

Natural Language Processing

Abhimanyu Chopra, Abhinav Prashar, Chandresh Sain

CSE Department, Dronacharya College of Engineering ,Gurgaon ,India
CSE Department, Dronacharya College of Engineering ,Gurgaon ,India
CSE Department, Dronacharya College of Engineering ,Gurgaon ,India
Email: abhinav.prashar@gmail.com, abhimanyuchopra@yahoo.com, chandresh.saini@gmail.com

ABSTRACT: Language is way of communicating your words Language helps in understanding the world ,we get a better insight of the world. Language helps speakers to be as vague or as precise as they like. NLP Stands for natural language processing. . Natural languages are those languages that are spoken by the people.Natural language processing girdles everything a computer needs to understand natural language and also generates natural language.Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics mainly focuses on the interactions between computers and human languages or natural languages. NLP is focussed on the area of human computer interaction. The need for natural language processing was also felt because there is a wide storage of information recorded or stored in natural language that could be accessible via computers. Information is constantly generated in the form of books, news, business and government reports, and scientific papers, many of which are available online or even in some reports. A system requiring a great deal of information must be able to process natural language to retrieve much of the information available on computers. Natural language processing is an interesting and difficult field in which we have to develop and evaluate or analyse representation and reasoning theories. All of the problems of AI arise in this domain; solving "the natural language problem" is as difficult as solving "the AI problem" because any field can be expressed or can be depicted in natural language.

Keywords: Naturallanguageprocessing (NLP), Syntactic, Symantic, Pragmatic, DiscourseIntegration, Morphological, Lexical, Linguistics, Generation, Machine Learning.

1. INTRODUCTION

Natural languages are those languages that are spoken by the people.Natural language processing girdles everything a computer needs to understand natural language and also generates natural language. **Natural Language Processing** is a subfield of Artificial Intelligence and linguistic,devoted to make computers understand the statements or words written in human languages.A Natural language also known as ordinary language that is spoken or written by people(humans) for general purpose communication. Natural language came into existence because when user wishes to communicate with the computer we cant force the users to learn machine specific langauge so this basically caters to managers or childrens who do not have enough time to learn new specific langauges or get skilled in them.Languages can be any like Hindi,French,english,chinese etc.A language is a system, a set of rules or set of symbols.

1. Symbols are combined and used for conveying information or broadcasting the information.
2. Rules tyrannize handling of symbols.NLP Besets anything a computer or machine needs to understand typed or spoken (natural language).

2. NATURALLANGUAGE PROCESSING

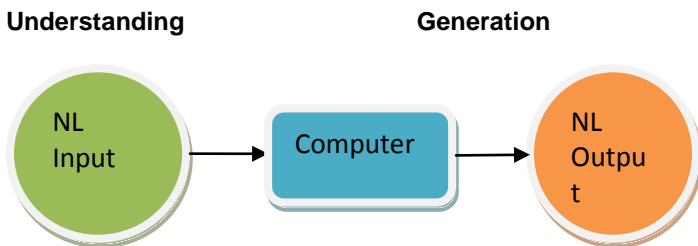


Fig 1.1 Natural Language Processing

2.1 NATURAL LANGUAGE UNDERSTANDING:

Its task is to understand and reason where input is a natural language.

2.2 NATURAL LANGUAGE GENERATION:

It is a sub generation of natural language processing.It is also referred to as text generation

3. HISTORY OF NLP

The history of NLP generally starts in the year 1950s. In 1950, Alan Turing published an article titled "Machine and Intelligence" which advertised what is now called the Turing test as a subfield of intelligence. Some beneficial and successful Natural language systems were developed in the 1960s were SHRDLU, a natural language system working in restricted "blocks worlds" withrestrictedvocabularies was written between 1964 to 1966.

4. LINGUISTICSAND LANGUAGE PROCESSING:

Linguistics is the science of language.Its study includes:

1. Sounds which refers to phonology
2. Wordformation refers to morphology
3. Sentence structure refers to syntax
4. Meaning refers to semantics
5. Understanding refers to pragmatics

5. PHASES OF LINGUISTIC ANALYSIS

- I. Higher level corresponds to **SPEECH RECOGNITION**
- II. Lower level corresponds to **NATURALLANGUAGE PROCESSING**

