International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

Survey on Natural Language Processing and its Applications

Aiyesha Sadiya
Department of Information Science Engineering
Vidya Vardhaka College of Engineering
Mysuru, India

Archana R Hegde
Department of Information Science Engineering
Vidya Vardhaka College of Engineering
Mysuru, India

Under the guidance of,
Asst. Prof. Chandini S B
Department of Information Science Engineering
Vidya Vardhaka College of Engineering
Mysuru, India

Anusha V Illur
Department of Information Science Engineering
Vidya Vardhaka College of Engineering
Mysuru, India

Ashwini R
Department of Information Science Engineering
Vidya Vardhaka College of Engineering
Mysuru, India

***Abstract* - Processing Natural Language such as English has always been one of the central research issues of Artificial Intelligence, both because of the key-role language plays in human intelligence & because of the wealth of potential applications. Natural Language Processing techniques can make possible the use of Natural Language to express programming ideas, this in turn increases the accessibility of programming to non-expert users. NLP holds a great promise for making computer interfaces easier to use for people. NLP is used to analyze text, hence allowing machines to understand how humans speak. In this paper, we give an overview of NLP from the scratch. We also briefly discuss some of its major applications.**

Keywords – Natural Language Processing, Machine Learning, Speech Recognition, Natural Language Understanding, Natural Language Generation, Context Free Grammar, Spam, And Summarization.

I. INTRODUCTION

Natural Language Processing is an area of research and application that explores how computer can be used to understand and manipulate Natural Language text or speech to do useful things.NLP is multi-disciplinary, it is closely related to linguistics. It also has links to computer and information sciences, psychology, electric and electronic engineering. Of course, it is also related to Artificial Intelligence.

Applications of NLP include a number of fields of studies such as machine translation, speech recognition, text processing and summarization and so on.

II. DESCRIPTION

Natural language processing (NLP) is a field of computer science, artificial intelligence & computational linguistics which deals with the study of interactions between computers & human (natural) languages. In other words, teaching/programming computers how to understand & generate human language (natural language) such as English.

NLP is the field of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. NLP can be defined as the automatic processing of human language. Alternate terms that are often preferred are, 'Language Technology' or 'Language Engineering'. It is closely related to linguistics.

To summarize, NLP is a discipline which is concerned with the interaction between natural human languages and computing devices.

It is a way in which computers analyze, understand and derive meaning from human language in a smart and useful way. This human-computer interaction enables real-world applications like Search engines (Google), Translation systems (Google Translate), automated question answering, Text categorization, Spelling and Grammar checking and more. NLP is integrated widely in a large number of contexts such as evaluation systems, e-learning, research, machine translation, multilingual and cross-language information retrieval (CLIR), speech recognition.

As computers play a major role in preparation, storage, analysis and transfer of information, endowing them with the ability to understand and to generate information expressed in natural language becomes important.

Hence, the goal of NLP is to design and build computer systems which have the ability to analyze natural languages (like English, German, etc.) and to generate output in natural language. Human linguistic communication occurs as speech and as written language.

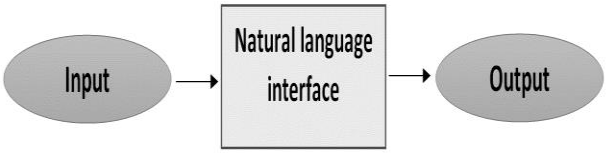
NLP's language processing problem can be divided into two chores:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

1) Processing the written text, this can be done using lexical, syntactic and semantic knowledge of the language.
2) Processing spoken language, this can be done using all the information needed and additional knowledge about phonology.

NLP system includes:

1) User input
2) It goes to the natural language interface
3) Output obtained in a language that is understood by the application program



The input and output of an NLP system can be of two types: speech and written text.

A. Components of NLP

There are two components of NLP.

1. Natural Language Understanding (NLU): It involves mapping the input into useful representations and analyzing different aspects of the language.
2. Natural Language Generation (NLG): It involves producing meaningful sentences in natural language form from the representations.

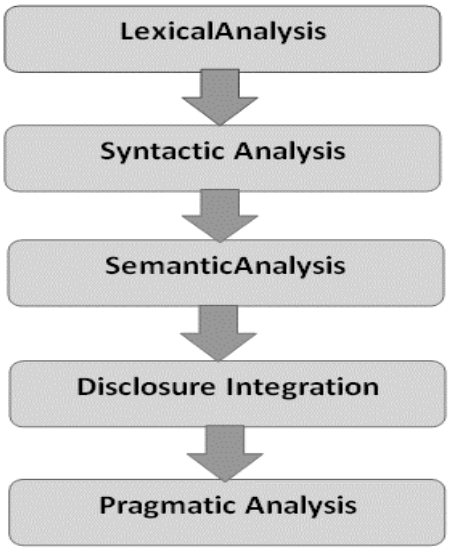
B. NLP Terminology

1) *Phonology*: Study of how sounds are organized and used in natural languages.
2) *Morphology*: Study of words, how they are formed and their relationship to other words in the same language.
3) *Syntax*: Arrangement of words and phrases to create well-formed sentences in a language.
4) *Semantics*: Study of meanings of words and phrases in a language. Has two main areas: lexical semantics and logical semantics
5) *Pragmatics*: Deals with using and understanding sentences in different situations and how the interpretation of sentence is affected.

