NL for Communication & Perception

Natural Language for communication: Phrase structure Grammars, syntactic Analysis, Augmented grammers and semantic Interpretation, Machine Translation, Speech Recognition.

Perception: Image Formation, Early Image Processing Operations, Object Recognition by appearance, Reconstructing the 3D world, Object recognition from structural Information, Using Vision.

---

1. Natural Language for Communication:-

Natural language is a form of communication used by humans to convey information and express thoughts and ideas.

It refers to the way people typically speak and write in their everyday interactions.

Natural language encompasses various components, including grammar, vocabulary, syntax, semantics and pragmatics.

Natural language allows people to convey complex concepts, ask questions, make requests, express emotions, engage in debates and much more.

NL for communication in AI refers to the ability of AI systems to understand and generate language like human to facilitate communication b/w machines and humans.

There are two main aspects of Natural language for communication in AI

(i) NLU - Natural language Understanding.

NLU is the process by which AI systems analyze and interpret human language input. It involves various tasks such as

(a) Text Understanding:
Extracting meaning, entities, and relationships from text.

(b) Speech Recognition:
Converting spoken language into written text.

(c) Intent Recognition:
Identifying the intent behind a user's input (what actions the user wants).

(d) Sentiment Analysis:
Determining the emotional tone of the text. whether +ve, -ve or neutral.

(e) NER (Named Entity Recognition):
Identifying and classifying entities such as names of places, peoples organizations etc.

(ii) NLG - Natural Language Generation.

NLG is the process by which AI systems produce human language output.

It involves tasks such as

(a) Text Generation:
creating coherent and contextually appropriate responses to user queries.

2. Phrase Structure Grammar :-

PSG is a type of syntactic grammar used in linguistic to describe the structure of sentences and phrases in a language.

It is also known as Constituency Grammar or Tree Grammar because it represents the hierarchical structure of a sentence using tree diagrams.

In PSG, a sentence is divided into smaller units called constituents, which can be further broken down into other constituents until the basic grammatical words are reached.

The key elements in Phrase structure Grammar are :

Phrase :-

A group of words functioning as a single unit within a sentence. Phrases can be nouns, verbs, adjectives, adverbs etc...

Eg :- the big brown dog, quickly ran.

Constituents :-

The smaller units that make up phrases. Constituents are either single words or other phrases.

Eg :- She is reading a book.

The constituents are (She, is reading, a book.

Hierarchical Structure :-

PSG represents the relationship b/w constituents in a hierarchical manner.

Each level of hierarchy is called a "node", and the nodes are connected in a tree-like structure.
Sentence forming the highest level, individual words forming the lowest level.

Phrase Rules:-

PSG uses rules to describe how phrases can be combined to form larger constituents.

S → NP VP

Sentence (S) is composed of NP - Noun phrase and VP - verb phrase.

Ex:- The cat is sleeping on the mat.

Det	N	V	V	P	Det	N
NP	NP	VP	VP	PP	NP	NP

PSG forms the basis for various natural language processing tasks, such as parsing, sentence generation, and machine translation.

Syntactic Analysis:-

There are number of algorithms researchers have developed for syntactic analysis, but we consider the following simple methods.

→ Context free grammar
→ Top down parser.

Context Free Grammar:-

It is the grammar that consists rules with

single symbol on the left hand side of the rewrite rules.

Ex:- The bird pecks the grains.

Articles (DET) = a, an, the

Nouns = bird, grains.

NP-Noun Phrase = Article + Noun | Article + adj + N

= Det. N | Det Adj N.

Verbs = pecks

VP-Verb Phrase = NP V | V NP

Adjectives = beautiful, small ...

The parse tree breaks down the sentence into structured parts so that the computer can easily understand and process it.

