



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 1, January 2019

An Epigrammatic Study on NLP -Tools, Challenges and it's Applications

Kavyashree E D¹, Sowmya Shree P², Anil Kumar C J³, Kiran B⁴

Assistant Professor, Dept. of Computer Science and Engineering, A T M E College of Engineering, Mysuru,
Karnataka, India^{1,2,4}

Associate Professor, Dept. of Computer Science and Engineering, A T M E College of Engineering, Mysuru,
Karnataka, India³

ABSTRACT: Natural Language Processing (NLP) is one of the emergent technologies in computer science, artificial intelligence, and information engineering. NLP focuses on interaction between machine and human (natural) languages. Concerned in how to program computers to process and analyze large amounts of natural language data. NLP is the ability of a computer program to understand human language as it is spoken. Data are not processed directly in a machine unless it is read and understood or processed manually or analyzed by automatic machines. In this paper we focus on NLP- tools/ libraries, challenges such as understanding, generation and speech recognition, and applications.

KEYWORDS: NLP, Text summarization, Machine translation, Natural language understanding, Natural language generation, Speech recognition.

I. INTRODUCTION

NLP was started in 1950's. In For many problems in real world various solutions are provided by Natural Language Processing. NLP is an efficient approach helps humans in assisting, writing, and assist in many ways. Text analytics is one of the NLP techniques. Human interact with the computers in their own language called natural language or human language such as English, French, Chinese and Spanish etc. eg of NLP is an automated online assistant providing customer service on a web page.[1][2][3]

NLP is computers reading **language**. **NLG** is computers writing **language**. **NLU** is computers **understanding language**. Most of the data are text form which is unstructured in nature. According to estimation in industry structured data is of only 21%. Data generate as we verbalize, message, tweet, also in blogs, articles, services review and patient records in medical sector etc. Text we see each day, It may be easier to learn to speak than to write.

In order to get the right data it is efficient to use right technology known as NLP. It is the branch of data science. NLP derive the data in more efficient manner by analyzing, understanding, and deriving the information from the text data in a smart and effective manner. By using NLP and its components user can organize the huge chunk of data i.e text, perform automatic task and solve problems like machine translation, relationship extraction, speech recognition, topic segmentation, automatic summarization, entity recognition, and sentiment analysis. Tokenization- method of converting a text into tokens. Tokens –words or entities present in the text. Text object- a phrase or a sentence or a word or an article.

Text processing : Text is the unstructured data as noise are present in data, it is not easily analyzable and difficult to process. Method of making data noise free and ready for analysis is called as text preprocessing. A way of computers to analyze understand and derive meaning from human language in a smart and useful way in 1950: machine translating from Russian to English.[4][5][6]

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 1, January 2019

3 steps to perform this are

1. Removal of noise
2. Lexicon normalization
3. Object standardization

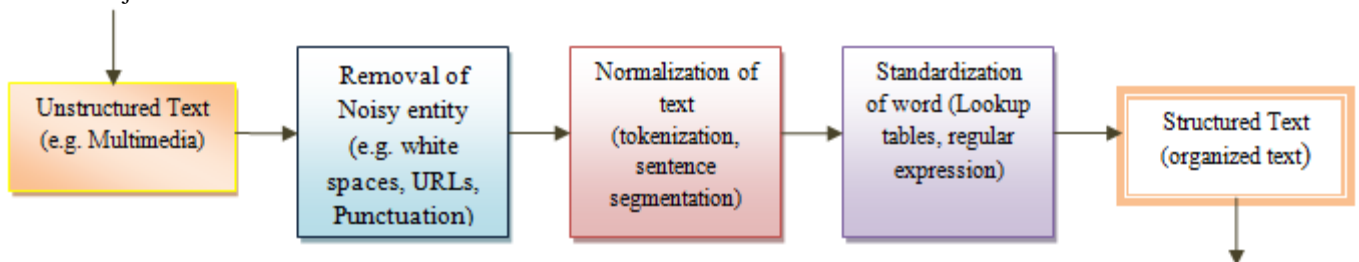


Fig 1.1: Text Cleaning Pipeline [3]

Why NLP?

Can carry out automated speech and text in less time.

