h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

6. Applications of Artificial Intelligence

(Artificial Intelligence and Neural Networks)

Natural Language Processing

Dr. Udaya Raj Dhungana

Assist. Professor

Pokhara University, Nepal

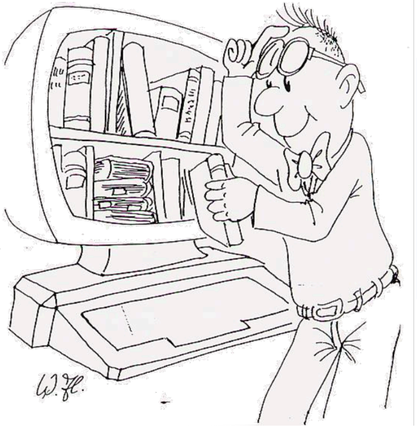
Guest Faculty

Hochschule Darmstadt University of Applied Sciences, Germany

E-mail: udaya@pu.edu.np and udayas.epost@gmail.com

Overview

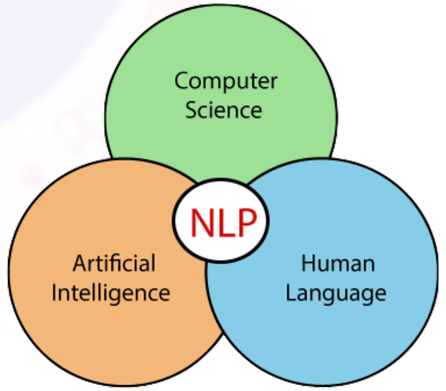
- NLP Levels of analysis
  - Phonetic
  - Syntactic
  - Semantic
  - Pragmatic
- Machine Vision



Natural Language Processing

- Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language.

- **Natural language processing** (**NLP**) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data.



Natural Language Processing

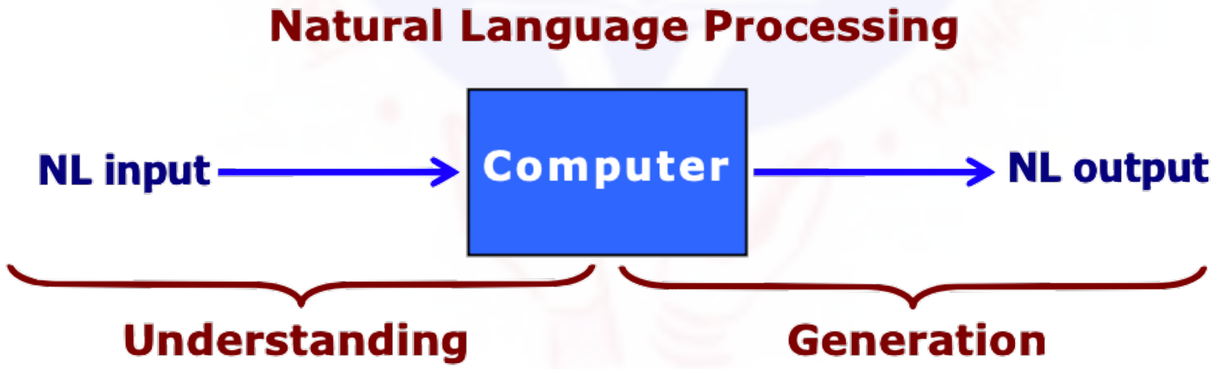
- Natural language processing is an interdisciplinary field / art / science:
  - computer science (A.I.)
  - linguistics (language independent)
  - mathematics (logics, predicate logic, knowledge based systems, statistics, ...)
  - psychology (cognitive science)
  - physics (speech recognition, spoken language)
  - ...

Natural Language Processing

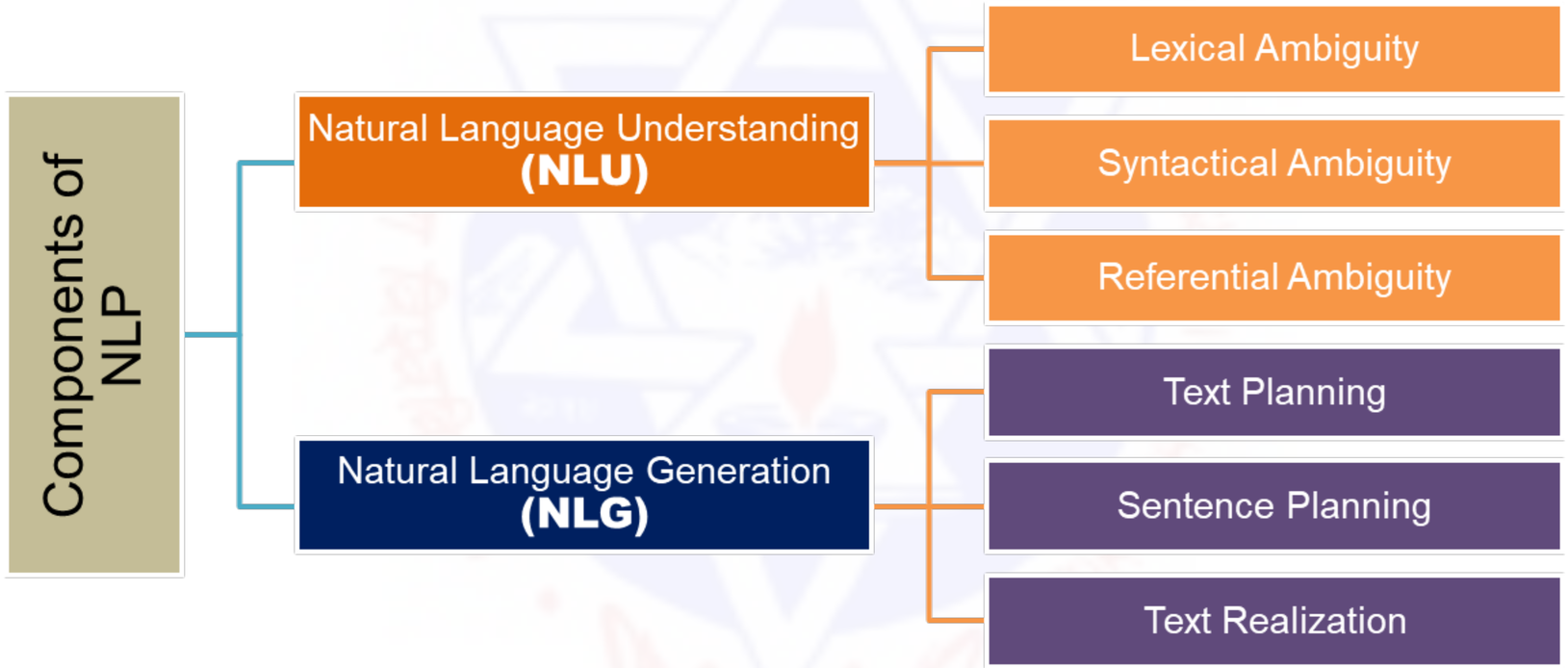
- NLP combines the power of linguistics and computer science to study the rules and structure of language, and create intelligent systems (run on machine learning and NLP algorithms) capable of understanding, analyzing, and extracting meaning from text and speech.
- NLP is used to understand the structure and meaning of human language by analyzing different aspects like syntax, semantics, pragmatics, and morphology. Then, computer science transforms this linguistic knowledge into rule-based, machine learning algorithms that can solve specific problems and perform desired tasks.
- Example: (Gmail) Emails are automatically categorized as *Promotions*, *Social*, *Primary*, or *Spam*, thanks to an NLP task called keyword extraction. By "reading" words in subject lines and associating them with predetermined tags, machines automatically learn which category to assign emails.

