Beginner's Guide to Vector Databases

AI by Hand ✍️

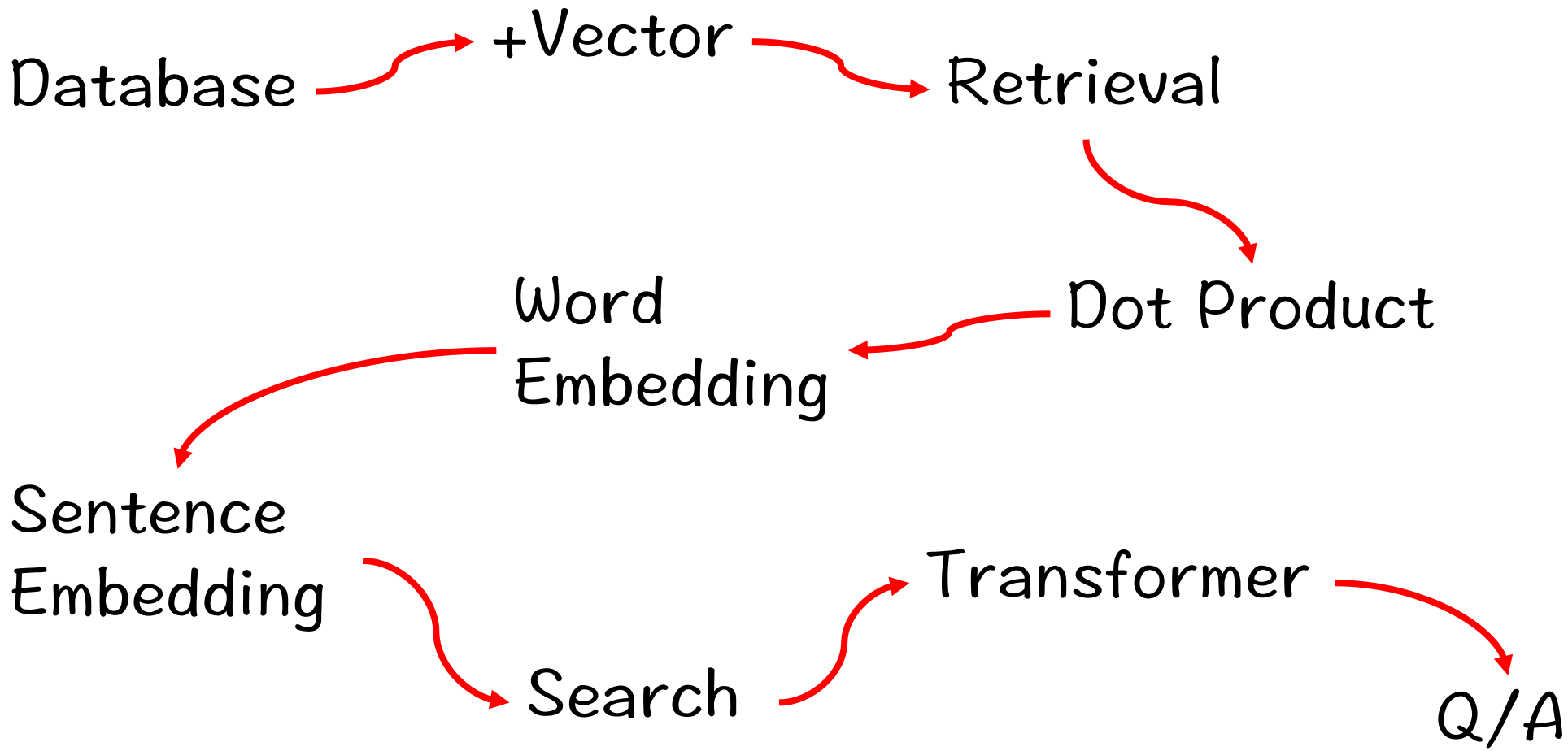
Prof. Tom Yeh

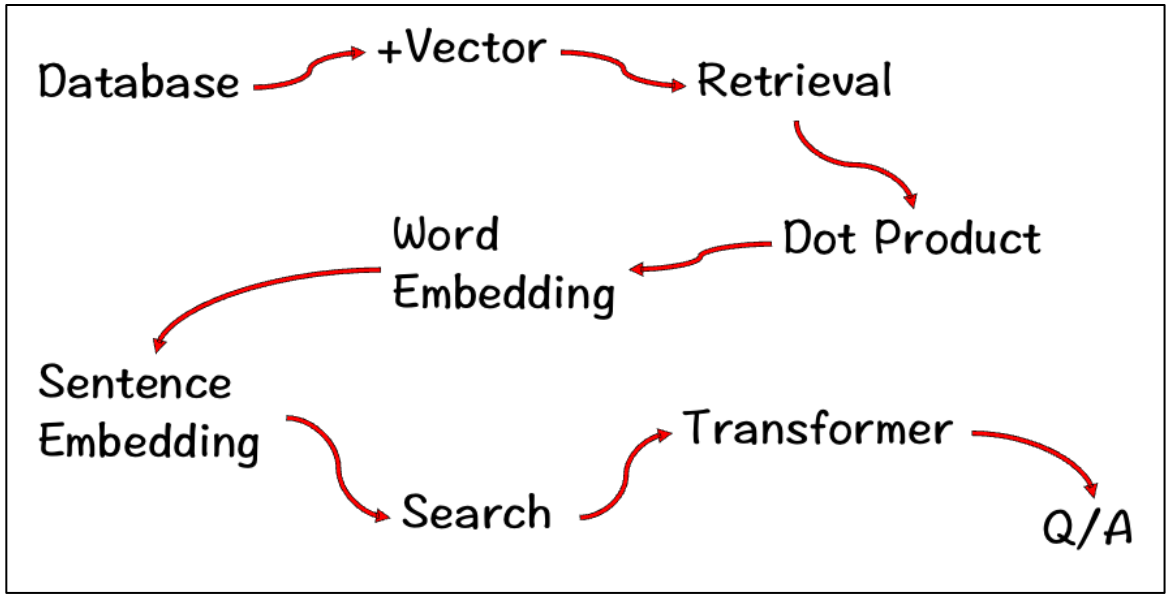
University of Colorado Boulder

Hosted by:

SingleStore

Roadmap


Database
+Vector
Retrieval
Dot Product
Word Embedding
Sentence Embedding
Search
Transformer
Q/A



Database

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

Fun fact

There are _____ millions dogs in the world!

How to create a table?

SQL:

_____ TABLE _____

(id _____,

name _____,

size _____,

pop _____)

id	name	size	pop

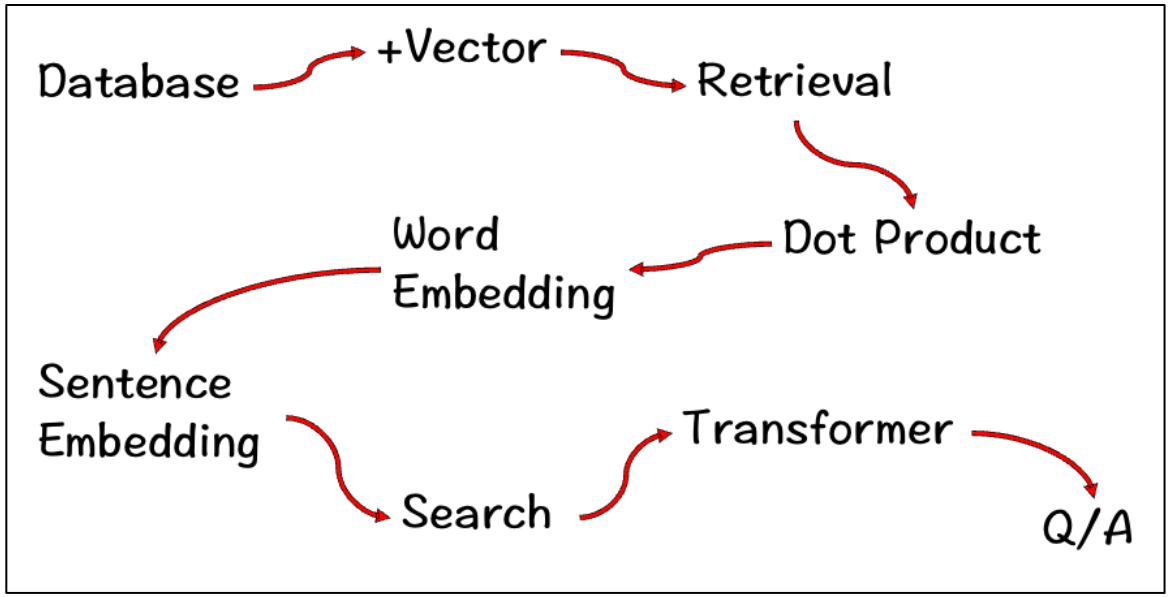
How to insert a record?