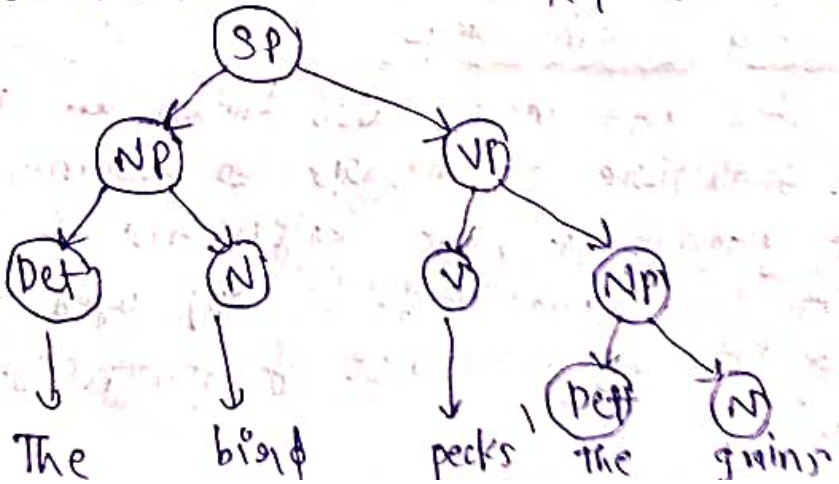
A/c to first order logic rule, if there are 2 strings NP and VP, then the string combined by NP followed by VP is a sentence.

$S \rightarrow NP\ VP$

$NP \rightarrow Det\ N \mid Det\ adj\ N$

$VP \rightarrow V\ NP$

The parse tree can be created as shown.



Top down parser :-

The top-down parsing uses the left-most derivation approach to construct the parsing tree of the text.

The top-down parser in NLP does not support the grammar having common prefixes.

We can perform the top-down parsing using two ways.

→ Using a backtracking
→ Without using backtracking.

The recursive descent parser in NLP follows the top-down parsing approach. This parser checks the syntax of the input stream of text by reading it from left to right. Hence, it is also known as left-Right parser.

---

3. Augmented Grammars and Semantic Interpretation

Augmented grammars and semantic interpretation are concepts used in NLP and computational linguistics to improve the understanding of human language by machines.

Augmented Grammars :-

Augmented grammars also known as augmented phase structure grammars & augmented context free grammars, are extensions of traditional context free grammars (CFGs) used to describe the syntax & structure of sentences in natural language.

CFGs are widely used to represent the hierarchical structure of sentences, but they have limitations in capturing certain linguistic phenomena like long distance dependencies.

Augmented grammars adds some of these limitations by introducing additional features.

One common extension is the use of feature structure, another approach is the use of augmented transition networks.

These augmented grammars are used as a foundation for natural language parsers, which analyze sentences and derive their syntactic structure based on the grammar rules.

Semantic Interpretation:-

Semantic interpretation is the process of assigning meaning to sentences or phrases in natural language.

While syntax deals with the structure of sentences, semantics focuses on the meaning of words.

Natural language is inherently ambiguous, and the same sentence can have different interpretations based on the content.

Semantic Interpretation aims to disambiguate the meaning of sentences and understand the intended meaning of the speaker or writer.

It involves mapping words and phrases to their corresponding concepts or entities in the real world and capturing the relationships b/w them.

There are various approaches to semantic interpretation, including rule based systems, statistical methods, and NL techniques.

Semantic interpretation is crucial for many NLP applications, such as question answering, sentiment analysis, information retrieval, and machine translation.

---

4. Machine Translation:

Machine translation is the process of using AI that automatically translate text or speech from one language to another without any human input.

The goal of machine translation is to enable communication and understanding b/w people who speak different languages.

There are several approaches to machine translation, the 3 most common types of machine translation include:

RBMT:

RBMT - Rule Based Machine Translation.
RBMT relies on a set of linguistic rules and grammar to perform translation. Linguistics and language experts manually create these rules, which are used to analyze the source text and generate the corresponding translation.

SMT:

SMT - Statistical Machine Translation.