Components of NLP

- NLP encompasses anything a computer needs to understand natural language (text or speech) and also generate the natural language.
- There are two components of NLP:
- Natural Language Understanding (NLU):
  - Taking some spoken/typed sentence and working out what it means
- Natural Language Generation (NLG):
  - Taking some formal representation of what you want to say and working out a way to express it in a natural (human) language



Components of NLP



Natural Language Understanding (NLU)

- NLU helps the machine to understand the data.
- It is used to interpret data to understand the meaning of data to be processed accordingly.
- It solves it by understanding the context, semantic, syntax, intent, and sentiment of the text.
- Mapping the given input in the natural language into a useful representation
- For this purpose, various rules, techniques, and models are used.
- There are three linguistic levels to understand language.
    - **Morphological analysis**
    - **Syntax**: It understands sentences and phrases. It checks the grammar and syntax of the text.
    - **Semantic**: It checks the meaning of the text.
    - **Pragmatic**: It understands context to know what the text is trying to achieve.

Natural Language Understanding (NLU)

- Let's take an example for more clarity.
  - If you asked: "How's today ?".
  - Now, what if the system answers, “Today is Feb 1, 2022.”
  - Is the system providing you the correct answer?
  - No, Because here, users want to know about the weather.
  - Therefore, we use NLU to learn the text's right meaning of some errors.

Natural Language Understanding (NLU)

- NLU is naturally harder than NLG tasks.
- Let’s see what are all challenges faced by a machine while understanding.
- There are lot of ambiguity while learning or trying to interpret a language.
    - Lexical Ambiguity
    - Syntactical Ambiguity
    - Referential Ambiguity

Difficulties in NLU

- Lexical Ambiguity: can occur when a word carries different sense, i.e. having more than one meaning and the sentence in which it is contained can be interpreted differently depending on its correct sense. Lexical ambiguity can be resolved to some extent using parts-of-speech tagging techniques.

He is looking for a **match**

What do you understand by 'match'?

Partner

Or **Cricket/Football Match**

Difficulties in NLU

- Syntactical Ambiguity means when we see more than one meaning in a sequence of words. It is also termed as grammatical ambiguity.

The chicken **is ready** to eat.

Is the chicken ready to eat his food or the chicken is ready for someone else to eat ? You never know.

Difficulties in NLU

- Referential Ambiguity: Very often a text mentions as entity (something/someone), and then refers to it again, possibly in a different sentence, using another word. Pronoun causing ambiguiyty when it is not clear which noun it is referring to

Feluda met Topse and Jotayu. They went to restaurant

They refer to Topse and Jotayu or all?

Natural Language Understanding (NLU)

- NLU System would be able to:
  - Paraphrase (express the meaning of) an input text
  - Translate the text into another language
  - Answer questions about the contents of the text
  - Draw inferences from the text

Natural Language Generation (NLG)

- NLG is the process of producing meaningful phrases and sentences in natural language.
- It involves –
  - **Text planning** – It includes retrieving the relevant content from knowledge base.
  - **Sentence planning** – It includes choosing required words, forming meaningful phrases, setting tone of the sentence.
  - **Text Realization** – It is mapping sentence plan into sentence structure.
- Producing output in the natural language from some internal representation
- Different level of synthesis required:
  - deep planning (what to say)
  - syntactic generation

Natural Language Generation (NLG)

NLU	NLP	NLG
NLU is a narrow concept.	NLP is a wider concept.	NLU is a narrow concept.
If we only talk about an understanding text, then NLU is enough.	But if we want more than understanding, such as decision making, then NLP comes into play.	It generates a human-like manner text based on the structured data.
NLU is a subset of NLP.	NLP is a combination of NLU and NLG for conversational Artificial Intelligence problems.	NLU is a subset of NLP.
It is not necessarily that what is written or said is meant to be the same. There can be flaws and mistakes. NLU makes sure that it will infer correct intent and meaning even data is spoken and written with some errors. It is the ability to understand the text.	But, if we talk about NLP, it is about how the machine processes the given data. Such as make decisions, take actions, and respond to the system. It contains the whole End to End process. Every time NLP doesn't need to contain NLU.	NLU generates structured data, but it is not necessarily that the generated text is easy to understand for humans. Thus NLG makes sure that it will be human-understandable.
It reads data and converts it to structured data.	NLP converts unstructured data to structured data.	NLG writes structured data.

NLP- Terminologies

- **Phonology** – It is study of organizing sound systematically.
- **Morphology** – It is a study of construction of words from primitive meaningful units.
- **Morpheme** – It is primitive unit of meaning in a language.
- **Syntax** – It refers to arranging words to make a sentence. It also involves determining the structural role of words in the sentence and in phrases.
- **Semantics** – It is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.
- **Pragmatics** – It deals with using and understanding sentences in different situations and how the interpretation of the sentence is affected.
- **Discourse** – It deals with how the immediately preceding sentence can affect the interpretation of the next sentence.
- **World Knowledge** – It includes the general knowledge about the world.

