IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

A Review Paper on Natural Language Processing (NLP)

Adam M Sanadi[1] and Venkatesh Bhat[2]
Student, Department of Computer Science and Engineering[1]
Senior Associate Professor, Department of Computer Science and Engineering[2]
Alva's Institute of Engineering and Technology, Tenkamijar, Karnataka, India

Abstract: *Natural Language Processing (NLP) is a technology that allows machines to become more human-like, narrowing the gap between humans and machines. In a nutshell, NLP allows humans to converse with machines more readily. NLP has a wide range of applications that have been developed during the last few decades. The majority of these are really helpful in everyday life, such as a machine that accepts voice commands. Many research organisations are working on this problem in order to produce more practical and usable solutions. Natural Language Processing has a lot of potential for producing computer interfaces that are easier to use for humans, because people will be able to communicate with computers in their own language rather than having to learn a computer language.*

Keywords: Natural Language Processing

I. INTRODUCTION

Natural language processing (NLP) is an artificial intelligence area in which computers intelligently analyse, comprehend, and infer meaning from human language. Developers may use NLP to organise and arrange knowledge so that they can accomplish tasks like automatic summarization, translation, named entity identification, connection extraction, sentiment analysis, audio recognition, and topic segmentation.

"Apart from common word processor operation that treat text as a series of symbols," John Rehling, an NLP expert at Meltwater Group, writes in How Natural Language Processing Helps Uncover Social Media Sentiment, "NLP considers the hierarchical structure of language: several words make a phrase, several phrases make a sentence, and, ultimately, sentences convey ideas.NLP systems have long served a beneficial purpose by evaluating language for meaning."Such as correcting grammar, converting speech to text and automatically translating between languages."

NLP is a text analysis technique that allows robots to comprehend human speech. Automated text summarization, sentiment analysis, topic extraction, named entity recognition, parts-of-speech tagging, connection extraction, stemming, and other real-world applications are possible because to this human-computer interaction. Text mining, machine translation, and automated question answering are just a few of the applications for NLP.

In computer science, NLP is regarded as a challenging topic. Rarely is human language exact or straightforward. To comprehend human language, one must comprehend not just the words themselves, but also the concepts and how they are related to generate meaning. Despite the fact that language is one of the most simple things for the human mind to understand, the ambiguity of language makes natural language processing a challenging subject for computers to grasp.

II. HISTORY OF NATURAL LANGUAGE PROCESSING

As previously stated, the notion sprang from a necessity for Machine Translation in the 1940s. The original languages were English and Russian at the time. However, other words, such as Chinese, were also used in the early 1960s. Then, in 1966, MT/NLP had a bad year, as evidenced by an ALPAC study claiming that MT/NLP was on the verge of extinction due to a lack of progress in this field at the time. This situation improved in the 1980s, when products connected to MT/NLP began to provide clients with some results. When the idea and need for Artificial Intelligence arose in the 1960s, the NLP/MT received a new lease on life.

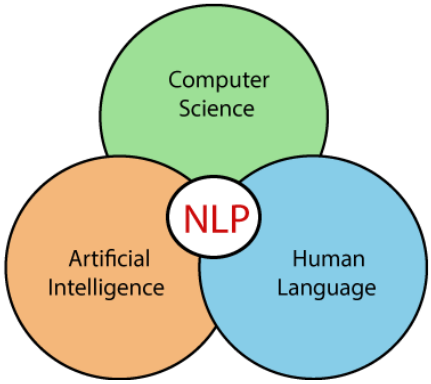
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

W.A. Woods created LUNAR in 1978 with the goal of analysing, comparing, and evaluating chemical data on lunar rock and soil composition that was accumulating as a result of Apollo moon missions and answering the related question. Computational grammar became a highly active subject of research in the 1980s, and it was related to the science of reasoning for meaning and taking into account the user's views and intentions. The rate of growth of NLP/MT accelerated in the 1990s. With the parsers, grammars, tools, and practical materials linked to NLP/MT became available.

The work on the lexicon received a research direction, while research on fundamental and futuristic problems such as word sense disambiguation and statistically coloured NLP received a research direction. Other crucial subjects such as statistical language processing, information extraction, and automated summarization were added to this search for the creation of NLP. The history of NLP would be incomplete without mentioning ELIZA, a chatbot software developed at MIT's Artificial Intelligence Laboratory between 1964 and 1966. Joseph Weizenbaum designed it. It was a programme that was built on a script called DOCTOR, which was arranged to Rogerian Psychotherapist and employed rules to respond to users' psychometric-based queries. It was one of those times. It was one of the few chatbots that could pass the Turing test at the time.