II. TOOLS FOR NLP/ LIBRARIES FOR NLP PYTHON [2][7][8]

NLTK: It short for Natural Language ToolKit best tools to learn the NLP domain. Helps in solving problems of NLP. An absolute toolkit for NLP techniques

TextBlob: Interface for NLTK convert text process into detail and understandable documentation. Supports the addition of components ex like sentiment analyzer and other related tools. It is also used for building prototype for various NLP models and easily grow into full scale projects.

Gensim: One of the potential fields in NLP science.

CoreNLP: Supports huge function like Part of Speech (PoS) tagging, titled entity recognition, pattern learning parsing and many more. Originally written in Java and supports various languages like python, by using specialized wrappers. CoreNLP provide accuracy results and high in speed. They are basically used in production environment.

SpaCy: Programmed in cython. Provide only one solution for the problem thus eliminating the problem for choosing best route. Tool is robust and new features are added on a regular basis. SpaCy provides a compelling approach to NLP. NLP with python and cython. Cython is a c extension of python.

Pattern: A web mining module tool for NLP and machine learning.

Software's for NLP [9]

1. Natural language toolkit
2. Apache Open NLP
3. SpaCy
4. apache cTAKES
5. General architecture for text engineering
6. Mallet

III. NLP TASKS, CHALLENGES, EXAMPLES AND APPLICATIONS [3][10][11]

1. **Text summarization:** The text article or paragraphs are organized in a relevant order.
2. **Document to information:** Parsing documents consists of textual data into analyzable form.
3. **Optical character recognition:** An image representing a printed text, determine the corresponding text.
4. **Natural language generation and understanding:** Translate information from computer database into readable human language called natural language generation. Converting chunk of text into the logical structure that are easy to manipulate known as language understanding.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 1, January 2019

5. **Machine translation:** Automatically translate the text from one person to another by taking care of semantics, syntax and information on the real world basis.

Challenges in NLP

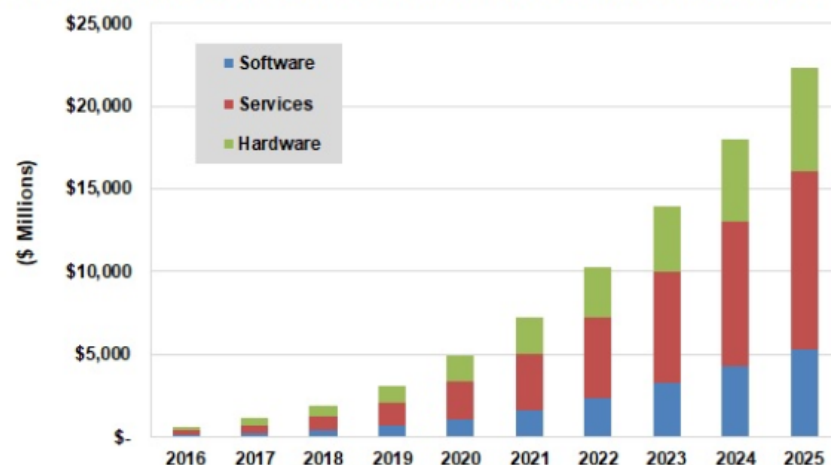
To develop natural language understanding model, it is required to use knowledge on various discipline such as linguistics, philosophy, psycholinguistics and computational linguistics. NLP focuses on acquiring, understanding and generating the human languages such as English, hindi, french, spanish etc. It faces semantics problems such as

1. **Natural language understanding:** A more precise language should be used without any ambiguity. Judgment on grammars is not a goal of language understanding.
2. **Natural language generation:** Generating natural language from machine representation system in a logical form or from a knowledge base. It is a translator converts data to a natural language representation.
3. **Speech Recognition:** It is the sub field of computational linguistics that develop methodologies and techniques, enables the translation of spoken language into text by computers. Also called as automatic speech recognition (ASR), computer speech recognition, or speech to text (STT). Speech recognition system requires training where speaker reads the text or vocabulary into the system. system analyze person voice and fine tune the voice for more accuracy.

Total Revenue of NLP in market :2016-2025

Tractica Report on the NLP in 2017, Estimates the markets revenue from 2016-2025, NLP hardware, software and services. Report also forecasts on NLP software solution Leveraging AI will see a market growth from \$ 135 million in 2016 and \$ 5.3 billion by 2025. By the fig. there is \$22.3 billion total revenue segment of NLP by 2025 according to market growth.