6. LEVELS ARE

Fig 6.1 Speech Recognition

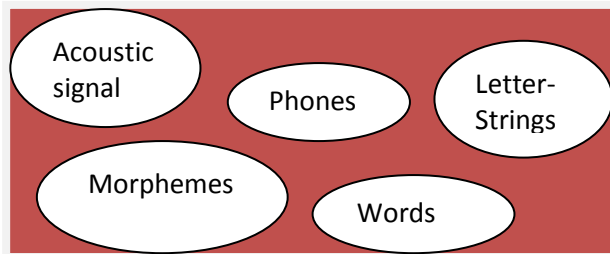
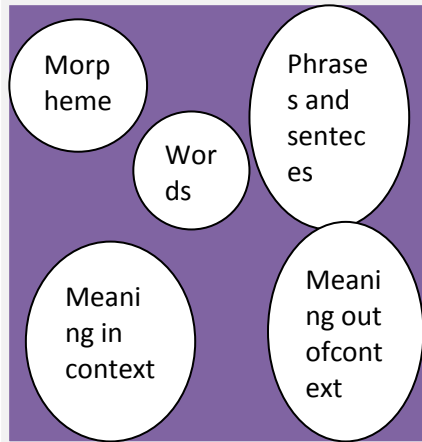


Fig 6.2 Natural language Processing



7. STEPS OF NATURAL LANGUAGE PROCESSING

There are 5 phases involved in natural language processing

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

7.1 Morphological and Lexical Analysis

The lexicon of a language is its vocabulary that includes its words and expressions. Morphology depicts analysing, identifying and description of structure of words.

7.1.1 Lexical Analysis

It involves dividing a text into paragraphs, words and the sentences

7.2 Syntactic Analysis

This involves analysis of the words in a sentence to depict the grammatical structure of the sentence. The words are transformed into structure that shows how the words are related to each other. Eg. "the girl the go to the school". This would definitely be rejected by the English syntactic analyser.

7.3 Semantic Analysis

This abstracts the dictionary meaning or the exact meaning from context. The structures which are created by the syntactic analyser are assigned meaning. There is a mapping between the syntactic structures and the objects in

task domain. Eg. "colorless blue idea". This would be rejected by the analyser as colorless blue do not make any sense together.

7.4 Discourse Integration

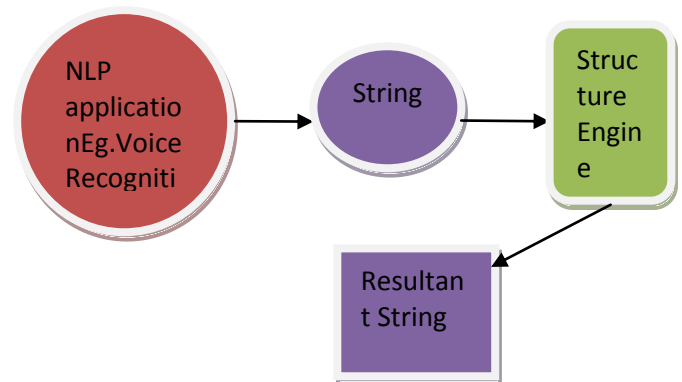
The meaning of any single sentence depends upon the sentences that precedes it and also invokes the meaning of the sentences that follow it. Eg the word "it" in the sentence "she wanted it" depends upon the prior discourse context.

7.5 Pragmatic Analysis

It means abstracting or deriving the purposeful use of the language in situations importantly those aspects of language which require world knowledge the main focus is on what was said is reinterpreted on what it actually means. Eg "closethe window?" should have been interpreted as a request rather than an order.

8. NATURAL LANGUAGE PROCESSING USING MACHINE LEARNING

Fig.8.1 NLP Procedure



Many NLP algorithms are based on machine learning, mainly statistical machine learning. The working of machine learning is different from attempts at language processing. Prior implementations of language-processing tasks typically involved the direct hand coding of large sets of rules. The machine-learning calls instead for using general learning algorithm to automatically learn such rules through the analysis of large ocean of typical real-world examples. Many different classes of machine learning algorithms have been applied to NLP tasks. These algorithms take as input a large set of characteristics that are generated from the input source. Some earlier produced algorithms such as decision trees, produced systems of hard if-then rules similar to the systems of hand-written rules that were then common. Increasingly, however, research has focused on statistical models, which make soft, probabilistic decisions based on attaching real-valued weights to each input. Various Systems based on machine-learning algorithms have many advantages over hand-produced rules:

8.1 The learning procedures used during machine learning automatically focuses on the most common cases, whereas when we write rules by hand it is often not correct at all where the effort should be deviated.

8.2 Automatic learning procedures can make use of statistical inference algorithms to produce models that are robust (means strength) to unfamiliar input e.g. containing words or structures that have not been seen before

8.3 Systems based on automatically learning the rules can be made more accurate simply by supplying more input data or source to it. However, systems based on hand-written rules can only be made more accurate by increasing the complexity of the rules, which is a much more difficult task.

9. Major tasks in NLP

This lists some of the different researches done in NLP.