Linguistic and Language Processing

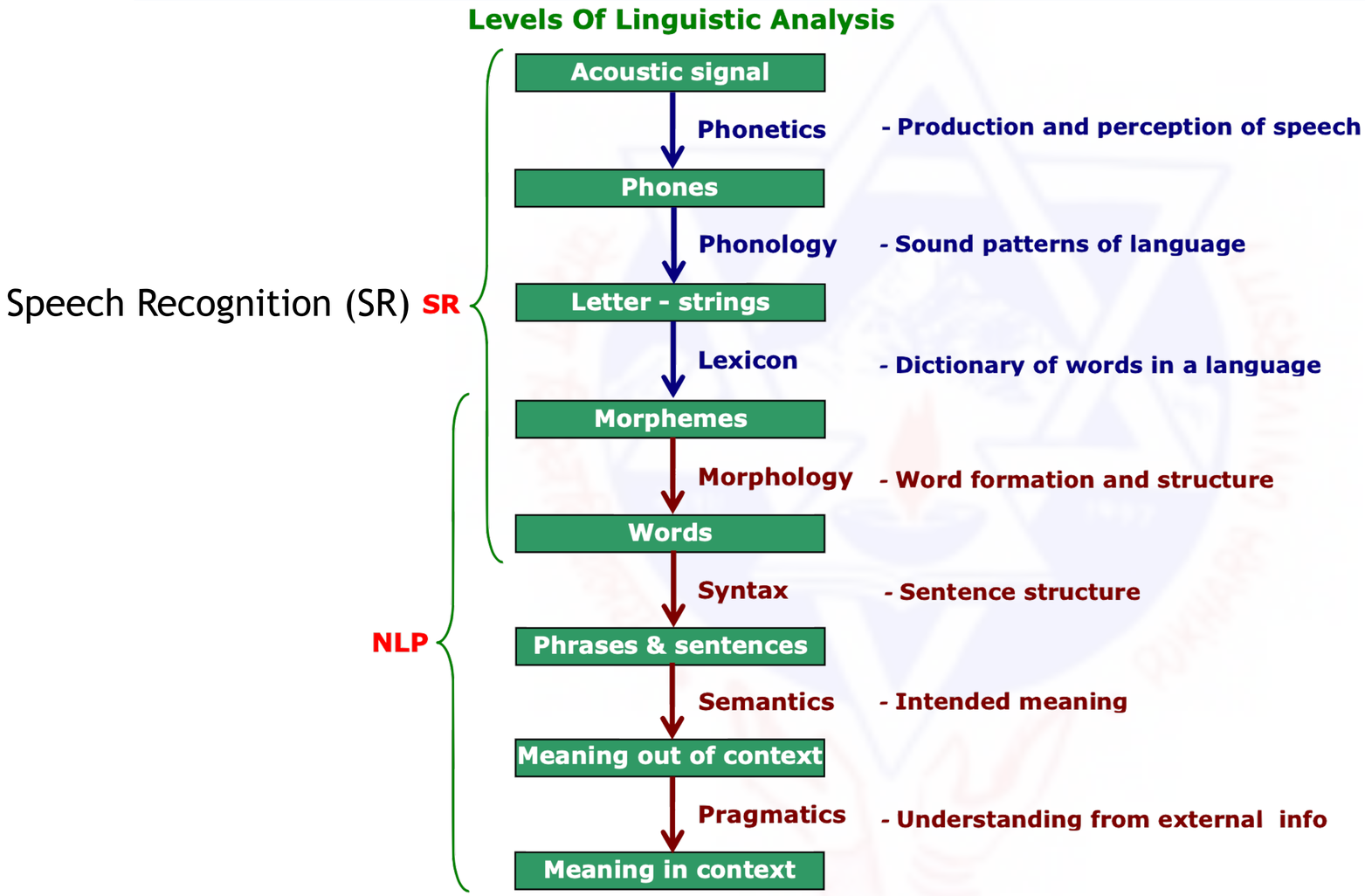
Linguistics is the science of language. Its study includes :

- sounds (phonology),
- word formation (morphology),
- sentence structure (syntax),
- meaning (semantics), and understanding (pragmatics) etc.

The levels of linguistic analysis are shown below.

- higher level corresponds to Speech Recognition (SR)
- lower levels corresponds to Natural Language Processing (NLP).