C. Steps in NLP

There are generally five steps:

a) *Lexical analysis*: It involves identifying and analyzing the structure of words. It is basically dividing the whole chunk of text into paragraphs, sentences and words.
b) *Syntactic analysis (parsing):* This involves analysis of words in the sentence for grammar and arranging words to show the relationship among the words.
c) *Semantic analysis:* It draws the exact meaning or the dictionary meaning of the text. The text is checked for meaningfulness by mapping syntactic structures and objects in the task domain.
d) *Discourse integration*: the meaning of any sentence depends upon the meaning of the sentence preceding and immediately succeeding it.
e) *Pragmatic analysis*: it involves deriving those aspects of language which require real world knowledge. During this, what was said is re-interpreted on what it actually meant.



III. BACKGROUND

NLP is incredibly old. It has over fifty years of history as a scientific discipline. The history of NLP generally started in the 1950s, although work can be found from earlier periods. Early work in NLP was targeted on generation of complete computer programs that would compile and run.

Dictionary look-up system developed in 1948 was the first recognizable NLP application. In 1949, Warren Weaver involved in code breaking during the Second World War. He viewed German as English in code.

During the 1950s most NLP researchers were concentrating on Machine Translation (Russian to English) as an application. Alan Turing published an article 'Computing Machinery and Intelligence' which proposed the 'Turing Test'. Turing test is a test of a machine's ability to exhibit intelligent behavior indistinguishable from that of a human. In 1957, Noam Chomsky, a young American Linguist, introduced the concept of generative grammar rule-based description of syntactic structures. Almost all work in NLP since then, has been marked by his influence.

In the mid-60s Machine Translation worked only word-by-word and funding got drastically reduced. In 1960s SHRDLU and ELIZA were some successful NLP system developed.

During the 1970s many programmers began to write "conceptual ontologies (SAM, PAM)", structured real-world information into machine understandable data.

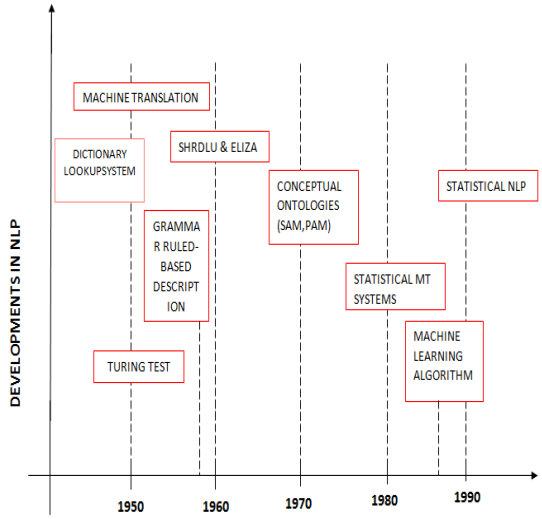
Up to the 1980s, NLP systems were based on complex sets of handwritten-rules. But in the late 1980s there was a revolution in NLP with introduction of machine learning

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

algorithms such as decision trees for language processing. The first Statistical Machine Translation Systems were developed in 1980s. And there were also some key developments such as Augmented Transition Networks, cost grammar, Semantic representations.

In the 1990s Statistical NLP became the most common paradigm.

Recent research has focused on unsupervised and semi-supervised learning algorithms. Such algorithms are able to learn from data that has not been navel-annotated with the desired answers. By 2025, industry experts expect NLP to be able to process all human languages at an accuracy of 99%.



IV. NLP ARCHITECHTURE

We need natural language as an input for NLP and to get the natural language as output we need the NLP system. NLP system has two parts, understanding part (input side processing) and generating part (output side processing). The general NLP system is shown in fig.4.

The NLP system includes:

1) A parser
2) Semantic Interpreter
3) Conversion Rules
4) Data Base Handlers

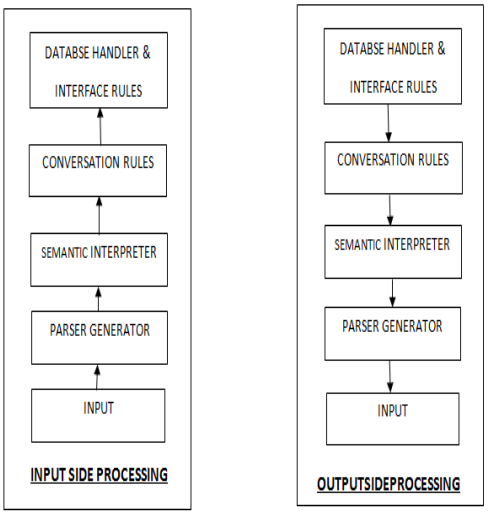
Each of the components are explained as follows:

1) *Parser*: Once the input is given to the parser it generalizes a syntactic structure in the form of parse tree. The purpose of parsing is to structure unstructured text. The fundamental idea of parsing is to group words together to form phrases, which behave as a single unit. These phrases can be combined together to form bigger phrases and eventually sentences. A parser is used to determine whether a given string belongs to the language and maps a string of words to its parse tree.