SQL:

_____ INTO animals

_____ (1, dog, 2, 900)

id	name	size	pop
1	dog	2	900



Vector Database

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

How to create a vector database?

SQL:

```
CREATE TABLE animals
( id INT,
  name VARCHAR(10),
  size INT,
  pop INT,
  emb _____   not null  )
```

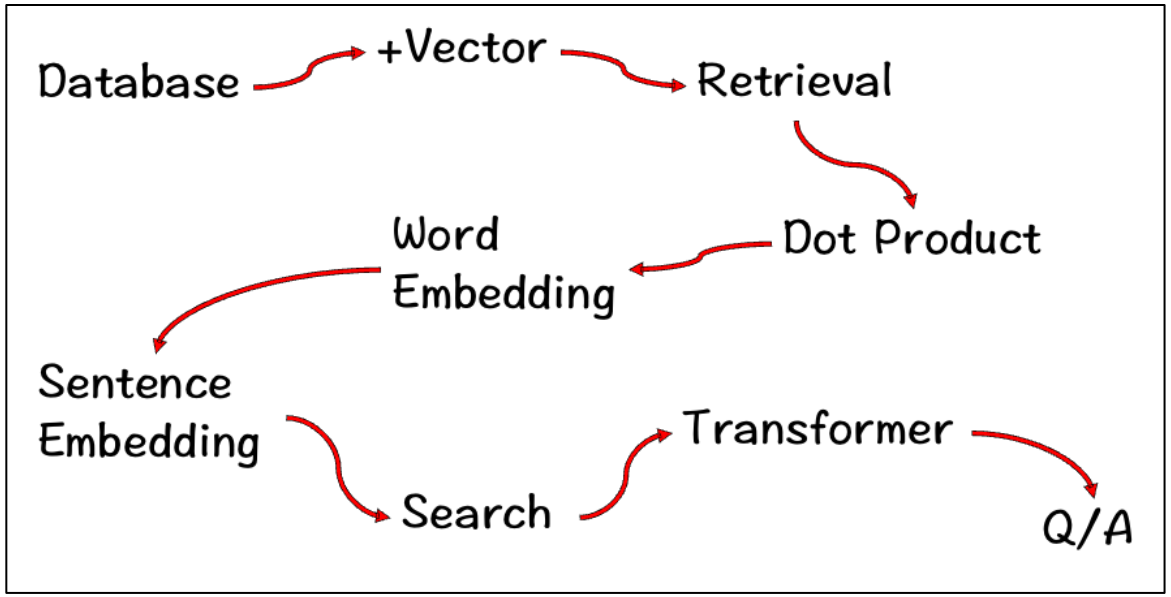
id	name	size	pop	

How to insert a record with a vector?

SQL:

```
INSERT INTO animals
VAUES (1, dog, 2, 900, _____________________)
```

id	name	size	pop	emb
1	dog	2	900	



Retrieval

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

Which record is relevant to the query “cat”?

Query

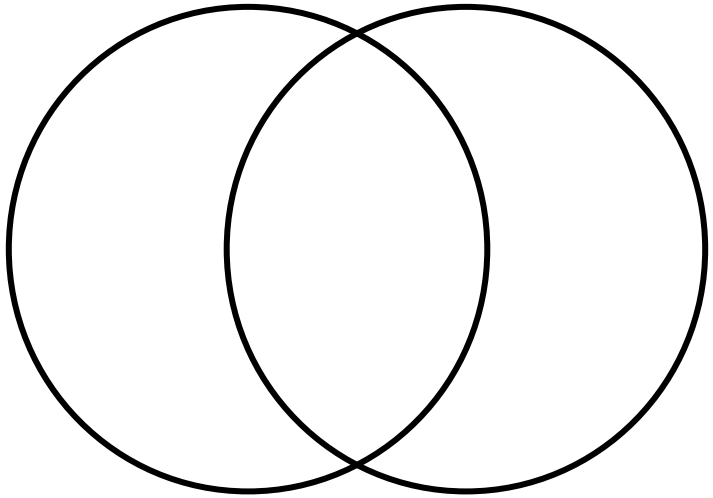
cat

1	2	0

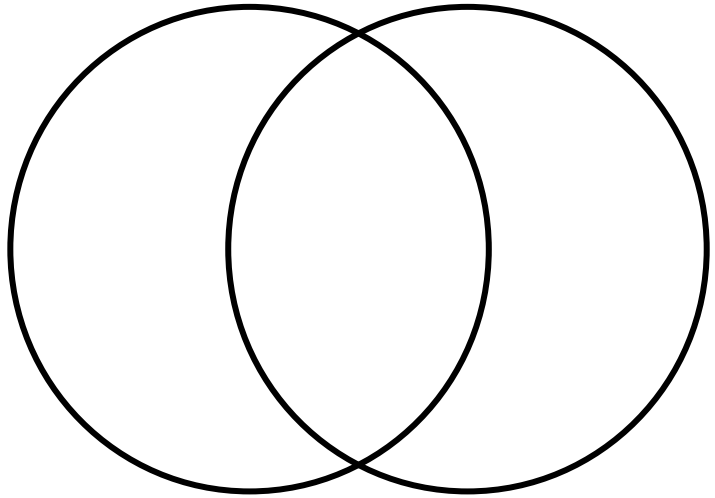
id	name	size	pop	emb
1	dog	2	900	2 1 0
2	bat	1	10000	0 1 2

