

Sentence Transformers

Cosine similarity:

→ Takes embedding of 2 sentences
and tells cosine similarity

↳ If we get matrix, then it
means input = mult. sentences

Ex:

(st 1) A man is eating food

(st 2) A man eating pieces of bread

(st 3) The girl is a hoe.

Opp

[1.000, 0.7553, -0.1050, ...]

st 1 (1st word) ≈ st 2 (1st word)

st 1 (2nd word) ≈ st 2 (2nd word)

st 3 ≠ st 1

↳ so we have one value

To find out sentence top similarity

↳ iterate, matrix

↳ 2 for loops

↳ collect all possible values.

↳ collecting pair value similarity.

SEMANTIC SEARCH

↳ searching technique where
it finds the intent & content
meaning of words.

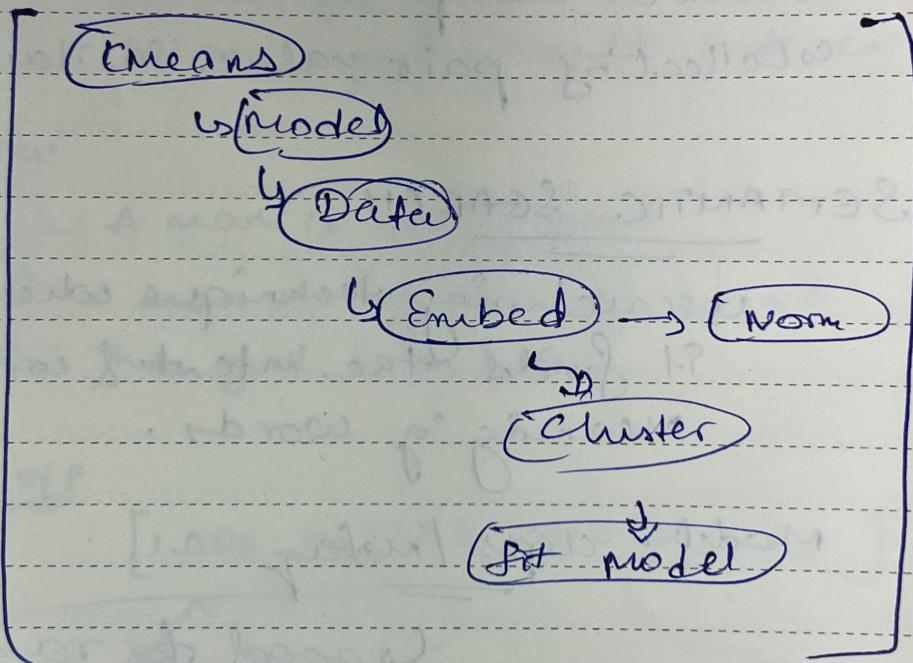
Model = clips / mafag

↳ good for rag
and trained on
MFAQ dataset

Clustering:

- use kmeans
- minmax used
- bunch of sent.
- embed → encode
- normalise embed to unit length

If we normalize vectors, it becomes easy for distance based metric to be optimized.



Clustering