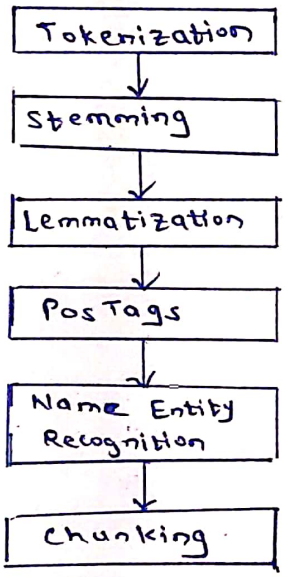
NLP
Module 1 - Introduction

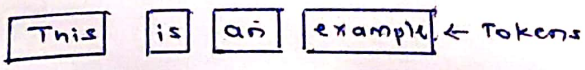
## Natural Language Processing

- Human language is used for communication.
- Human language is used to express their emotions.
- Language is used to shape thoughts.
- It carries a meaning.
- NLP is concerned with development of computational models of aspects aspects of human language.
- NLP is a process of analysis of input provided in human language and converting it into useful form of representation.
- Input & output of NLP is text or speech.
- Need of NLP
  1. Automatic tool for NLP
  2. Better understanding of human communication

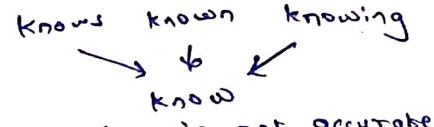
## Steps in NLP



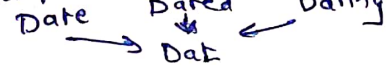

### ① Tokenization
- Break sentence into small tokens.
- Example:
  This is an example.



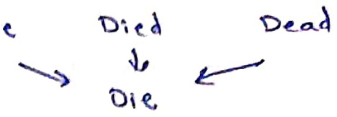

### ② Stemming
- Normalize words in their root form.
- Example:



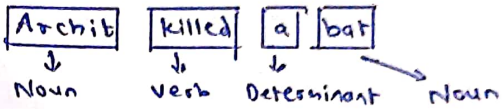

- stemming is not accurate and efficient
- Example:



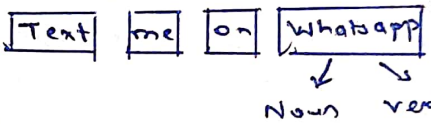

### ③ Lemmatization
- It is an extension of stemming.
- Output of lemmatization is accurate.
- It combines similar words to generate output (Lemma).
- Example:




### ④ Pos Tags
- Part of speech Tags.
- It tags all the tokens with Pos Tags.
- Example:




Disadvantage:
- There could be multiple POS Tags to single token.




### ⑤ Name Entity Recognition
- It overcomes drawback of Pos Tag.

### ⑥ Chunking
- It groups individual tokens and forms a chunk.
- It helps in getting insightful and meaningful information from text.

## Levels of NLP

1. Phonology Level
2. Morphological level
3. Lexical level
4. Syntactic level
5. semantic level
6. Discourse level
7. Pragmatic level.

## Applications of NLP

1. Machine Translation
2. speech Recognition
3. speech synthesis
4. Information retrieval
5. Information extraction
6. Question Answering
7. Text summarization
8. Sentiment Analysis

## Ambiguity in NLP

1. Lexical Ambiguity
   - When words have multiple assertions.
   - Example:
     Back stage (Noun)
     Back Door (Adjective)
2. Syntactic Ambiguity
   - Sentence can be parsed in different ways
   - Example:
     "I saw the girl on the beach with my binoculars."

   → Confusion in meaning is created.
3. Semantic Ambiguity
   - It is related to sentence interpretation
4. Metonymy Ambiguity
   - It deals with phrases where literal meaning is different from the figurative assertion.

## Challenges of NLP

1. Contextual words and phrases and homonyms.
2. Synonyms
3. Irony and sarcasm
4. Ambiguity
5. Errors in text or speech.
6. Idioms and slang
7. Domain specific language.
8. Low Resource Languages

## Stages in NLP

1. Lexical Analysis
   - Also known as morphological analysis
   - In this, structure of words is identified and analyzed
   - It divides whole text into para, sentences and words.
2. Syntactic Analysis
   - Also known as Parsing.
   - It involves analysis of words in the sentence for grammar and ordering words in a way that shows relationship among words.
3. Semantic Analysis
   - It draws exact dictionary meaning from text.
   - Text is checked for meaningfulness.
   - It is done by mapping syntactic structure.
4. Discourse Integration
   - Meaning of sentence depends upon meaning of sentence just before it.
   - It brings meaning for immediate following sentence.
5. Pragmatic Analysis
   - In this, what was said is re-interpreted on what is truly meant.
   - It derives real world knowledge.