2) *Semantic Interpreter*: Semantic interpreter captures the semantic details of the parse tree and generates a deeper structure of the parse tree.

3) *Conversion Rules*: the conversion rules accept the deep structure of sentence from the semantic interpreter and make it compatible to be stored in the database.

4) *Database Handler*: the database handler works on the modified deep structure and generates a processed form for storage.



V.CONTEXT FREE GRAMMAR

Context-free grammar is a commonly used mathematical system for modeling phrase structure in Natural Language.It was first defined for Natural Language in 1957 by Chomsky.Context-Free grammar belongs to the field of FLT where a natural language is viewed as a set of sentences; a sentence as string of words from the vocabulary of the language & a grammar as a finite.

So, a Context-free grammar consists of 4 components:

a) ∑, the terminal vocabulary: the words of the language being defined.
b) N, the non-terminal vocabulary: a set of symbols disjoint from T.
c) R, a set of rules of the form
   $X \rightarrow Y_1, Y2,...,Yn$
   $N \rightarrow 0$, X€N, Yi € (NUS).
d) S, a start symbol, a member from N.

Example: A Context-free grammar for English

N={S, NP, VP, PP, DT, V_I, V_T, NN, IN}

S=S (Start symbol)

∑ = {words in the language}

R= Set of rules

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

S → NP VP
VP → V_i VP → V_t NP VP → VP PP
NP → DT NN NP → NP PP
PP → IN NP

S= Sentence, VP= verb phrase, NP= noun phrase, PP= prepositional phrase, V_I= intransitive verb, V_T= transitive verb, DT= determiner.

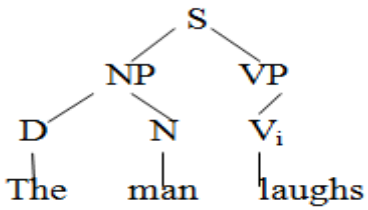
A. Left Most Derivation

A left most derivation is a sequence of strings, S_1,....., Sn, where S_1=S, the start symbol. $S_n \in \sum^*$, i.e.; S_n is made up of only terminal symbols.

For e.g.: [S], [NP VP], [D N VP], [the N VP], [the man VP], [the man V_i], [the man laughs]

Representation of derivation as a tree:

An Example: The man laughs.



A language is defined via derivation & a set of rules.

Derivation	Rules used
S	S → NP VP
NP VP	NP → DT N
DT N VP	DT → the
The N VP	N → man
The man VP	VP → V_i
The man V_i	V_i → laughs

In the derivation, we start with the start symbol 'S', and in each step we define a rule for the derivation. i.e., S goes to NP VP (S →NP VP).

And then, in the next step we replace 'S' in the derivation with NP VP. The basic idea here is that at each step we pick the left most non terminal symbol in the derivation & replace it with the non terminals defined in the rules, and it is continued to get a sequence of words. So a complete derivation always ends with a string where every word in a string is a word in the language defined. Then finally, these derivations are represented as a parse tree.

So, a Context-Free Grammar basically defines a set of possible derivations

VI. ENGLISH SYNTAX

A. Parts of Speech

Tags	Types	Examples
NN	Singular noun	Man, dog, park.
NNS	Plural noun	Flowers, houses, cars.
NND	Proper noun	Sita, IBM.

a) Nouns:

b) Determiners: Determiners come before nouns.

Tag	Examples
DT	The, a, some, every.

c) Adjective: Adjectives usually come between determiners and nouns.

Tag	Examples
JJ	red, green, large, small etc.

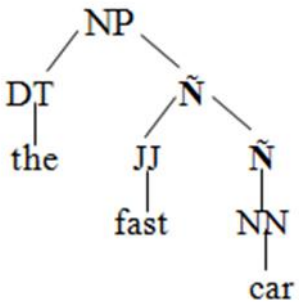
B. Noun Phrase Grammar

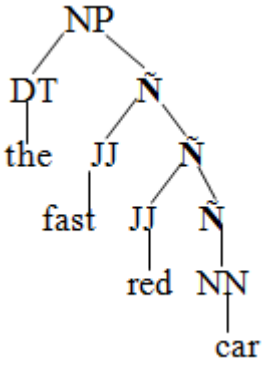
A noun phrase is formed by a determiner (DT) followed by a noun.

