NATURAL LANGUAGE PROCESSING

Presented by:
Prof Nirali Arora

CHAPTER 1
INTRODUCTION TO NLP

TOPICS TO BE COVERED

- INTRODUCTION
- ORIGIN & HISTORY OF NLP
- KNOWLEDGE AND GRAMMER IN LANGUAGE PROCESSING
- STAGES IN NLP
- AMBIGUITIES AND ITS TYPES IN ENGLISH AND INDIAN REGIONAL LANGAUGES
- CHALLENGES OF NLP
- APPLICATIONS OF NLP

Introduction to Natural Language processing

Natural Language Processing is a field of research and application that determines the way computers can be used to understand and manage natural language text or speech to do useful things.

NLP is a range of computational techniques for analyzing and representing naturally occurring text at one or more levels of linguistic analysis for the purpose of achieving human like processing for range of tasks or applications.

INTRODUCTION

- Natural Language Processing is a method of a computer program to understand human language, which is either written or spoken.
- It is the component of Artificial Intelligence that enables the computer to understand the human language process.
- The main purpose of NLP is to read and understand the human language and deliver the output accordingly.

- Need of Natural Language Processing
- It saves time to perform certain tasks like automated text writing and automated speech.
- It also helps the people who can’t write but can share their queries by speaking.
- A computer can be more useful if it is capable to communicate with human beings.
- The main use of the NLP application is machine translation, It helps to convert the information from one language to another.
- It also helps non-programmers to interact with the computer system and access information from it.

History of NLP

Phase I : 1940 to 1960

- **Phase II : 1960 to 1970**
- **Phase III : 1970 to 1980**
- **Phase IV : 1990 to 2000**
- **Phase V : 2000 to present**

History of NLP

Phase 1 : 1940-1960

Research on NLP started with Machine Translation (MT)

- **Automatic Machine Translation from Russian to English.**
- **It was very basic and rudimentary form.**
- **Limited experiments were carried out in IBM Gorgetown demonstrations.**
- **First International conference on MT was held in 1952 and Second International conference was held in 1956 at Washington.**
- **Most of the research work during this period made on Syntax and partially semantic processing.**

History of NLP

Phase 2

Second phase work was AI flavoured.

- **BASEBALL Question Answering System (1961).**
- **Emphasis on world knowledge**
- **Manipulation of meaning representations.**

Phase 3

SHRDLU (1973) an early natural-language understanding computer program, developed by Terry Winograd at MIT in 1968–1970.

- Woods (1978) Answering Questions about ATN Grammars.
- ARPA Speech Understanding Project (1980).
- Front end projects were used for interfacing with large automatic databases.
- INTELLECT was first commercial front end

History of NLP

- Research on NLP began in 1950, after the Booth and Richen investigation, when Alan Turing announced an article titled “Machine and Intelligence.”
- It was tried to automate translation from Russian to English in 1954.
- The first international meeting on machine translation was held in 1952, and the second was in 1956.
- At the beginning of 1961, the work started on the problems of creating data or knowledge base influenced by AI. A BASEBALL question-answering system was also introduced in the same year.

- An advanced system was **described in 1968. After comparing this system with the BASEBALL question-answering system, it was recognized and then provided to fulfill the need for inference on the knowledge base to interpret and respond to language input.**
- **In 1980, the work on the lexicon also supported the grammatico-logical approach.**
- **In 1990, the data-driven and probabilistic became fully standard.**

Generic NLP system

Any natural language processing should start with some input and end with effective and accurate output.

The inputs for natural language processor can be text or speech

The outputs that can be generated by the system could be form of answer when the input is the question.

Outputs can be database update,spoken response,semantics,part of speech or semantics of word/sentences etc.

Levels of NLP

Phonology it deals with interpretation of speech sounds within and across words.

Morphology it is the study of the way words are built up from smaller meaning-bearing units called as morphemes.

Semantics it is the study of the meaning of the words that are associated with a grammatical structure.It consist of two types of approaches syntax driven semantic analysis and semantic grammar.