Draw distance vs similarity

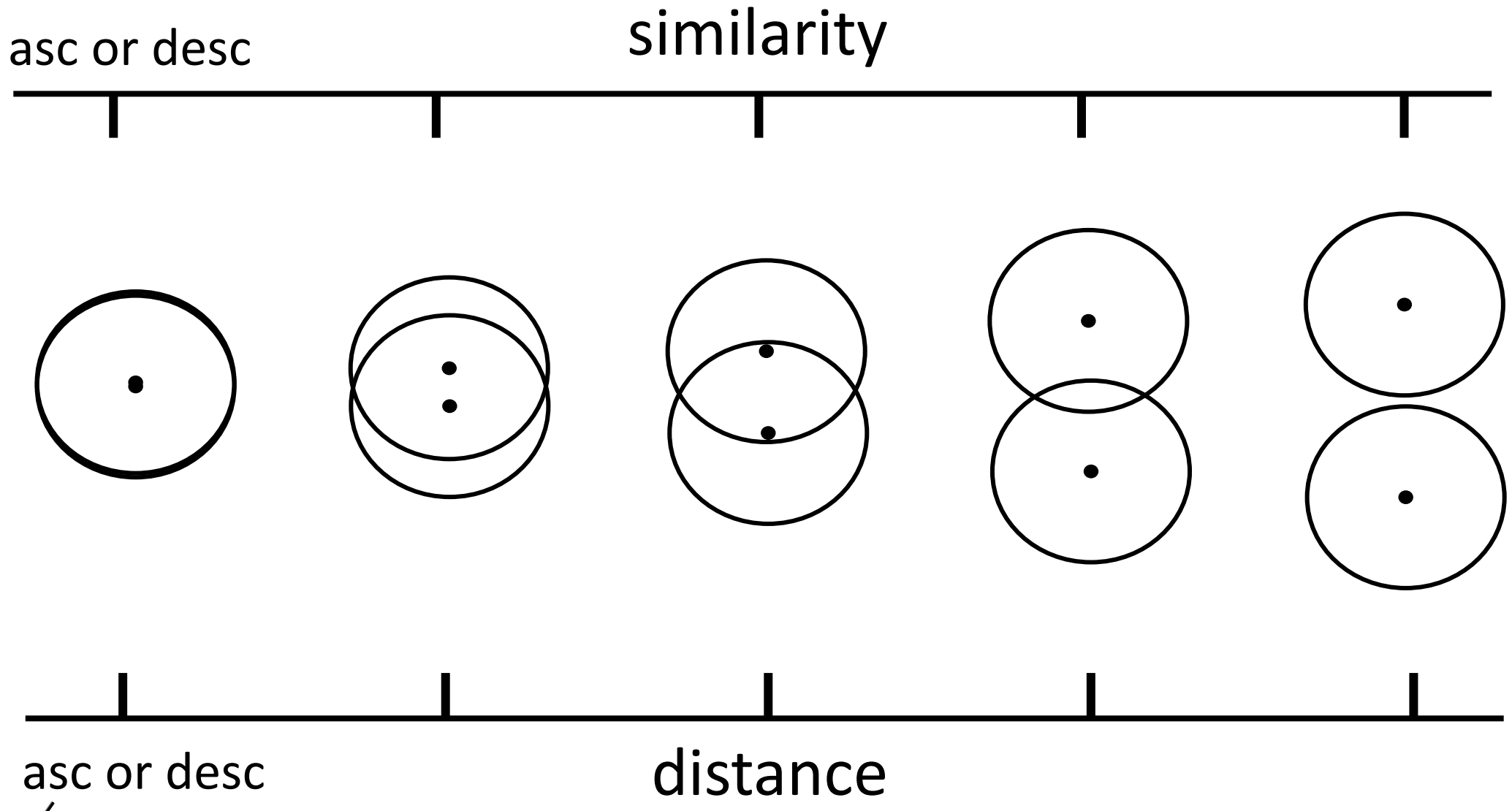
distance



similarity



Distance vs similarity on a scale of 1 to 5



How to retrieve by similarity? (dot product)

_____ name, emb<___>[__,__,__] AS score

FROM animals

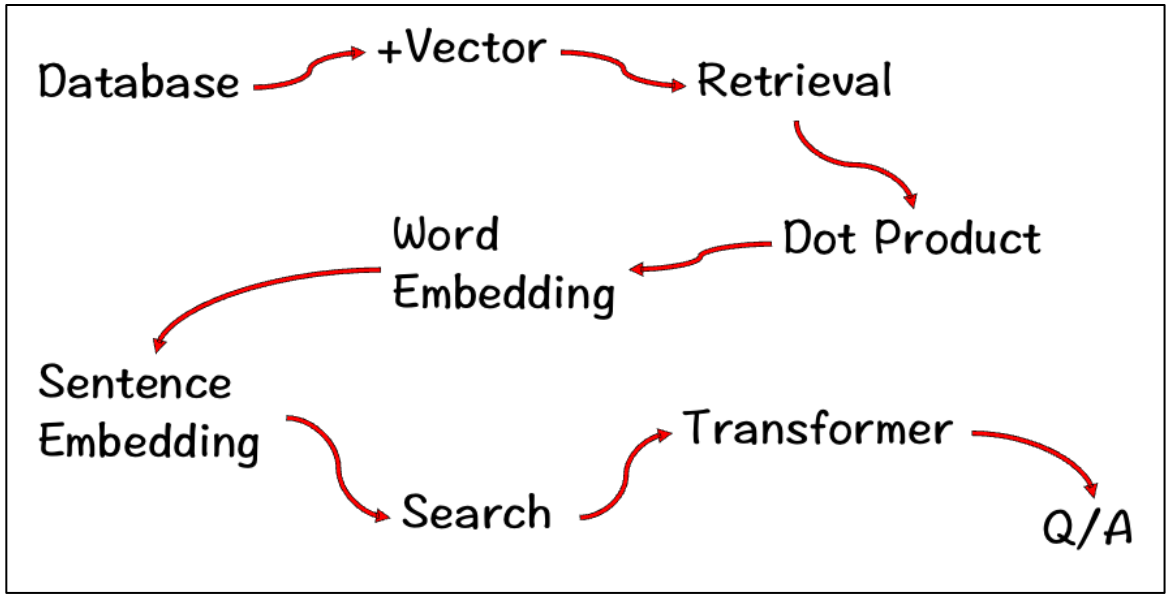
_____ BY _____ ASC | DESC ;l

How to retrieve by distance? (Euclidean)

```
SELECT name, emb<*>[1, 2, 0] AS score

FROM animals

ORDER BY score  DESC;
```