Rules:

NP => DT Ñ
Ñ => NN
Ñ => NN Ñ
Ñ => JJ Ñ
Ñ -> ÑÑ

Example:





International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

C. Prepositional phrase Grammar

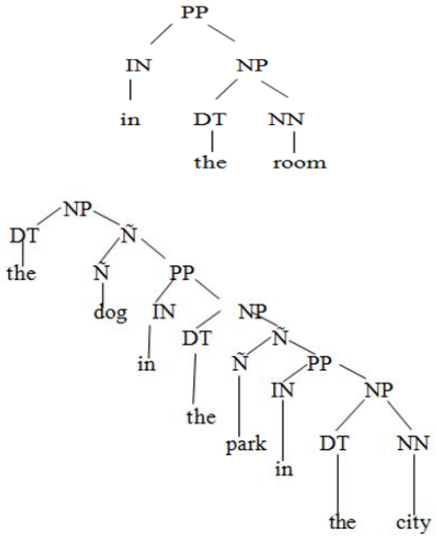
Prepositions usually come in front of a noun phrase.
Preposition phrase can be formed by the preposition tag IN

Tag	Examples
IN	in, of, out, beside, as.

followed by NP
Rules:

PP => IN NP
Ñ => Ñ PP

Example:



The rules can be recursively applied to build the larger sequences of prepositional phrase modifiers as in fig.2

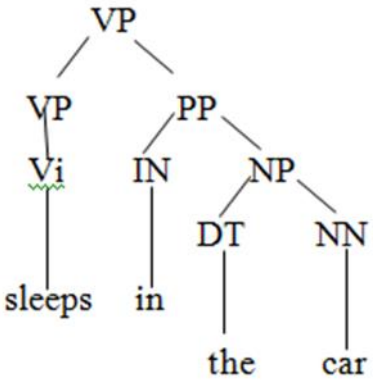
D. Verb phrases Grammar

Basic verb types

Tags	Verb types	Examples
Vi	Intransitive verb	Sleeps, walks, laughs.
Vt	Transitive verb	Sees, saws, likes.
Vd	Di-transitive verb	gave

Rules:
Verb phrases can be made up of transitive verb followed by noun phrase. Verbs can also be formed by noun phrases or preposition phrases.

VP => Vi
VP => Vt NP
VP => Vd NP NP
VP => VP PP



Example:

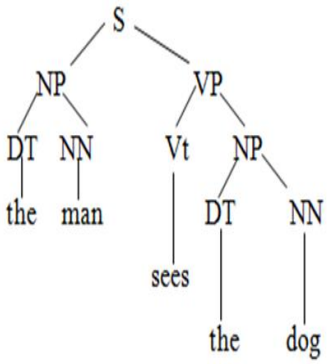
Tag	Example
S	The man sleeps.

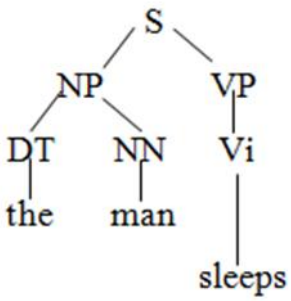
E. Sentences

Basic Sentence rule: The sentence can be formed by a noun phrase.(NP) followed by a verb phrase (VP).

S => NP VP

Example:



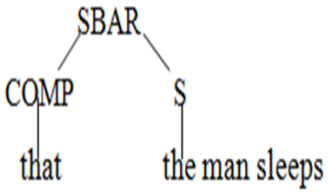


International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

F. Complimentizers and SBARS:

Tag	Example
COMP	that
SBAR	that the man sleeps

SBAR can be formed by a COMP followed by a sentence (S).



SBAR => COMP S

G. More verbs

Tags	Examples
V[5]	said, reported.
V[6]	told, informed.

New verb types:

VP => V [5] SBAR

VP => V [6] NP SBAR

Rules:

Example:said that the man sleeps.
told the man that she likes a rose.

H. Coordination

Another parts of speech is a coordinator that has the tag CC.

Tag	Examples
CC	And, or, but

Rules:

NP => NP CC NP

Ñ => Ñ CC Ñ

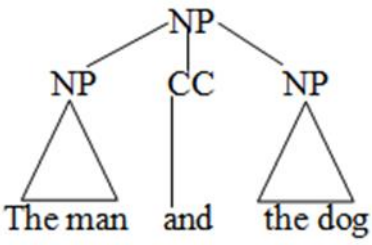
VP => NP CC VP

VP => VP CC VP

S => S CC S

SBAR => SBAR CC SBAR

Example:



VII. PARSING

Parsing is a technique in NLP that checks if the sentence is correct according with the grammar and if so returns a parse tree representing the structure of the sentence.

So, a parse tree is a tree structure with the words of sentence as the leaves of the tree.

The fundamental idea of parsing is that words group together to form phrases, which behave as a single unit. These phrases can combine together to form bigger phrases and eventually sentences.

And basically, a Parser is used to determine whether a given string belongs to the language and, maps a string of words to its parse tree.

