



# SURVEY ON NATURAL LANGUAGE PROCESSING

<sup>1</sup>Divyansh Dhukarya, <sup>2</sup>Malvika Sharma, <sup>3</sup>Ananya Anand, <sup>4</sup>Dr. Manish Bhardwaj

<sup>1</sup>Department of Computer Science, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India.

<sup>2</sup>Department of Electrical and Electronics Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India.

<sup>3,4</sup>Department of Computer Science and Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India.

## Abstract

Natural Language Process delineates a predominant part in the field of emerging technologies that allow the computer to transfer the data with humans in their language. The machine indulges in reading texts, hearing speeches, interpreting IT, measuring sentiments, etc. It is also known as a language-centric task scaler. With the changing world where technology becomes a crucial part of our lives, so does NLP. By utilizing Natural Language Processing methods, you can make it feasible for non-expert users to express their programming concepts through natural language. NLP is arguably the most convenient and conventional type of technology field today which fosters Siri, Alexa, and Cortana that has become a unified part of our lives. Natural Language is a subset of artificial intelligence whose primary purpose is to comprehend and produce meaningful human language utterances. It is salient because it tends to resolve the ambiguity in language. Over a decade the high-profile researchers had put in efforts- for social media monitoring, the opportunity of customer service and chatbots, fake news detection, generative pre-training model (GPT), etc. Therefore, NLP provides you with the usage of Artificial Intelligence and is also used to recover the information in data mining. This technology is going to take over the market soon as it is in high demand because of its easier approach and accessibility. Natural Language Processing (NLP) and its many techniques are examined in this study.

5812

**KeyWords:** Natural Language Processing (NLP), Natural Language, Artificial Intelligence, Technology, Data Science, Machine Learning, RNN (Recurrent Neural Network)

DOI Number: 10.14704/nq.2022.20.11.NQ66580

NeuroQuantology 2022; 20(11): 5812-5818

## INTRODUCTION

**Natural language processing (NLP)** is a sub-discipline of artificial intelligence providing a bridge between natural human languages and computer. In other words you can say that it is related to the field of computer - human interaction. It translates human language to computers language to makes users work easier. It is also known by a name "**Computational Linguistics**" [1]. Natural Language Processing is an area of research and application that explores how computer can be used to understand and manipulate Natural Language text or speech to do useful things.

Natural language processing has filtered its roots in the 1950 with help of Richard Bandler and John Grinder. The foundation of NLP is basically lies on many disciplines, such as linguistics,

mathematics, computer sciences, psychology, and so on. The goal of NLP researchers is to create a software program that enables computers to understand or generate language used by humans [2].

NLP is a branch of computer science that deals with the interplay between human languages and computers. It is a method through which computers intelligently and usefully analyze, comprehend, and derive meaning from human language. Developers can organize and arrange knowledge to do tasks such as automatic summarization, translation, named entity identification, relationship extraction, sentiment analysis, speech recognition and themes and segmentation.

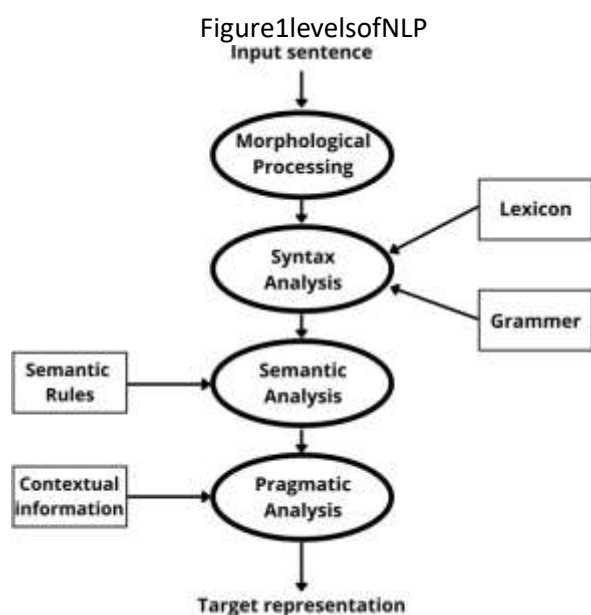
Input from the user is sent to the natural language interface software, and outputs in a



language recognized by the application programs are generated.

