

A Study on Historical Perspective of Natural Language Processing

AARTI¹, Dr. NEETU SHARMA²

¹DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, GITAM, Kablana

malhotra.malhotra5@gmail.com

²DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, GITAM, Kablana

neetush75@gmail.com

Abstract

A language is a system, a set of rules or set of symbols. Symbols are combined and used for conveying information or broadcasting the information. The Language is the way of communicating your words. The Language helps in understanding the world in better way. NLP Stands for Natural Language Processing. Natural languages are those languages that are spoken by the persons in daily routine. Natural language processing is the field of computer science (CS), artificial intelligence (AI), and linguistics predominantly focuses on intercommunications between computers and the human languages (natural languages). NLP is focused on the area of human computer intercommunication. The exigency for the natural language processing was felt because there is an immense storage of information stored in natural language that could be accessible through the computers. Information is generated in the form of books, business and government reports news, scientific papers, many of which are available online. A system requiring a great understanding of information must be able to process natural language to fetch much of the information available or present on the computers. Natural Language Processing is provocative and tough field in which we have to develop and analyze the representation and the reasoning theories. All of the problems of Artificial Intelligence arise in this domain; solving "the natural language problem" is as complex as solving "the AI problem" because any field can be expressed or can be represented in natural language.

Keywords: *Naturallanguageprocessing (NLP), Syntactic, Semantic, Pragmatic, Lexical, Linguistics, Morphological, Discourse Integration, Machine Learning.*

1. Introduction

Natural language processing (NLP) can be defined as the automatic (or semi-automatic) processing of human language. NLP is sometimes contrasted with computational linguistics. NLP is essentially multidisciplinary: it is closely related to linguistics. It also has links to research in cognitive science, psychology, philosophy and math (especially logic). Within CS, it relates to formal language theory, compiler techniques, theorem proving, machine learning and human-computer interaction. Of course it is also related to AI, though nowadays it's not generally thought of as part of AI.

Natural Language Processing (NLP) is the field of research and applications that analyze how computers can be used to understand and manipulate the natural language text or speech to do advantageous things. Natural Language

Processing is the artificial intelligent concept where the machines understand Natural Languages like English, Hindi, Korean, French and Telugu etc. The support of NLP lie in a number of disciplines like mathematics, artificial intelligence and robotics, psychology, electronic engineering, computer science, linguistics etc. NLP need a tool that will take the Database queries in the form of natural language and then processes it and gives the desired result. It includes many small components like Language Analyzer, Query Builder and Viewer. Firstly, the system will parse the query in the natural language and then finds the major parts from the string. Then it will look for the table name and then it parses the string using the WHERE clause and ORDER BY clause. Once the process of parsing completes it will construct the query string based on the data available. Then the generated

Structured Query Language (SQL) query is posted to the database to retrieve or fetch the results.

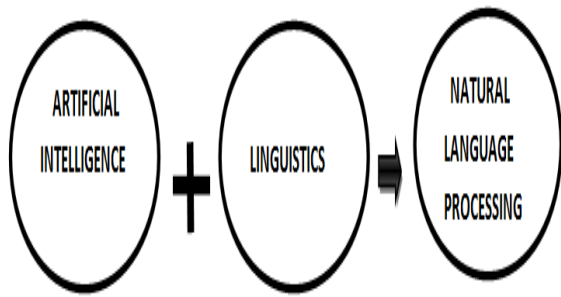


Figure 1.1: Natural Language Processing

2. LITERATURE REVIEW

The research work in the NLP i.e. Natural Language Processing has been increasing day by day. The NLP is computerized method for analyzing the text and being very active area in research and development. The literature distinguishes the main application of NLP and the methods to describe it.

1) NLP for Speech Synthesis: This approach is based on text to speech (TTS) conversation in which the text data is the firstly entered into the system. It uses high level modules and also it uses the sentence segmentation in which the punctuation marks are used with a decision tree.

2) NLP for Speech Recognition: The speech recognition system makes use of NLP techniques based on grammars. It uses the CFG i.e. Context Free Grammar for representing the syntax of the language through the automatic summarization including indexing.

3. NATURAL LANGUAGE PROCESSING

Understanding

Generation

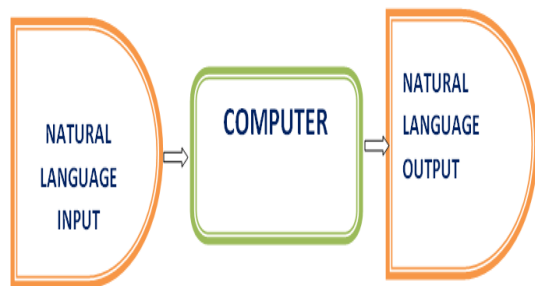


Figure 3.1: Natural Language Processing

3.1 NATURAL LANGUAGE UNDERSTANDING

NLU is the task is to understand and the input is a natural language.

