A Survey on Natural Language Processing and It's Applications

Abhilasha K V, Sahana S K, Hemanth Kumar A

Department of CSE, G. Madegowda Institute of Technology (GMIT), Maddur, Mandya District, Karnataka, India

ABSTRACT

NLP concerns with a design and implementation of computational machinery that communicates with humans using natural language. Humans can communicate easily and efficiently with one another using a natural language. Natural languages are those languages that are spoken by the peoples. Natural language aims to design the language input- output components of artificial intelligence system that are capable of using language flexibly as humans do. NLP is focused on human computer interaction. NLP is an interesting and difficult field in which we have to develop and evaluate or analyse a representation and reasoning theories. In this paper, we have carried out a survey on NLP, it's available tools and application.

Keywords. Linguistic, Natural language processing, Artificial intelligence, Data analytics

I. INTRODUCTION

NLP is a sub field of computer science artificial intelligence and inter disciplinary subject. Its aim is to build intelligence computation that can interact with human beings like human beings. NLP refers to the language spoken by the people example English, Kannada, Tamil and Japanese as opposed to artificial language like C++ and Java. [1] NLP is theoretically motivated range of computation and technique for analysing and representing naturally occurring text at one more levels of linguistic analysis it focuses on statistical methods. By utilising NLP, [2] [3]developers can organise and structure knowledge to perform tasks such as automatic summarization, sentimental analysis, speech recognition and topic segmentation. Apart from common word processor operation that treat text like a mere sequence of symbols, NLP considers the hierarchical structure of languages. NLP is used to analyse texts allowing machines to understand how humans speak. NLP is characterised as a hard problem in computer science.

Human language is rarely precised or plainly spoken to understand. Human language is not only to understand the words but the concepts and how they are link together to create a meaning.
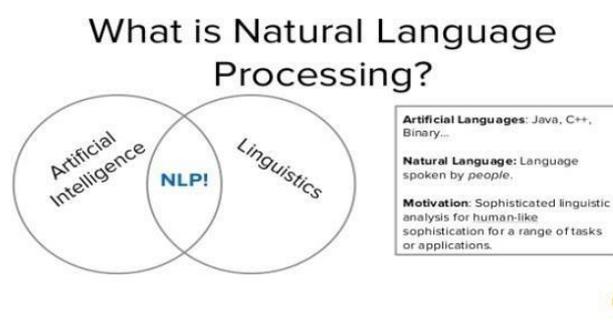

Figure 1a. Natural Language Processing

II. APPLICATIONS

NLP is everywhere even if you don't realise it. Does your e-mail application automatically correct you when you try to send an e-mail without the attachment that you referenced in text of e-mail? This is NLP applications at work. [5]

1.Machine translation
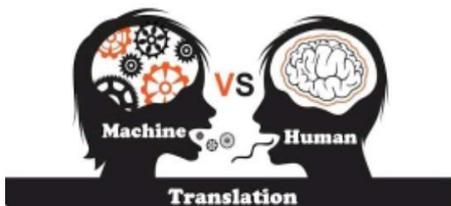
Figure 1b. Machine Translation

Amount of Information available online is growing the need to access it becomes increasingly important and the value of NLP applications becomes clear. Machine translation helps us to conquer language barriers that are often encounter by translating technical manuals. The challenge with machine translations is not in translating but in understanding the meaning of sentences to provide true translation.

1. Automatic Summarization
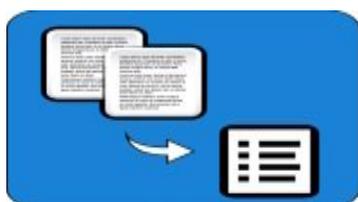
Figure 1c. Automatic Summarization

Information over loaded is a real problem when we need to access a specific important piece of information from huge known base. Automatic summarizing the meaning of document and information but also for understanding the emotional meaning inside the information such as collecting the data from social media.

2. Sentimental Analysis

The goal of sentimental analysis is to identify sentiment among several posts or even in the same post where emotion is not always explicitly expressed. Companies like NLP applications such as sentimental analysis to identify the opinions and sentiment online to help them understand what customers think about their product and services.

(i.e., "I love the new Iphone;" and few lines later "But sometimes it doesn't work well where the person is still thinking about the Iphone").

Figure 1c. Sentimental Analysis

3. Question Answering

As speech understanding technology and voice input applications improves, the need for NLP will only increase. Question answering is becoming more and more popular. Thanks to applications such as Siri, Ok Google and Virtual Assistance. A Question answering application is a system applicable of answering human request.

Figure 1d. Question Answering

4.Language Modelling

Language modelling is really a subtask of more interesting natural language problems, specifically those that condition the language models on some other input. The problem is to predict the next word given the previous words. The task is the fundamental to speech or optical character recognition, handwriting recognition, and statistical machine translation it is the key component of many deep learning natural language processing architectures.

5. Caption Generation

Caption generation is the problem of describing the contents of a given digital image, such as photo. It generates a textual description of the contents of

the image, a language model is used to create the caption that is conditioned on the image Ex., Describing the contents of a scene, creating a caption for a photograph and describing a video. This is not just an application for the hearing impaired, but it is also in generating human readable text for image and video data that can be searched as web.

