International Journal of Scientific Research & Engineering Trends
Volume 4, Issue 4, July-Aug-2018, ISSN (Online): 2395-566X

Natural Language Processing (Nlp) : A Comprehensive Study

M.Tech. Scholar Rajat Verma
Department of Computer Science & Engineering
Amity School of Engineering & Technology
Amity University, Lucknow
Uttar Pradesh, India
rajatverma310795@gmail.com

***Abstract* – we the people, are living in this modern era of Artificial Intelligence. This is clearly depicted in this paper. The combination of lexical, syntactic, semantic usage in the domain of social networking which is an application of NLP is also mentioned in this paper. Turing Test is also mentioned in this paper. The individual functionalities of lexical, syntactic and semantic analysis is also illustrated in this paper. These all features will clearly increase the usage of all 3 will bring a better tomorrow in the field of Natural Language Processing**.

***Keywords* – Natural Language, Processing, Lexical, Syntactic, Semantic, Turing Test.**

I. INTRODUCTION

Natural Language Processing [1] is associated to computer science abbreviated as 'CS' as well as Artificial Intelligence commonly abbreviated as 'AI'. AI [2] is particularly famous as a superset of machine learning. NLP is related to the communication between the computer (artificial) language as well as the language used by the homo-sapiens that is very original as well as natural.

The antiquity of Natural Language Processing, commenced from the 1950s. A well renowned computer scientist, Alan Turing [3] developed a test in the year 1950. This test was termed as a "Turing test". Fig.1. depicts a portrait of Alan Turing.

Fig.1. Alan Turing.

The target of the Alan Turing Test was to illustrate that the machines can think cleverly like homo-sapiens or not?

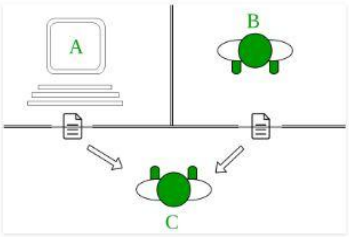


Fig.2. Turing test.

In the given figure above, A is a computer. B is a human being but as a player. C is the interrogator and he is a human being also.

NLP is linked with other concepts such as Machine Learning and deep learning, this is illustrated in Fig.3.

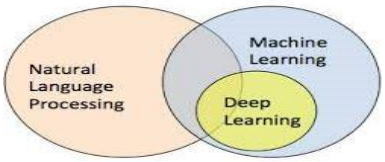


Fig.3 Relation among NLP, Machine Learning and deep learning.

Machine learning [4] is a subset of artificial intelligence, and helps in the learning procedures for machine aspect with the usage of statistics.

Fig.4. a view to statistics (sample).

Deep Learning [5] is concerned with wider aspect of machine learning. For example, deep neural Networks, deep belief network is a part of deep learning. In machine learning, it only had neural networks and not deep neural networks.

An Automated Online Assistant is an example of natural language processing and it is depicted in Fig.5.

Fig.5 an automated online assistant.

In the given figure above, Maya is the online assistant that will answer the questions, the user puts in the message box and then sends it. Two components will be done here one is understanding of the question and the other is answering of the question.

II.EVOLUTION OF NATURAL LANGUAGE PROCESSING

The Evolution of Natural Language Processing is depicted in Fig. 6.

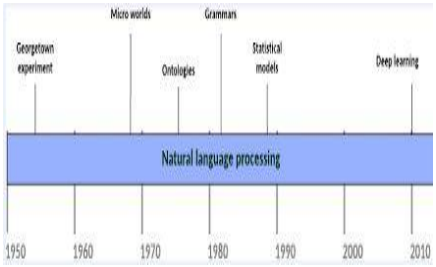


Fig.6 Evolution of Natural Language Processing

Between 1950 and 2010, many popular experiments and tests were taken and new technologies were developed. Few examples are Georgetown Experiment, Micro worlds, Ontologies, grammars, statistical models and deep learning etc.

III.ELEMENTS OF NATURAL LANGUAGE PROCESSING

There are 2 elements of NLP that are as follows-

- Understanding of Natural Language
- Generations of Natural Language

In the first phase, the inputs get converted into essential and helpful formats or representations, to examine the different components of speech.

As the real world's unprocessed raw facts and figures are very noisy, the language can be very confusing that means one word can have many meanings.

These confusions can be of 3 types-

- Lexical confusion
- Syntactic confusion
- Referential Confusion

The Lexical confusion [6] deals with the word confusions, the syntactical confusions [7] deals in the parsing procedure and the referential confusion [8] deals with the meaning confusions.

An example of Referential confusion can be Rajat went to Rajan and said 'I am angry'. Now in this sentence a human mind can judge that Rajat is angry but the computer cannot check it properly. In terms of computer it should be Rajat went to Rajan and Rajat said 'I am angry'.

The 2nd phase involves the conversion of the representations to the normal language that is to be understood by humans or homo-sapiens.

This procedure can be done in a sequence of 3 states-

- Text Planning
- Sentence Planning
- Text Realization

These 3 steps are the solutions of the 3 confusions mentioned above. Text planning deals with the extraction of mastery from the mastery base.