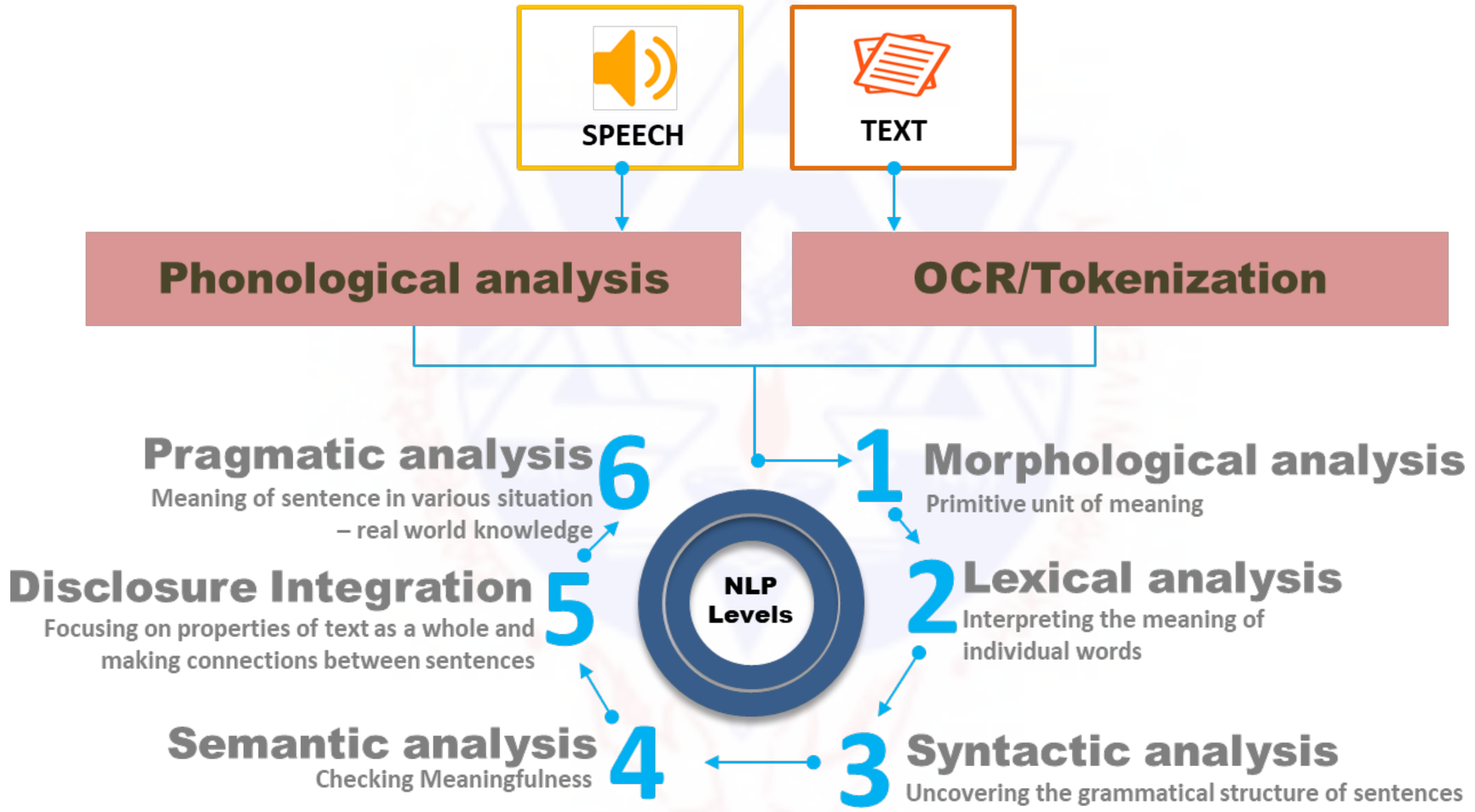
Linguistic and Language Processing



Levels in NLP

- The typical steps involved while performing NLP tasks:
    1. Phonological Analysis
    2. Morphological Analysis
    3. Lexical Analysis
    4. Syntactic Analysis
    5. Semantic Analysis
    6. Discourse Integration
    7. Pragmatic Analysis

Levels in NLP



Levels in NLP

- Phonological Analysis:
  - This level deals with the interpretation of speech sounds within and across words.
  - There are three types of rules used in phonological analysis:
    - 1) phonetic rules – for sounds within words;
    - 2) phonemic rules – for variations of pronunciation when words are spoken together, and;
    - 3) prosodic rules – for fluctuation in stress and intonation across a sentence.
  - In an NLP system that accepts spoken input, the sound waves are analyzed and encoded into a digitized signal for interpretation by various rules or by comparison to the particular language model being utilized.

Levels in NLP

- Morphological Analysis:
  - This level deals with the componential nature of words, which are composed of morphemes – the smallest units of meaning.
  - For example, the word preregistration can be morphologically analyzed into three separate morphemes:
    - the prefix pre,
    - the root registra, and
    - the suffix tion.
  - Since the meaning of each morpheme remains the same across words, humans can break down an unknown word into its constituent morphemes in order to understand its meaning.
  - Similarly, an NLP system can recognize the meaning conveyed by each morpheme in order to gain and represent meaning. For example, adding the suffix –ed to a verb, conveys that the action of the verb took place in the past.

Levels in NLP

- Lexical Analysis:
  - At this level, humans, as well as NLP systems, interpret the meaning of individual words
  - Several types of processing contribute to word-level understanding – the first of these being assignment of a single part-of-speech tag to each word.
  - It involves identifying and analysing the structure of words.
  - Lexicon of a language means the collection of words and phrases in a language.
  - Lexical analysis is dividing the whole chunk of test into paragraphs, sentences, and words.
  - In order to deal with lexical analysis, we often need to perform **Lexicon Normalization**.

Levels in NLP

- Lexical Analysis:
  - **Lexicon Normalization:** Lexical normalization is the task of translating/transforming a non standard text to a standard register.
  - Example: new pix comming tomoroe new pictures coming tomorrow
  - The most common lexicon normalization practices are
    - Stemming: Stemming is a rudimentary rule-based process of stripping the suffixes (“ing”, “ly”, “es”, “s” etc) from a word.
    - Lemmatization: Lemmatization, on the other hand, is an organized & step by step procedure of obtaining the root form of the word, it makes use of vocabulary (dictionary importance of words) and morphological analysis (word structure and grammar relations).

Levels in NLP

- Syntactic Analysis
  - This level focuses on analyzing the words in a sentence so as to uncover the grammatical structure of the sentence.
  - This requires both a grammar and a parser.
  - The output of this level of processing is a (possibly delinearized) representation of the sentence that reveals the structural dependency relationships between the words.
  - Some word sequences may be rejected if they violate the language's rules for how the words can be combined. For example, an English syntactic analyzer would reject the sentence "Boy the go the to store."
  - Syntax conveys meaning in most languages because order and dependency contribute to meaning. For example the two sentences: 'The dog chased the cat.' and 'The cat chased the dog.' differ only in terms of syntax, yet convey quite different meanings

Levels in NLP

- Semantic Analysis:
  - Semantic processing determines the possible meanings of a sentence by focusing on the interactions among word-level meanings in the sentence.
  - The structures created by the syntactic analyzer are assigned the meaning. In other words, a mapping is made between the syntactic structures and objects in the task domain. Structures for which no such mapping is possible may be rejected. For example, “Colorless green ideas sleep furiously” would be rejected as semantically anomalous.
  - This level of processing can include the semantic disambiguation of words with multiple senses

Levels in NLP

- Discourse Integration:
  - While syntax and semantics work with sentence-length units, the discourse level of NLP works with units of text longer than a sentence.
  - The meaning of an individual sentence may depend on the sentences that precede it and may influence the meaning of the sentences that follow it.
  - For example, the word “it” in the sentence, “Ram wanted it.” Depends on the prior discourse context, while the word “Ram” may influence the meaning of later sentence such as “he always plays.”
  - The discourse focuses on the properties of the text as a whole that convey meaning by making connections between component sentences.

Levels in NLP

- Pragmatic Analysis:
  - The structure representing what was said is reinterpreted to determine what was actually meant.
  - The sentence “Do you know what time it is?” should be interpreted as request to be told the time.

Levels in NLP

- Pragmatic Analysis:
  - Pragmatics is the study of language usage from a functional perspective and is concerned with the principles that account for how meaning is communicated by the speaker (writer) and interpreted by the listener (reader) in a certain context.
  - Different from semantics, pragmatics studies the contextual meaning.
  - This distinction can be seen in the following example:
    - **Mike**: What happened to that bowl of cream?
    - **Annie**: Cats drink cream.

Levels in NLP

- Pragmatic Analysis:
  - Pragmatics concentrates on those aspects of meaning that cannot be predicted by linguistic knowledge alone and takes into account our knowledge about the physical and social world.
  - The four utterances in the following dialogue are all syntactically incomplete, but pragmatically they are all "appropriate" in the particular context.
    - **Jane**: Coffee?
    - **Steve**: Sure.
    - **Jane**: White?
    - **Steve**: White.

