

# Ambiguity Issues in Different Applications of Natural Language Processing

Rakesh Kumar

*Department of Computer Science, University College Miranpur, Patiala, India.*

*rakesh1404@gmail.com*

## Abstract

Natural language processing (NLP) can be defined as the ability of computer systems to better comprehend the human speaking terms and written text data. Natural Language processing is utilized by several domains to draw solution-based conclusions with help of text analytics where computer systems performs interactions using textual phrases for automating and structuring text to retrieve and analyze. The Natural Language Processing (NLP) community has experience of a large number of methods and techniques that can be applicable to the problem of delivering effective results in different applications of natural language processing. This research paper reports the different ambiguities issues present in different kind of knowledge while handling processing of natural language. This paper also discusses the different widely used applications of natural language processing along with different kind of knowledge involved in these kind of NLP applications. Several issues and challenges are addressed and their potential solutions are discussed.

**Keywords:** Natural Language Processing, Text Entailment, Information Retrieval, Ambiguity, Parse Trees.

## I. Introduction

Natural Language processing is a subfield of artificial intelligence that requires the expert skills in multiple inter-disciplinary areas, such as linguistics, statistics, expert systems etc. Natural language processing (NLP) can be defined as the capability of computing machines to better comprehend the human speaking terms and written text data. Natural Language processing utilized by several domains to draw solution based conclusions with help of text analytics where computer performs interactions using textual phrases for automating and structuring text to retrieve and analyze [3]. Natural language processing is exhaustively utilized in recent technologies to render help for privacy of spam emails, personal voice assistants and application of language translation [4]. Natural language processing is very much in demand and exciting research area in linguistic informatics in current times.



The term Natural language processing is often utilized to include a collection of methods and techniques involved in the processing of text of unstructured type, although the techniques and methods themselves covering a large range in their application of language knowledge. However, some techniques and methods utilize reduced semantic knowledge, and are based on the appearance of words in textual data [5]. For these techniques and methods, the only semantic knowledge needed is the knowledge of what makes a word; these techniques and methods often count on the bag-of-words approach or a keyword approach. One example of a method is a search engine, which utilizes only words, retrieving all the records data containing the appearance of a combination of words in a cluster, although these words may not totally identical to each other in the documents that are extracted [6].

The rest of the paper is organized as follows. In the subsequent Section II, different applications of Natural Language Processing are discussed. Section III presents different kinds of knowledge involved in NLP applications. In Section IV Methodology adopted in NLP process is outlined. Section V, VI and VII provides the detailed information about Syntactic, Semantic and Pragmatic Knowledge. The Ambiguity Issues and Challenges present in different kinds of knowledge in NLP applications are discussed in section VIII. Finally, in Section IX, the conclusion of research findings on ambiguity issues in NLP applications is presented along with its future directions.

## II. Applications of Natural Language Processing

Natural language processing (NLP) has a very wider range of potential applications [3]. The following are the some of applications of NLP:

### 1. Machine Translation

Machine Translation is the process of translating textual form of one natural language to another natural language. It is extensively utilized to gain more knowledge about the human beings such as finding out their semantics, factoids and real life events of the world.

### 2. Named Entity Recognition

Named Entity Recognition is the process in which specified match of the textual form is mapped to the designated names and is conveyed the type of name and the use of capitalization provide assistance in identifying out the named entities in natural languages like English. Although, the use of capital words is not carried out in some languages like German language, Spanish language and French language but names in theses languages are served as the adjectives.

### 3. Lexical Semantics

Lexical Semantic is the process of studying meaning of words. It conveys about either internal



structure of the words or structure that could occur in the words or particular context.

#### **4. Natural Language Understanding**

Natural language understanding is the process of converting the blocks of data into human comprehensible form as First order logic structures which computers are capable of manipulating easily and very fast.

#### **5. Optical Character Recognition**

Optical character recognition is the process of recognizing text format from an image and determining its actual contents.

#### **6. Question Answering**

Question Answering Systems provides a list of possible applicable summary reports in response to a question asked by user, question answering [2] provides the user with either just the answer in the form of text itself or a pathway leading to answer. Provided a question and a collection of documents, a QA system attempts to search the exact response or answer, or at slightest the exact part of text in which the answer is contained [7].

#### **7. Text Entailment Recognition**

Text Entailment Recognition is the job of determining, when two textual fragments are provided, whether the meaning of one text can be inferred (can be entailed) from another text. There could be are two text fragments available from which one could be true and the other text fragment could entail false.

#### **8. Relationship Extraction**

Relationship Extraction is the process of identifying relationship present among the named entities when blocks of texts are provided.

#### **9. Sentiment Analysis**

Sentiment Analysis is the process of identifying trends that are going to be clouded over the social media in the future. This process is extensively used to find the trends so could be prove beneficial in marketing field and also serve as a platform delivering giving online reviews about humans and other associated factors.

