



## ① Tokenization:-

→ Process.

→ Break.

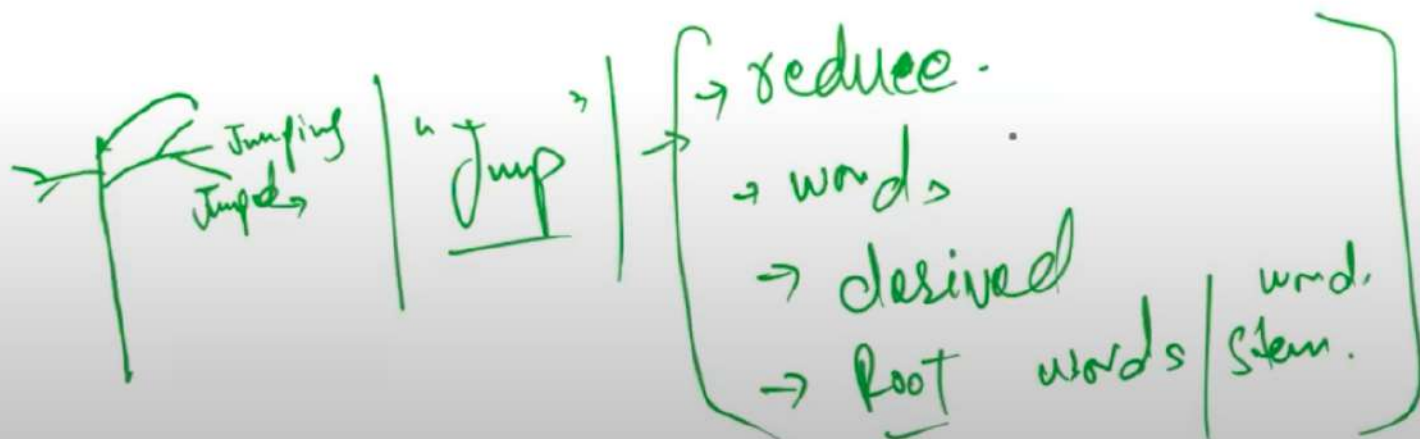
→ Sentences.

→ word/terms.

The cat sat on mat.

"The", "cat", "sat", "on", "mat"

## ② Stemming:-



(3) Lemmatization → similar to stemming  
But → retain the context.  
→ meaningful base form.

["good", "better", "best"] → Lemmas  
["good", "good", "good"]



④ Stop Words:- [a, an, the, of, in,  
filter out .

"The cat sat on a mat" → ["cat", "sat", "mat"]



④ Stop Words:- [a, an, the, of, in]  
filter out.

"The cat sat on a mat" → ["cat", "sat", "mat"]

⑤ Bag of (BOW)  
words = [a: 12, The: 31]



④ Stop Words:- [a, an, the, of, in]  
filter out.

"The cat sat on a mat" → ["cat", "sat", "mat"]

⑤ Bag of words = (BoW)  
[a: 12, The: 31] (cat: 12, sat: 13)  
(frequency of each token)





## ⑥ TF-IDF (Term Frequency-Inverse Document frequency)

- Numerical Statistic
- Reflect
- Importance of a word / term
- Document

## ⑥ TF-IDF (Term Frequency-Inverse Document frequency)

- Numerical Statistic
- Reflect
- Importance of a word / term
- Document

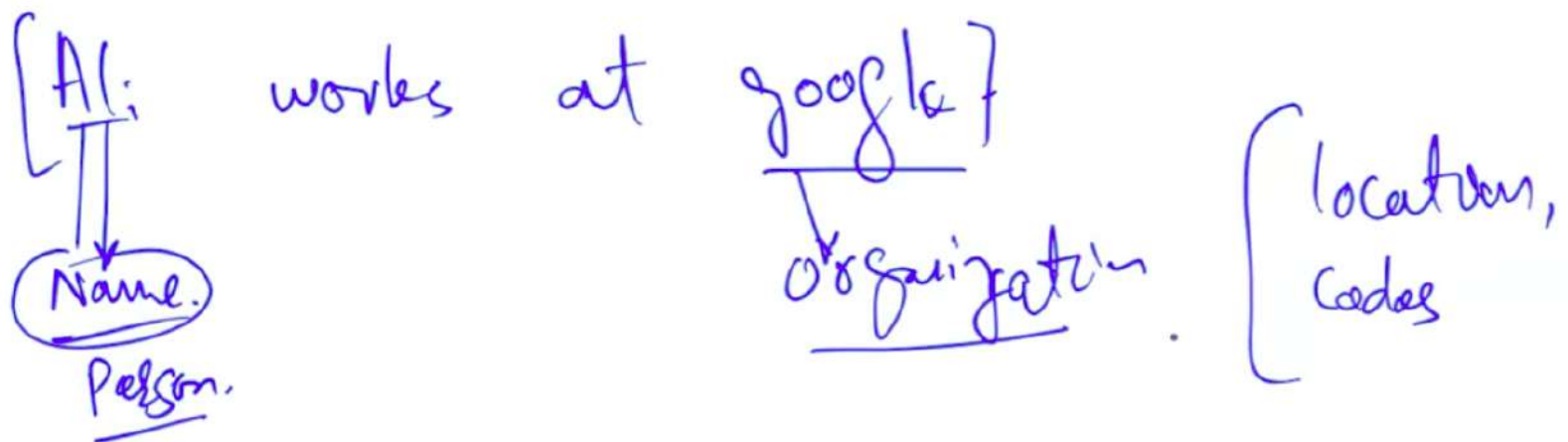
[The "cat" "sat" "on" "a" "mat"]