Levels in NLP

- Pragmatic Analysis:

Levels in NLP

- Pragmatic Analysis:
  - Pragmatics concerns the overall communicative and social context and its effect on interpretation
  - It means abstracting or deriving the purposeful use of the language in situations
  - Importantly those aspects of language which require world knowledge
  - The main focus is on what was said is reinterpreted on what it actually means
  - E.g. “close the window?” should have been interpreted as a request rather than an order

Levels in NLP

Phonology	• Interpretation of speech sounds within and across words
Morphology	• Deals with morphemes – the smallest units of meaning (affixes)
Lexical	• Interpret the meaning of individual words • Divides txt into paragraphs, sentences, and words
Synctactic (Parsing)	• Analysis of words in the sentence for grammar and arrange words in a manner that shows the relationship among the words
Semantic	• The text is checked for meaningfulness • It is done by mapping syntactic structures and objects in the task domain
Discourse	• 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
Pragmatic	• What was said is re-interpreted on what it actually meant. • Derive those aspects of language which require real world knowledge

Syntactic Analysis

- A number of algorithms are developed for syntactic analysis,
- Two of them are:
  - Context-Free Grammar
  - Top-Down Parser

Syntactic Analysis

- Context-Free Grammar
  - It is the grammar that consists rules with a single symbol on the left-hand side of the rewrite rules.
  - Let us create grammar to parse a sentence – “The bird pecks the grains”
    - **Articles (DET)** – a | an | the
    - **Nouns** – bird | birds | grain | grains
    - **Noun Phrase (NP)** – Article + Noun | Article + Adjective + Noun
      - = DET N | DET ADJ N
    - **Verbs** – pecks | pecking | pecked
    - **Verb Phrase** (VP) – NP V | V NP
    - **Adjectives (ADJ)** – beautiful | small | chirping

Syntactic Analysis

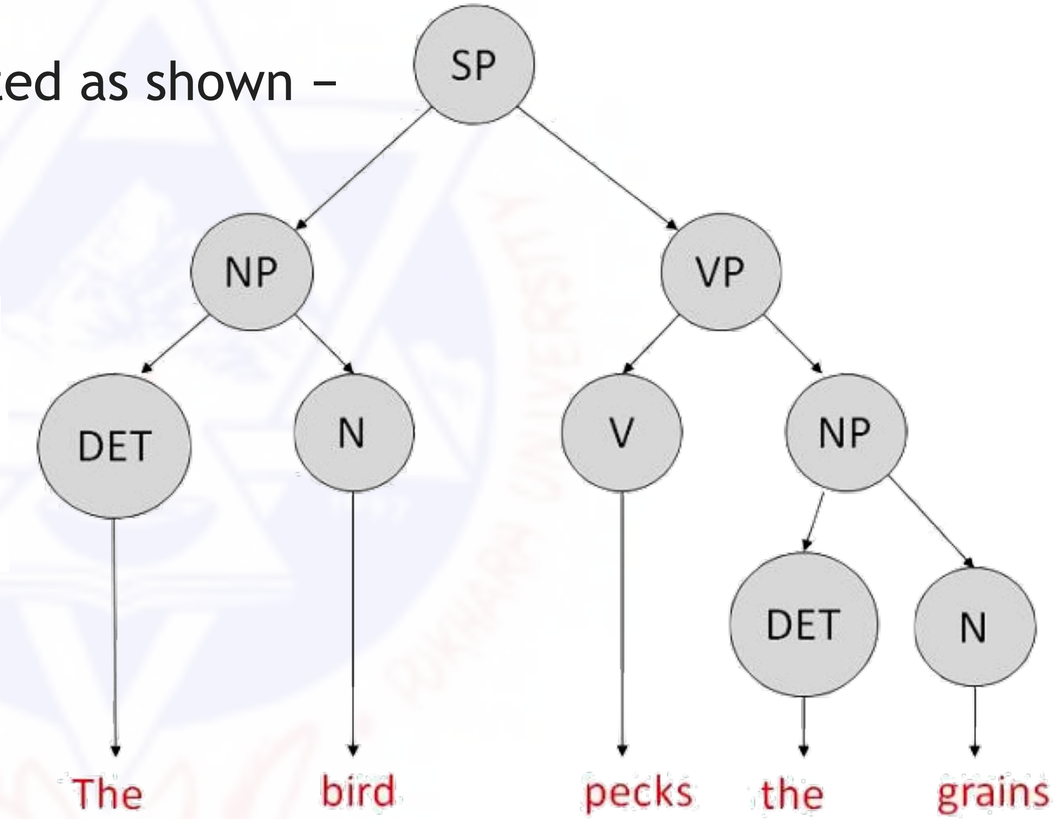
- Context-Free Grammar
  - The parse tree breaks down the sentence into structured parts so that the computer can easily understand and process it.
  - In order for the parsing algorithm to construct this parse tree, a set of rewrite rules, which describe what tree structures are legal, need to be constructed.
  - These rules say that a certain symbol may be expanded in the tree by a sequence of other symbols.
  - According to first order logic rule,
    - if there are two strings Noun Phrase (NP) and Verb Phrase (VP), then the string combined by NP followed by VP is a sentence. The rewrite rules for the sentence are as follows −
    - S → NP VP
    - NP → DET N | DET ADJ N
    - VP → V NP

Syntactic Analysis

- Context-Free Grammar
  - Lexocon –
    - DET → a | the
    - ADJ → beautiful | perching
    - N → bird | birds | grain | grains
    - V → peck | pecks | pecking

