

## Lecture 38 : Embedding Models

- ... review of importance of RAG
- ... importance of jargons, vocabulary - confidence
- ... loaders ✓ (1st step)

-- splitting (breaking large into small manageable for llm)

... embedding (No model can't understand text rather

mathematically way)  
representations which ai can understand.

we use embedding model.

There are multiple embedding models available.

embedding actually numerical representation.  
(vector form)

• DDM Model

• Urdu voice bot ⇒ Assignment

... review of RAG Pipeline

Ingestion

Document → chunks → Embedding → index

## Embeddings

⇒ closely related words numerical representation also closer... their difference minimum.

(vector)

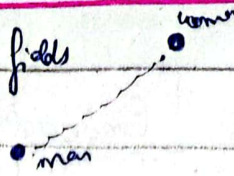
man	0.6	-0.2	0.8	0.9	-0.1	-0.7	-0.7
woman	0.7	0.3	0.9	-0.7	0.1	-0.8	-0.4

difference less ⇒ semantically same

difference greater ⇒ semantically not same



- widely used in NLP and other fields

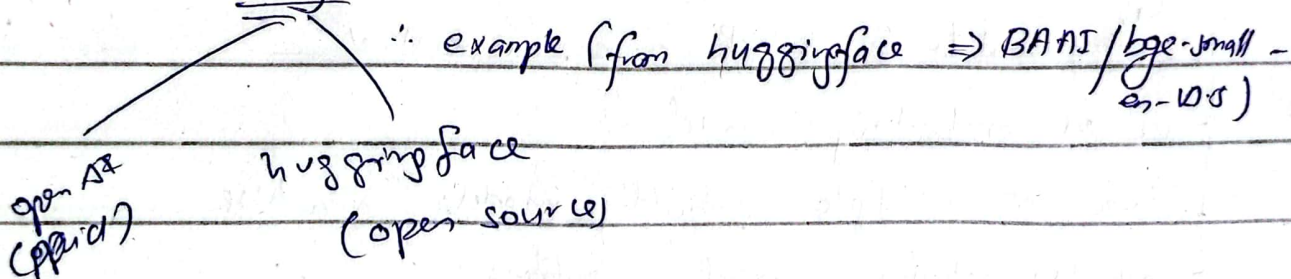


## Practically demo

- loading pdf document
- splitting into chunks [recursive relative spliter]

## Embeddings

can use any model



## Semantic Splitters

(split the text based on semantic similarity)

split into sentences → groups into group of 3

sentences → merge one that are similar

doc → embeddings → similar embedding into chunks

breakhold threshold-type ← "percentile"