# Advances in Natural Language Processing – A Survey of Current Research Trends, Development Tools and Industry Applications

### Krishna Prakash Kalyanathaya, D. Akila and P. Rajesh

*Abstract--- Natural Language Processing (NLP) is a subfield of Artificial Intelligence and getting lot of focus on research and development due to emergence of its applications. The research areas in focus are conversation systems, Language processing, Machine Translation, Deep learning. The researches in these areas lead to development of many tools to build industrial applications. Combining Deep Learning techniques with Natural Language Processing is finding lot of applications in domains such as Healthcare, Finance, Manufacturing, Education, Retail and customer service. This paper provides bird's view of advancement in research, development and application areas of Natural Language Processing. This paper captures21research focus areas, 22 development tools and 6 domains where Natural Language Processing are making rapid advancements.*

*Key terms used in this text: NLP, Natural Language Processing, Deep Learning, Sentiment Analysis, Question Answering, Dialogue Systems, Parsing, Named-Entity Recognition, POS Tagging, Chatbots, Human-Computer-Interface.*

## I. INTRODUCTION

Natural Language processing (NLP) is a subfield of artificial intelligence dealing with computational algorithms to automatically represent and process various forms of human (natural) language inputs and communicate with Human-Computer-Interface (HCI). It is also known by a name "Computational Linguistics".

Natural Language Processing involves following stages of processing namely, lexical (structure) analysis, parsing, semantic analysis, discourse integration, and pragmatic analysis. Some well-known application areas of NLP are Speech Recognition, Optical Character Recognition (OCR), Machine Translation, and Chatbots.

Recently, Machine Learning algorithms are used to process Natural Language input by studying millions of examples of text — words, sentences, and paragraphs — written by humans. By studying these samples, training algorithms gain an understanding of the "context" of human speech, writing, and other modes of communication. The machine learning and deep learning algorithms are widely is used to develop frameworks for NLP and efficiently perform common NLP tasks.

## II. BRIEF HISTORY

Although the work on NLP dates back to around 1950 with the development of what is called as "Turing Test" and in 1957 a rule based system of syntactic structures. The progress was slow until 1990 due to limited computational power and systems were based on complex sets of hand written rules and limited vocabulary. With the introduction of machine learning and steady increase in computational power, recently interest on research and applications are growing. The recent major breakthrough areas of NLP are: speech recognition, language processing, dialog systems and applying deep learning techniques.

While NLP is still facing lot of challenges (like human computer interfaces), there has been lot of research interests and it has opened to many opportunities for using the techniques in robotics, automation and digital transformation.

## III. RESEARCH WORKS ON NLP

Until 1990, most of the research work was done on the NLP concepts and machine translation. Most recent research work on NLP have harnessed the power of statistical models, machine learning, Deep learning technologies that are using data driven approach.

The research topics on Natural Language Processing sometimes overlap with some artificial intelligence and Deep Learning topics. These approaches generally adopted recently to perform NLP tasks in most efficient way. The ACL 2018 Main Conference invited papers in 21areas which are Dialogue and Interactive Systems, Discourse and Pragmatics, Document Analysis, Generation, Information Extraction and Text Mining, Linguistic Theories, Cognitive Modeling and Psycholinguistics, Machine Learning, Machine Translation, Multidisciplinary, Multilinguality, Phonology, Morphology and Word Segmentation, Question Answering, Resources and Evaluation, Sentence-level Semantics, Sentiment Analysis and Argument Mining, Social Media, Summarization,

---

**Revised Version Manuscript Received on 22 February, 2019**

**Krishna Prakash Kalyanathaya,** M.Phil, Research Scholar, Department of Computer Science, VELS Institute of Science, Technology & Advanced Studies, Chennai, Tamilnadu, India. (e-mail:krishna.prakash.kk@ gmail.com)

**D. Akila,** Associate Professor, Department of Information Technology, School of Computing Sciences, VELS Institute of Science, Technology & Advanced Studies, Chennai, Tamilnadu, India. (e-mail:akiindia@yahoo.com)

**P. Rajesh,** Assistant Professor, Department of Information Technology, School of Computing Sciences, VELS Institute of Science, Technology & Advanced Studies, Chennai, Tamilnadu, India. (e-mail:itsrajesh91@gmail.com)

Tagging, Chunking, Syntax and Parsing, Textual Inference and Other Areas of Semantics, Vision, Robotics, Multimodal, Grounding and Speech, Word-level Semantics [3][4]. The conference accepted 258 long papers and 156 short papers on the above areas.