Statistical MT builds a statistical model of the relationships b/w words, phrases and sentences in a text.
It applies the model to a second language to convert those elements to the new language.
The models use probabilities to determine the most likely translation for a given sentence or phrase.

NMT:-
NMT - Neural Machine Translation.
NMT is the latest and most prominent approach to machine translation.
It relies on deep learning neural networks, specifically seq-to-seq models.
It is more accurate, easier to add languages, and much faster once trained.

The process of MT typically involves the following steps:

(i) Data collection:
Gather large parallel corpora containing texts in both the source and target languages.

(ii) Training:
Use the collected data to train the MT model

(iii) Inference:
During inference, the trained model takes a sentence or text in the source language as input and generates the corresponding translation in the target language.

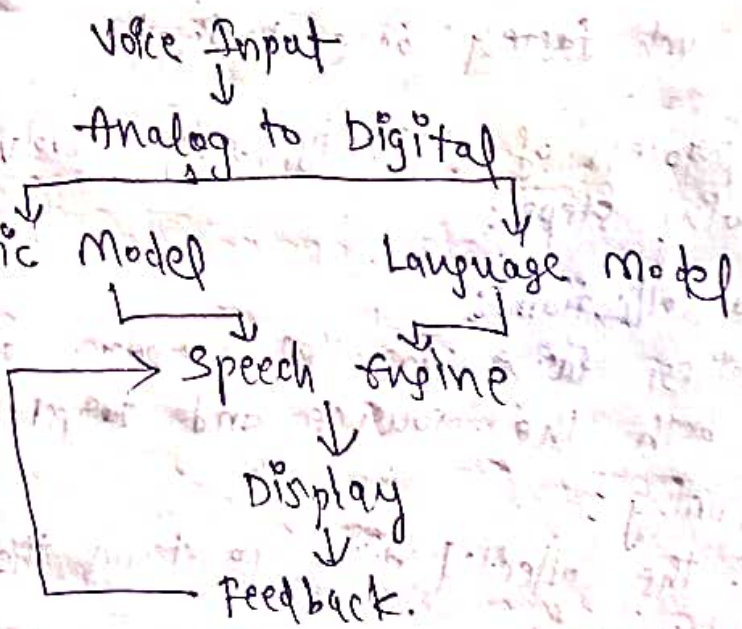
Machine translation is widely used in various applications, such as multilingual customer support, cross-border communication, website translation, and language learning.

5. Speech Recognition :-

Speech recognition also known as ASR - Automatic Speech Recognition (or STT - Speech To Text) is a technology that enables computers to convert spoken language into written text.

It is essential component of NLP and has various applications.

Speech recognition systems allow users to interact ~~systems~~ with devices, applications, and services using spoken language instead of typing.



Uses :-

→ system control navigation.
Eg: GPS connected digital maps.

→ Commercial | Industrial applications in the car steering system.

→ Voice dialling hands-free use of mobile in the car.

The process of speech recognition involves several steps.

(i) Acoustic Signal Processing:
The speech recognition system receives an audio signal as input. The raw audio data is pre-processed to remove background noise and normalize the volume.

(ii) Feature Extraction.
From the pre-processed audio, relevant features are extracted to represent the speech signal.

(iii) Acoustic Modeling:
Acoustic models are used to match the extracted features to phonetic units.

(iv) Language Modeling:
Language models are used to estimate the likelihood of word sequences in the given language.

(v) Decoding:
The speech recognition system uses the acoustic and language models to decode the most likely sequence of words that best match the input speech.

(vi) Post-processing.
The output of the decoding process may undergo post processing to correct errors and improve the accuracy of the recognized text.

Types of SR :-

Speaker Dependent :-

Those individuals who will be using the system train the speaker dependent systems. These systems are capable of achieving a high better command count than 95% accuracy for word recognition.

Speaker Independent :-

Speaker Independent is a system trained to respond to a word regardless of who is speaking.

The command word count is generally lower than the speaker dependent, where as high accuracy can still be maintained within processing limits.