III. CHALLENGES

Many challenges facing the practical application of natural language processing, there is one in particular that has seen especially little serious computational attention.namely the problem of metaphor.[8] Although many researches not in the metaphor area may be tempted to think of metaphor as a peripheral problem (perhaps mostly to do with poetry and other literary language) it is in-fact a pervasive feature of mundane language (conversation, newspaper articles, popular science writing, etc.) Providing adequate unified solutions for the "Microscopic" (short messages) and "Macroscopic" (really long sentences). The proliferation of social media suddenly made the "New Lang" of Twitter and Facebook more relevant than before for the real word applications. Most NLP application start with basic situation.simple but grammatically correct sentences. Now simply the sentences some more but make them grammatically incorrect, possibly with lacking punctuation, spelling errors. This is much more difficult. On the other side, we have the "Whales". Very long sentences with sub-ordinates clauses, detours and complex hierarchy. [6] They very relevant, too even in the same scenario of media monitoring, if we include traditional media, or social media for people with higher attention span.

Generic Solutions
A language is an infinite set by definition, even if we look at the lexicon and not the utterances produced. Different techniques work with different success, but it is inconceivable to build a different tool for every new environment the availability of

solutions also varies greatly depending on how mainstream a language is. Finally, the classic issues of the natural language processing are still not completely resolved today. There is no tool that can provide an expert human quality word-sense disambiguation yet.

IV. TOOLS

NLTK. It is a general library for NLP written in python. If your language of choice is Python, [4] then look no further than NLTK for many of your NLP needs. Similar to the Stanford library, it includes capabilities for tokenizing , parsing, and identifying named entities as well as many more features.
https://www.nltk.org/

Open NLP. It is library written in Java that implements different NLP tools [4]. http:j//opennlp.sourceforge.net/projects.html

Stanford Core NLP. It is a library including many of NLP tools develop at Stanford. [4] A GPL licensed framework of tools for processing English, Chinese and Spanish . It includes tokenization, part of speech tagging, named entity recognition and more. https://stanfordnlp.github.io/coreNLP

Weka. It is a machine learning library [4] spplying a large number of machine learning algorithm. https://www.cs.waikato.ac.nz/ml/weka/

Apache Open NLP. Using different underlying approach than Stanford library,[4] the open NLP project is an Apache licenced suite of tools to do tasks like tokenizing, part of speech tagging, parsing and named entity recognition. It remains solid choice.
https://opennlp.apache.org/docs/1.8.0/manual /opennlp.html

NLP tool kit. It is similar to Stanford library, it includes capabilities for tokenizing, [4] parsing, identifying named entities as well as many more features.

http://www.phontron.com/nlptools.php

RapidMiner. RapidMiner Studio is a powerful data mining tool for rapidly building predictive analytic workflows. This all-in-one tool features hundreds of data preparation and machine learning algorithms to support all your data mining projects. RapidMiner Studio is a visual workflow designer that lets data scientist use machine learning to produce insights on any data at any scale. [4] Replace multiple IBM products with one. It blends data, and deploy into production-all in single tool. No coding required.

GATE and Apache UIMA. As your processing capabilities evolve, you may find yourself building, complex NLP [4] workflows which need to integrate several different processing steps in these cases you may want to work with a frame work like GATE or UIMA that standardizes and abstracts much of the repetitive work that goes into building a complex NLP applications http://uima.apache.org/http://gate.ac.uk/

Apache Lucene and Solr. While not technically targeted at solving NLP problems, [4] Lucene and Solr contain a powerful number of tools for working with text ranging from advanced string manipulation utilities to powerful and flexible tokenization libraries to blazing fast libraries for working with finite state automatons. lucene.apache.org/solr

MALLET.a java based package for statistical natural language processing, [4] document classification, clustering.

KYTea.A toolkit for word segmentation and pronunciation [4] estimation www.diylife.com/Tool
ScalaNLP. Is the umbrella project for several libraries, including Breeze and epic[4].
Breeze is a set of libraries for machine learning and numerical computing.

Epic is a high -performance statistical parser and structured prediction library. www.scalanlp.org

V. CONCLUSION

Natural language processing in short NLP, focuses on the interactions between human language and computers. It is the intersection of computer science, artificial intelligence and computational linguistics. NLP is the way for computers to analyse, understand, and derive meaning from human language. By utilizing NLP, developers can perform tasks such as, summarization, translation, entity recognition, relationship extraction, sentiment analysis, speech recognition. NLP analyses language by it's meaning and also perform roles like correcting grammar, converting speech to text and automatically translating between languages. It allows machines to understand natural language. This human and computer interaction enables real-world applications like machine translation, automatic summarization, sentimental analysis, question answering, language modelling and more. NLP is commonly used for text mining.

VI. ACKNOWLEDGEMENT

VII. REFERENCES

[1] JaganandhGopinadhan "natural language processing" - slideshare October 2012
[2] Prashant Dohake "Natural Language Processing" in Artificial Intelligence November 2013
[3] Jason Brownlee in "Natural Language Processing" September 2017

[4] Grantingersoll "Open Sources Tools For Taming Text"

[5] July 2015

[6] Expert System "Examples of Natural Language Processing" May 2016

[7] Jason Brownlee "challenges in natural language processing" 20 September 2017

[8] International Journal of Speech Technology December 2008

[9] Barden J.A (2007) Metaphor, Semantic preferences and context sensitivity in K. Ahmad C. Brewster and M. Stevenson (Eds) . Words and Intelligence II:essay in honor of Yorick Wilks (pp 39-62) Berlin Springe