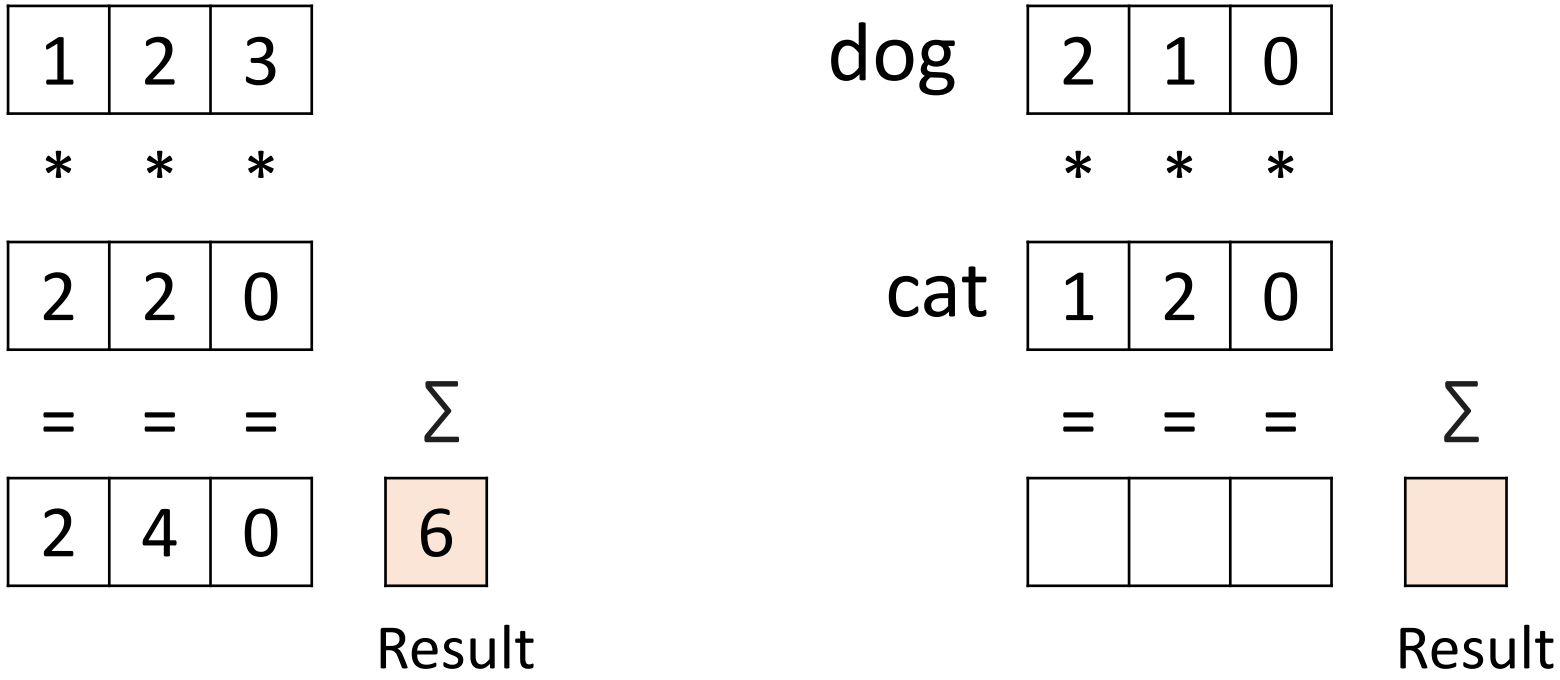
Dot Product

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

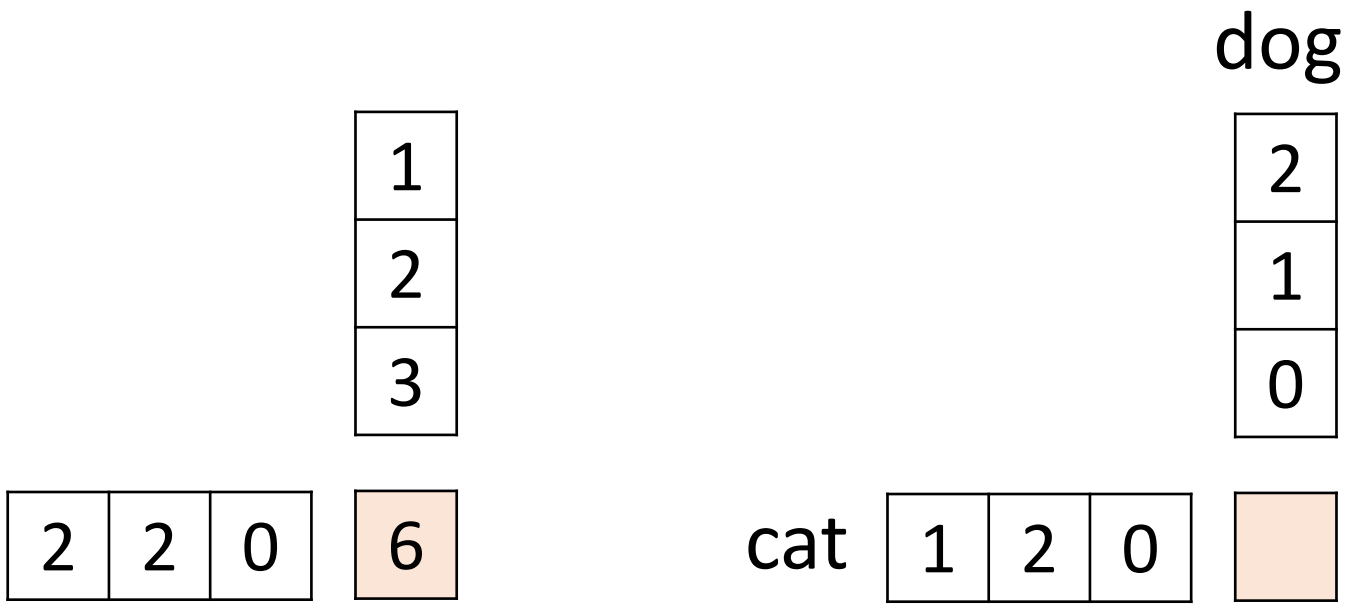
How to compute dot product?

Example:



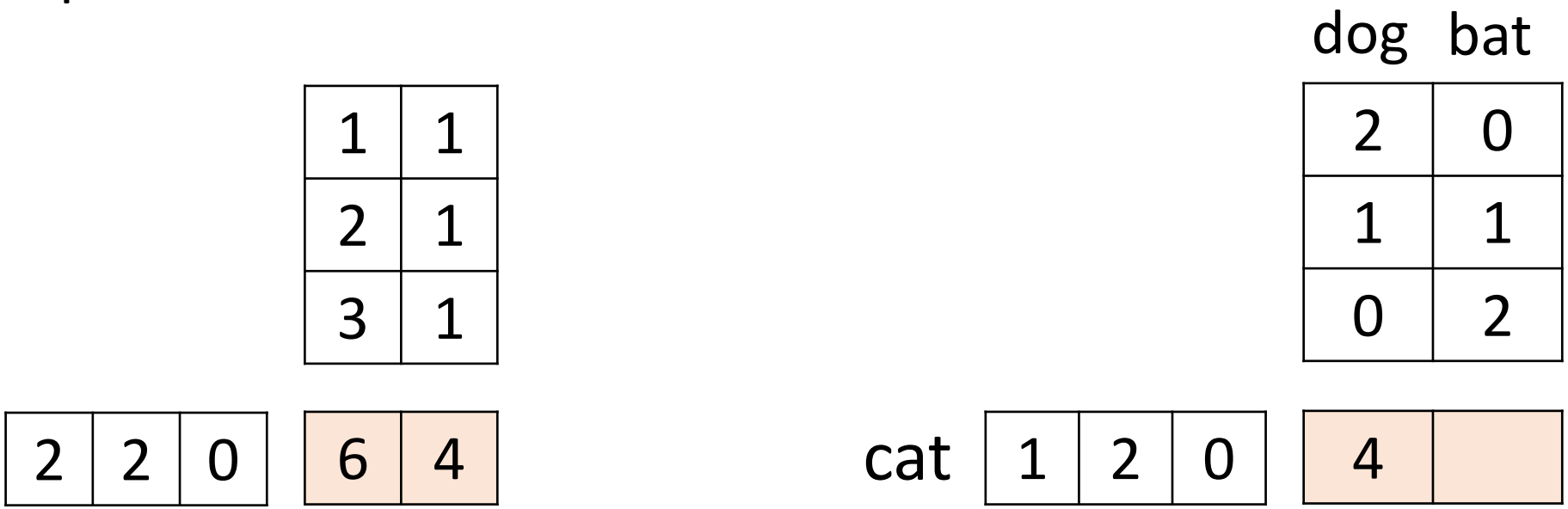
How to compute dot product using matrix multiplication?

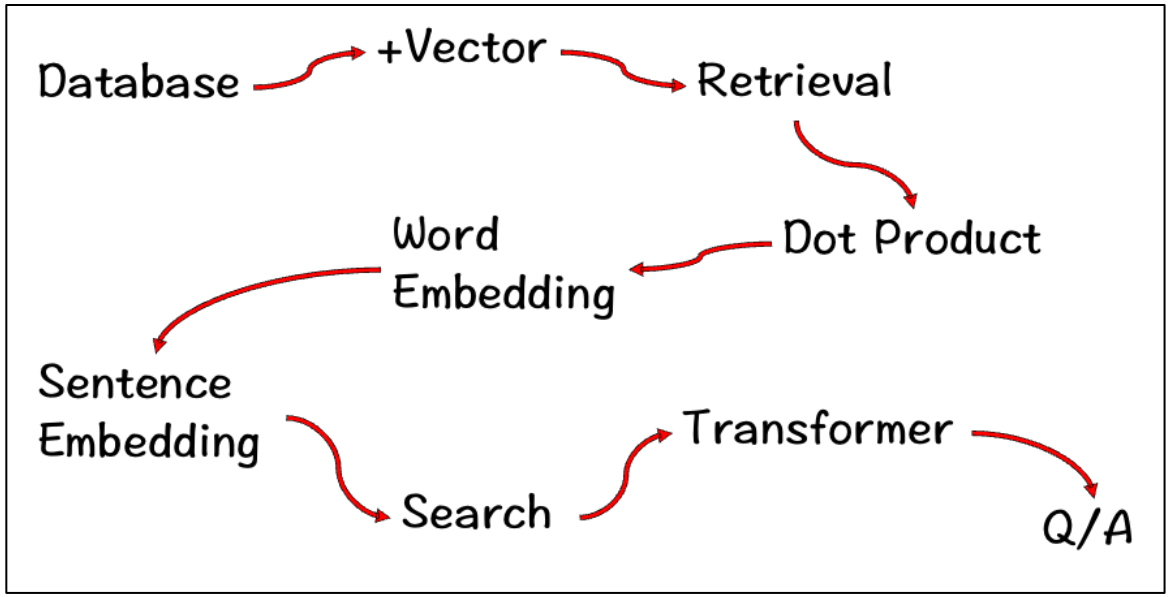
Example:



How to compute dot product with multiple vectors?

Example:



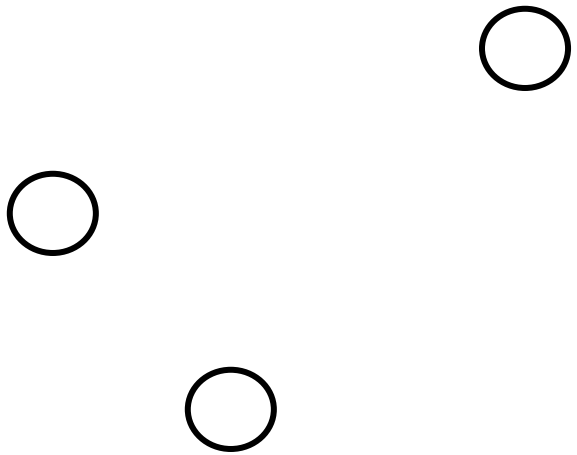


Word Embedding

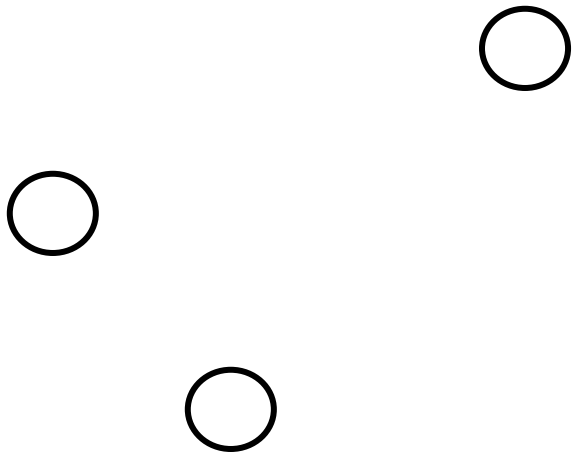
Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

Where are dog, cat and bat in the "name" space?