#### **10. Topic Recognition and Segmentation**

Topic Recognition and Segmentation is the process of segmenting a block of text into several segments and then to look out for best possible the topic for each segmented block.



## 11. Word Sense Disambiguation

As multiple words in a block of text will have similar meaning and thus one have to choose the best sensible meaning for the word. Word Sense Disambiguation is the process of identifying which sense of word is best suitable in given context. Nowadays, the focus of many researchers is oriented towards topics like natural language processing or understanding, machine translation, voice or speech recognition and grammatical and linguistic problems.

## III. Discovering Knowledge in NLP

Knowledge could be in any form like data, information, facts, descriptive knowledge, skill information etc., [1]. Knowledge can be termed as information, understanding of anything which is learned or gained through experience. The knowledge can be categorized into many levels.

### 1. Phonetic and Phonological knowledge

Phonetic and Phonological knowledge conveys the realization of words in manual way similar to sounds.

### 2. Morphological Knowledge

Morphological type of knowledge is capable of telling how the words are derived out from fundamental and meaningful units that are being considered as phonemes.

### 3. Syntactic Knowledge

Syntactical knowledge conveys the process of clubbing of letters and the words to be formed together in unique sentences.

### 4. Semantic Knowledge

Semantic Knowledge refers to the knowledge about common things. It conveys that some related words and their organization and best possible meaning extraction to form the meaningful and sensible sentences.

### 5. Pragmatic and Discourse Knowledge

Pragmatic and Discourse Knowledge refers to the knowledge in which language can be incorporated or adopted in different settings or contexts which could greatly influence interpretation of the spoken sentences.

### 6. World Knowledge

World Knowledge refers to the knowledge of includes worldly affairs about the different type of



happenings, events and even outline of the world in current scenario.

#### IV. NLP Methodology

Natural language processing (NLP) follows the same process methodology as the other computational machines do. Firstly, data in natural language is inputted, which is then processed and finally output is generated from processing the data provided in natural language [8]. The flow chart of NLP methodology is given in Figure 1.

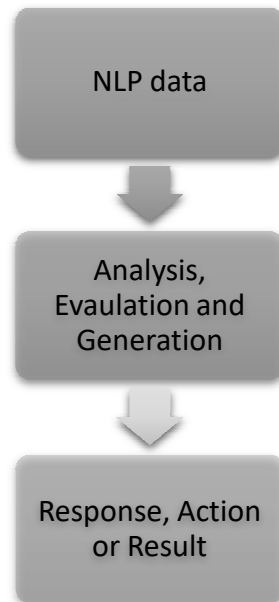


Fig 1. NLP Methodology

#### V. Semantic Knowledge

Semantic knowledge is the collection of routines that are capable of classifying content because it is easily accessible at any time and obtained in a stipulated time period [1]. Semantic information act as a representative form of the information or information about business or commercial associations. A few of the particular practices are keeping up the cluster records into a single record , transforming the data's or substance into several file formats, any keyword exploration about specified word or doing it viably in exceptional manner, including more setting based substance in creation of that substance, diminishing the rates of the generation fetched, diminish the generation costs for the record designs, diminish the manual representation of shaping right information at right time. This knowledge provides the study of meaning of linguistic utterances. This may too be utilized as to know well versed almost how the meaning of the happening is closely related to

meaning of words, expressions and after that morphemes. Their lies numerous issues in it as with information based representation of semantic information as unquestionable status, unambiguous representation and dubiousness.

## VI. Syntactic Knowledge

Syntactic knowledge is approximately portraying how the single words can be integrated into a legitimate meaningful and sensible sentences, utterances and phrases [1]. Computers are really exceptionally much speedy in their work and it's moreover the foremost capable machines which can entirety up and execute the content that's composed by human minds in a diverse way and treating them in a diverse way. The most objective of syntactic evaluation is to find whether the given information in an input frame could be a idealize sentence in a given characteristic dialect [8]. A few depiction of syntactic structure of a sentence is given as an outlines within the frame of induction tree and such constructions are basically pointing at more center on the comprehending capability of the computers. Moreover, searching and comprehending the association between the words additionally between the comparing individuals, activity and things that are to be carried out. Syntactic analysis can be primarily oriented towards creating on punctuation sentences conjointly discourse arrangement with having common dialect interface or as too a building piece in a machine translation framework. The language of Czech Republic is very rich inflectional that needs much knowledge in grammatical rules rather than any other languages. According to the research about this syntactic type of comprehension is exceptionally intense and difficult to analyze. The Natural Language Processing research facility is as of now on creating location of syntactic, syntactical analyzer. Agreeing to the test execution done the syntactical comes to the review of 92% and exactness of approximately 84% and basically seeing each and every sorts of determination tree. This information provides you the understandable vision of formal connections between the words. In 1956, popular linguist, Noam Chomsky who to begin with made the CFG structure parsing trees. Following are the most adopted methods or adopted ways in which task of parse sentences is carried out:

- **Top-Down Approach:** In it, the task to parse sentences is started with root node and then proceeded towards to the bottom of the leaves nodes.
- **Bottom-Up Approach:** In it, the task to parse sentences is commenced with words as input and endeavors to construct the words from these tree structures and is conducted by application of grammatical rules and regulations performed at a given time.
- **Depth First Approach:** In it, the task to parse sentences is performed expanding the search area



in incremental manner by one stage at a given time.

