

Unit 7:

Leveraging Linguistics and Computer Science

1. Understanding Human Language Complexity

Linguistics is a field of study that focuses on the strategic application of linguistic principles and practices to meet specific objectives or goals.

The field of linguistics is often used in marketing, advertising, communications, education, natural language processing, etc.

- Understanding how language works (including its structure, semantics and pragmatics, as well as sociolinguistics) allows individuals and organizations to tailor their messages, improve communication and influence behavior.
- Machines require systems that research scientists call Natural Language Processing, or NLP, to understand human language.

- NLP is a branch of Artificial Intelligence (AI) that allows computers to understand, create, and manipulate human speech. NLP has the capability to query the data with natural-language text or voice. It is also known as “language in”.
- For example, NLP is at the core of the technology behind virtual assistants like ODA, Siri, Cortana, Alexa, etc. NLP can be applied to both written text and speech. Some examples of tools that are powered by NLP are: Web search, Email spam filtering, Auto-translate text or speech Document summarization Sentiment analysis Grammar/spell checking.

2. Introduction to Natural Language Processing

In NLP, machines segment sentences and extract meaning from “tokens” of human language.

- Human language is unstructured.
- Although it is loosely held together by rules of grammar, our language expresses information in many confusing ways.
- Unlike structured information, which can be arranged in tables or matrices with neatly labeled rows and columns, unstructured information is messy and difficult to understand.

To see why, consider this famous joke by Groucho Marx.

“One morning I shot an elephant in my pajamas. How he got in my pajamas, I don’t know.”

-Adapted from Groucho Marx, 20th century comedian and movie star

- To deal with the “messiness” of unstructured information, computers begin with one sentence at a time. This is called sentence segmentation.
- Computers then break the information into small chunks of information, called tokens, that can be individually classified.
- Once the tokens in text have been sorted into a structure based on what they mean, NLP can work with them.

The following activities show you how Groucho Marx's joke can be tokenized into useful categories called entities and relationships.

- An entity is a noun representing a person, place, or thing. It is not an adjective, verb, or other article of speech.

Activity 2:

Now keeping this in mind, identify the entities in the sentence “**I shot an elephant in my pajamas.**”

Answer: I, elephant, pajamas

- A relationship is a group of two or more entities that have a strong connection to one another.

Activity 3:

Recall Groucho Marx's quote:

"One morning, I shot an elephant in my pajamas. How he got my pajamas, I don't know."

Now keeping this in mind, identify the relationship between the entities in Groucho Marx's statement:

I + elephant, I + pajamas, in + pajamas, elephant + pajamas, I + shot

Relationship between two entities

I + elephant, I + pajamas

elephant + pajamas

Not a relationship between two entities

I + shot

in + pajamas

- Once an AI has classified entities and relationships in text or speech, the AI can begin structuring the information as a step toward understanding it.
- For example, consider the following two sentences: “Manu broke the toy. He always breaks toys.” Notice that there is a relationship between the two sentences: the word he is related to the word Manu.

- A **concept** is something implied in a sentence but not actually stated. This is trickier because it involves matching ideas rather than the specific words present in the sentence.

Activity 4:

Recall, once more, Groucho Marx's statement:

"One morning, I shot an elephant in my pajamas. How he got my pajamas, I don't know."

Based on this statement, select the words that are concepts implied, but not stated.

- ☐ Safari
- ☐ Rifle
- ☐ Photographed
- ☐ Pajamas

Answers: Rifle, Photographed, Safari

Emotion Detection and Sentiment Analysis



Aspect	Emotion Detection	Sentiment Analysis
Definition	Identifies distinct human emotion types.	Measures the strength of an emotion.
Examples	Determining if an expression is anger, happiness, etc.	Assessing if data is positive, negative, or neutral.
Use Cases	Analyzing user ratings, comments in surveys, etc.	Reading social media posts, customer service chats, etc.
AI Training	Can be trained to classify emotions.	Utilizes a sliding scale between positive and negative.
Purpose	Identifying emotional tokens to understand context.	Assessing the overall tone or sentiment of text data.