Where are dog, cat and bat in the "name" space?



Which embedding is better?

Embedding 1

dog	cat	bat
2	1	0
1	2	1
0	0	2

Embedding 2

dog	cat	bat
2	0	1
1	1	0
0	2	2

Which embedding is better?

Desired dot product similarity

dog	cat	bat
2	1	0
1	2	1
0	0	2

dog	2	1	0
cat	1	2	0
bat	0	1	2

	H	L
H		L
L	L	

Embedding 1

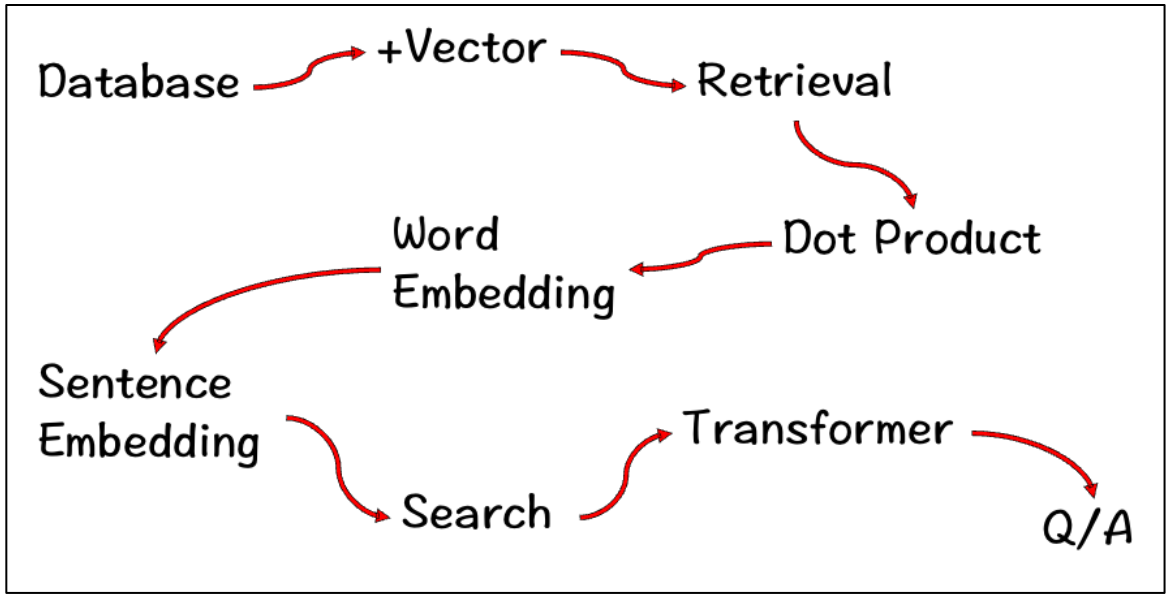
dog	cat	bat
2	1	0
1	2	1
0	0	2

dog	2	1	0
cat	1	2	0
bat	0	1	2

Embedding 2

dog	cat	bat
2	0	1
1	1	0
0	2	2

dog	2	1	0
cat	0	1	2
bat	1	0	2



Sentence Embedding

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

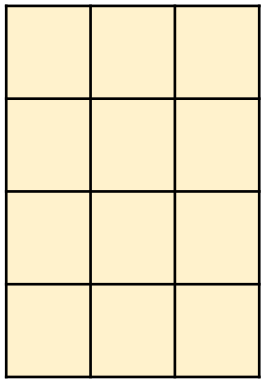
How to embed sentences?

id	comment	user	emb
1	How are you?	John	?
2	Who are you?	Mary	?

“How are you” → word embedding vectors

how	are	you

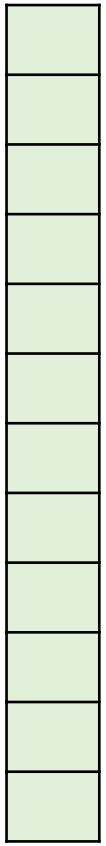
a	an	the	how	why	who	what	are	is	am	be	was	you	we	I	they	she	he	she	me	him	her
0	-1	0	1	0	1	0	0	-1	1	0	0	0	3	1	0	-1	0	0	0	-1	0
2	0	2	0	0	0	-1	1	0	0	0	2	1	0	2	0	2	0	0	2	0	0
-1	0	-1	1	2	0	0	1	0	1	-1	0	0	-1	0	3	0	0	-1	0	2	-1
0	1	0	0	1	0	1	0	1	0	1	-2	0	0	0	1	0	1	0	1	0	1



Word vectors → Sentence vector

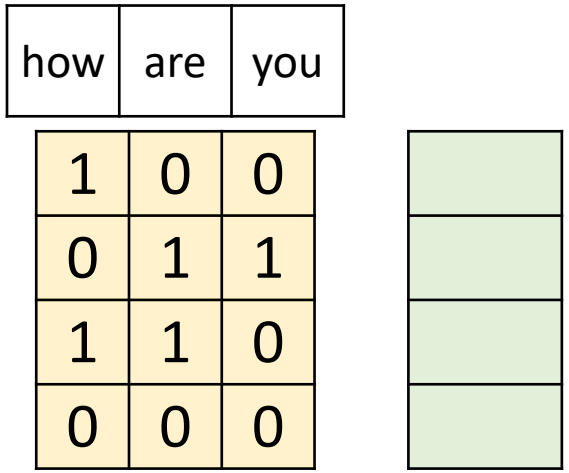
Method 1: Concatenate

how	are	you
1	0	0
0	1	1
1	1	0
0	0	0



Word vectors → Sentence vector

Method 2: Average



id	comment	user	emb
1	How are you?	John	
2	Who are you?	Mary	

“Who are you” → word embedding vectors

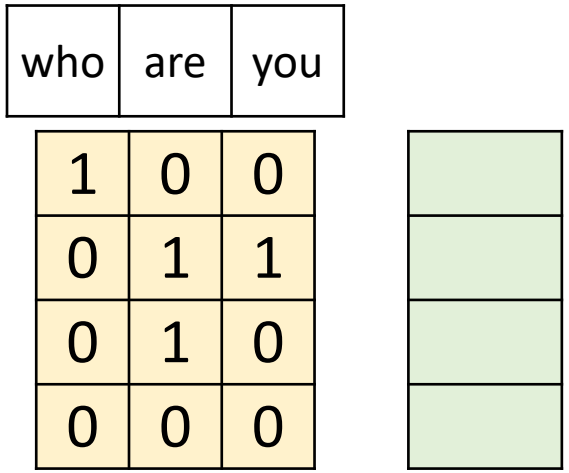
who	are	you

a	an	the	how	why	who	what	are	is	am	be	was	you	we	I	they	she	he	she	me	him	her
0	-1	0	1	0	1	0	0	-1	1	0	0	0	3	1	0	-1	0	0	0	-1	0
2	0	2	0	0	0	-1	1	0	0	0	2	1	0	2	0	2	0	0	2	0	0
-1	0	-1	1	2	0	0	1	0	1	-1	0	0	-1	0	3	0	0	-1	0	2	-1
0	1	0	0	1	0	1	0	1	0	1	-2	0	0	0	1	0	1	0	1	0	1

	0	0
	1	1
	1	0
	0	0

Word vectors → Sentence vector

Method 2: Average



id	comment	user	emb
1	How are you?	John	[1/3, 2/3, 2/3, 0]
2	Who are you?	Mary	

How to query by SQL?