- **Repeat Parsing of Subtrees:** In it, the task to parse sentences is dealt with the results of inefficiencies present in other parse procedures.
- **Programming Approach:** In it, the task to parse sentences is utilized to believe on resolution of ambiguities.

## VII. Pragmatical Knowledge

Pragmatical knowledge is the idea of how the people are capable of communicating meanings for the words and the working procedure of delivering the sentences or articulations in an appropriate way. Pragmatical knowledge incorporates of conveyance in social manner conjointly useful information. Pragmatics is the Sub field containing of two major parts as linguistics and semiotics. This pragmatics contributes the setting substance to the characterized meaning. Pragmatics incorporates the speech act theory, talks in interactions and other approaches like philosophy, sociology, linguistics and anthropology. For instance in the case of semantics that is capable of examining meanings those are traditional or that are programmed in a provided language, and pragmatics examination the transmission procedure of meaning is dependent on structured and lingual knowledge also. Pragmatics clarifies how dialect clients are capable to cope up with each concerns and since it depends as it were on the expressions, way, time and put. Words may be converted conjunctively into appropriate sensible sentences, utterances and expressions. Computing machines are exceptionally much fast in their assignments and it's moreover the foremost effective machines which can whole up and evaluate the content that's composed by minds of the human in a distinctive pattern.

## VIII. Ambiguity Issues in NLP

Applications of Natural Language Processing faces different kinds of ambiguities issues and challenges while handling of different kind of knowledge involved in NLP applications [8]. In Recent time the research trend is tilting towards in resolving these kind of ambiguities issues and challenges. Some of the issues and challenges and their resolution are as follows:

- The ambiguities present in the morphological and syntactical knowledge could are the main contributing factor in these kind of issues.
- Semantic Knowledge can also give birth to ambiguity issues in natural language processing while the time of citing an instance as “[Making a verb or verb that may be either “devise” or “develop”]”.



- Parts-of-Speech (POS) could render a greater help in solving these kind of ambiguity issues in Natural Language Processing.
- Clarity in words can also help in solving these kind of ambiguities present in natural language.

## IX. Conclusion

In this paper, different issues and challenges related to ambiguity present in the different kind of knowledge involved in different applications of Natural Language Processing are addressed and some suggestion are provided to resolve these ambiguity issues. Discovery of different kind of knowledge in natural language processing is also discussed. Syntactic, semantic and pragmatic knowledge and their issues and challenges are thoroughly elaborated in this paper. In future, these NLP applications can contribute a lot in several fields like banking, healthcare, commercial and academics domains if these ambiguity issues are challenges are resolved.

## References

- [1] Sowa, John F. Knowledge representation: logical, philosophical, and computational foundations. Vol. 13. Pacific Grove: Brooks/Cole, 2000.
- [2] Hirschman, Lynette, and Robert Gaizauskas. "Natural language question answering: the view from here." *natural language engi-neering* 7.4 (2001): 275-300.
- [3] Chowdhury, Gobinda G. "Natural language processing." *Annual review of information science and technology* 37.1 (2003): 51-89.
- [4] Nadkarni PM, Ohno-Machado L, Chapman WW. (2011). Natural language processing: an introduction. *J Am Med Inform Assoc.* 2011 Sep-Oct,18(5),544-551. doi: 10.1136/amiajnl-2011-000464.
- [5] Friedman C, Rindflesch TC, Corn M. (2013) Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. *J Biomed Inform* 46(5), 765–773.
- [6] Lehnert, Wendy G. *Strategies for natural language processing.* Psy-chology Press, 2014.
- [7] Yogish, D., N, P. M. T., & Hegadi, P. R. S. (2016). A Survey of Intelligent Question Answering System Using NLP and Information Retrieval Techniques. *International Journal*, 5(5), 536–540. <https://doi.org/10.17148/IJARCCE.2016.55134>
- [8] Clementeena, A., & Sripriya, P. (2018). A literature survey on question answering system in Natural Language Processing. *International Journal of Engineering and Technology(UAE)*, 7(2.33 Special Issue 33), 452–455. <https://doi.org/10.14419/ijet.v7i2.33.14209>