Classification Problem

- Here is an old-fashioned riddle: Why does your nose run and your feet smell?
- Human language is full of terms that are vague or have double meanings. This is called a classification problem.
- Similarly, in everyday language, phrases like "shipping a box by train" or "filling in a form by filling it out" may seem contradictory or confusing due to the double meanings of the words used. While humans can quickly grasp the intended meaning based on context, AI systems may struggle to accurately classify such phrases without a comprehensive understanding of language nuances and context.

To address this problem,

- An AI system utilizes machine learning techniques such as supervised learning.
- By feeding the system with a large dataset containing examples of language usage and their corresponding classifications, the AI system learns patterns and relationships between words, phrases, and their meanings.
- Over time and with exposure to more data, the AI system improves its classification accuracy by adjusting its internal parameters based on the observed patterns.
- AI systems may not achieve perfect classification accuracy. There will always be some degree of uncertainty or error associated with the system's classifications. To address this, well-designed AI systems not only provide a response but also a confidence value, indicating the system's level of certainty in its classification.

Chatbots

1. Imagine you are playing a game with a friend where you can only give pre-programmed responses. What are some challenges you might face in communicating with your friend?
2. Have you ever interacted with a voice assistant or messaging bot? What did you find interesting or challenging about your experience?

- Chatbots are software applications or computer programs designed to simulate conversation with human users, typically through text-based or voice-based interactions.
- They use artificial intelligence (AI), natural language processing (NLP), and machine learning techniques to understand user queries and provide appropriate responses.
- Chatbots can be integrated into various platforms such as websites, messaging apps, and voice assistants, and they serve a wide range of purposes including customer support, information retrieval, task automation, and entertainment.

- Chatbots, utilized in various fields from retail to healthcare, offer constant availability for online engagement.
- While not always capable of providing answers, they serve to listen and respond to repetitive inquiries, relieving businesses from the need to allocate human resources for such tasks.
- Chatbots operate effectively with small-scale data, catering to specific needs like movie queries for a cinema chain or broader inquiries for AI systems scanning social media.

Chatbots	Rule-based Chatbots	AI-powered Chatbots
Description	Operate on predefined rules and decision trees. Follow programmed rules to respond to user input.	Utilize natural language processing (NLP) and machine learning algorithms. Also known as chat agents or virtual assistants.
Advantages	<ul style="list-style-type: none"> - Easy to develop and maintain. - Provide consistent and accurate answers to specific questions. 	<ul style="list-style-type: none"> - 24/7 availability for immediate and consistent support. - Offer personalized interactions based on user preferences and history. - Improve efficiency and cost savings by automating tasks and reducing service costs.
Limitations	<ul style="list-style-type: none"> - Struggle with understanding complex language. - Unable to adapt to situations beyond programmed rules. 	<ul style="list-style-type: none"> - High development costs and resource requirements. - Prone to biases from training data and lack of transparency in decision-making. - Ethical considerations regarding privacy, manipulation, and responsible use.
Use Cases	<ul style="list-style-type: none"> - Customer service tasks like answering common questions and 	<ul style="list-style-type: none"> - Entertainment and Gaming: Engage users with interactive stories and personalized gaming experiences.

Chatbots	Rule-based Chatbots	AI-powered Chatbots
	providing order updates. - Guiding users through specific processes.	- Finance and Banking: Answer queries about accounts, transactions, and financial products, and process simple requests.

Structure of a chatbot

- A chatbot has a “frontend” and a “backend”.
- The frontend of a chatbot serves as the messaging channel through which users interact, providing a user-friendly interface. However, one limitation of the frontend is that it may lack contextual understanding, meaning it might struggle to grasp the full meaning or context of user messages beyond the immediate input.
- The backend of a chatbot is where the hard work takes place. The backend operates application logic and has enough memory to remember earlier parts of a conversation as dialog continues.

- Chatbots understand a question by breaking it into parts and relating those parts to things in its memory. A chatbot's goal is to identify entities and intents, then use what it is found to trigger a dialog.