_____ comment, emb<___>[__,__,__,__] AS score

FROM posts

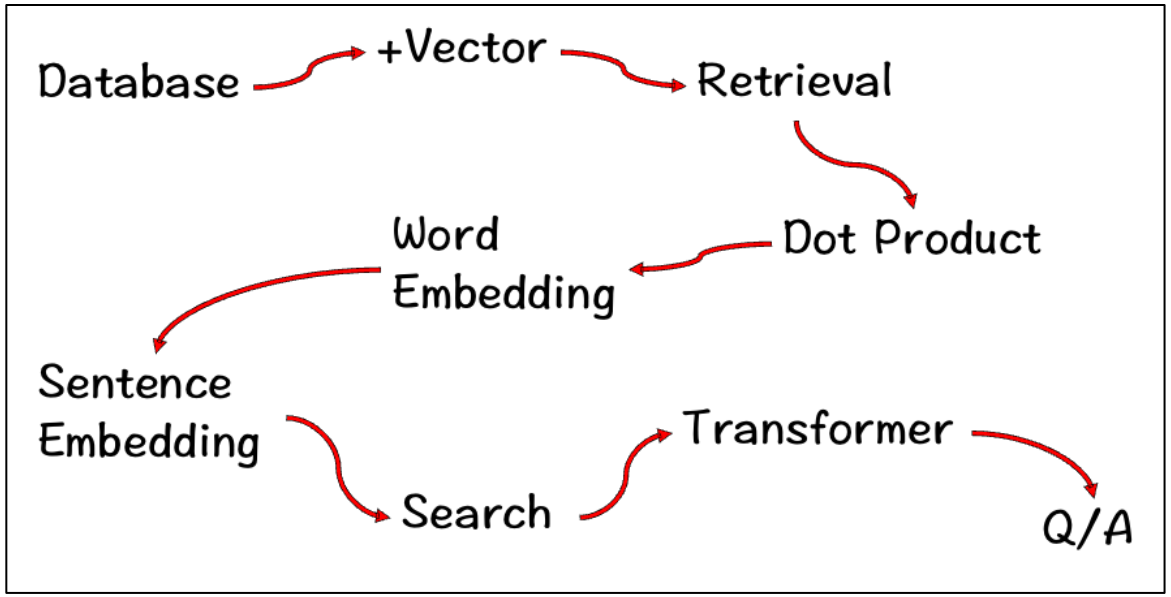
ORDER BY _____ ASC | DESC ;

How to query using a high-level API?

```
query = Query(post_index)
    ._________(post)
    ._________(relevance_space.text, Param("________"))

app.query(query,  _________ = "who are you?" )
```

Source: Superlinked.com

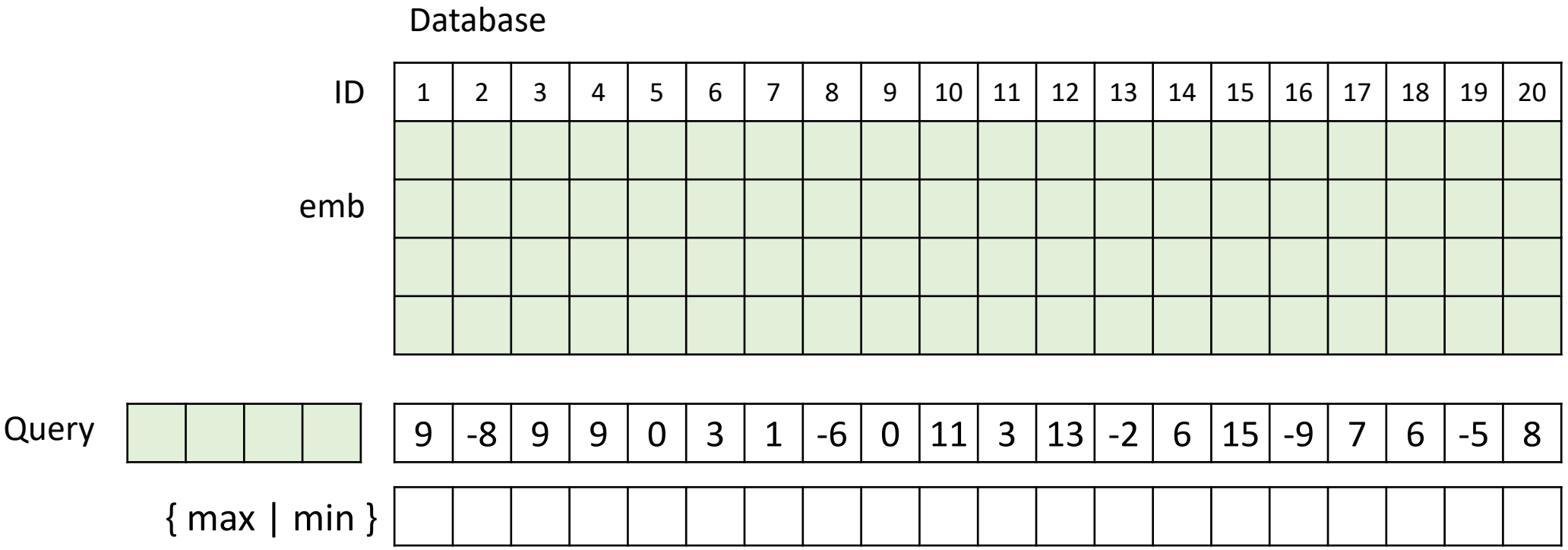


Search

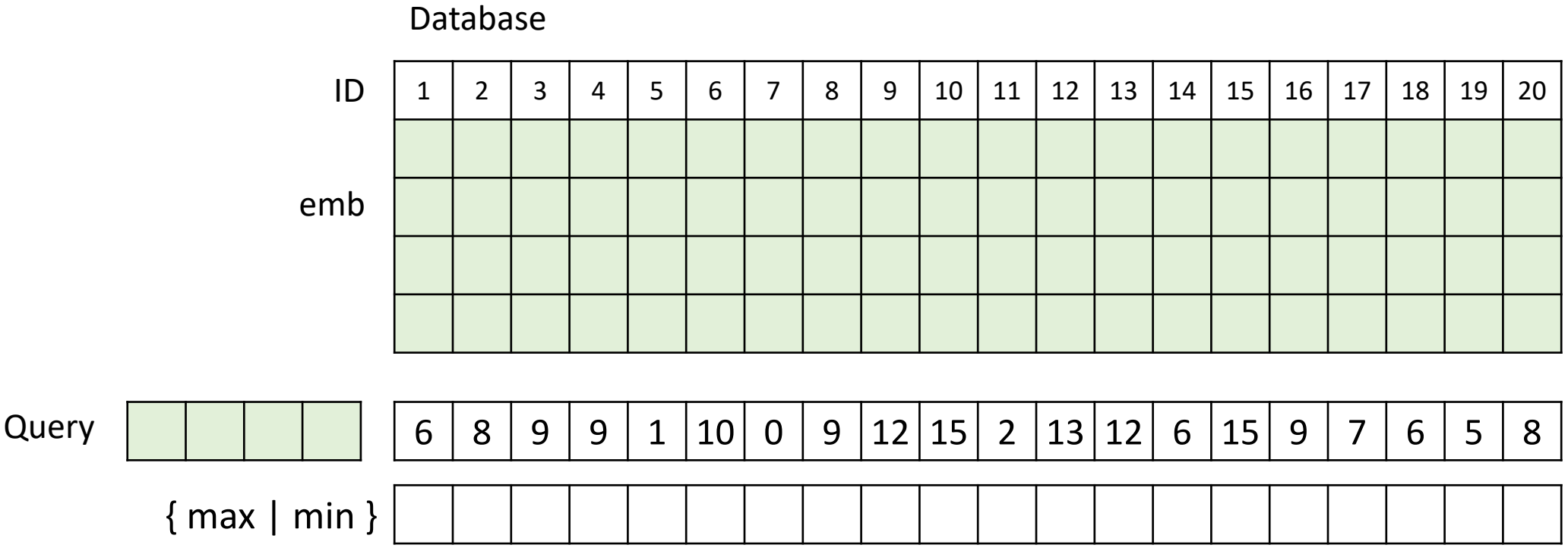
Beginner's Guide to Vector Databases - AI by Hand ✍️