## IV. INDUSTRIAL APPLICATIONS OF NLP

NLP aims to dominate human-to-machine interaction to the point where talking to a machine is as easy as talking to a human. NLP still continues to harness unstructured data and make it meaningful to a machine. IDC recently forecasted that the quantity of analyzed data by cognitive systems will grow by a factor of 100 to 1.4 ZB by 2025 impacting thousands of industries and companies around the globe [2]. Robotics, health care, financial services, connected auto and smart homes are some of the sectors that will continue to be advanced by NLP.

One of initial applications of NLP in early years of 2000 was on machine translation to work as translator from one human language to another. But it rapidly found its acceptance in customer service industry. Most popular application of NLP in customer service is called as "Chatbots" or Virtual assistant.

Industrial applications of NLP can be broadly classified into 3 categories:Conversational systems, Text Analytics, Machine translation

### A. Conversational Systems

Conversational system allows us to make the conversation with the automated system in a natural language via voice or text interface. They help to automate the complex workflows in an organization with 24X7 support to its users.

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

### B. Text Analytics

Text Analytics also called as Text mining which aims to extract meaningful content from text, either in documents, emails or short-form communications such as tweets and SMS texts [2].

Most common use cases of text analytics on social media analytics.

### C. Machine Translation

Machine translation is the task of automatically translating one natural language into another, preserving the meaning of the input text [4].

Most popular application of machine translation is Google translator. Other machine translation software's are also used in speech translation and teaching.

Now, we will look at some industrial applications in following domain areas: Healthcare, Automotive, Finance, Manufacturing, Retail, Education and customer service.

### D. Healthcare

Hospitals are deploying Virtual Assistants developed with combination of Natural Language Processing, Computer Vision and Machine learning which will automatically create and retrieve patient history by interacting with the Patients. Virtual Assistant handle routine tasks such as scheduling appointments and registration of patients.

### E. Automotive

Self driving cars are one of 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 industry.

### F. Finance

NLP based solutions are developed in applications such as credit scoring, sentiment analysis and document search. Credit scoring application help the banks and financial institutions assess an individual's creditworthiness and provide credit score using NLP and machine learning. In sentiment analysis applications automate the tasks of text mining for real time information on market from news site and social media and then perform document classification and named entity recognition to filter out most relevant information to the investors needs.In document search applications, the banks or financial institutions uses chatbot interface that enables its customers to search for information and get answers to basic transactional queries.

### G. Manufacturing

Robotics and Process Automation are two very promising applications areas that will 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. The industries have already begun to use NLP based solution to automate their process workflow for improving efficiency of operations.

### H. Retail

Using Natural Language, Computer Vision and Machine learning technologies, a Virtual Assistant placed in front of the Retail shop can identify and know what customer wants and provide them quick information and promotional offers. The specialized virtual assistant also understand the shoppers comments and complaints regarding the products and provide a quick direction for resolution. Many e-commerce platforms are currently using conversational agents for customer support.

### I. Education

A combination of Natural Language Processing and Computer vision enable a platform to deliver virtual classroom for students. We have already seen the digital assistants helping students solving the problems from

expert knowledge available in digital libraries.

The top vendors delivering NLP solutions are IBM, Google, Amazon, Cerner, Nuance Communications and Microsoft.