Intent

- An intent is a purpose: the reason why a user is contacting the chatbot.
- Think of it as something like a verb: a kind of action. Users may have various intents when interacting with a chatbot, such as filing a complaint, asking for directions, or speaking to a salesperson.
- Institutions often have multiple intents that they want their chatbots to address.

Intent	Possible user inputs
Open	When do you open?
	What are your hours?
	You open now?
	How late are you open?
	Can I walk in at 7 pm?

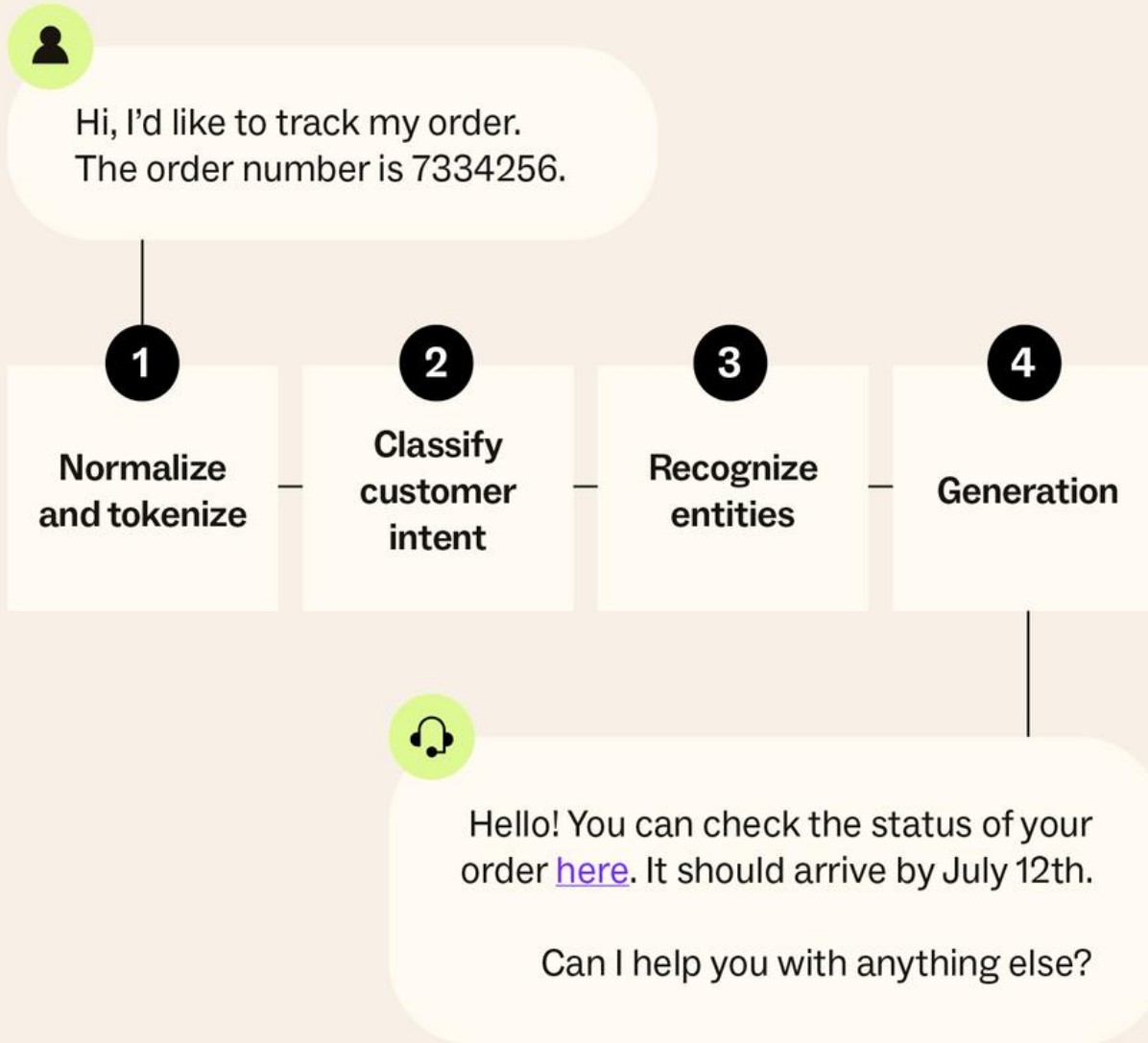
Entity

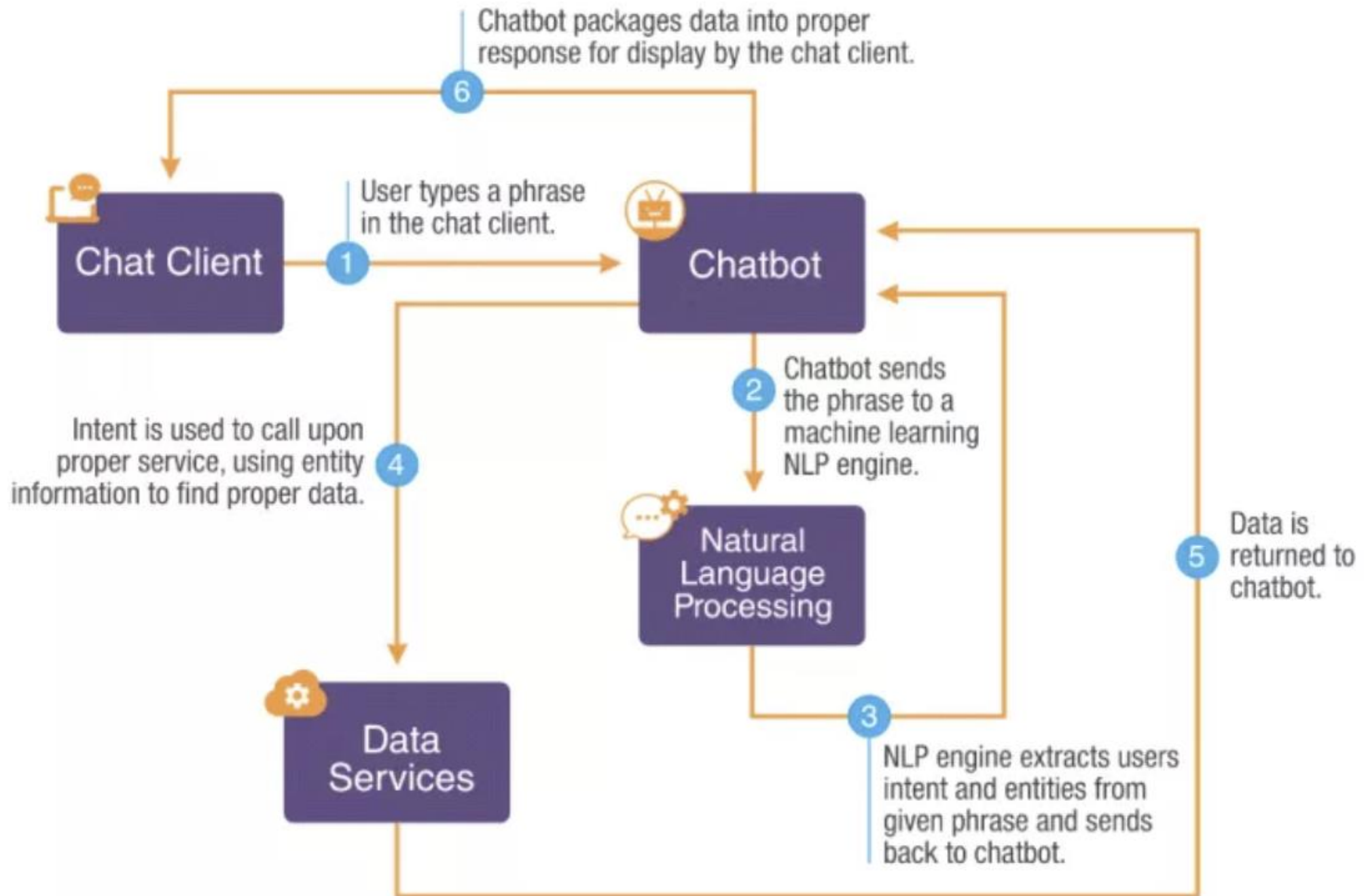
- An entity is a noun: a person, place, or thing. Once you have a list of the intents you want your chatbot to fulfill, you are ready to continue.
- If a user asks, “What are the hours for the Bangalore office?”, then providing business hours is the intent and Bangalore is the entity.
- A chatbot needs a full list of entities to be helpful.