Reasoning in order to produce an answer to the question which is not explicitly stored in a database NLIDB carries out the reasoning based on the data stored in the database.

KNOWLEDGE AND GRAMMER IN LANGUAGE PROCESSING

- Natural language understanding system must have knowledge about what the words mean, how words combine to form sentences, how word meanings combine to from sentence meanings and so on.
- The different Levels (forms) of knowledge required for natural language understanding are given below.
- **1. Phonetic Knowledge-** Phonetics is the study of language at the level of sounds while phonology is the study of combination of sounds into organized units of speech.
- Phonetic and phonological knowledge are essential for speech based systems as they deal with how words are related to the sounds that realize them.

- **2. Morphological Knowledge-** Morphology concerns word formation.
- It is a study of the patterns of formation of words by the combination of sounds into minimal distinctive units of meaning called mophemes.
- Morphological knowledge concerns how words are constructed from morphemes.
- **3. Syntactic Knowledge-** Syntax is the level at which we study how words combine to form phrases, phrases combine to form clauses and clauses join to make sentences.
- Syntactic analysis concerns sentence formation. It deals with how words can be put together to form correct sentences.

4. Semantic Knowledge- It concerns meanings of the words and sentences.

- This is the study of context independent meaning that is the meaning a sentence has, no matter in which context it is used.
- Defining the meaning of a sentence is very difficult due to the ambiguities involved.

5. Pragmatic Knowledge- Pragmatics is the extension of the meaning or semantics.

- Pragmatics deals with the contextual aspects of meaning in particular situations.
- It concerns how sentences are used in different situations and how use affects the interpretation of the sentence.

6. Discourse Knowledge- Discourse concerns connected sentences.

- It is a study of chunks of language which are bigger than a single sentence.
- Discourse language concerns inter-sentential links that is how the immediately preceding sentences affect the interpretation of the next sentence.
- **7. World Knowledge-** World knowledge is nothing but everyday knowledge that all speakers share about the world.
- It includes the general knowledge about the structure of the world and what each language user must know about the other user's beliefs and goals.
- This essential to make the language understanding much better.