The second one deals with the selection of appropriate words and forming sequences of words. The last one deals with the conversion of all process discussed above into reality.

IV.FLOW CHART (NLP STEPS)

It can be illustrated with the following figure mentioned below as Fig.7.

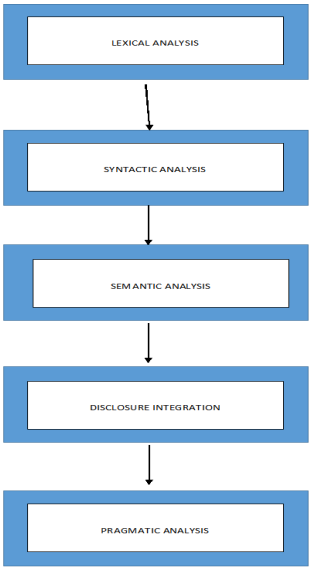


Fig.7 NLP steps.

V.RESEARCH BACKGROUND

1. Lexical Analysis

The term "Lexical" [9] is concerned with the dimension of concordance. Directly coming to the point that is by

taking an example in the field of social networking [10], let's take the case of Facebook, people use to post statuses on their wall so that their public can watch that. It could be happy, sad and angry or any kind of emotion or sentiment, we are free to post any feeling that we feel. This can be expressed by the means of words. Some example could be "I bought a new IPhone, it is amazing and I am happy", "Alas! My car broke down in an accident". In the first example, two words that are amazing and happy expresses happiness whereas in the second example, alas and accident expresses a feeling of sadness. People use this technique in the caption of their pictures other than statuses.

Coming to the aspect of twitter, **"#"** is used to indicate something. Let's take an example of a recent tournament that happened in Dubai. It was "Kabaddi Masters", if a person has to express a sentiment then he/she must use #KabaddiKurfew in his/her tweet. An example could be, a tweet used by a legend player "Virendra Sehwag", he tweeted "Another demolition.

What a great performance by India to beat Pakistan once again. Ek aur jeet dila Diya humka Thakur, Ajay Thakur. 41-17 this time after 36-20 a few days ago. Well done, Team India #IndvPak #KabaddiKurfew". This was done when India had beaten Pakistan in a match. It was a sentiment of happiness. In the field of compiler also, lexical analysis is the first thing to do. Sentiment Analysis is an application of Natural Language Processing.

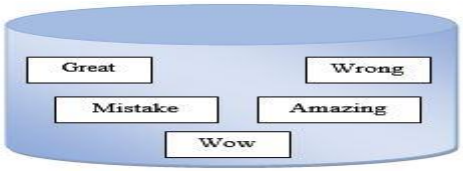


Fig.8 An example of a lexicon of sentiments.

In the illustration mentioned above a sample of sentiment lexicons are mentioned. Great, Amazing and Wow expresses joy and happiness whereas Wrong and Mistake expresses bad feelings or sentiments.

2. Syntactic Analysis

This term is responsible for the syntactic connections among locutions. The Parts of Speech commonly abbreviated as (POS) [11] plays a deep impact role in this syntactic scenario. Let's take the case of Verb that are basically of two types that are active verb and passive verb. If we consider a passive verb then an example of it could be "Person was good". If we consider an active verb then the example of it could be "He protested". They are syntactically correct but are semantically weak.

This syntactic approach [12] also has positive and negative viewpoints. An example of positive view point can be "Many Congratulations" and an example of negative viewpoint in this dimension "He Killed". All these examples can be posted as facebook statuses or as a caption of a picture. In the criteria of compiler phases, syntactic analysis comes in 2nd position after the lexical analysis.

Examining a sequence of string can be considered as syntactic analysis. Parsing is just a synonym of it.

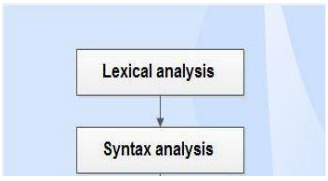


Fig.9 illustrating the appropriate use of syntax analysis that is after lexical analysis.

Coming to the aspect of twitter, it is not an obvious thing that each and every time the tweets are syntactically correct that is Subject then Verb then Object. Many a times, tweets can be noisy and messy as well as not-structured that becomes a problem in maintaining a syntactic approach. This statement becomes true when we think of real world data (unprocessed raw facts and figures) is highly inconsistent.

In twitter also, like facebook can happen that syntactically a statement is correct but not semantically. And if a statement is not syntactically correct, there is very less probability that semantically it would be correct!

3. Semantic Analysis

The Third component that comes in this amalgamation [13] process is semantics. Its meaning is concerned with rationality along with sense. It should be done after the syntax analysis have already been done.

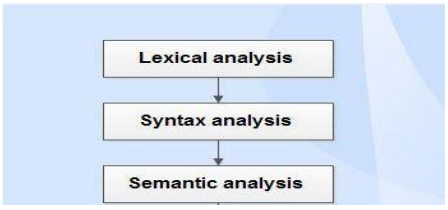


Fig.10 illustrating the correct stage of semantic analysis that is after syntactic analysis.

In the dimension of social networking basically talking about facebook and twitter, the semantics plays a major role.