III. What is Natural Language Processing?

Natural Language Processing (NLP) is a branch of computer science that combines human language with artificial intelligence. It is the technology that allows machines to comprehend, analyse, manipulate, and interpret human speech. It aids developers in organising their information for tasks like translation, automated summarization, Named Entity Recognition (NER), audio recognition, relationship extraction, and topic segmentation.



IV. Natural Language Processing Use Cases

In many current real-world applications, natural language processing is the driving force behind machine intelligence. Listed below are a few examples:

- **Spam Detection:** You might not think of spam detection as an NLP solution, yet the top spam detection solutions examine emails for language that commonly indicates spam or phishing. Overuse of financial words, typical bad grammar, threatening tone, improper haste, misspelt corporate names, and other signs can all be used as indicators. Although you may object that this does not reflect your email experience, spam detection is one of a handful of NLP problems that experts consider' mostly solved.'
- **Machine Translation:** Google Translate is a good example of NLP technology at action. More than just substituting words in one language with words in another is required for really useful machine translation. Effective translation must precisely capture the meaning and tone of the input language and convert it to text in the output language with the same meaning and desired impact. In terms of accuracy, machine translation systems are making good progress. Translating text to one language and then back to the original is an excellent approach to test any machine translation algorithm. A well-known classic example: "The spirit is willing but the flesh is weak," as translated from English to Russian not long ago, resulted in "The vodka is wonderful but the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

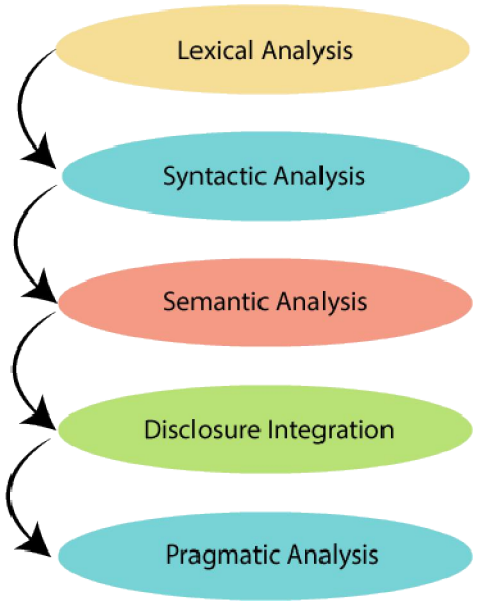
meat is rotting." Today, the outcome is "The spirit craves, but the flesh is weak," which isn't flawless but gives the English-to-Russian translation a lot more confidence.

- **Virtual Agents and Chatbots:** Virtual agents like Apple's Siri and Amazon's Alexa identify patterns in voice requests and reply with appropriate action or helpful comments using speech recognition and natural language generation. In response to written text entries, chatbots work the same magic. The best of them also learn to understand contextual hints in human requests and utilise them to improve their replies or possibilities over time. The capacity to react to our questions—anticipated or not—with meaningful and useful replies in their own words is the next development for these apps.
- **Social Media Sentiment Analysis:** Natural language processing (NLP) has evolved into a critical business tool for identifying hidden data insights from social media channels. Sentiment analysis may extract attitudes and emotions in reaction to goods, promotions, and events by analysing language used in social media postings, answers, reviews, and more–information that organisations can utilise in product design, advertising campaigns, and more.
- **Text Summarising:** Text summarization employs natural language processing (NLP) techniques to digest large amounts of digital text and provide summaries and synopses for indexes, research databases, and busy users who don't have time to read the entire document. To provide valuable context and conclusions to summaries, the finest text summarising programmes utilise semantic reasoning and natural language generation (NLG).

4.1 Different levels of NLP

NLP has a very complex shape and organisation. It's really vague. There are many degrees of analysis-

1. Lexical Analysis
2. Syntactic Analysis
3. Semantic Analysis
4. Disclosure Integration
5. Pragmatic Analysis



A. Lexical Analysis

Lexical analysis is a branch of NLP that studies words at the level of their lexical meaning and part-of-speech. The lexicon of a language, which is a collection of individual lexemes, is used at this level of linguistic processing. A lexeme

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

is a fundamental element of lexical meaning; it is an abstract unit of morphological analysis that reflects a single morpheme's set of forms or "senses."