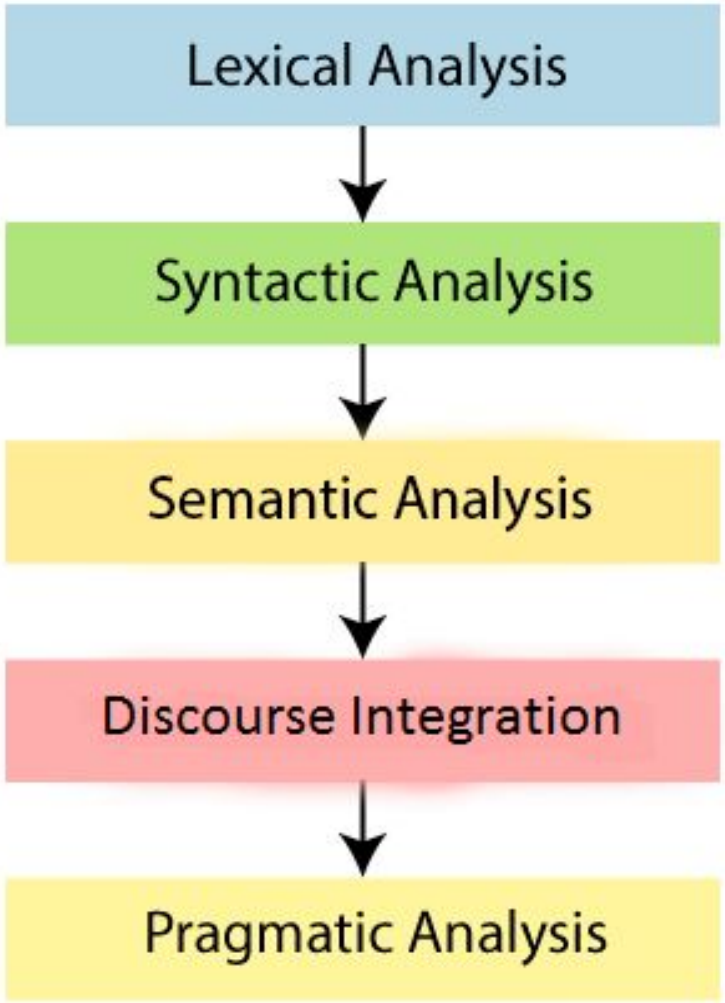
KNOWLEDGE AND GRAMMER IN LANGUAGE PROCESSING

- **Components of NLP**
- There are two components of NLP
- 1. Natural Language Understanding (NLU)
- 2. Natural Language Generation (NLG).
- Natural Language Understanding (NLU) which involves transforming human language into a machine-readable format.
- It helps the machine to understand and analyse human language by extracting the text from large data such as keywords, emotions, relations, and semantics.
- Natural Language Generation (NLG) acts as a translator that converts the computerized data into natural

Levels and Tasks of NLP

- 1. Morphology- It is the analysis of individual words that consist of morphemes (smallest gramatical unit).
- 2. Syntax- It is concerned with rules. It includes legal formation of the sentences to check the structures.
- 3.Semantic- During this phase, meaning check is carried.
- 4. Discourse Intergration- It deals the meaning of the current sentence is dependent on the previous sentence.
- 5. Pragmatic- It deals with mapping of what the has interpreted from conveye part and what was actually expected.
- 6. Prosody- It handled the rhythm.
- 7. Phonology- This involves analyisis of different

Stages in NLP



- 1. Lexical Analysis/Morphology-
- The first phase of NLP is the Lexical Analysis.
- This phase scans the source code as a stream of characters and converts it into meaningful lexemes.
- It divides the whole text into paragraphs, sentences, and words.

- **2. Syntactic Analysis (Parsing) –**Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.
- The sentence such as “The school goes to boy” is rejected by English syntactic analyzer.
- **3. Semantic Analysis –** It draws the exact meaning or the dictionary meaning from the text.
- The text is checked for meaningfulness. The semantic analyzer disregards sentence such as “hot ice-cream”.
- **4. Discourse Integration-** The meaning of any sentence depends upon the meaning of the sentence just before it.
- In addition, it also brings about the meaning of immediately succeeding sentence.
- **5. Pragmatic Analysis –** During this, what was said is

Ambiguity and Uncertainty in Language

- Ambiguity, generally used in natural language processing, can be referred as the ability of being understood in more than one way.
- NLP has the following types of ambiguities –
- **1. Lexical Ambiguity-** The ambiguity of a single word is called lexical ambiguity.
- For example, the word “back” can be a noun (back stage), an adjective (back door) or an adverb (back away).
- **2. Syntactic Ambiguity-** This kind of ambiguity occurs when a sentence is parsed in different ways.
- For example, the sentence “The man saw the girl with the telescope”. It is ambiguous whether the man saw

- **3. Semantic Ambiguity-** This kind of ambiguity occurs when the meaning of the words themselves can be misinterpreted.
- In other words, semantic ambiguity happens when a sentence contains an ambiguous word or phrase.
- For example, the sentence “The car hit the pole while it was moving” is having semantic ambiguity because the interpretations can be “The car, while moving, hit the pole” and “The car hit the pole while the pole was moving”.
- **4. Anaphoric Ambiguity-** This kind of ambiguity arises due to the use of anaphora entities in discourse.
- For example, the horse ran up the hill. It was very steep. It soon got tired. Here, the anaphoric reference of “it” in two situations cause ambiguity.

- **5. Pragmatic ambiguity-**Such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations.
- In simple words, we can say that pragmatic ambiguity arises when the statement is not specific.
- For example, the sentence “I like you too” can have multiple interpretations like I like you (just like you like me), I like you (just like someone else dose).

