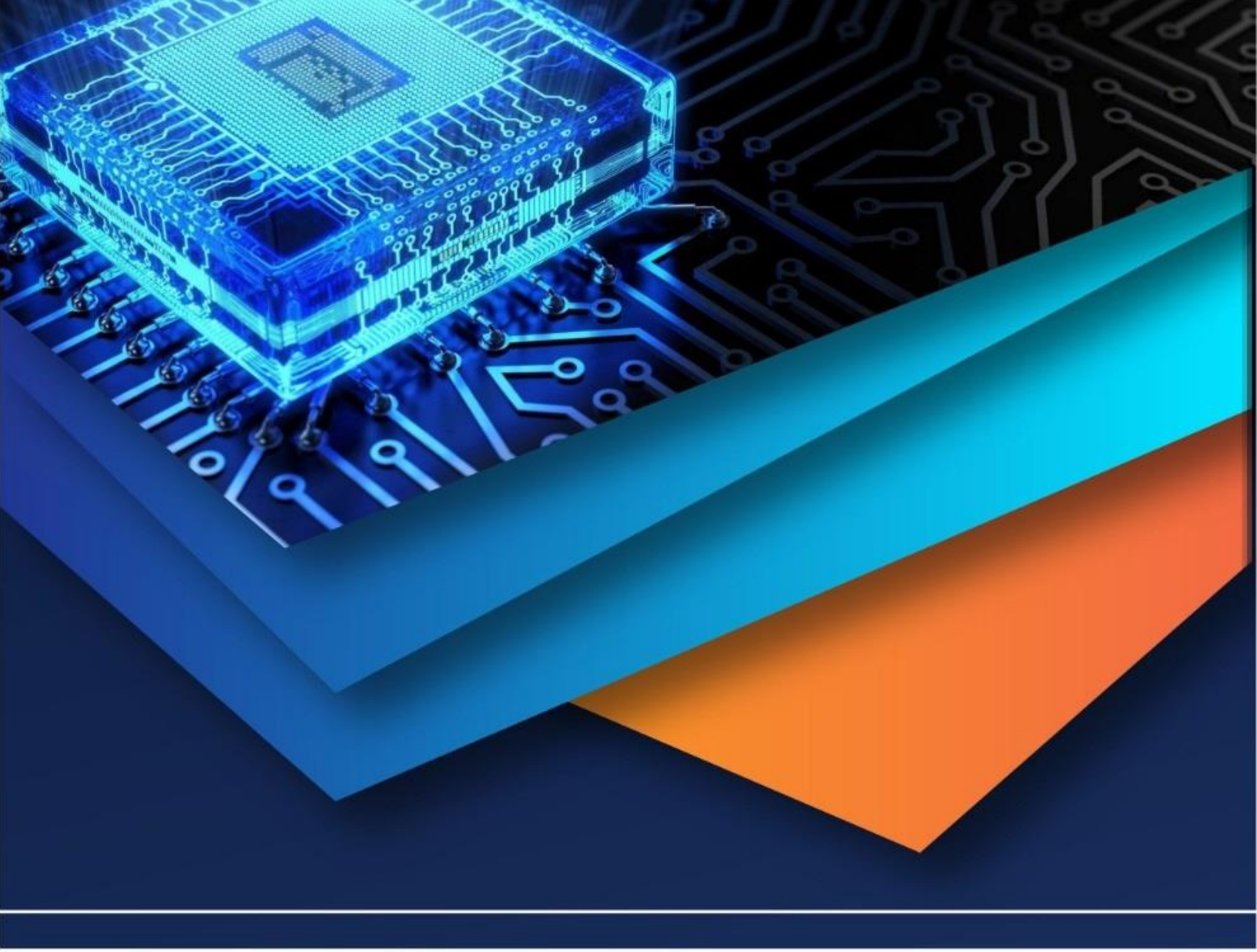
In the last, in NLP talked about the applications of nlp in different sectors like about how it aims to develop machine to human interaction and it faced so many difficulties but with the help of AI and robotics still it is continuing its growth to the various enhancement. Then in NLP discussed about the types and how nlp works on these sectors like healthcare, automotive, finance, manufacturing, retail and last education. Still the working and enhancement of this is going on. It is believed that with the help of nlp, can achieve many things that didn't consider in this world.

All these points and topics discussed above help us to learn a lot about NLP and How NLP ease the works of users and satisfy them to communicate in human language.

REFERENCES

[1] Advances in natural language processing- A survey of current research trends, development tools and industry applications.
[2] Svetlana Sicular, Kenneth Brant 2018, Hype Cycle for Artificial Intelligence, 2018 Gartner report July 2018.
[3] http://mallet.cs. umass.edu. 2002.
[4] Quarteroni, Silvia. (2018). Natural Language Processing for Industry: ELCA's experience. Informatik-Spektrum. 41.10.1007/s00287-018-1094- 1.
[5] Young, Tom & Hazarika, Devamanyu & Porias, Soujanya & Cambria, Erik. (2018). Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. IEEE Computational Intelligence Magazine. 13. 55-75. 10.1109/ MCI.2018.2840738.
[6] Amirhosseini, Mohammad Hossein, Kazemian, Hassan, Ouazzane, Karim and Chandler, Chris (2018) Natural language processing approach to NLP Meta model automation. In: International Joint Conference on Neural Networks (IJCNN), 8-13 July 2018, Rio de Janeiro, Brazil.
[7] Wikidata , www.wikidata.org
[8] https://www.ijrte.org
[9] https://cs224d.standford.edu/papers/advanc es.pdf

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)