Applications :-

→ Voice assistants like Siri, Google Assistant, Alexa, and Cortona. etc.
→ Speech to text conversion for transcription services.
→ Voice command systems for hands free operation of devices.
→ Voice based search and navigation systems
→ Voice authentication and security applications.

1. Perception :-

Perception is a process to interpret, acquire, select and then organize the sensory information that is captured from the real world.

Eg :- Human beings have sensory receptors such as touch, taste, smell, sight and hearing. So, the information received from these receptors is transmitted to human brain to organize the received information.

A/c to the received information, action is taken by interacting with the environment.

Perception and action are very important concepts in the field of Robotics.

There is important difference b/w the AI ~~and~~ program and robot. The AI program performs in a computer stimulated environment, while the robot performs in the physical world.

The process of perception can be described in several stages.

(i) Sensation :
Where sensory receptors detect and respond to environmental stimuli. These stimuli are converted into electrical signals that are sent to the brain for processing.

(ii) Selection :
The brain filters and selects relevant sensory

information from the vast amount of incoming stimuli.

(iii) Organization:
The selected sensory information is organized and grouped into meaningful patterns and structures. The ~~brain uses previous experie~~

(iv) Interpretation:
Once the information is organized, the brain interprets its meaning and significance.

(v) Perceptual Constancy:
Despite changes in sensory input, the brain maintains a consistent perception of objects and the environment.

(vi) Perceptual Illusions:
These are situations where our brain misinterprets sensory input, leading us to perceive something differently from reality.

---

2. Image formation:

Image formation in a digital camera is the process by which the camera captures light from the scene and converts it into a digital image that can be stored and displayed.

This process involves several steps, including light-gathering, focusing, and converting analog signals to digital data.

Light Gathering:
The camera's lens collects and focuses light onto the camera's sensor.

Lens Focus:-
The lens of the camera adjusts its shape to focus light rays onto the camera's image sensor.

Image Sensor:-
It is an electronic device that captures the light and converts it into an electrical signal. There are 2 common types of image sensors used in digital cameras.
→ CCD - Charge Coupled Device
→ CMOS - Complementary Metal-Oxide Semiconductor.

Photo Detection:-
When light strikes the pixels on the image sensor, the pixels generate electrical charges proportional to the amount of light received.

Analog to Digital Conversion:-
An analog-to-digital converter (ADC) samples the analog signal and assigns digital values to represent the pixel's brightness and color intensity.

Image Processing:-
The camera image processor processes the digital data. Image processing involves tasks such as white balance adjustment, color correction, noise reduction, and compression to create the final digital image file.

Image Storage:-
The processed digital image is then stored in a

memory card within the camera. The images can be stored in various file formats such as JPEG, RAW and others.

Display & Output:-

The stored digital image can be displayed on the camera's LCD screen.

---

3. Early Image Processing Operations:-

Early image processing operations refer to the basic techniques and methods that were developed in the early days of digital image processing.

Some of the early image processing operations include:

(a) Image Enhancement:

Image enhancement techniques aim to improve the visual quality of an image for better perception or analysis. (Contrast stretching, histogram equalization etc).

Image Filtering:

Filtering operations involve applying convolution kernels or masks to an image to highlight certain image features.

Ex:- Low pass filtering for noise reduction.
high pass filtering for edge detection.

Image Restoration:-

Image restoration techniques aim to recover the original image from a degraded version. Early methods used basic filters and deblurring techniques to restore images.

Geometric Transformations:-

Geometric transformations include operations like image rotation, scaling, and translation.

Thresholding:-

It is used to convert a grayscale image into a binary image.

Edge-detection:-

Edge detection techniques aim to identify the boundaries b/w different regions in an image.

Morphological Operations:

Morphological operations involve processing binary images using operations like dilation and erosion.

Color Space Conversion:-

Basic color space conversion techniques, such as converting RGB images to grayscale or other color representations.

Image Compression:

Simple methods like run-length encoding and Huffman coding to reduce the storage space required for images.

