A Review on Natural Language Processing

[1]Vishant Kumar, [2] Rajkumar Singh Rathore
[1]Student, [2]Assistant Professor
[1] Department Of Computer Science And Engineering
[1] Galgotias Educational Institutions, Greater Noida, Uttar Pradesh, India

***Abstract* - The field of natural language processing is an offshoot of areas like linguistics, artificial intelligence & computer science. The basis of this field is to enable communication between human beings natural language and the computers. One can say that it is related to the field of human-computer interaction. Like any other research field this field has its own challenges. In this paper we discuss about various challenges, applications and techniques used in Natural language processing.**

***Index Terms* – Discourse Analysis, Machine Translation, Natural Language Processing, Optical Character Recognition**

I. INTRODUCTION

Natural Language Processing (NLP) can be defined as a process that enables a machine to become more like a human , hence slashing the distance between machines and humans. In this field challenges like understanding the natural language i.e. permitting the machines to understand natural language of humans are faced. Some popular tasks of NLP are : discourse analysis, morphological separation, machine translation, understanding and generating natural language, recognizing named entities, part of speech tagging, recognizing optical characters, speech recognition and sentiments analysis etc. More interest in learning algorithms that are either semi-supervised or unsupervised in nature can be witnessed in the current researches. These techniques can efficiently perform the task of learning data that is not annotated manually with required answers or by employing a concoction of data that is neither annotated nor non-annotated. Usually, this task is hard in comparison to the task of learning which is supervised and displays little correct results for particular amount of data as input. Less accurate results are produced by it because of a large amount of non-annotated data available online.

Section II of paper discusses a few challenges of NLP, section III gives a detail of various applications of NLP and techniques of NLP are discussed in section IV.

II. CHALLENGES IN NATURAL LANGUAGE PROCESSING

A number of times a sentence is understood differently because of the mixing of the word boundaries. At the other level the syntax of the language is helpful in deciding the appropriate combination of words so as to make larger meanings. These are the major challenges faced in the NLP systems:

- Development of a program for understanding natural language
- A large number of natural languages that further contain infinite number of sentences.
- A large amount of ambiguity in natural languages
- Number of meanings of a single word
- Different meanings of sentences in different contexts

The above problems makes it difficult to design programs that understand a natural language, a major challenge.

III. APPLICATIONS OF NATURAL LANGUAGE PROCESSING

The data available on internet is uncountable but still an estimated figure of the same can be equalized to 20 billion pages. NLP expertise are required in applications for processing large quantity of data. Following is the list of such requirements:

- Classification of texts into categories
- Indexation and searching of large texts
- Automatic translation
- Understanding speech like understanding the telephonic conversations.
- Extracting relevant and useful information from the resumes
- Automatic summarization
- Question answering
- Acquiring knowledge
- Generating text or dialogues

IV. TECHNIQUES OF NATURAL LANGUAGE PROCESSING

- *Machine Translation*

The process of translating [1] text automatically from one human language to another human language. It is a very critical problem which is further associated with problems called as AI- complete. To completely solve the translation problem, it is necessary to

acquire various types of knowledge that human beings have i.e. knowledge of grammar, semantics and concepts related to real world.

- *Discourse Analysis*

This task is related to as many jobs to do. Structure determination of discourse of text that is connected and is a type of relation between discourse among line like: explanation and contrast is one such job.
Another job is to identify and categorize speech acts in the particular text. For example, questions with options like true or false, yes or no etc.

- *Morphological Splitting*

Splitting the terms in order to separate morphemes and the recognize categories for corresponding morphemes. The main issue with this kind of job is compilation of team structures in the language considered by us.

- *Generation and Understanding of Natural Language*

The process of generating natural language involves translating the information into easily readable language of human beings from the computerized databases. Natural language understanding involves change of text sections into much formal notations like the structure related to logistics in the first order that can be easily manipulated by the programs.
Semantics recognition using feasible semantics that are obtained using natural language expressions that are normally disguised as organized notations present in natural language concepts are dealt by it.

- *Named Entities Identification*

Identifying terms that can be labeled as named entities from the input text and then identifying the types to which these named entities might belong. Named entities can be identified in English language because of capitalization but this isn't helpful in identifying the name type.

- *Marking Part of Speech*

In case of every term input a line, identify and mark part of speech. A number of terms with same spellings but different meanings might be treated as more than one part of speech. For instance the term book might either be treated as a noun or might be treated as a verb. H. Recognizing Boundary of Sentences It deals with finding boundary of lines in given input text. We can normally mark line boundaries by full stop character or other characters like ?, ! etc but problem is that same type of characters may also be used for other motives like forming abbreviations.

- *Optical Character Recognition (OCR)*

Identifying text from images and then denoting the same in printed form is called as *Optical Character Recognition.* Quality of images plays a vital role in identifying the text and then denoting it in a printed text.

- *Recognizing Boundary of Sentences*

This technique deals with finding the boundaries of lines in given input text. Boundaries are normally marked using symbols like full stop, comma or ? character. The main problem is that similar types of characters might also be used for some other purposes.

V. CONCLUSION

The field of natural language processing is an offshoot of areas like linguistics, artificial intelligence & computer science. The basis of this field is to enable communication between human beings natural language and the computers. One can say that it is related to the field of human-computer interaction. Like any other research field this field has its own challenges. Mostly popular tasks of NLP are: analysis of discourse, morphological separation, machine translation, generation and understanding of natural language, recognition of named entities, part of speech tagging, recognition of optical characters, recognition of speech and analysis of sentiments etc. more interest in learning algorithms that are either supervised or semi-supervised in nature is highlighted by the current research studies.

References

[1] A. Lopez, "Statistical Machine Translation," Iternational Journal of ACM Computing Surveys, vol. 40, 2008.
[2] R. Mihalcea, H. Liu, and H. Lieberman, "NLP (Natural Language Processing) for NLP (Natural Language Programming), " In Proceedings of CICLING'06, LNCS, Springer, 2006, pp. 319-330.
[3] B. Manaris, "Natural Language Processing: A Human–Computer Interaction Perspective," Appears in Advances in Computers (Marvin V. Zelkowitz, ed.), Academic Press, New York, vol. 47, 1998, pp. 1-66.
[4] W. Fan, L. Wallace, S. Rich and Z. Zhang, "Tapping into the Power of Text Mining", International Journal of ACM, Blacksburg, 2005.
[5] P. M Nadkarni, L. O. Machado and W. W. Chapman, "Natural Language Processing: An Introduction,"J Am Med Inform Assoc, vol.18, 2011, pp. 544-551
[6] C. Wu and Y. Chen, "A Survey of Researches on the Application of Natural Language Processing in Internet Public Opinion Monitor," Proceedings of International Confernce on Computer Science and Service System (CSSS), IEEE, 2011, pp. 1035-1038.
[7] Z. B. Wu, L. S. Hsu, and C. L. Tan, "A survey on statistical approaches to natural language processing," Technical Report, 1992.