For example, "duck" can be a noun or a verb, but its part-of-speech and lexical meaning can only be deduced from the context of the other words in the phrase/sentence. This is, in reality, a first step toward a more complex Information Retrieval system that uses part-of-speech tagging to increase accuracy.

B. Syntactic Analysis

The lexical analysis' part-of-speech tagging output can be used to group words into phrase and clause brackets at the syntactic level of linguistic processing. Syntactic analysis, often known as "parsing," allows for the extraction of phrases that communicate more meaning than individual words alone, such as a noun phrase.

Parsing may be used to enhance indexing in information retrieval since phrases can be utilised as representations of texts, providing more information than single-word indices. Similarly, terms produced syntactically from the question provide stronger search keys for matching with documents that have been processed similarly.

Nonetheless, syntax can be ambiguous at times, as in the case of the news headline: "Boy paralysed after tumour fights back to gain black belt" — which actually refers to a boy who was paralysed due to a tumour but persevered in fighting the disease and eventually achieved a high level of martial arts competence.

C. Semantic Analysis

The semantic level of language processing is concerned with determining what a phrase truly means by linking syntactic elements to the context and disambiguating terms having different definitions. Rather than analysing individual words or phrases, this level requires the accurate interpretation of the meaning of sentences. The query and document matching procedure in Information Retrieval can be done on a conceptual level rather than in basic words, boosting system accuracy. Furthermore, by using semantic analysis on the question, term extension with the use of lexical sources would be conceivable, allowing for better retrieval of relevant materials even if precise phrases are not used in the query. Precision may improve as a result of query expansion, and recall is likely to improve as well.

D. Disclosure Integration

The discourse level of language processing is concerned with the examination of text structure and meaning beyond a single phrase, as well as the formation of links between words and sentences. Anaphora Resolution is also accomplished at this level by recognising the entity referred to by an anaphor (most commonly in the form of, but not limited to, a pronoun).

Document and query representations are improved as a result of the ability to recognise and resolve anaphora relationships, because the implicit presence of concepts is accounted for throughout the document as well as in the query at the lexical level, and an integrated content representation of the documents and queries is generated at the semantic and discourse levels.

Because parts may be split down into (1) Title, (2) Abstract, (3) Introduction, (4) Body, (5) Findings, (6) Analysis, (7) Conclusion, and (8) References, structured texts benefit from discourse analysis. The exact responsibilities of bits of information are established as to whether it is a conclusion, an opinion, a forecast, or a fact, which considerably improves information retrieval systems.

E. Pragmatic Analysis

The pragmatic level of language processing is concerned with the application of real-world information and the comprehension of how this affects the meaning of what is being conveyed. A more complete representation is generated by examining the contextual dimension of the documents and queries.

This level of Natural Language Processing in Information Retrieval is largely concerned with query processing and comprehension by incorporating the user's history and intentions, as well as the context in which the query is being made. Contexts can include things like time and place.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

This level of analysis allows the interaction between the IR system and the users, allowing the elicitation of the purpose for which the information being sought is intended to be utilised, ensuring that the information retrieval system is fit for purpose.

4.2 Components of NLP

NLP is broken down into two parts.

1. Natural Language Understanding
2. Natural Language Generation

A. Natural Language Understanding (NLU)

The field of natural language processing (NLP) is concerned with how computers are designed to analyse language and promote "natural" back-and-forth communication between people and computers

On the other hand, natural language comprehension is concerned with a machine's capacity to comprehend human language. NLU is the process of rearranging unstructured data so that machines can "understand" and evaluate it.

Consider it this way: Before a computer can convert unstructured text into a machine-readable format, it must first comprehend the nuances of human speech.

Natural Language Understanding Examples:

- Automatic Ticket Routing
- Machine Tranclation (MT)
- Automated Reasoning
- Automatic Ticket Tagging and Reasoning
- Question Answering

B. Natural Language Generation (NLG)

NLG is a software technique that automatically turns data into plain-English content. It is a subfield of artificial intelligence (AI). By composing the phrases and paragraphs for you, the technology may truly present a tale that is identical to that of a human analyst. NLG is one of the most rapidly gaining traction in the business world. NLG has a wide range of applications, but it is most successful when used to automate time-consuming data processing and reporting tasks.