### **TERMINOLOGY OF NLP**

- Phonology–itmaybeconsideredasstudyof different sounds and the way they came together to form speech or words.
- Morphology- simplyitisstudyofwordsandtheirparts.
- Syntax- order or arrangement of wordsand phrases toform proper sentences. Further divided into lexicon and grammar
- Semantics– studyofmeaningofwordsandphrasesinlanguage.
- Pragmatics– studytofocusing on therelationship betweennatural language the language used by user.



### **NLP WITH MACHINE LEARNING**

Text analysis and natural language processing rely on machine learning techniques to decipher the content of long text texts quickly and accurately. When it comes to identifying entities and general sentiment, machine learning models excel. However, when it comes to discovering themes and subjects, some algorithms suffer. Text data may be analyzed using machine learning. Text analytics and natural language processing (NLP) rely on a collection of statistical approaches for distinguishing the various parts of speech, titles, and moods in text. Supervised machine learning is another name for these approaches, which may be described as a model and then applied to additional texts. An unsupervised machine learning technique might potentially be used to extract meaning from a big amount of data [3].

Speech recognition, document categorization, segmentation of documents, word sense clarification, and machine translation are all examples of machine learning techniques being applied.

### **NLP with Deep Learning**

Deep learning refers to applying deep neural networks to massive amounts of data to learn a procedure aimed at handling a task. Deep learning for NLP is a branch of artificial intelligence that aids the computer in manipulating and interpreting human language [4]. It is an extended field of machine learning that has proven to be highly useful in the domains of text, image, and speech, primarily. NLP enables computers to perform a wide range of natural language related tasks at all levels, ranging from parsing and part-of-speech (POS) tagging to machine translation and dialogue systems. Deep learning algorithms help to solve problems in Natural Language Processing (NLP) that machine learning algorithms couldn't solve.

### **Literature Review**

In the current technical environment, Natural Language Processing (NLP) has become an immensely convenient and beneficial source of Data Analysis. It helps the computer to read and interpret the human language that is specified by the user. This process involves various high-

level features of Natural Language Processing such as sentimental analysis, stemming, named entity recognition, text classification, text-to-speech conversion, analysis of the document, Sentence Segmentation [5]. NLP was pioneered back in the year 1950 by **Alan Turing** when Turing published an article named "**Computing Machinery and Intelligence**" in the same year which was also recognized as the "**Turing test**". The article made him contemplate things like "Can machines think?" and the two words "machine" and "think" together sounded unfamiliar.

In 1960, **SHRDLU** was developed because of some Natural Language Processing Systems. SHRDLU was a computer program that was used to understand computer programs developed by **Terry Winograd** at MIT in 1968–1970. It was inspected as a tremendously successful invention of Artificial Intelligence (AI). It was considered the first example of interactive fiction as the user can just move the objects in a virtual environment by their input. In days gone by, Natural Language Processing was concluded various times, but it recovered again in the 1980s along with Artificial Intelligence to recuperate its edge. An NLP revolution was made because it was now using Machine Learning Algorithms which made it up to the field. The pure NLP statistical methods became a major technology with a massive flow of online texts. The idea of a **Recurrent Neural Network (RNN)** was raised in 1986. RNN is the only one with internal memory and is a powerful type of neural network. Because of their internal memory, RNN's can be very precise about predicting upcoming data. As an enhancement to intermittent neural networks, long-term memory (LSTMs) substantially enhances memory capacity. The units of an LSTM are utilized as building units for the layers of an RNN, regularly called an LSTM network. These networks are a complex area of Deep Learning. LSTM plays a significant role in solving complex problems in domains like speech recognition, machine translation, and more which makes it suitable for Natural Language Processing and LSTMs are also capable of learning order dependence in sequence prediction problems.

There are three gates in a Long Short-Term



Memory: Input, output, and forget gates. These gates decide if to let a new contribution to (input gate), let it impact the result at the current timestep (output gate), or erase the data since it isn't significant (forget gate).

Natural Language Processing has also been applied in **Log Analysis** and **Log Mining**, the technique of the computer itself to make sense through the given data. Both technologies involve the extraction of data and interpreting it into its intelligence using the process of correlations and extracting patterns to predict any unexpected information and reveal the information. Artificial intelligence, according to Google's Director of Engineering, Ray Kurzweil, will "reach human levels of intellect" by the year 2029. The **Natural Language Toolkit (NLTK)** is a set-up of libraries and projects that can be utilized for emblematic and measurable regular

language handling in English, written in Python.