One term that meant here is "Contextual Semantics". It particularly means that "Terms that happen in homogeneous situations basically meant similar". In the year 1953, it was told by "Wittgenstein". An Illustration of it can be depicted with a figure given below-

In the positive manner, if the word "Trojan Horse" is used, in social network then as a status or post it man refer to Greek Culture, A Tale, History, Class,, Wooden, Troy etc.

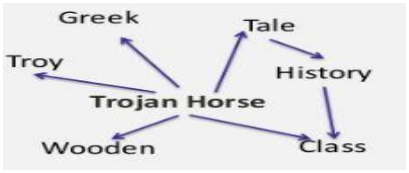


Fig.11 Positive meaning of words (Trojan horse)

If taken in a negative manner, it may be used for some code, malware, threat, hack, program etc.

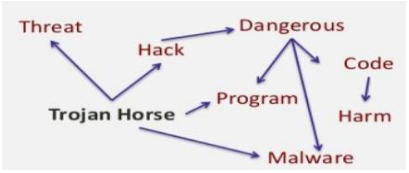


Fig.12 Negative Meaning of words (Trojan horse)

The point is that, on social media positive things may become viral or not but negative things become viral at a very alarming rate!!.

This word can be used as a post, status or a tweet on social media and can be used in both ways that is in favorable or non-favorable aspect.

Some structure that belongs to Contextual Meaning Emotions/Sentiments also plays a vital role in this scenario. It can be depicted as few terms that are mentioned in variable tweets with similar meaning, logics and emotions tend to form clusters. An illustration of this is depicted in Fig.13.

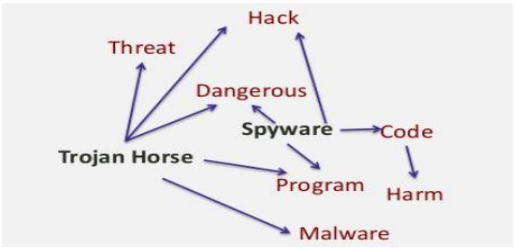


Fig.13 Trojan horse and Spyware are combined to form a new scenario.

The Fig.13 follows a negative contextual pattern and if presented on social media, it will expand exponentially!

1.Disclosure Integtartion

In this scenario, the sense of a sequence of words is validated with the sequence of words above them.

2.Pragmatic Analysis

In this case, the real world knowledge in concern to linguistics [14] is necessary. The sequences of words are re-interpreted to check the accuracy of meaning in the mentioned area.

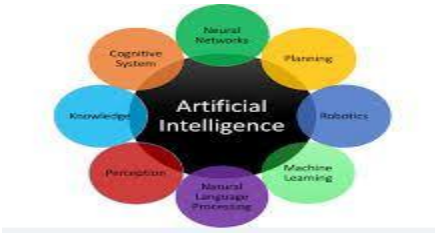
3. Applications

Some of the applications are as follows [15]-

- Natural language generation.
- Speech Recognition.
- Sentiment Analysis.
- Question Answering.

VI.CONCLUSION

The Natural Language Processing is one of the dimensions of Artificial Intelligence.



It has contributed in the research sector alone as well as when in combination with Machine Learning and Deep Learning. The individual aspects of Lexical, Syntactic and Semantic analysis is mentioned in this paper. Disclosure Integration and Pragmatic Analysis is also mentioned in this paper. Elements as well as evolution of NLP is also mentioned in this paper. Introduction as well as Turing Test is also depicted in this paper. All this will definitely contribute in the research and development sector.

ACKNOWLEDGMENT

I would like to thank my family and friends for always supporting me in the research dimension.

REFERENCES

[1]. Chowdhury, G. G. (2003). Natural language processing. Annual review of information science and technology, 37(1), 51-89.
[2]. De Kleer, J. (1986). An assumption-based TMS. Artificial intelligence, 28(2), 127-162.
[3]. Turing, A. M. (2009). Computing machinery and intelligence. In Parsing the Turing Test (pp. 23-65). Springer, Dordrecht.

[4]. Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. Machine learning, 3(2), 95-99.
[5]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436.
[6]. Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & cognition, 14(3), 191-201.
[7]. MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological review, 101(4), 676.
[8]. Maes, S. H., & Neti, C. V. (2005). U.S. Patent No. 6,964,023. Washington, DC: U.S. Patent and Trademark Office.
[9]. Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.
[10]. Lenhart, A., & Madden, M. (2007). Social networking websites and teens: An overview.
[11]. Hengeveld, K. (1992). Parts of speech. Layered structure and reference in a functional perspective, 29-55.
[12]. Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. Journal of cognitive neuroscience, 11(2), 194-205.
[13]. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391-407.
[14]. Attardo, S. (2010). Humorous texts: A semantic and pragmatic analysis (Vol. 6). Walter de Gruyter.
[15]. Budanitsky, A. (1999). Lexical semantic relatedness and its application in natural language processing.

AUTHOR PROFILE

Rajat Verma
I am pursuing Master of Technology in Computer Science & Engineering from Amity University, Lucknow.
Contact No: +91-8808223952
Email Id: rajatverma310795@gmail.com