9.1 Automatic summarization

It produces an understandable summary of a set of text. It is used to provide summaries or detailed information of text of a known type.

9.2 Coreference resolution

It refers to a sentence or larger set of text that determines which words refer to the same objects, example of this is concerned with matching up pronouns with the nouns or names that they link to.

9.3 Discourse analysis

The task is identifying the discourse structure of connected text, i.e. the nature of the discourse relationships between sentences e.g. elaboration, explanation, contrast. Another possible task is recognizing and classifying the speech acts in a large set of text e.g. yes and no questions, content question, statements, assertion etc.

9.4 Machine translation

Automatically translates text from one human language to another.

9.4.1 Morphological segmentation

Separate words into individual morphemes and identify the class of the morphemes. The difficulty of this task depends greatly on the complexity of the morphology i.e. the structure of words of the language being considered.

9.5 Named entity recognition (NER)

It describes a stream of text, determine which items in the text relates to proper names, such as people or places, and what the type of each such name or place we are referring to is.

9.5.1 Natural language understanding

It Converts large set of text into more formal representations such as first-order logic structures that are easier for computer programs to manipulate notations of natural languages concepts.

10. Other tasks includes

- a. **Optical character recognition (OCR):** Given an image representing printed text, helps in determining the corresponding or related text.

- b. **Part-of-speech tagging:** It describes a sentence, determines the part of speech for each word.
- c. **Parsing:** It refers to the parse tree (grammatical analysis or evaluation) of a given sentence
- d. **Question answering:** It answers a given human language question and determines its answer.

11. Statistical NLP:

Statistical natural-language processing using random, probabilistic and statistical methods to settle some of the difficulties especially the ones which arise because longer sentences are highly equivocal when processed with realistic grammars. NLP comprises all quantitative approaches to.

12. FUTURE OF NLP

Human level or human readable natural language processing is an AI-complete problem. It is equivalent to solving the central artificial intelligence problem and making computers as intelligent as people so that they can solve problems like humans and think like humans as well as perform activities that humans can't perform and making it more efficient than humans. NLP's future is closely linked to the growth of Artificial intelligence. As natural language understanding or readability improves, computers or machines or devices will be able to learn from the information online and apply what they learned in the real world. Combined with natural language generation, computers will become more and more capable of receiving and giving useful and resourceful information or data.

13. CONCLUSION

The strength or the capability to use natural language for query specification and retrieval bags over the keyword, keyphrase approaches. The believe that the restricted use of natural language in captions for multimedia data abstraction is a less cumbersome task than full natural language fact abstraction, and feel that we have a system that can be judged and built upon not only for abstracting images but also the form so multimedia (audio, video, text, data etc) data or input sources as well.

14. REFERENCES

- [1]. Artificial Neural Network, http://en.wikipedia.org/wiki/Artificial_neural_network
- [2]. Neural network, www.en.wikipedia.org/wiki/Neural_network
- [3]. Guoqiang Zhang, B. Eddy Patuwo, Michael Y. Hu* (July 1997), Forecasting with the Artificial Neural Networks, Kent State University, Kent, USA.
- [4]. http://www.slideshare.net/jhonrehmat/natural_language_processing.
- [5]. Artificial Intelligence by Elaine Rich and Kevin Knight, 1991, McGraw-Hill.

- [6]. Artificial Intelligence-A Modern Approach by Russel and Norvig, 1995, Prentice Hall
- [7]. Natural Language Processing an Introduction www.ncbi.nlm.nih.gov v.18(5); Sep-Oct 2011.
- [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]. NLP, <https://www.coursera.org/course/nlp>
- [12]. NLP, research.microsoft.com/en-us/groups/nlp/
- [13]. NLP, see.stanford.edu/see/courseinfo.aspx?coll=63480b48-8819-4efd.
- [14]. NLP, nlp.stanford.edu/
- [15]. NLP, nptel.iitm.ac.in/courses/106101007/
- [16]. www.autonomy.com/content/Functionality/idol...nlp/index.en.html
- [17]. web.cs.wpi.edu/~cs534/f06/LectureNotes/Slides/natural_lang_processing.ppt
- [18]. www9.georgetown.edu/faculty/mad87/05/nlp/slides/intro.ppt
- [19]. www.pages.drexel.edu/~rw37/courses/info629/slides629/629ppt7.ppt
- [20]. www.city.academic.gr/s3ctit03/presentations/nenad.ppt
- [21]. www.site.uottawa.ca/~nat/Courses/csi4106_2008/..../AI05_10b_NLP.ppt
- [22]. www.pes.edu/mcnc/AI/data/.../Rich.../Chapter15_NaturalLanguage.ppt
- [23]. www.authorstream.com/.../aSGuest26001-239367-natural-language-proc...