39 1 60



39 60

# ⑦ NER: Named Entity Recognition:







⑧ (POS) Parts of Speech → Tagging:-  
(noun, verbs, adjectives)---



# ⑨ Syntax and Parsing:-

Rules, principal.

(is sitting).

"The cat is sitting on mat".



⑩ Semantic Analysis:  
→ meaning of a sentence  
"Apple has high stock rate".

⑪ Sentiment Analysis:  
→ computationally determine.  
→ meaning of Sentic.



(Apple) has high stock sat

## ② Sentiment Analysis

- Computationally determine.
- meaning of sentence.
- positive, -ve, neutral, Sad, happy.



→ Computationally determine.  
→ meaning of sentence.  
→ positive, -ve, neutral, Sad, happy

"This movie is great."

(positive)





## ①① Sentiment Analysis

→ Computationally determine.

→ meaning of sentence

→ positive, -ve, neutral, Sad, happy

"This movie is great."

(positive)



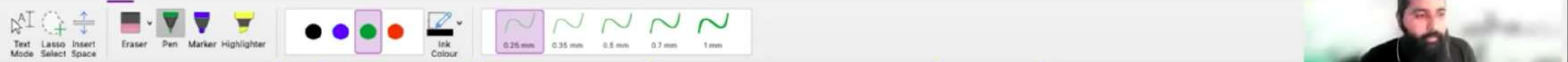
This movie is great,. (positive)



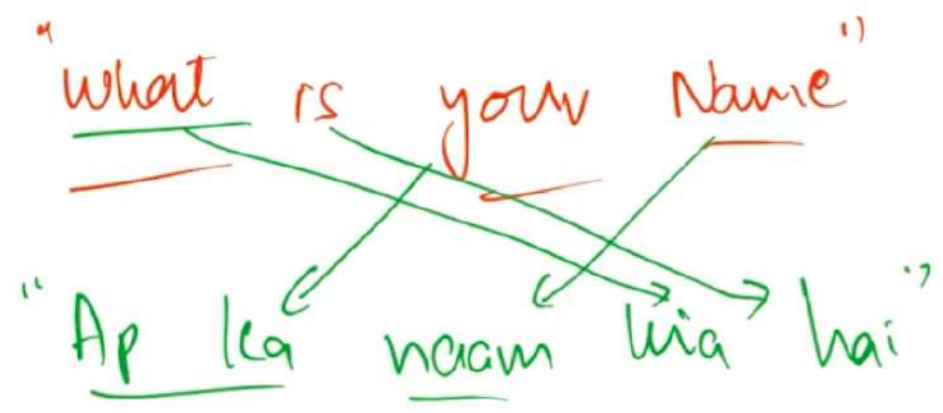
(12) Word Embedding:-

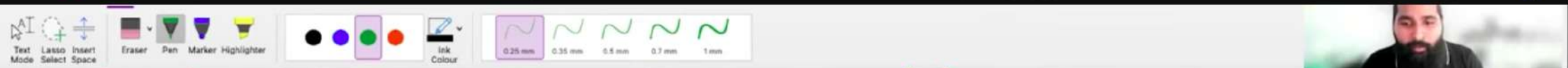


king = man , Queen = female.