### **NLP MARKET VALUE**

The natural language processing industry is expected to grow substantially over the next several years, benefiting people all over the world. Nearly 14 times greater than it was in 2020, the NLP market is expected to grow by 2027. It's projected to grow from \$3 billion in 2020 to \$43 billion by 2027. The field of artificial intelligence known as "natural language processing" aims to help computers better interpret, interrupt, and control human speech. The global Natural language processing market is expected to develop at a CAGR of 20.2 percent from 2021 to 2030, with an estimated market value of US\$ 10 billion in 2020. By 2030, the market is predicted to be worth \$63 billion.

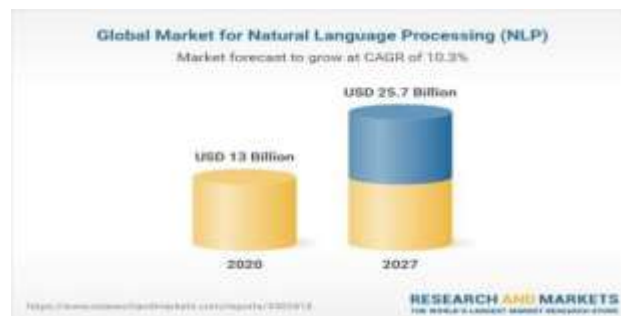


Figure 2. NLP market level

### **Future Implementations**

There are many future implementations of Natural Language Processing in our daily lives that help us make our life easier and better. Data Science Machine Learning is included while making these, it makes the working easier. Stated below are some uses of NLP for the same-

- **The Bots**

The bot is an electronic device that is used to perform the task assigned by the user and it is operated with the help of the internet. Bots are automated, which means they cannot operate on their own, every time you must start it manually. They usually do repetitive tasks and are faster than humans as they give the response in the blink of

an eye. When you visit any website, you see a box that appears on the screen asking you certain questions about the search you've made or about the website you currently are in. In response to it when you choose the answer to the options given on the screen, they display the reply to that question and this process is a continuous process that you will never feel that it is a one-way conversation. It is fast, faster than a normal human being, it is reliable as it gives you the appropriate answers to all the questions, and it is available 24x7 i.e., you access it any time and from any where across the globe all you need is the internet for it. It is accessible in the language that you speak so there is no language barrier for it [7].





Figure3.humanmachineconversations

- **Thesmartersearch**

The smarter search engines are the simplest way of searching the text and other information more smartly. It includes the algorithms that help to analyze the results, it uses your history results for the search as well. Works on the analogy, data, and the results to give the required data on the screen. Be it graphs or some information that is required. It includes various ways to search for a result on the internet in easier and faster ways. You can easily find the answer from a text scan or picture scan. All the users must do is scan that particular by bringing it in the scanning range of the frame and it will process the results itself. After that many results come on the screen and you just must choose from all of them [8]. Well, there is a con of it that sometimes the search fails as the results you are searching for are either not available or the content that you demanded is not sufficient for it. Besides everything, it is the

fastest way of getting your route, scanning a number, finding a location, finding a dress, etc.

- **Autocorrection**

Autocorrection is a way of correcting something to make it right, it is usually done with the spellings and the wrong grammatical sentences and syntaxes. One of the day-to-day tools used is the internet 'Autocorrection'. It is required by everyone as everyone makes mistakes while writing, due to which this feature was invented to rectify the mistake. While typing the mistakes change themselves which is called autocorrection. These are inbuilt features in the keyboard that can be switched on and off by going into the settings of a particular device [9]. It is also called the **replace- as-you-type** feature of the keyboard using Natural Language Processing. It not only corrects the texts and grammar but also gives you suggestions for a better word or phrase that can be replaced easily and will suit better on the sentence.