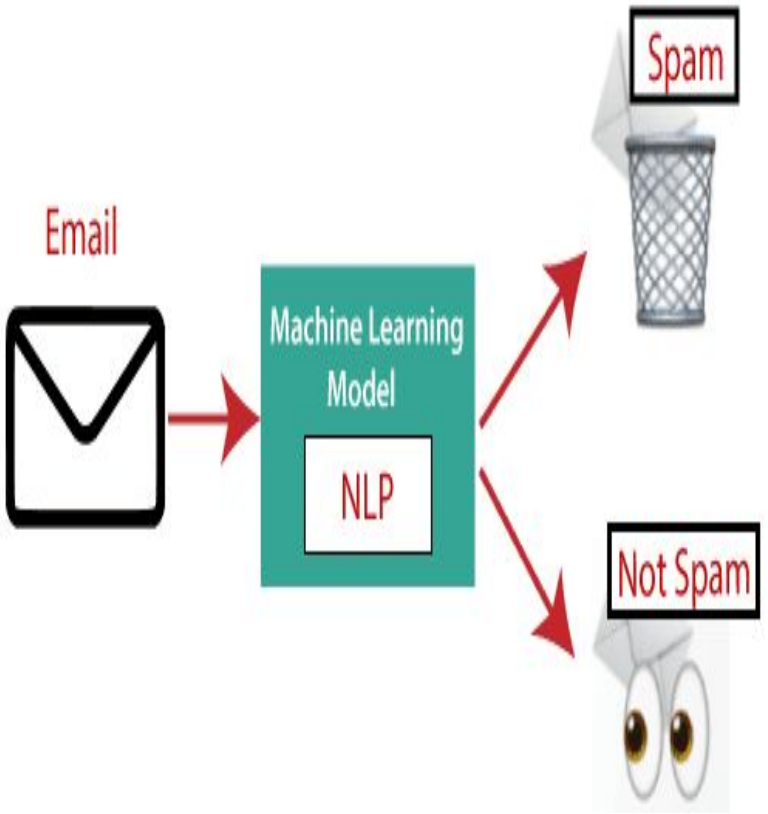
CHALLEGES OF NLP

- Language Differences
- Training Data
- Development Time
- Phrasing Ambiguities
- Missplelling
- Words with Multiple Meaning
- Phrases with Multiple Intensions

APPLICATIONS OF NLP

- **1. Question Answering**
- Question Answering focuses on building systems that automatically answer the questions asked by humans in a natural language.

- **2. Spam Detection**
- Spam detection is used to detect unwanted e-mails getting to a user's inbox.
- We receive all kinds of emails from various sources; some are work-related , while others are spam or promotional emails



- **3. Sentiment Analysis**
- Sentiment Analysis is also known as opinion mining.
- It is used on the web to analyse the attitude, behaviour, and emotional state of the sender.
- This application is implemented through a combination of NLP.
- It also identify the mood of the context (happy, sad, angry, etc.)

- **4. Machine Translation**
- Machine translation is used to translate text or speech from one natural language to another natural language.
- Example:
- Google Translator

- **5. Spelling and Grammmer Correction**
- Microsoft Corporation provides word processor software like MS-word, PowerPoint for the spelling and grammer correction.

- **6. Speech Recognition-**
- Speech recognition is used for converting spoken words into text.
- It is used in applications, such as mobile, home automation, video recovery, dictating to Microsoft Word, voice biometrics, voice user interface, and so on.
- **7. Chatbots**
- With the increase in technology, everything has been digitalized, from studying to shopping, booking tickets, and customer service.
- Instead of waiting a long time to get some short and instant answers, the chatbot replies instantly and accurately.
- NLP gives these chatbots conversational capabilities

- **Advantages of NLP**
- NLP helps users to ask questions about any subject and get a direct response within seconds.
- NLP offers exact answers to the question means it does not offer unnecessary and unwanted information.
- NLP helps computers to communicate with humans in their languages.
- It is very time efficient.
- Most of the companies use NLP to improve the efficiency of documentation processes, accuracy of documentation, and identify the information from large databases.

- **Disadvantages of NLP**
- NLP may not show context.
- NLP is unpredictable
- NLP may require more keystrokes.
- NLP is unable to adapt to the new domain, and it has a limited function that's why NLP is built for a single and specific task only.