## V.    DEVELOPMENT FRAMEWORKS AND TOOLS FOR NLP

The development frameworks and tools will help to build industrial applications discussed in the previous section. There has been numerous development tools available today due to significant interest shown by open source communities around the world. These frameworks and tools provide built in libraries and also customizable to adapt specific needs of the industry.
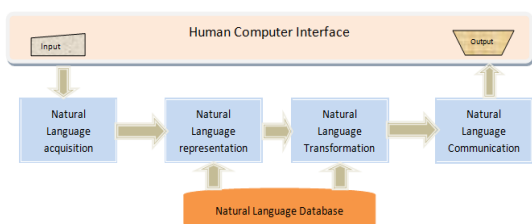


**Fig 1: Block representation of stages in the development of NLP tools**

Fig 1 shows the block representation of various stages in NLP application development. The Natural Language acquisition block built with speech processing, Computer vision or any data acquisition tools to inject Natural language text into the system.

**Table 1: Natural Language Processing Tools**

| Tool | Developer/ Supported by | Native Language Support | Usage | Application areas |
|------|------------------------|-------------------------|-------|-------------------|
| CoreNLP | Stanford Group | Java | Open source GPL[5] | Common NLP tasks and Analytics |
| NLTK[2] | Steven Bird and Edward Loper | Python | Apache License | Common NLP tasks[1] 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[5] | 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[3] | 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[5] | 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[5] | Supports common NLP tasks[1] |
| Dynet[6] | 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 | FracoisChollet | Python | MIT License | Supports easy prototyping, various deep neural networks |
| Lingpipe | Alias-iInc | 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 |

Natural Language representation block uses structured, tree or graph models to represent the Natural Language understanding. Natural Language database is a repository of Natural Language data like MNIST or similar databases which are then used by machine learning algorithms to perform other NLP tasks.

This database is accessed by representation and transformation blocks to perform their tasks. Natural

Language transformation will consists of suite of various learning, extraction algorithms to extract meaningful actions from the NLP tasks. Natural Language communication is a presentation of the desired actions as an outcome of the NLP tasks. The generated outcome may be either in Natural Language or in the form of computer action such as robot arm movement.

The following table gives an overview of the popular tools currently available in the market.

## Abbreviations

[1]**Common NLP tasks**: Lexical analysis, Parsing, semantic analysis, POS tags, Chunking, Name Entity Recognition (NER), Semantic Role Labeling (SRL)and Syntactic parsing.
[2]**NLTK**: Natural Language Toolkit
[3]**UIMA**: Unstructured Information Management Architecture
[4]**GATE**: General Architecture for Text Processing
[5]**GPL**: General Public License
[6]**Dynet**: Dynamic Neural Network
[7]**MALLET**: Machine learning for Language Toolkit
[8]**LUIS**: Language Understanding Intelligent Service

## REFERENCES

1. Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita. 2018. A Survey of the Usages of Deep Learning inNatural Language Processing. 1, 1 (July 2018), 35 pages.
2. ROBERT DALE. "The commercial NLP Landscape in 2017", Article in Natural LanguageEngineering, July2017
3. ACL 2018: 56th Annual Meeting of Association for Computational Linguistics https://acl2018.org
4. Predictive Analytics Today: www.predictiveanalyticstoday.com[accessed in Dec 2018]
5. Ali Shatnawi, Ghadeer Al-Bdour, Raffi Al-Qurran and Mahmoud Al-Ayyoub 2018. A Comparative Study of Open Source Deep Learning Frameworks. 2018 9th International Conference on Information and Communication Systems (ICICS)
6. Intelligent automation: Making cognitive real Knowledge Series I Chapter 2. 2018, EY report.
7. Jacques Bughin, Eric Hazan, SreeRamaswamy, Michael Chui , TeraAllas, Peter Dahlström, Nicolaus Henke, Monica Trench, 2017. MGI ARTIFICIAL INTELLIGENCE THE NEXT DIGITAL FRONTIER? McKinsey & Company McKinsey & Company report  July 2017
8. Svetlana Sicular, Kenneth Brant 2018, Hype Cycle for ArtificialIntelligence, 2018 Gartner report July 2018.
9. McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.
10. Quarteroni, Silvia. (2018). Natural Language Processing for Industry: ELCA's experience. Informatik-Spektrum. 41. 10.1007/s00287-018-1094-1.
11. Young, Tom &Hazarika, Devamanyu&Poria, 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.
12. 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.