S: Sentence

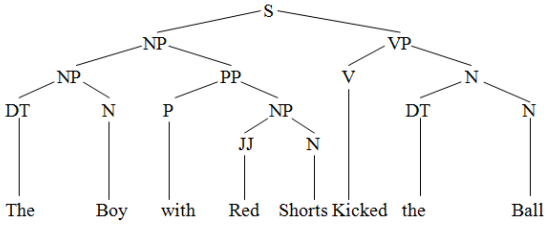
NP: Noun Phrase

VP: Verb Phrase

DT: Determiner

JJ: Adjective

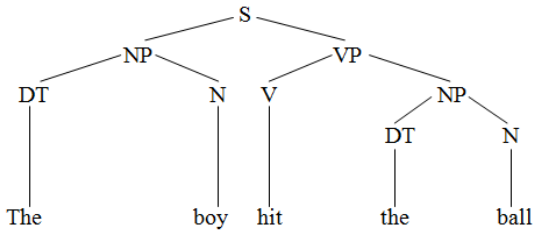
Output:



The information represented by the parse trees:

A. Parts of speech for each word.

N = noun
V = verb
DT = determiners

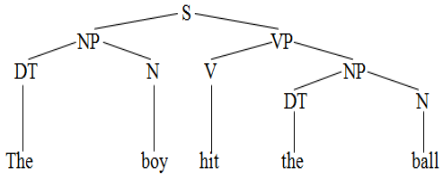


The first level of the tree encodes the parts of speech tag sequence for the input sentence.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

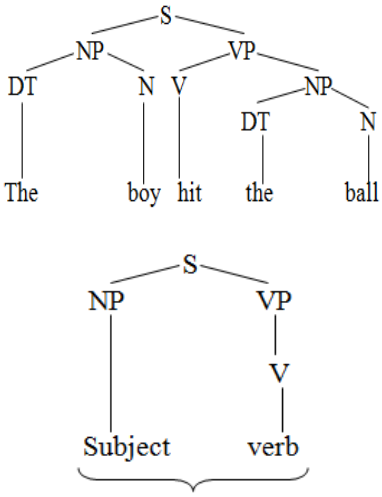
B. Phrases

At the next level, there is a hierarchical grouping of words into phrases (constituents)



Noun phrases (NP): "The boy", "the ball"
Verb phrases (VP): "hit the ball"
Sentences(S): "The boy hit the ball"

C. Grammatical Relationships



Subject-verb relationship

"The boy" is the subject of "hit".
These grammatical relationships allow us to identify who did what to whom.

VIII. APPLICATIONS OF NLP

Humans perform majority of activities through language either by direct communication or by using natural language. So, to communicate, there is a need to understand the language we use. As the study of human languages developed, the concept of communicating with no-human devices was investigated. This idea gave rise to the Natural language processing. Thus, NLP helps machines to read text by simulating the human ability to understand the language. We use NLP applications everyday and everywhere directly or indirectly even if we don't realize it.

The applications of NLP can be divided into two categories as follows:

1) Text-based applications:

As the name says, text based applications are applications that deal with processing of written text such as book, newspapers, e-mail messages, manuals, reports, research papers and many more. These applications such as extracting information from messages, articles, documents, machine translation and summarizing texts and so on.

2) Dialogue-based applications:

Dialogue based applications deal with spoken language. This involves applications such as question-answering systems, tutoring systems, interactive problem solving, voice controlled machines, automated customer service over telephone and so on.

A. Machine Translation:

Machine translation is the automatic translation of text or speech from one language to another. It is the problem of converting a source text to another language by using automated computing without human intervention. Developed in the 1950s, machine translation is also referred to as automatic translation.

Today, a large amount of data is generated online and the need to access it becomes very important. Different countries have different languages and different cultures. Hence, there is a need of inter-language translation to share ideas, to transfer information and communicate with one another.

Machine translation helps us conquer language barriers that we often encounter by translating technical manuals, content or catalogs at a significantly reduced cost. It offers a mere scalable alternative to harmonizing the world's information.

There are three types of machine translation:

a) Rules based systems:They use a combination of grammatical and language rules and dictionaries for words.
b) Statistical systems:They analyze the large amount of data for each language pair and learn to translate.
c) Neural systems:They incorporate an approach which makes the machines learn to translate the text through a large neural network.

B. Fighting spam:

With the popularity of the internet, e-mail has become a part of our daily life. Often emails are prone to spam emails. Spam emails are unsolicited bulk email. These emails advertise products and services or they lead to malicious sites. Thus, spam filters have become a must these days, important as the first line of defuse against the ever increasing problem of unwanted e-mail. Although numerous techniques have been developed, we still receive them quite often as the spammers manipulate the filters. Also, the important mails may accidentally get caught in the filter.

Natural language processing is used for the filtering of spam emails in order to enhance online security. There are various approaches in NLP to solve this problem. In one such approach, NLP engine takes the unclassified email and its category as an input and then processes it further using the statistical NLP approach. The Bayesian spam filtering is considered to be the best approach.

C. Text Classification:

Text classification is basically, the classification of text based on the content of the text. Given an input text, it predicts a pre-defined class label for it.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

A popular example is spam filtering. Some other examples include classifying the language of the source text & genre of a fictional story.