Natural Language Processing Total Revenue by Segment, World Markets: 2016-2025



Source 3.1 :Tractica [12]

Examples of NLP and it's applications [13]

1. Speech Recognition
2. Machine Translation
3. Optical character recognition
4. Sentiment analysis

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 1, January 2019

5. Semantic search
6. Machine Learning
7. Affective Computing
8. Natural language programming
9. Natural Language Generation
10. Text analytics

Applications: [14][15][16]

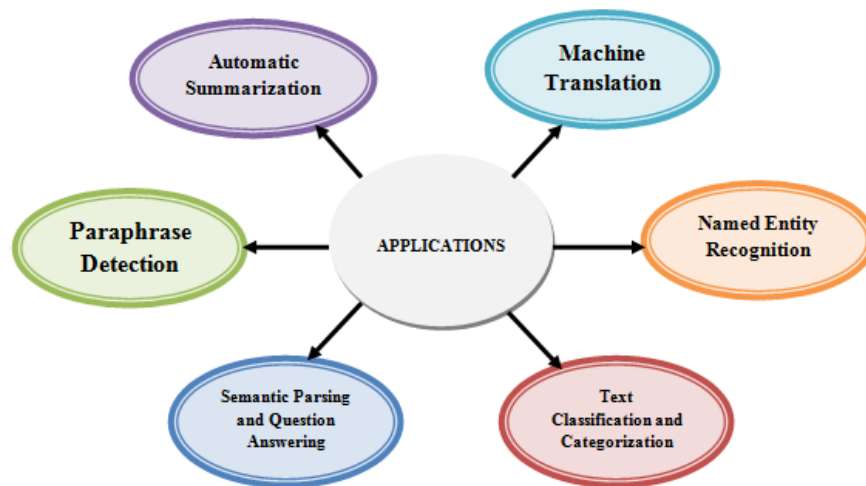


Fig 3.2: Applications of NLP

IV. CONCLUSION

NLP fills the gap between human communication and system understanding; it is increasing in communication between human and intelligent systems rapidly with advanced algorithms and efficient computations. Though computer native language is machine code, it understands the data or information in the form of zeroes and ones and produce relevant data. Understanding Human language such as English, Spanish, is complex for a computer and producing relevant output is remarkable. NLP makes it feasible for computers to hear speech, reading the text, and interpret it, measure the data and determine which part is critical within data.

REFERENCES

- [1] AfsaneMoharamkhani, LotfollahKarimi, SeyyedDariushAhmadi, The Impact of Neuro Linguistic Programming (NLP) on EFL Learners' Vocabulary Achievement, IOSR Journal Of Humanities And Social Science (IOSR-JHSS) Volume 21, Issue11, Ver. 11 (Nov. 2016) PP 27-37.
- [2] <https://towardsdatascience.com> [Accessed 28 June 2018]
- [3] <https://www.analyticsvidhya.com> [Accessed 29 June 2018]
- [4] <http://www.the-wow-collection.com> [Accessed 30 June 2018]
- [5] Dr. Khaled M. Alhawiti, Natural Language Processing and its Use in Education, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 12, 2014
- [6] DU Jia-li , YU Ping-fang , A computational linguistic approach to natural language processing with applications to garden path sentences analysis, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 9, 2012
- [7] Giovanni Guida And Ciancarlo Mauri, Evaluation Of Natural Language Processing Systems: Issues And Approaches, Proceedings Of 1-Hc Ieee, Vol 74, No 7, July 1986.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

- [8] Neelima Bhatia , Arumma J aiswal ,Automatic Text Summarization and it's Methods - A Review, 2016 IEEE.
- [9] [ONLINE] Availabel:<https://blog.algorithmia.com> [Accessed 13 July 2018]
- [10] <https://searchbusinessanalytics.techtarget.com>[Accessed 13 July 2018]
- [11] <https://machinelearningmastery.com> [Accessed 13 July 2018]
- [12] <https://www.techemergence.com> [Accessed 29 July 2018]
- [13] <https://simplicable.com/new/natural-language-processing> [Accessed 13 July 2018]
- [14] <https://www.sas.com> [Accessed 15 July 2018]
- [15] NLP Videos [Accessed 15 July 2018]
- [16] <https://www.lifewire.com> [Accessed 15 July 2018]