5816



Figure4.AutocorrectusingNLP

- **Virtual Assistance**

The most used feature of Natural Language

Processing is Virtual Assistance. Virtual Assistance is a unique voice detection attribute



that analyses what you are saying. It can also work on your command like it can switch off and, on the lights, it can play music, it can switch on the fan, etc. The various examples of this fascinating technology that has attracted millions are Siri, Cortana, Alexa, Ok Google, and many more. Each of them has a special feature for themselves and is built in different technologies like, iOS has Siri, Android has Alexa and Microsoft has Cortana [10]. All of them work on the voice command of a human. You can mak

them pick a call or make a call, text someone, play music from any app you like, set reminders, set alarms, it is a tool to make your lives easier. It saves time and money both that make it a useful tool. It is also user-friendly, anyone of any age can use it. There are various drawbacks for it as well, in many surveys, it was found that people are tending towards this feature of Natural Language Processing that it's becoming an alarming cause for the health of human beings [11].



Figure 5. Human versus artificial intelligence

5817

### Email spam

Email is an electronic way of transferring information. Spam emails are very common in your email account. [12] Spam email can also disturb your routine. Due to this, the email accounts are installed with an email filter. This takes a long time to delete all the spam emails. According to studies it has been

observed that there is a rapid increase in spam emails, almost 89% of the total email traffic is of spam emails. It can cause financial loss or identity theft of the user. Spammers are used to by pass the manual filters like the [13] misspelled words by adding extra letters to it. Using machine learning in this can also help to detect it faster. [14]



Figure 6. restrict spam mails

### CONCLUSION

The consideration is made that Natural

Language Processing (NLP) is a branch of computer science that studies how computers



can be used to comprehend and modify Natural Language text or voice to accomplish meaningful tasks. And its major protocol is that' to convert human language to computer language and make conversation easy between machine and humans. But after surveying in literature review, the study came to broaden about the other applications of NLP in other fields too. Although NLP plays major role in linguistics problems but now after so many evolutions in the field of artificial intelligences, there were new inventions for NLP also. and now NLP performs various functions with various domains.

In current era of new technology, NLP plays a significant part in **MACHINE LEARNING** be it document segmentation, part-of-speech tagging, and word sense clarification are all included in this process. Also, NLP performs specific tasks with deep learning too. And after this survey you gained knowledge for other applications of NLP which can be used in future for safety as well as of contentment purpose. Some of these were Email spam, Virtual Assistance, Chatbots etc. Earlier before this survey few protocols of natural language processing were in sight But and after this numerous Innovations for and with NLP can be studied.

### References

[1] D. A. Krishna Prakash Kalyanathaya, "Survey paper," *Advances in Natural Language Processing –A Survey of Current Research Trends, Development Tools, and Industry*, vol. 7, issue 5, p. 13, February 2019.

[2] Aiysha Sadi. Archana R., "Survey Paper," *Survey on Natural Processing and its applications*, vol. 7, no. 1, p. 11, January 2018.

[3] S. L. Martin Emms, " Research Paper," *Machine Learning for Natural Language*

*Processing*, p. 103, August 2007.

[4] Priya B.J.Nandhini "research paper," *An Analysis of the Applications of Natural Language Processing in Various Sectors*, p. 5, October 2021.

[5] V. Gupta, "survey paper," *A Survey of Natural Language Processing*, vol. 5, no. 1, p. 3, January 2014.

[6] A. Shaikh. D. More, "survey paper," *A Survey Paper on Chatbots*, vol. 6, no. 4, p. 5, April 2019.

[7] DOMINIK CVETEK, "conference paper," *APPLICATION OF NLP ALGORITHMS IN ITS MARKET ANALYSIS*, p. 11, May 2018.

[8] Diksha Khurana. A. kholi, "survey paper," *Natural Language Processing: State of The Art, Current Trends and challenges*, p. 26, August 2017.

[9] R. Bose, *Natural Language Processing: Current state and future directions*, p. 11, January 2004.

[10] J. Q. W. Y. W. Ray Oshikawa, "survey paper," *A Survey on Natural Language Processing for Fake News Detection*, p. 8, 2020.

[11] S. Jusoh, "research paper," *A study on nlp applications and ambiguity problems*, vol. 96, no. 6, p. 12, March 2018.

[12] Dongyeop Kang. W Ammar "research paper," *A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications*, p. 13, June 2018.

[13] Paulo Silva. C. Godhino, "survey paper," *Using Natural Language Processing to Detect Privacy Violations in Online Contracts*, p. 3, March 2020.

[14] Zhiqiang Zeng. Hua. Shi, "survey paper," *Survey of Natural Language Processing Techniques in Bioinformatics*, vol. 2015, p. 10, October 2015.