The following table lists examples of entities that map to the intent and possible user inputs of the previous restaurant chain example.

Intent	Possible user inputs	Entities
Open	When do you open?	Bangalore
	What are your hours?	Schedule
	You open now?	Time
	How late are you open??	Time
	And so forth; there are many inputs mapping to this intent.	And so forth; there are many entities implied by this intent.

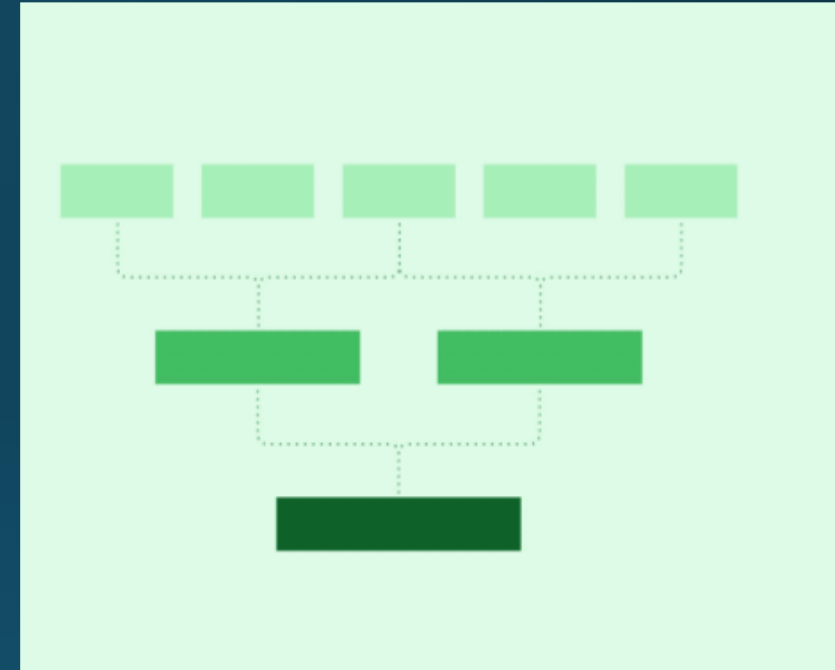
How NLP works





Dialog

- A dialog is a flowchart—an IF / THEN tree structure that illustrates how a machine will respond to user intents.
- A dialog is what the machine replies after a human asks a question.
- Even if a human uses run-on sentences, poor grammar, chat messaging expressions, and so on, artificial intelligence allows the NLP to understand well enough to provide a response.



- The dialog represents each possible word or phrase a user might enter, the matched response for the chatbot, and the many possible subsequent replies a user might make next.
- That is too much for an ordinary flowchart to show (you might need three or four dimensions!), so chatbot software condenses each moment of the conversation into a node.
- A node contains a statement by the chatbot and a long, expandable list of possible replies.

- Creating the flowchart for a chatbot's responses entails a comprehensive process.
- Each potential user input following the chatbot's initial greeting must be assigned a corresponding reply.
- For instance, in the case of inquiries about a restaurant's operating hours, all conceivable questions would trigger a single response.
- This mapping of a vast array of possible questions to a limited number of responses persists throughout the conversation until its conclusion.

- Natural Language Processing
 - –
- Converting Speech to Text & analysing its intent

- Natural language processing (NLP) involves a series of five phases that enable machines to analyse, categorize, and understand both spoken and written language.
- These steps utilize deep neural network-style machine learning techniques to mimic the brain's ability to process data accurately.
- Businesses leverage tools and algorithms aligned with these NLP stages to extract insights from vast amounts of data and facilitate informed decision-making.
- Effective comprehension of these NLP stages is essential for optimizing the use of NLP in text and voice applications.