Types of text classification:

- *Binaryclassification*: If there are exactly two classes to which a document belongs to, then it is a binary classification.
- *Multi-classclassification*: If there are more than two classes & each document belongs to exactly one class, it is a multi-class classification.
- *Multi-labelclassification*: In this, a document has more than one associated category in the classification scheme.

D. Speech Recognition:

A field of computational linguistics which involves developing methodologies & technologies that enables recognition of an acoustic signal containing spoken language & translates it into text.

In simple words, it is the technique of converting spoken language into text. A language model is used to create the text output that is conditioned on the audio data. When a person reads text into the system, the system analyzes the person's specific voice & uses it to fine-tune the recognition of that person's speech, thus maintaining accuracy. Speech recognition is alternatively called voice recognition.

Advances in deep learning over the last 10 years have allowed major players to deploy systems that involve:

a) Transcribing a speech.
b) Creating text captions for a movie or T.V show.
c) Issuing commands on radio while driving.
d) Voice search.(Google, Siri)
e) Call steering.
f) Automated identification using voice biometrics.

E. Sentiment Analysis:

Goal of sentiment analysis is to identify the sentiment among several posts or even in the same post where emotion is not always explicitly expressed. Sentiment analysis determines the attitude, emotional state, judgment or intent of the writer. Companies use sentiment analysis to identify options and sentiment online to understand what their customers think about their product and services and overall indicators of their reputation.

It is done by assigning a polarity to the text – positive or negative – or trying to recognize the underlying mood.

Sentiment analysis understands sentiment in context to help us better understand what's behind an expressed opinion.

F. Document Summarization:

Information overload is a real problem when we need to access am important piece of information, and already our access to knowledge and information far exceeds our capacity to understand it. Hence an ability to summarize the meaning of documents and information is becoming increasingly important.

Automatic summarization is also relevant to collecting data from social media and also when used to provide an overview of a news item or blog post. Another desired outcome is to understand deeper emotional meaning. It reduces the redundancy from multiple sources.

G. Question Answering:

As speech understanding technology and voice input applications improve, the need for NLP will only increase.

Search engines put the world's information at our fingertips, but still are primitive when comes to actually answering specific questions posted by humans.

Question answering is becoming more and more popular, with the rising of applications like Siri, OK Google, chat bots and virtual assistants. Question answering application is a system capable of coherently answering a human request in natural language. It may be used both as a text-only interface and as a spoken-dialogue system.

Google has seen the frustration it has caused in users, who often need to try a number of different search results to find the answer they are looking for. Though certainly improving, this remains a relevant challenge for search engines, hence becoming one of the main applications of NLP research.

H. Co-reference Resolution:

Co-reference resolution connects pronouns to the right objects. It is essential to interpret the text correctly. It is an important step for NLP applications like document summarization, question answering, etc.

It occurs when two or more expressions in a text refer to the same person or thing.

E.g. *Bill said he would come*, the proper noun *Bill* and the pronoun *he* refers to the same person, i.e. Bill.

Sense Disambiguation: Sense disambiguation is the problem of determining which sense (meaning) of a word is activated by the use of the word in a particular context. While the human brain is pretty good at this task, a computer won't find it easy to recognize it.

E.g. the computer won't be able to recognize that the term *pounds* in the sentence *I gained 20 pounds since the wedding!* , most likely refers to the unit of mass rather than the currency.

I. Named Entity Recognition (NER):

Named entity recognition is a subtask of information extraction that locates and classifies named entities in text into predefined categories. These categories may be names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

Applying NER to a sentence will be able to convert it from, *Valium makes me sleepy*, to *{drug} makes me {symptom}.*

J. Language Modeling:

A language model is a function that puts a probability measure over strings drawn from some vocabulary. It is a conditional distribution of determining ithword in a sequence of text, given all the previous word identities.
Language modeling is used in speech recognition, machine translation, part-of-speech tagging, information retrieval, parsing, etc.

IX. MAJOR CHALLENGES IN NLP

The most natural means of communication between humans is Natural Language, spoken, written or typed. The dominance of natural language as a means of communication among humans suggests that it would be an agreeable medium in human-computer interaction. Thus, the major goal of NLP would be the ability to use natural language as effectively as humans do. There is no tool that can provide an expert human quality word-sense disambiguation.

Also, the goal of NLP is to enable computers to engage themselves in communication using natural human speech and language, so that non-programmers can interact with the computers easily and effectively.

When this goal is achieved, computer systems will be able to understand, analyze, summarize, translate and generate accurate human text and language, which is the most natural means of communication between humans.

NLP considers hierarchical structure of language instead of treating text as a sequence of symbols i.e., in NLP several words make a phrase, several phases make a sentence. So, to achieve this, goal computers must be endowed with natural language processing capabilities, and these provide the following major challenges to the NLP systems:

A. Machine Translation:

Machine Translation is the task of automatically converting one natural language into another natural language, preserving the meaning of the input text and producing fluent text in the output language. Correct translation requires not only the ability to analyze and generate sentences in human languages but also human like understanding of world knowledge and context, despite the ambiguities of languages i.e., computers should be able to understand input in more than one language, provide output in more than one language and translate between languages.
Words and phrases must be passed and interpreted so that their intended meaning (as command, query, or assertion) maybe determined and an appropriate response is formulated and expressed. Application areas include science, diplomacy, multinational commerce and intelligence.