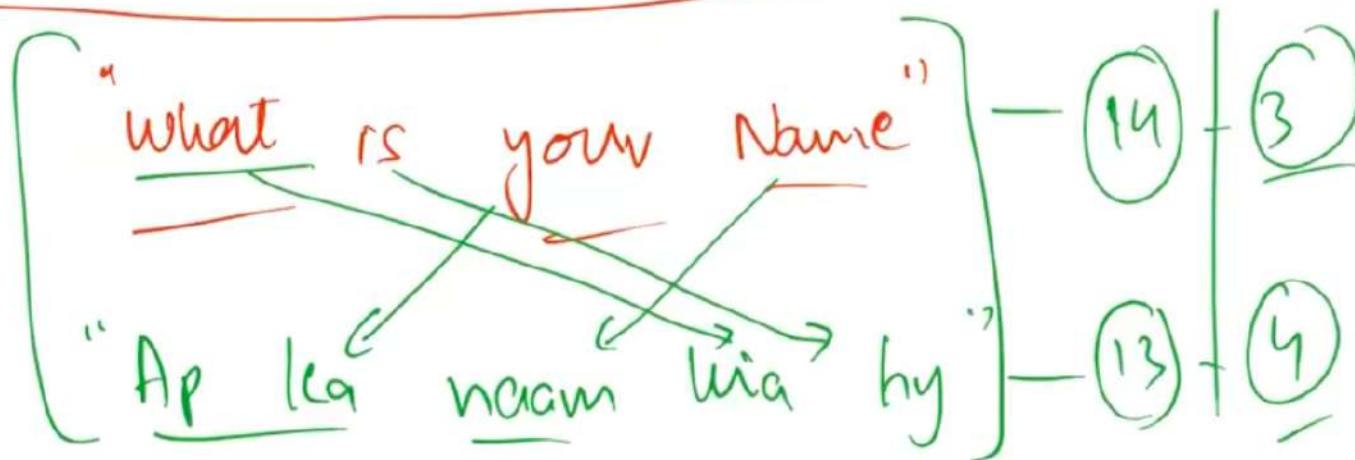


# (13) Sequence - to - Sequence Models





# (13) Sequence-to-Sequence Models-



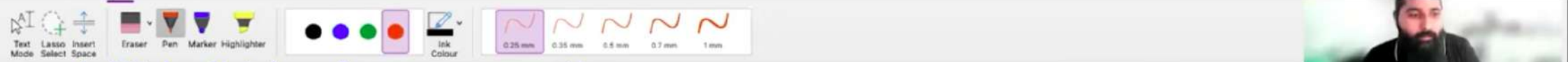


"Ap ka naam kya hai?" - (13) - (4)

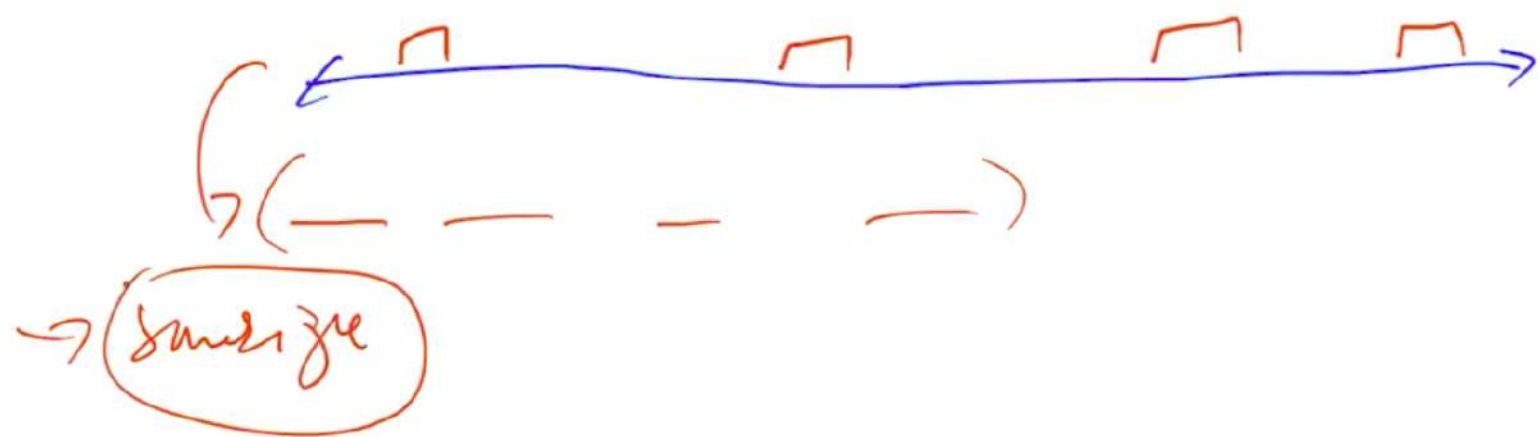
Italia, "Do you speak English?"

"Parla inglese?"





# (14) Attention Mechanisms





⑮ Transformers:- Self attention mechanism

→ State of the art.

[ → Font Classification, NER,  
→ Question answer? ]

→ BERT  
→ GPT-2



⑮ Transformers:- Self attention mechanism

→ State of the art.

[ → Text Classification, NER,  
→ Question answer? ]

→ BERT  
→ GPT-2



⑮ Transformers:- Self attention mechanism

→ State of the art.

[ → Font Classification, NER,  
→ Question answer? ]

→ BERT  
→ GPT-2



(16)

# Language Model:-

→ fish

→ machine learning → vocabulary.

→





(16)

# Language Model:-

- fish
- machine learning → vocabulary.
- 

"The weather is . . . . ."



(17) BERT:..  
= Bidirectional. Encodes Representations  
from Transformers -

18 GPT :- Generative Pretrained Trans former.

→ Autoregressive

→ Human like text.

→ Verify 95%

18) GPT :- Generative Pre-trained Transformer.

→ Autoregressive

→ Human like text.

→ Verify 95% ✓ - Big AI, Resplawit



NLP → python, c++, Java

N (Ty — — — — —)

Music LM

log — — — — —