3.2 NATURAL LANGUAGE GENERATION

NLG issue generation of the natural language processing. It is also called as Text Generation

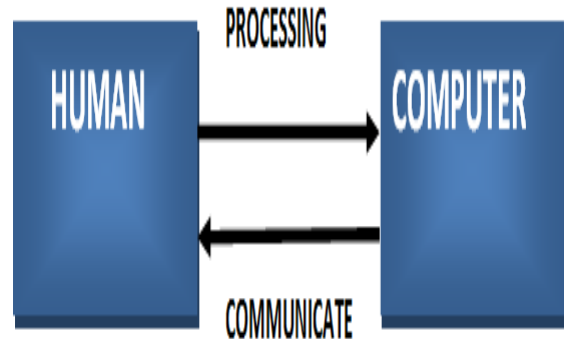


Figure 3.2: Human and Computer Interaction

3.3 HISTORY OF NLP

The history of NLP was started in the year 1950s. In early 1950, Alan Turing published an article named as "Machine and Intelligence" which was very much advertised and is now called the Turing test as a subfield of intelligence. Most useful Natural language systems were developed in the 1960s and were called SHRDLU.

4. FEW LINGUISTIC TERMINOLOGY AND LP:

Linguistics is the science of language. Linguistic includes following study:

1. **Phonology:** Sounds refers to phonology.
2. **Morphology:** The method of Word Formation is known as Morphology. For example, *unexpectedly* can be thought of as composed of a pre_x *un-*, a stem *expected*, and an af_x *-ly*.
3. **Syntax:** Meaning refers to semantics The procedure for building of the Sentence is called Syntax. For example, it is part of English syntax that a determiner such as *the* will come before a noun.
4. **Semantics.** It is the construction of meaning based on syntax (the meaning of individual words).
5. **Pragmatics:** Understanding or Meaning from the context refers to pragmatics.

4.1 PHASES OF LINGUISTIC

There are two phases of Linguistic:

- Higher Level Corresponds To Speech Recognition
- Lower Level Corresponds To Natural Language Processing

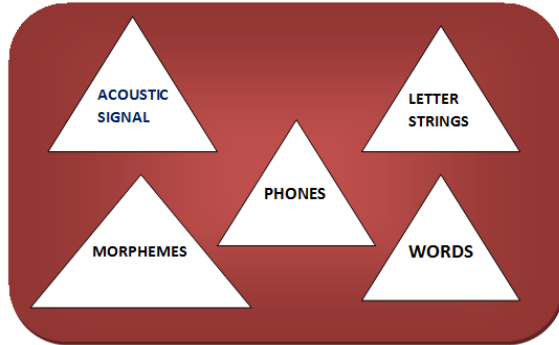


Figure 4.1: Speech Recognition

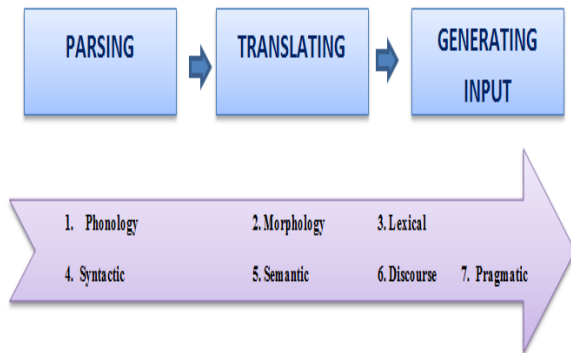


Figure 4.2: Natural Language Processing

5. STEPS OF NLP

There are five phases which are involved in NLP:

1. Morphological and Lexical Analysis
2. Syntactic Analysis
3. Semantic Analysis
4. Discourse Integration
5. Pragmatic Analysis

5.1 Morphological and Lexical Analysis

Morphology represents analyzing, identifying and also the description of structure of the words. The lexicon of the language is its vocabulary that comprises of its words and also expressions.

5.1.1 Lexical Analysis

It includes partition of a text into words, paragraphs, and sentences.

5.2 Syntactic Analysis

In this step the words are analyzed in a sentence to show the grammatical architecture of the sentence. The words are converted into the structure that depicts how the words are related to each other. Eg. "the girl the go to the college". This sentence would definitely be refused or rejected by the English syntactic analyser.