Today, most NLP resources and systems are available only for high resource languages (HRLs), such as English, French and German. Whereas many low resource languages (LRLs) such as Indonesian, Swahili-spoken and written by millions of people have no such resources or systems available. So a future major challenge for the NLP community is to develop resources and tools for hundreds and thousands of languages, not just a few.

B. Reading and Writing Text:

Text reading and writing is one of the major challenges in NLP. Machine reading is the idea that machines could become intelligent, integrate and summarize information for humans, by reading and understanding the text available i.e., computers should be able to understand and process the data.
Areas of applications include intelligence, logistics**,** office automation and libraries. With the emergence of the modern online world, we have huge storage of online information coded in human languages. E.g. scientific literature, where findings are still reported almost entirely in human language text. The quality of scientific literature is growing rapidly. So scientists are unable to keep up with the literature. Thus, we can say that there is an increased need for machine reading for the purpose of comprehending and summarizing the literature as well as extracting facts and hypothesis from this material. Also, machine reading has to provide question-answering systems, by which humans can get answers from constructed knowledge bases.

C. Interactive dialogue:

Since the 1980s, dialogue has been a popular topic in NLP research.
Interactive dialogue is allowing humans simple, effective access to computer systems, using natural language for problem solving, decision making and control. Application areas include database access, command and control, factory control, office automation, logistics and computer assisted instructions. Human machine interaction should be as natural, facile and multi-modal as interaction among humans.

Early work on text-based dialogue has now been expanded to include spoken dialogue systems on mobile devices for information access and task-based applications (SDS).

Although SDS work fairly well in limited domains, where the topics of interaction are known in advance and where the words people are likely to use can be predetermined, they are not yet very successful in open domain interaction, where users may talk about anything at all. The other challenges in building SDS are basic problems of recognizing and producing normal human conversational behaviors.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

D. Sentence Generation:

It is very non-trivial to come up with models that always generate grammatically correct and meaningful sentences.

X. LIMITATIONS OF NLP

In theory, NLP is a very attractive method of human-computer interaction. Natural language recognition requires extensive knowledge on outside world.

Systems such as SHRDLU, which were developed earlier had restricted "block words" with restricted vocabularies and worked extremely well. But it was soon failed when the systems were extended to more realistic situations with real world ambiguity and complexity.

Hence, the developing of NLP application is challenging as computers require humans to "speak" to them in a programming language that is precise, unambiguous and highly structured.

The major problems with natural language processing are ambiguity, vagueness and uncertainty.

These problems have to be faced whether one is dealing with a single sentence or discourse.

The critical problem associated with NLP is ambiguity. Ambiguity refers to an expression (word/ phrase/ sentence) having more than one interpretations. Ambiguity can occur at the levels of lexical, syntactic, semantic, discourse and pragmatic analysis.

Different types of Ambiguity are:

1) Lexical Ambiguity:

It is the ambiguity of a single word. A word can be ambiguous with respect to its syntactic class.

Ex: 1) She bagged a silver medal.
2) She made a silver speech.
3) His worries had silvered his hair.

The word 'silver' is used as a noun, an adjective, or a verb.

2) Syntactic ambiguity or Structural ambiguity:

It is of two kinds: Scope ambiguity and Attachment ambiguity.

a) *Scope ambiguity*: Scope ambiguity involves operators and quantifiers.
Ex: Old men and women were taken to safe locations.
The scope of the adjective (i.e., the amount of text it qualifies) is ambiguous. That is, whether the structure (old men and women) or ((old men) and women)?
The scope of quantifiers is often not clear and it creates ambiguity. Ex: Every man loves a woman.
The interpretations can be, for every man there is a woman and also it can be there is one particular woman who is loved by every man.

b) Attachment *ambiguity:* A sentence is said to have attachment ambiguity if a constituent fits more than one position in a parse tree. Attachment ambiguity arises from uncertainty of attaching a phrase or clause to a part of a sentence.
Ex: The man saw the girl with the binocular.
It is ambiguous whether the man saw a girl carrying a binocular, or he saw her through his binocular.
The meaning is dependent on whether the preposition 'with' is attached to the girl or the man.

3) Semantic Ambiguity:

This occurs when the meaning of the words themselves can be misinterpreted. There are two ways of reading the sentence,even after the syntax and the meanings of the individual words have been resolved. Ex: Seema loves her mother and Sriya does too.
The interpretations can be Sriya loves Seema's mother or Sriya likes her own mother.

4) Pragmatic Ambiguity:

Pragmatic ambiguity occurs when the context of a phrase gives it multiple interpretations.The problem involves processing of highly complex tasks such as user intention, sentiment, belief world, and modals etc.