1. Lexical analysis

- This step involves understanding and examining the structure of words in a language.
- It breaks down the text into paragraphs, phrases, and words.
- Lexical normalization techniques like stemming and lemmatization are commonly used to reduce words to their base forms.
- Stemming reduces words to their root form, such as removing suffixes like "ing", "ly", "es", and "s".
- Lemmatization reduces words to their dictionary form, considering factors like parts of speech (POS) to determine their meaning in context.

2. Syntactical Analysis

- Syntactic Analysis is used to check grammar, word layouts, and word relationships.
- Example: Mumbai travels to the Anuj.
- The line "Mumbai travels to Anuj" makes no sense, hence it is rejected by the Syntactic Analyzer.
- Syntactical parsing is the analysis of words in a sentence for grammar.
- Dependency Grammar and Part of Speech (POS) tags are significant syntactic elements.

3. Semantic Analysis

- Semantic analysis aims to understand the various meanings conveyed by a sentence in a clear and contextually appropriate manner.
- It extracts relevant insights from the text to comprehend its intended message.

3. Semantic Analysis

- Semantic analysis aims to understand the various meanings conveyed by a sentence in a clear and contextually appropriate manner.
- It extracts relevant insights from the text to comprehend its intended message.

4. Discourse Integration

- This involves understanding the context of a statement or word based on preceding sentences or words.
- It helps interpret references like pronouns and proper nouns by identifying their connections with earlier parts of the conversation.
- **Example- Arti wants it.**
- We can observe from the following sentence that the "it" keyword makes no sense. In reality, it applies to anything we don't know.
- That is all this "it" word depends on the prior sentence, which is not provided.
- So, if we know what "it" is, we can simply find the reference.

5. Pragmatic Analysis

- It denotes the study of meanings in a particular language.
- Process of extracting insights from a text. It involves verbal repetition, such as "who said what to whom?" It recognizes how individuals communicate with one another, the context in which they are speaking, and many other factors.

Applications of NLP

Businesses can use natural language processing tools to analyse data, discover insights, automate time-consuming operations, and gain a competitive advantage.

- **Sentiment Analysis:** Natural language processing (NLP) may evaluate consumer comments, social media posts, product reviews, and other text data to determine the sentiment (positive, negative, or neutral) associated with a specific brand, product or service. This data aids firms in understanding client impressions and sentiments.

- **Voice Assistants:** Voice assistants are becoming increasingly popular! Whether it's Siri, Alexa, or Google Assistant, practically everyone uses one to make calls, create reminders, plan meetings, set alarms, browse the internet, and so on. How do they work? They use a complicated blend of voice recognition, natural language interpretation, and natural language processing to understand and respond to what humans say. The long-term purpose of voice assistants is to serve as a bridge between humans and the internet, providing a variety of services through simple speech interaction. However, they are still a long way from achieving that aim, as Siri occasionally fails to grasp what you are saying!

- **Email Filtering:** Email is a part of our daily lives. We find ourselves bombarded with emails about job, study, and a variety of other topics. We receive emails from a variety of sources; some are work-related or from our dream school or institution, while others are spam or promotional in nature. Here, Natural Language Processing comes into play. It classifies incoming emails as "important" or "spam" and assigns them accordingly.

- **Document Analysis:** Document analysis is another use of natural language processing. Companies, institutions, and schools, among other places, are constantly inundated with data that must be properly organized, stored, and searched. All of this may be accomplished with NLP. It not only searches a keyword but also categorizes it according to the instructions, saving us from the tedious and time-consuming task of searching for a single person's information from a large number of files. It is not only restricted to this, but it also assists users in making informed decisions about claims and risk management.

- **Automatic Summarization:** Data has grown in line with technological advancements. This rise of data has broadened the scope of data processing. Still, manual data processing is time-consuming and error-prone. NLP provides a solution for this as well; it can not only summarize the meaning of information but also identify the emotional meaning hiding within it. As a result, the summary process becomes more efficient and precise.

THANK YOU