Syntactic Analysis

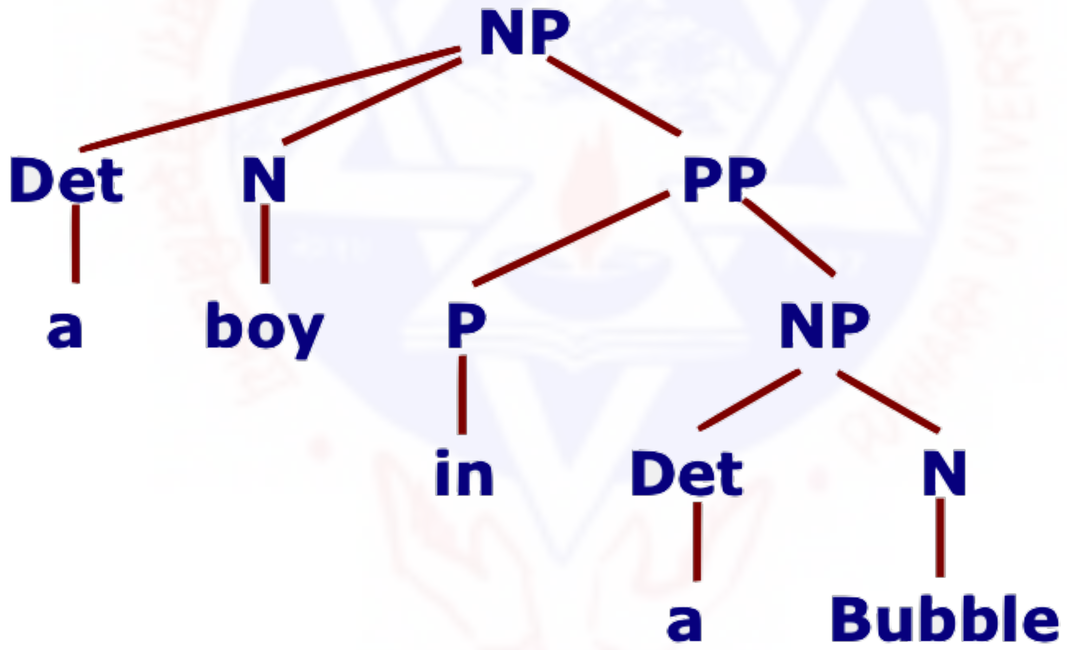
- Context-Free Grammar
  - The parse tree can be created as shown –



Now consider the above rewrite rules. Since V can be replaced by both, "peck" or "pecks", sentences such as "The bird peck the grains" can be wrongly permitted. i. e. the subject-verb agreement error is approved as correct.

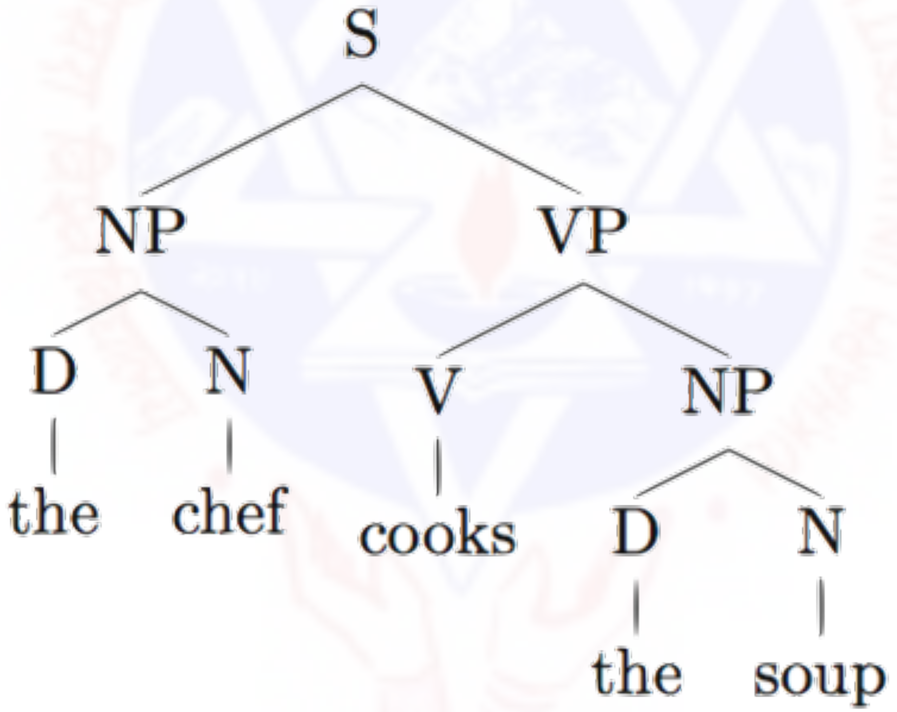
Syntactic Analysis

- Context-Free Grammar
  - The Sentence: A boy in a bubble.
  - The parse tree:



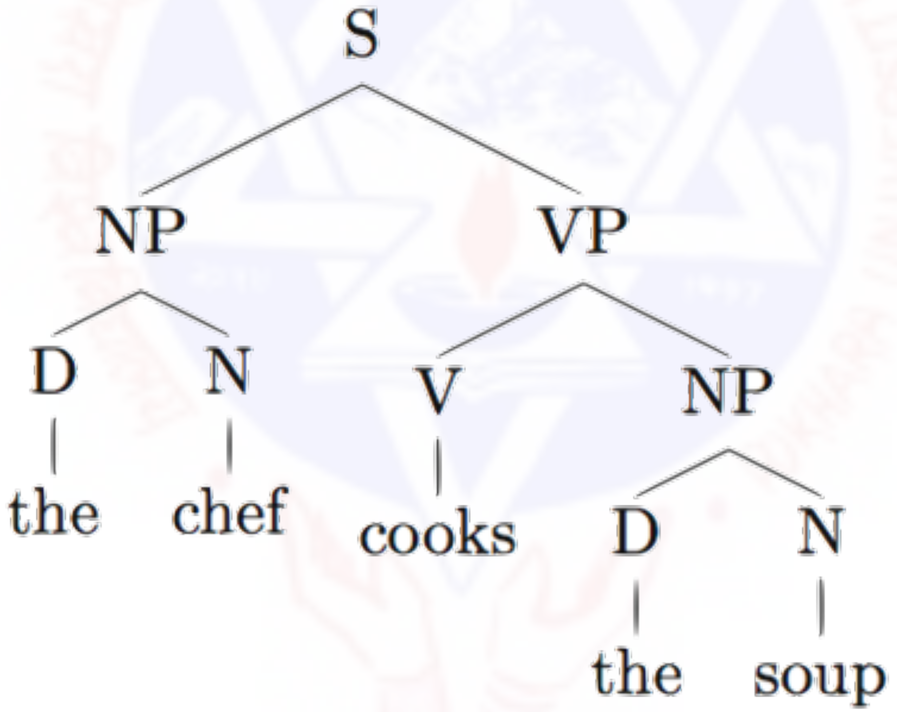
Syntactic Analysis

- Context-Free Grammar
  - The Sentence: The chef cooks the soup.
  - The parse tree:



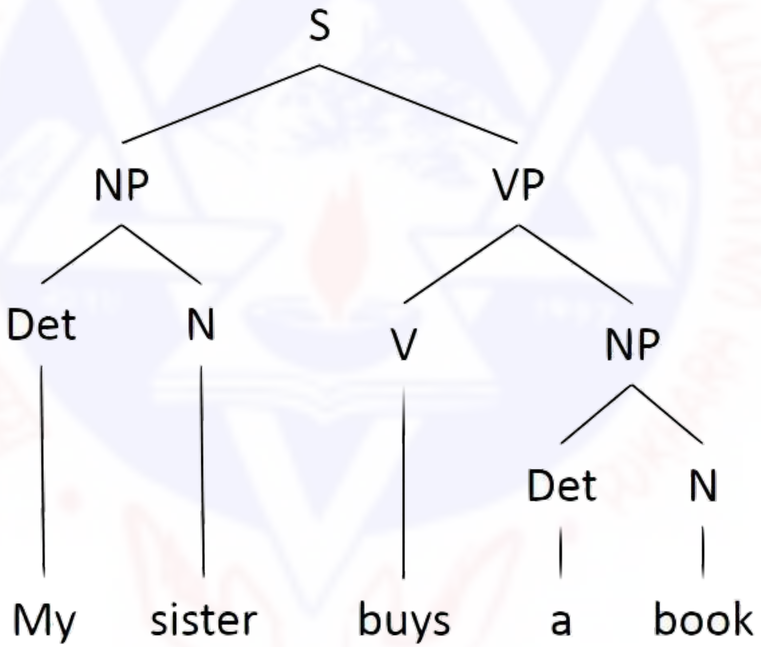
Syntactic Analysis

- Context-Free Grammar
  - The Sentence: The chef cooks the soup.
  - The parse tree:



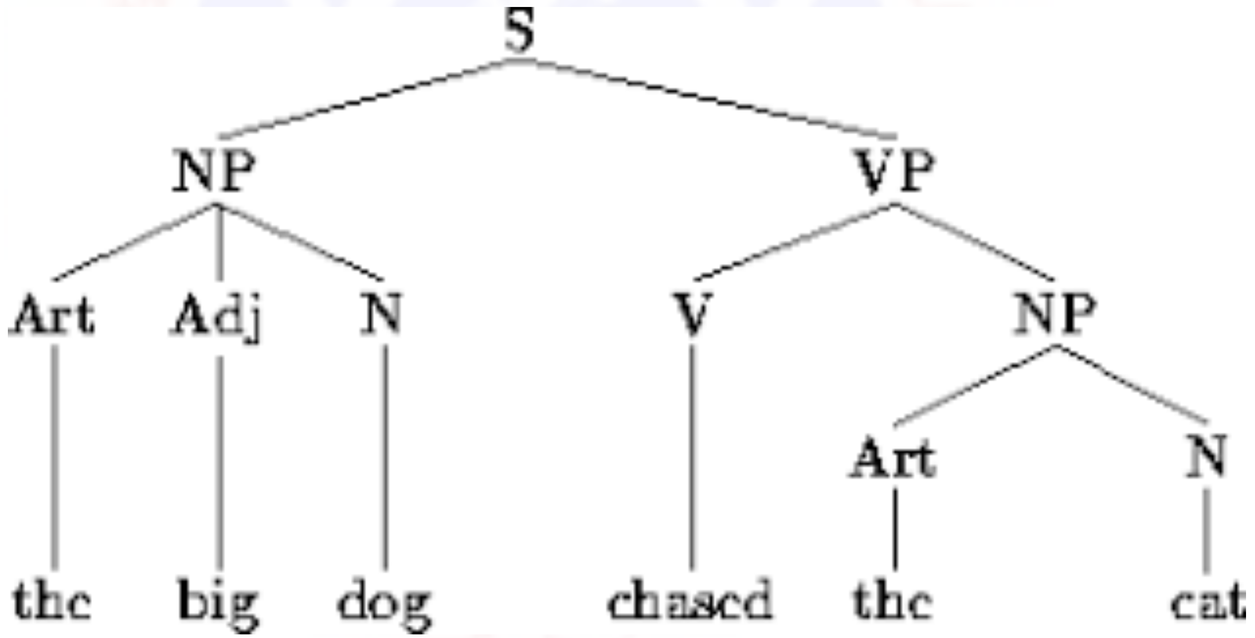
Syntactic Analysis

- Context-Free Grammar
  - The Sentence: My sister buys a book.
  - The parse tree:



Syntactic Analysis

- Context-Free Grammar
  - The Sentence: My sister buys a book.
  - The parse tree:



Some NLP Task

- **Speech recognition**, also called speech-to-text, is the task of reliably converting voice data into text data. Speech recognition is required for any application that follows voice commands or answers spoken questions.
- **Part of speech tagging**, also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context. Part of speech identifies ‘make’ as a verb in ‘I can make a paper plane,’ and as a noun in ‘What make of car do you own?’
- **Word sense disambiguation** is the selection of the meaning of a word with multiple meanings through a process of semantic analysis that determine the word that makes the most sense in the given context. For example, word sense disambiguation helps distinguish the meaning of the verb 'make' in ‘make the grade’ (achieve) vs. ‘make a bet’ (place)

Some NLP Task

- **Named entity recognition,** or NEM, identifies words or phrases as useful entities. NEM identifies ‘Kentucky’ as a location or ‘Fred’ as a man's name.
- **Co-reference resolution** is the task of identifying if and when two words refer to the same entity. The most common example is determining the person or object to which a certain pronoun refers (e.g., ‘she’ = ‘Mary’), but it can also involve identifying a metaphor or an idiom in the text (e.g., an instance in which 'bear' isn't an animal but a large hairy person).
- **Sentiment analysis** attempts to extract subjective qualities—attitudes, emotions, sarcasm, confusion, suspicion—from text.
- **Natural language generation** is sometimes described as the opposite of speech recognition or speech-to-text; it's the task of putting structured information into human language

NLP tools and approaches

- Python and the Natural Language Toolkit (NLTK):
  - The Python programing language provides a wide range of tools and libraries for attacking specific NLP tasks. Many of these are found in the Natural Language Toolkit, or NLTK, an open source collection of libraries, programs, and education resources for building NLP programs
  - The NLTK includes libraries for many of the NLP tasks listed above, plus libraries for subtasks, such as sentence parsing, word segmentation, stemming and lemmatization (methods of trimming words down to their roots), and tokenization (for breaking phrases, sentences, paragraphs and passages into tokens that help the computer better understand the text). It also includes libraries for implementing capabilities such as semantic reasoning, the ability to reach logical conclusions based on facts extracted from text.