Object Recognition by Appearance:-

Object recognition by appearance is a computer vision task that involves identifying and classifying objects in images or videos based on their visual appearance.

The goal is to teach a computer system to recognize objects in a similar way that humans do by their unique visual features and patterns.

Here are the key steps involved in object recognition by appearance:

Image Acquisition:-

The first step is to acquire images either from a video source. These images serve as input to the object recognition system.

Feature Extraction:

In this step, relevant features are extracted from the input images to represent the visual appearance of objects.

Feature Representation:-

The extracted features are represented in a way that the computer system can process. This representation could be a feature vector, a feature map.

Model Training:

A machine learning or deep learning model is trained on a labeled dataset of images.

Classification:

After training the model can classify new images into different object classes based on the visual appearance features extracted from the images.

Post Processing:-

After classification, post processing techniques may be applied to refine the object recognition results.

Applications:-

Object recognition by appearance finds applications in various domains including:

→ Object detection.
→ Image classification.
→ Semantic segmentation.
→ Instance segmentation.
→ Object Tracking

CNN - Convolutional Neural Networks have demonstrated exceptional performance in various object recognition tasks, ~~and these~~

---

5. Reconstructing the 3D World:-

Reconstructing the 3D world is a computer vision task that aims to create a 3D representation of the real world environment from 2D images or video sequences.

The process of reconstructing the 3D world typically involves several steps including:

Camera Calibration:

Camera calibration involves estimating the intrinsic parameters (focal length, principal point) and extrinsic parameters (position and orientation) of the camera relative to the scene.

Feature Extraction and Matching:

Feature extraction means, distinctive features are extracted from the images (or video frames. These features can be key points, corners (or edges.

Feature matching is performed to find corresponding features across multiple views or frames.

Stereo Vision:

Stereo vision can provide dense depth maps for the scene.

It is a common approach for 3D reconstruction using a pair of calibrated cameras.

SfM-Structure from Motion:

SfM is a technique used when multiple images or frames are available from different view points.

SfM uses feature matching and bundle adjustment algorithms to refine the camera poses and 3D point cloud.

Depth estimation:

Depth estimation techniques can be applied to monocular images (or) videos to estimate the depth of each pixel.

Point cloud Generation:

The 3D points obtained from either stereo vision (or) SfM are used to create a point cloud. which is a collection of 3D points representing the 3D structure of the scene.

Mesh Reconstruction:

A mesh is a surface representation consisting of interconnected triangles.

Reconstructing the 3D world has various applications such as augmented reality, virtual reality, autonomous navigation, 3D modeling, and scene understanding.

---

6. Object Recognition from Structural Information Using Vision:-

Object recognition from structural information using computer vision involves identifying and classifying objects based on their geometric shapes, spatial relationships and structural characteristics.

Here are some key points and techniques involved in object recognition from structural information.

Edge Detection:

Edge detection is a fundamental step in extracting structural information from an image. It involves identifying the boundaries (or edges) of objects and regions in the image.

Ex: Canny Edge Detector, Laplacian of Gaussian (LoG) filter.

Contour Detection:

Once edges are detected, contour detection algorithms are used to group adjacent edges into continuous contours that represents the object shape.

Shape Representation:-

Various methods can be employed to represent the shapes of objects using shape descriptors like Hu moments, zernike moments. These descriptors capture unique structural properties of the objects shape that can be used for recognition.

Graph Based Methods:-

Graph based approaches represent objects as graphs, where nodes represent key points or points on the objects contour, and edges represent spatial relationships b/w these key points.

Template Matching:-

It is another technique used to get the structural information.

It involves comparing a template (object shape) with different regions of the image to find matches.

Geometric Features:-

Geometric features like angles, lengths, and ratios b/w parts of the object can also be used to represent the objects structure.

Spatial Relations:-

Techniques like spatial pyramid matching or spatial relationship reasoning can be employed to capture the relative positions and orientations of multiple objects.