5.3 Semantic Analysis

This step included the abstraction of the dictionary meaning or the exact meaning from the context. The architecture which is formed by the syntactic analyser is assigned the proper meaning. There is linking between the syntactic architecture and the objects in task domain. Eg. "Colorless red thought". This would be cancelled or rejected by the analyser as the colorless thought does not make any sense with each other.

5.4 Discourse Integration

The meaning of separate sentence depends upon the sentences that comes before it and also invokes the meaning of the sentences that follow it .Eg the word "it" in the sentence "He wanted it" depends upon the previous discourse context.

5.5 Pragmatic Analysis

It means deriving the use of the language in some situations where the aspects of language require real world knowledge and the main target is on what was said and that is again interpreted so that one may know what it actually means. Eg "Open the Door?" This sentence should have been interpreted as a request rather than an order.

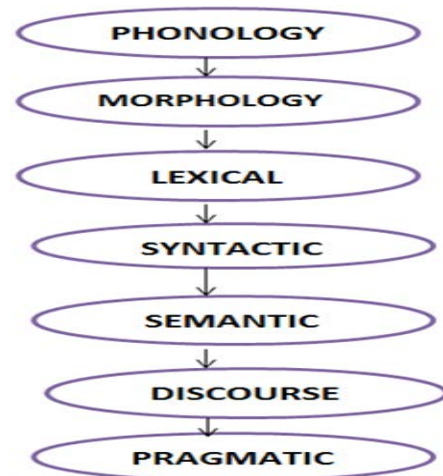


Figure 5: 1LEVELS OF NLP

6. NATURAL LANGUAGE PROCESSING USING MACHINE LEARNING

There are so many NLP algorithms which are based on machine learning, mainly the statistical machine learning (SML). Implementations of language-processing tasks involve the direct hand coding of large sets of rules. Many different kinds of classes ML algorithms have been applied to NLP tasks. Many Systems that are based on machine-learning algorithms have many advantages over hand-produced rules.

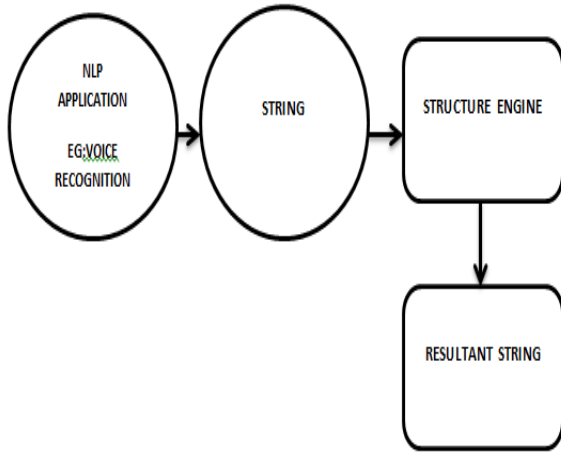


Figure 6.1: NLP Procedure

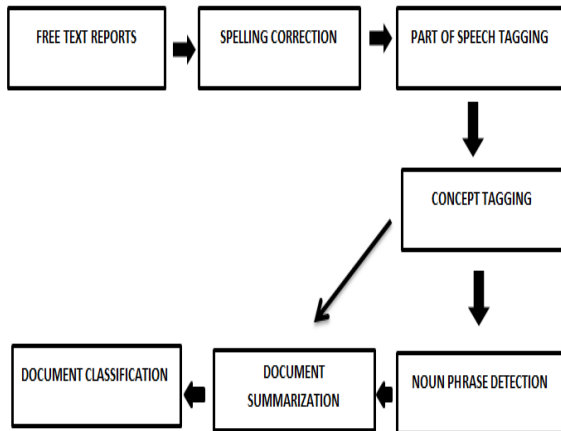


Figure 6.2: DFD of NLP

7. APPLICATIONS OF NLP

There are various applications of NLP. Some of the most popular applications of NLP are as under:-

- Screen Readers for the Blind and partially Sighted Users
- Alternative and Augmentative Interaction (the systems to aid people who have complexity in communicating because of the disability)

- Spelling and Grammar Checking
- Optical Character Recognition (OCR)
- Machine Aided Translation (the systems which help a human translator, For instance, by storing the translations of sentences and providing the online dictionaries integrated with the word processors)

Other useful applications of NLP are as follows:

- Information Retrieval
- Information Extraction
- Machine Translation
- Spelling correction, grammar checking
- Question Answering and Exam Marking
- Text Segmentation
- Document classification (Routing and Filtering)
- Document Clustering
- Summarization
- Report Generation(containing ManyLanguages)
- NLI(Natural Language Interfaces) to DB
- Better search engines
- Email Understanding
- Speech Recognition (ASR)

i. Machine Translation

Machine translation helps us to conquer the language barriers that we often found by the translation of technical guide manuals, catalogs at a significantly reduced cost.

ii. Automatic summarization

It produces an understandable summary of a set of text. It is used to provide the detailed information of text.

iii. Sentiment analysis

It is useful for identifying the trends of public opinion in the social media, for the purpose of marketing.

iv. Text classification

With the help of **Text classification** it become possible to assign the predefined categories to a document and organize it.

v. Question Answering

The need for NLP is increasing day by day as speech-understanding technology and voice-input applications are improving. QA is becoming more and more popular as we can see the functioning of

OK Google, Siri, virtual assistants etc. Question answering application is a system capable of answering a human's request.

vi. Fighting Spam

Spam filters have become important as the problem of unwanted email is increasing day by day. But almost everyone that uses email regularly has experienced over unwanted emails. Important emails that have been accidentally caught in the filter.

vii. Optical character recognition (OCR)

In OCR an image represents the printed text.

viii. Word sense disambiguation

As we know that many words have multiple meaning. We have to select the meaning which makes the most appropriate sense in the context.

8. FUTURESCOPE OF NLP

Human readable NLP is an Artificial Intelligence problem. It is equivalent to solving the AI problem and make the computers as brilliant as human beings so that they can solve the problems like humans and perform various activities that humans can't do and making it more effective than humans. NLP's future is closely related to the growth of AI. As NLU or readability improves, the computers will be able to learn from the information that are present online and apply whatever they learned in the real world. Combined with NLG (Natural Language Generation), computers will become more capable of receiving and giving the beneficial information.

As we know that there are so many applications that we can dream with Natural Language Processing Techniques. How about robots that understand and follow all the instructions that are given by human voice, driving the car by talking to the car like in some movies. All of these ideas will be real one day. Just Imagine we have the computer system that follow simple human general instructions and do whatever we want it to do. How effective and efficient will it be? But now leave all of these points to the Future.

9. CONCLUSION

The capability to use the natural language for query specification and fetching bags over the keyword, keyphrase approaches. The trust that the restricted use of natural language in multimedia data

abstraction is a less inconvenient task than full natural language fact abstraction and feels that we have a system that can be checked and built not only for abstracting images but also for the audio, video, text, data etc.

10. ACKNOWLEDGMENT

The author would like to thank **Dr. Neetu Sharma**, Head of Department (Computer Science), Ganga Institute of Technology and Management [GITAM], Haryana for his continuous support, help and suggestions to improve the eminence of this paper.

Author(s): Aarti, Graduated in Computer Science and Engineering from MDU University, Rohtak, India (2014), Student of M.Tech. 2 year, Computer Science and Engineering, Ganga Institute of Management of Technology [GITAM], Haryana.

REFERENCES

1. Winograd, Terry (1971). Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. <http://hci.stanford.edu/winograd/shrdlu/>
2. Liddy, E. D, "Natural language processing", In Encyclopedia of Library and Information Science, 2nd Ed, NY, Marcel Decker, Inc., 2001.
3. Hutchins, J. (2005). "The history of machine translation in a nutshell".
4. Goldberg, Yoav (2016). <https://www.jair.org/media/4992/live-4992-9623-jair.pdf> A Primer on Neural Network Models for Natural Language Processing. Journal of Artificial Intelligence Research 57 (2016) 345–420
5. Yucong Duan, Christophe Cruz (2011), *Formalizing Semantic of Natural Language through Conceptualization from Existence*. International Journal of Innovation, Management and Technology (2011) 2 (1), pp. 37-42.
6. <http://www.slideshare.net/jhonrehmat/natural-language-processing>.
7. www.authorstream.com/.../aSGuest26001-239-367-natural-language-proc
8. NaturalLanguageProcessing, www.myreaders.info/html/artificial_intelligence.html.
9. NaturalLanguageProcessing-Computer science and engineering www.cse.unt.edu/~rada/CSCE5290/Lectures/Intro.ppt
10. NLP, www9.georgetown.edu/faculty/mad87/05/nlp/slides/intro.ppt

11. www9.georgetown.edu/faculty/mad87/05/nlp/slides/intro.ppt

12. Ian Goodfellow, Yoshua Bengio and Aaron Courville. [http://www.deeplearningbook.org/Deep Learning](http://www.deeplearningbook.org/Deep%20Learning)]. MIT Press.