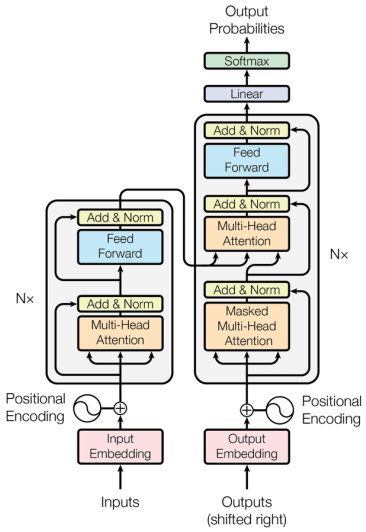
University of Colorado Boulder

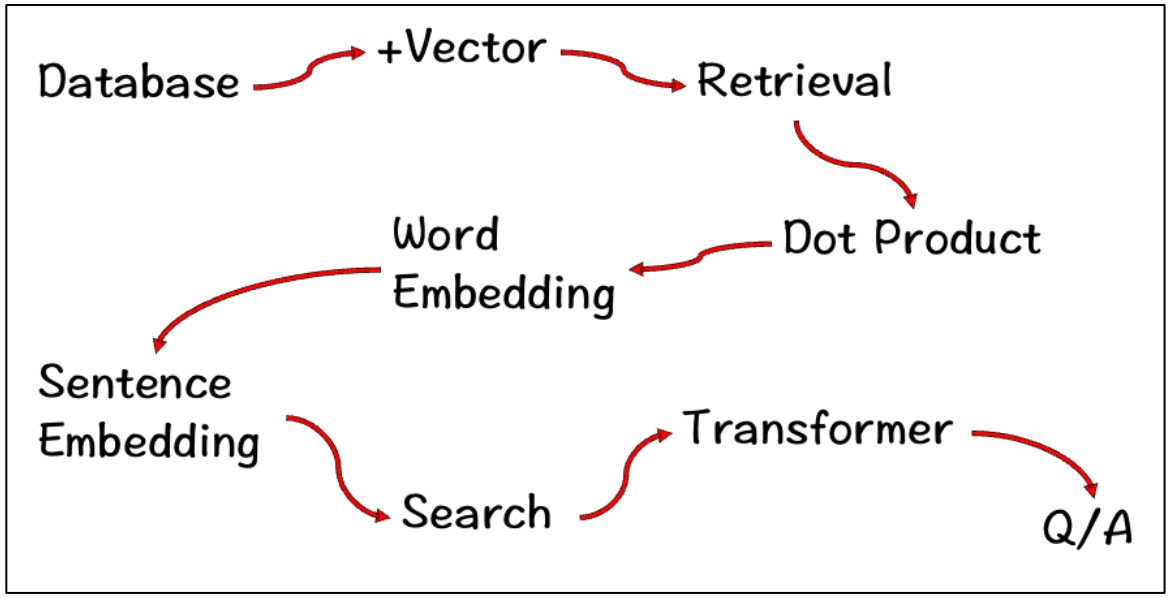
K-Nearest Neighbor, K=3, Dot-Product



K-Nearest Neighbor, K=3, L2







Transformer

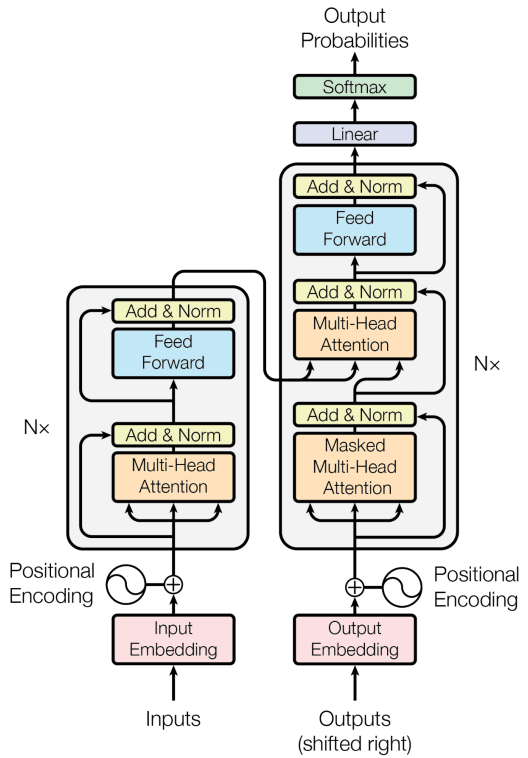
Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

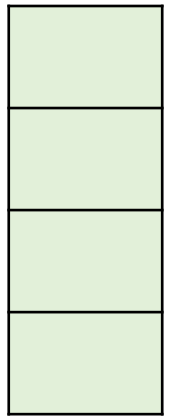
How to use a Transformer to get a sentence embedding vector?

Word Embedding Vectors

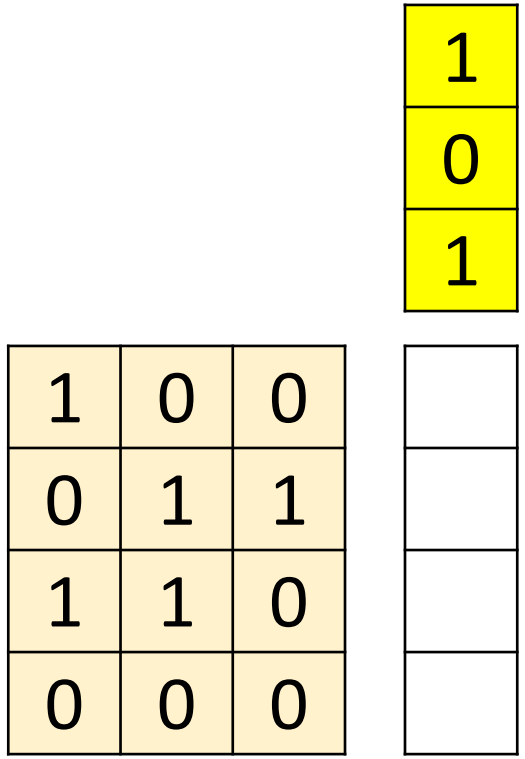
1	0	0
0	1	1
1	1	0
0	0	0



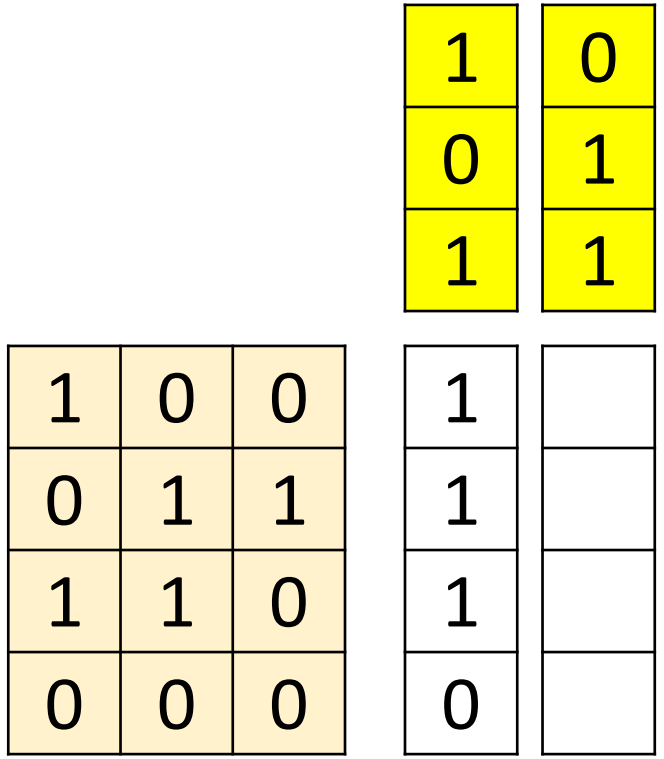
Sentence Embedding Vector



How to combine across positions?



How to combine across positions?



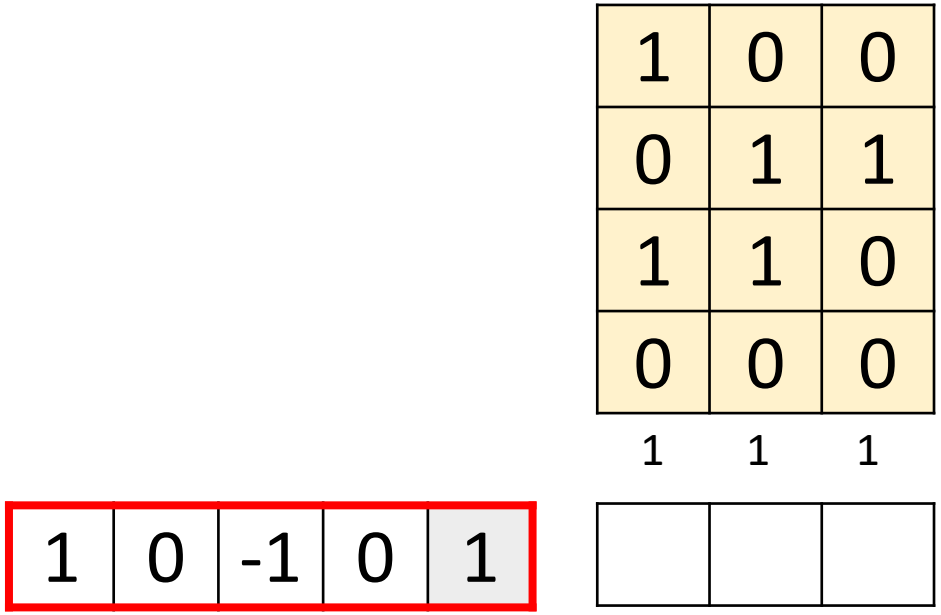
How to combine across positions?