NLP tools and approaches

- Statistical NLP, machine learning, and deep learning:
  - The **earliest NLP** applications were hand-coded, **rules-based** systems that could perform certain NLP tasks, but couldn't easily scale to accommodate a seemingly endless stream of exceptions or the increasing volumes of text and voice data
  - The **statistical NLP** combines computer algorithms with machine learning and deep learning models to automatically extract, classify, and label elements of text and voice data and then assign a statistical likelihood to each possible meaning of those elements.
  - Today, **deep learning models and learning** techniques based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs) enable NLP systems that 'learn' as they work and extract ever more accurate meaning from huge volumes of raw, unstructured, and unlabeled text and voice data sets.

NLP tools and approaches

Tools	Features
 NLTK	▪ The most well-known and full NLP library ▪ Plenty of approaches to each NLP task ▪ Supports large number of languages ▪ No integrated Word Vectors
spaCy	▪ Fastest NLP framework ▪ Easy to learn as it has one single highly optimized tool for each task ▪ Supports neural networks for training some models ▪ Lesser Language support
scikit learn NLP toolkit	▪ Most effective for Machine Learning implementation ▪ Good documentation available ▪ No neural network support for text processing
gensim	▪ Works with large datasets and processes data streams ▪ Supports Deep Learning ▪ Designed primarily of unsupervised text modeling

NLP tools and approaches

- NLTK: download/installation/Documentation
  - https://www.nltk.org/
- Book: Natural Language Processing with Python
  - https://www.nltk.org/book/

NLP API

- NLTK: download/installation/Documentation
  - https://www.nltk.org/
- Book: Natural Language Processing with Python
  - https://www.nltk.org/book/

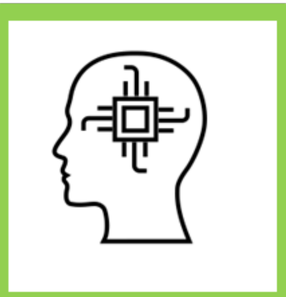
NLP- Real World Applications

- **Spam detection:** the spam detection technologies use NLP's text classification capabilities to scan emails for language that often indicates spam or phishing.
- **Machine translation:** Google Translate is an example of widely available NLP technology at work. Truly useful machine translation involves more than replacing words in one language with words of another.
- **Virtual agents and chatbots:** Virtual agents such as Apple's Siri and Amazon's Alexa use speech recognition to recognize patterns in voice commands and natural language generation to respond with appropriate action or helpful comments. Chatbots perform the same magic in response to typed text entries. The best of these also learn to recognize contextual clues about human requests and use them to provide even better responses or options over time. The next enhancement for these applications is question answering, the ability to respond to our questions—anticipated or not—with relevant and helpful answers in their own words.

NLP- Real World Applications

- **Social media sentiment analysis:** NLP has become an essential business tool for uncovering hidden data insights from social media channels. Sentiment analysis can analyze language used in social media posts, responses, reviews, and more to extract attitudes and emotions in response to products, promotions, and events–information companies can use in product designs, advertising campaigns, and more.
- **Text summarization:** Text summarization uses NLP techniques to digest huge volumes of digital text and create summaries and synopses for indexes, research databases, or busy readers who don't have time to read full text. The best text summarization applications use semantic reasoning and natural language generation (NLG) to add useful context and conclusions to summaries.

AI, ML and DL

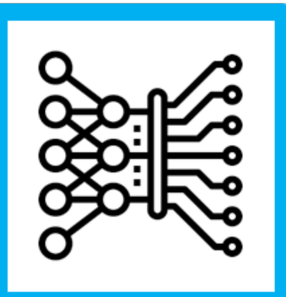


Artificial Intelligence

In short - incorporating human intelligence to machines. Whenever a machine completes tasks based on a set of stipulated rules that solve problems (algorithms), such an "intelligent" behavior can be termed as artificial intelligence.

Machine Learning

As the name suggests, machine learning can be loosely interpreted to mean empowering computer systems with the ability to "learn". The intention of ML is to enable machines to learn by themselves using the provided data and make accurate predictions. It is the field of study that gives computers the capability to learn without being explicitly programmed. ML is a subset of artificial intelligence; in fact, it's simply a technique for realizing AI.



Deep Learning

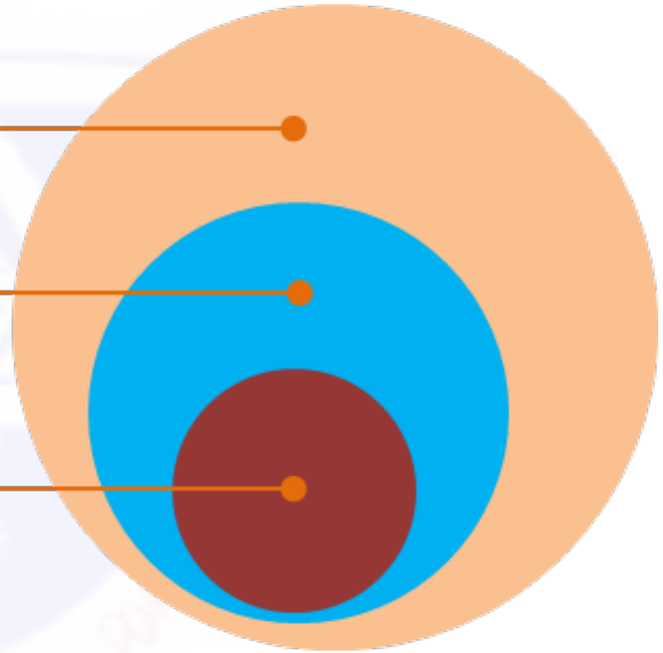
DL is the next evolution of machine learning. DL algorithms are roughly inspired by the information processing patterns found in the human brain. Just like we use our brains to identify patterns and classify various types of information, deep learning algorithms can be taught to accomplish the same tasks for machines.

AI, ML and DL

Artificial Intelligence
Enable Computers to mimic human behaviour

Machine Learning
Subset of AI, which use statistical methods to enable machines to improve with experiences

Deep Learning
Subset of ML, which use algorithms inspired by function of human brain



Machine Vision

- Assignment:
  - What the machine vision is.
  - Components of machine vision system.
  - Benefits of machine vision.
  - Application areas of machine vision.

  - (*Take reference of the document "Ch-6 Machine-Vision.pdf" which is uploaded in your TEAMS class's lecture slide folder.*)

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

THANK YOU

End of Chapter