Data has long been used to communicate ideas. However, the organisation must find inventive methods to keep up with the explosion of data that has to be evaluated and understood, as well as rising efforts to cut costs and fulfil consumer requests.

It turns out that a computer can transmit concepts from data at a massive scale and with incredible precision. It may also do it in a very eloquent manner. Productivity rises when routine analytical and communication chores are automated, allowing staff to focus on higher-value activities.

Natural Language Generation Examples:

- Deep, Automated Personalization
- AI-Generated Narrative Reporting
- Advanced Monitoring for the Interner of Things
- Content Creation at Scale
- Conversational AI

4.3 Pros of Natural Language Processing

1. The NLP system provides precise replies to inquiries, with no extraneous or undesired data.
2. The quantity of relevant information offered in the questions enhances the correctness of the answer.
3. Organizing a large amount of unstructured data.
4. Users may ask questions about any topic and receive an immediate response.
5. It is simple to put into practise.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

6. It is less expensive to use a software than to hire a human. The activities outlined can take a person two or three times longer than a machine to complete.
7. The Natural Language Processing (NLP) system responds to inquiries in natural language.
8. Allow you to compare more language-based data to a human without becoming fatigued and in a consistent and fair manner.
9. The NLP technique enables a computer to speak with a person in their own language while also allowing it to perform other language-related activities.
10. It has a quicker response time for customer support.

4.4 Cons of Natural Language Processing

1. The NLP system does not have a user interface that is devoid of elements that allow users to engage with it more deeply.
2. If a new model must be developed without utilising a pre-trained model, it might take up to a week to produce satisfactory results, depending on the amount of data.
3. Because of its restricted features, the system was designed for a particular and specialised goal and is unable to adapt to new areas and challenges.
4. If a question is poorly written or confusing, the system may not be able to offer the proper answer in sophisticated query language.
5. It's not 100 percent reliable, and it'll never be 100 percent trustworthy. There is a chance that its predictions and outcomes will be incorrect.

V. CONCLUSION

Natural Language Processing is a kind of Artificial Intelligence that allows robots to understand and comprehend human language. Machines can understand written or spoken content using natural language processing. The practise of teaching robots to read and interpret human conversational inputs is known as natural language processing. Machine Learning-based Natural Language Processing may be used to construct communication channels between humans and machines. NLP has already shown effective in a variety of disciplines, despite the fact that it is still growing. NLP can help organisations and people save time, enhance efficiency, and raise customer satisfaction through various applications. Translation, speech recognition, sentiment analysis, question/answer systems, automatic text summarization, chatbots, market intelligence, automatic text categorization, and automatic grammar checking are some of the available NLP-based technologies. Natural Language Processing has both advantages and cons when used. When businesses use NLP, they may save money, minimise client wait times, and improve customer satisfaction. However, training takes time, and machine learning is never 100 percent accurate.

REFERENCES

[1]. S. Chaturvedi, H. Peng, and D. Roth. Story comprehension for predicting what happens next. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1603–1614, 2017.
[2]. Z. Chen, H. Zhang, X. Zhang, and L. Zhao. Quora question pairs. https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs, 2018.
[3]. L. Logeswaran and H. Lee. An efficient framework for learning sentence representations. ICLR, 2018.
[4]. FangtaoLi§, YangGao† et al. “Deceptive Answer Prediction with User Preference Graph”.
[5]. Yonghui Wu, Mike Schuster, Zhifeng Chen et al. “Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation”.
[6]. Luisa Bentivogli, Arianna Bisazza, and Mauro Cettolo “Neural versus Phrase-Based Machine Translation Quality: a Case Study”
[7]. Dong Yu, Adam Eversole, Mike Seltzer, Kaisheng Yao, Zhiheng Huang, Brian Guenter, Oleksii Kuchaiev, Yu Zhang, Frank Seide, Huaming Wang, et al. 2014. An introduction to computational networks and the computational network toolkit. Microsoft Technical Report MSR-TR-2014–112.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, March 2022
Impact Factor: 6.252

[8]. Adam Paszke, Sam Gross, Soumith Chintala, Gre- gory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In NIPS-W.

[9]. Baud, R. H., Alpay, L., & Lovis, C. (1994). Let's Meet the Users with Natural Language Understanding. Knowledge and Decisions in Health Telematics: The Next Decade, 12, 103.

[10]. Wan, X. (2008). Using only cross-document relationships for both generic and topic- focused multi-document summarizations. Information Retrieval, 11(1), 25-49.