Ex: I love you too.
This can be interpreted as
I love you (just like you love me)
I love you (just like someone else does)
I love you (and I love someone else)
I love you (as well as liking you).

XII. FUTURE OF NLP

A. The bots:

Chat bots are able to translate and interpret human language input. This is done through a combination of natural language processing and machine learning. There are well known examples of AI and chat bots, for e.g. Clever bot, Crotona, Siri. Chat bots are frequently used in customer services, chat bots help customers get right to the point without the wait, answering customer questions and diverting the, to relevant resources. To be effective, chat bots must be fast, smart and easy to use. To accomplish this, chat bots employ NLP to understand language usually over text or voice recognition interactions.

The most glaring shortcoming of today's chat bots is that they don't understand what you are saying. They'll often misinterpret what you type, or ignore it completely. On the other hand, the chat bots that try to understand and respond to every word you say can fail even more awkwardly. Thus, overcoming this issue of chat bots becomes a major research area in NLP.

B. Support invisible UI:

An invisible UI (user interface) is where the user experience of a feature is so intuitive that a UI isn't really required. A real life example could be sliding doors, there is no signage required, no button necessary and the solution is invisible.

As software and internet becomes more useful every day there is ever the more need to design invisible UIs. This level of intuition helps the user achieve their goalfaster and easier, hence leaving the user using your application or

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 7 Issue 01, January-2018

website feeling satisfied. The concept of invisible UI relies on direct interaction between user and machine. NLP leverages greater contextual understanding of human language. It gets better understanding of us – no matter how we say it and what we are doing- which will be essential for any invisible UI application.

C. Smarter Search:

Future of NLP is also for smarter search. The same capabilities which allow a Chat bot to understand a customer's request can enable "search like you talk" functionality.

XIII. CONCLUSION

In our paper we have given an overview of NLP and its applications.

The impact of computer use of Natural Languages will have as profound an effect on society as would the breakthroughs in superconductors, inexpensive fusion or genetic engineering. The impact of NLP by machine will be greater than the impact of microprocessor technology in the last 20 years, because Natural Language is fundamental to almost all business, military & social activities. Therefore, the application of NLP has no end.

XIV. REFERENCES

[1] Ms. Rijuka pathak et al. Natural language processing approaches, application and limitations. In Natural language processing approaches, application and limitations (Vol. 1 Issue 7, September – 2012)
[2] Julia Hirschberg et al. Advances in natural language processing. In SCIENCE (17 JULY 2015 • VOL 349 ISSUE 6245)
[3] Anjali M K1 et al.Ambiguities in Natural Language Processing. In International Journal of Innovative Research in Computer and Communication Engineering (Vol.2, Special Issue 5, October 2014)
[4] http://www.expertsystem.com/natural-language-processing-applications/
[5] http://language.worldofcomputing.net/understanding/applications-of-natural-language-understanding.html
[6] http://www.ps.uni-saarland.de/~niehren/oz-natural-language-script.html/vorlesung/node49.html
[7] https://en.m.wikipedia.org/wiki/Natural_language_processing
[8] https://blog.algorithmia.com/introduction-natural-language-processing-nlp/
[9] Ralph Weiscbedel. White Paper on Natural Language Processing. In ResearchGate (November 2003)
[10] http://www.expertsystem.com/natural-language-processing-applications/
[11] https://www.techemergence.com/natural-language-processing-business-applications/
[12] ttps://www.cl.cam.ac.uk/teaching/2002/NatLangProc/revised.pdf
[13] http://www.glossary.sil.org/term/pragmatics
[14] https://en.wikipedia.org/wiki/Morphology_(linguistics)
[15] https://en.wikipedia.org/wiki/Phonology
[16] https://www.lifewire.com/applications-of-natural-language-processing-technology-2495544?print
[17] https://www.tutorialspoint.com/artificial_intelligence/artificial intelligence_natural_language_processing.htm
[18] http://language.worldofcomputing.net/machine-translation/machine-translation-process.html
[19] https://en.m.wikipedia.org/wiki/Language_model
[20] https://en.m.wikipedia.org/wiki/Speech_recognition
[21] Geetika Sethi et al. Role of Natural Language Processing in Education. In IJSRE (Volume 4 Issue 3 March 2016)
[22] https://machinelearningmastery.com/applications-of-deep-learning-for-natural-language-processing/
[23] https://en.m.wikipedia.org/wiki/Speech_recognition
[24] http://www.expertsystem.com/future-natural-language-processing/
[25] Dr. Khaled M Alhawiti Natural Language Processing and its Use in Education. In International Journal of Advanced Computer Science and application (Vol. 5 number 12, 2014)
[26] Ronan Collobert et al. Natural Language Processing (Almost) from Scratch. In Journal of Machine Learning Research
[27] Jill Brstein Opportunities of Natural Language Processing Research in Education. In Springer-Verlag Berlin Heidelberg 2009
[28] Dianne Litman Natural Language Processing for Enhancing Teaching and Learning. In AAAI Conference on Artificial Intelligence 2016