1	0	0
0	1	0
1	1	1

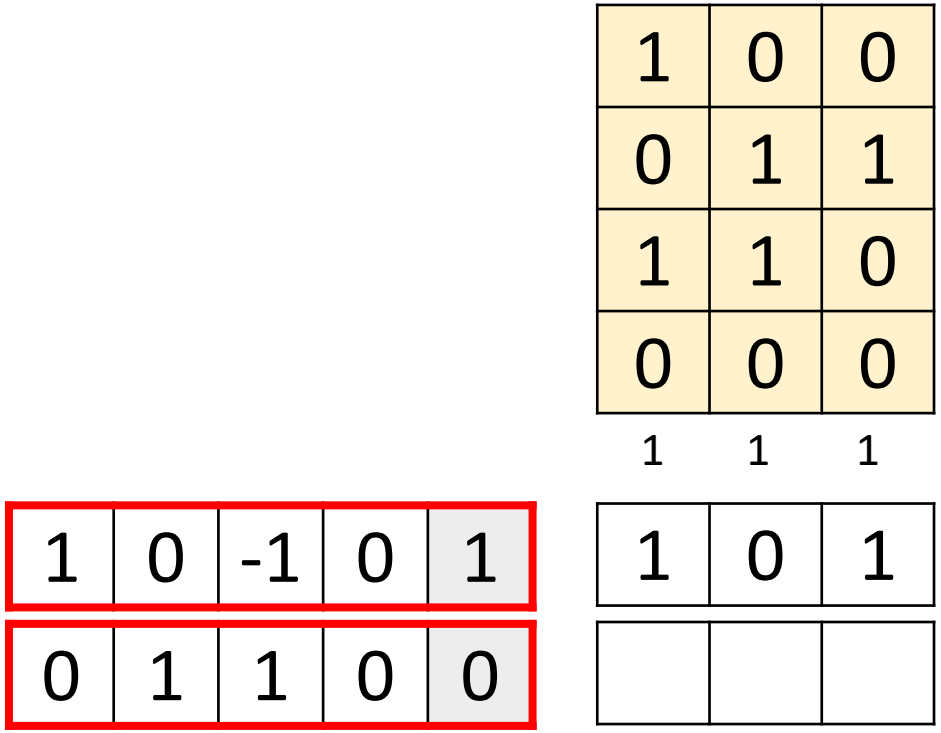
1	0	0
0	1	1
1	1	0
0	0	0

1	0	
1	2	
1	1	
0	0	

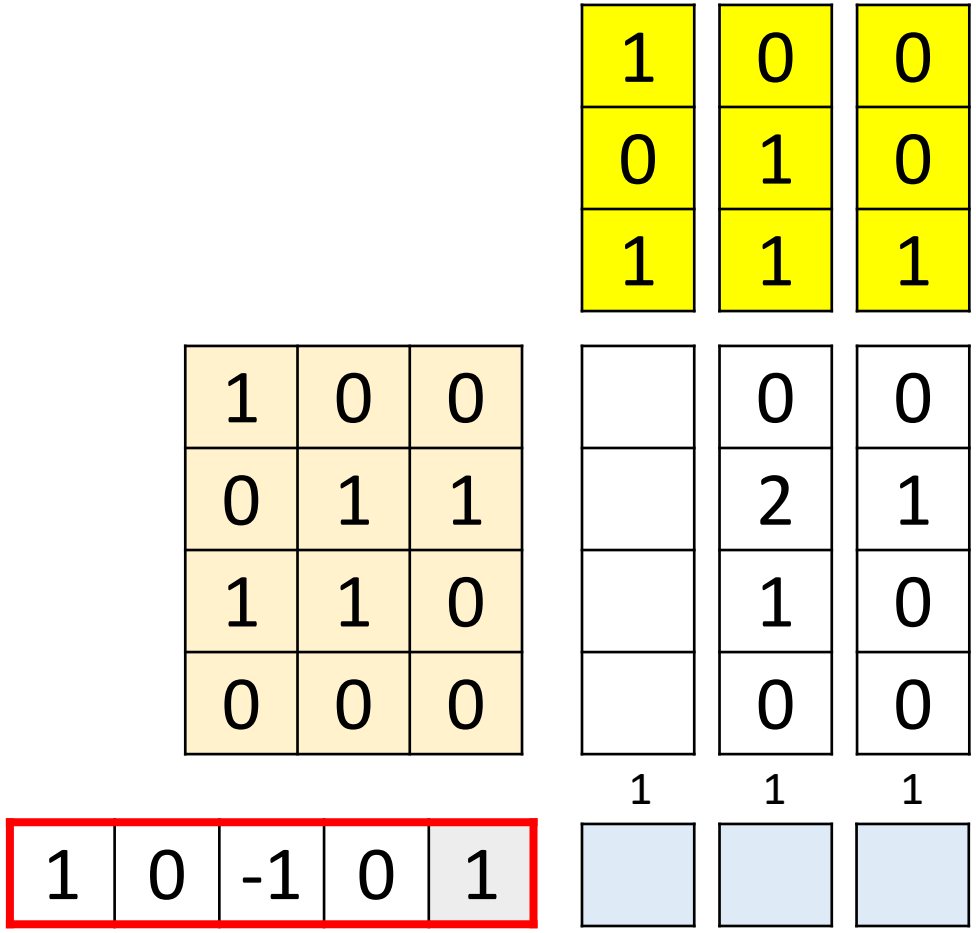
How to combine across features?



How to combine across features?



How to combine across positions and features?



How to use a Transformer to get a sentence embedding vector?

Word Embedding Vectors

1	0	0
0	1	0
1	1	1

1	0	0
0	1	1
1	1	0
0	0	0

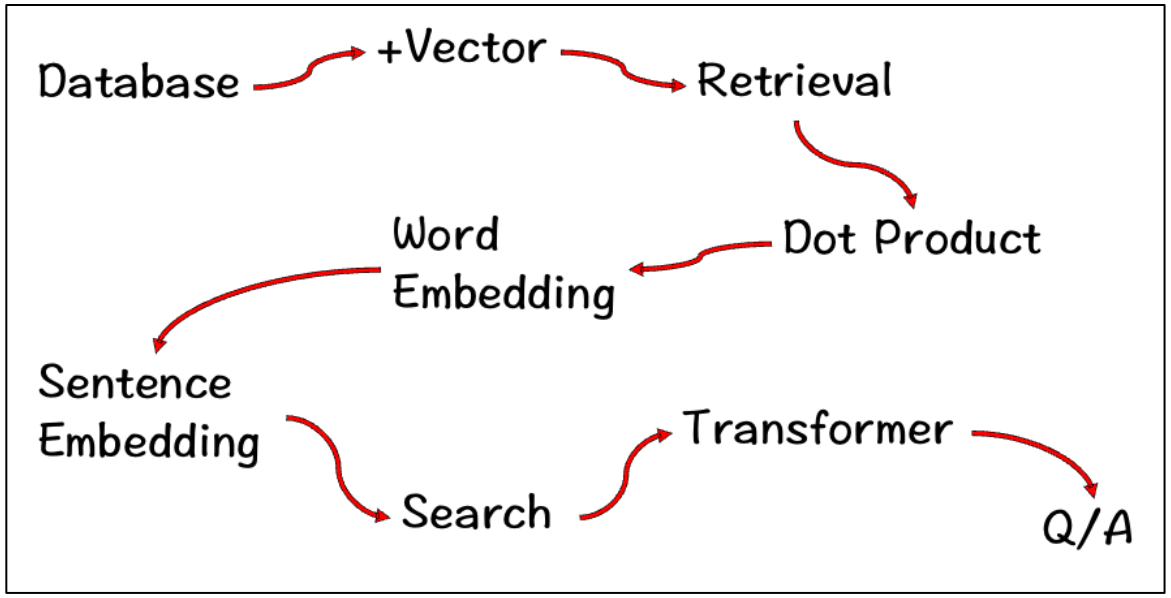
1	0	-1	0	1
0	1	1	0	0
0	0	0	1	1
0	0	1	1	0

	0	0
	2	1
	1	0
	0	0

1 1 1

2	3	1
1	1	1
1	1	0

Sentence Embedding Vector



Q/A

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder