

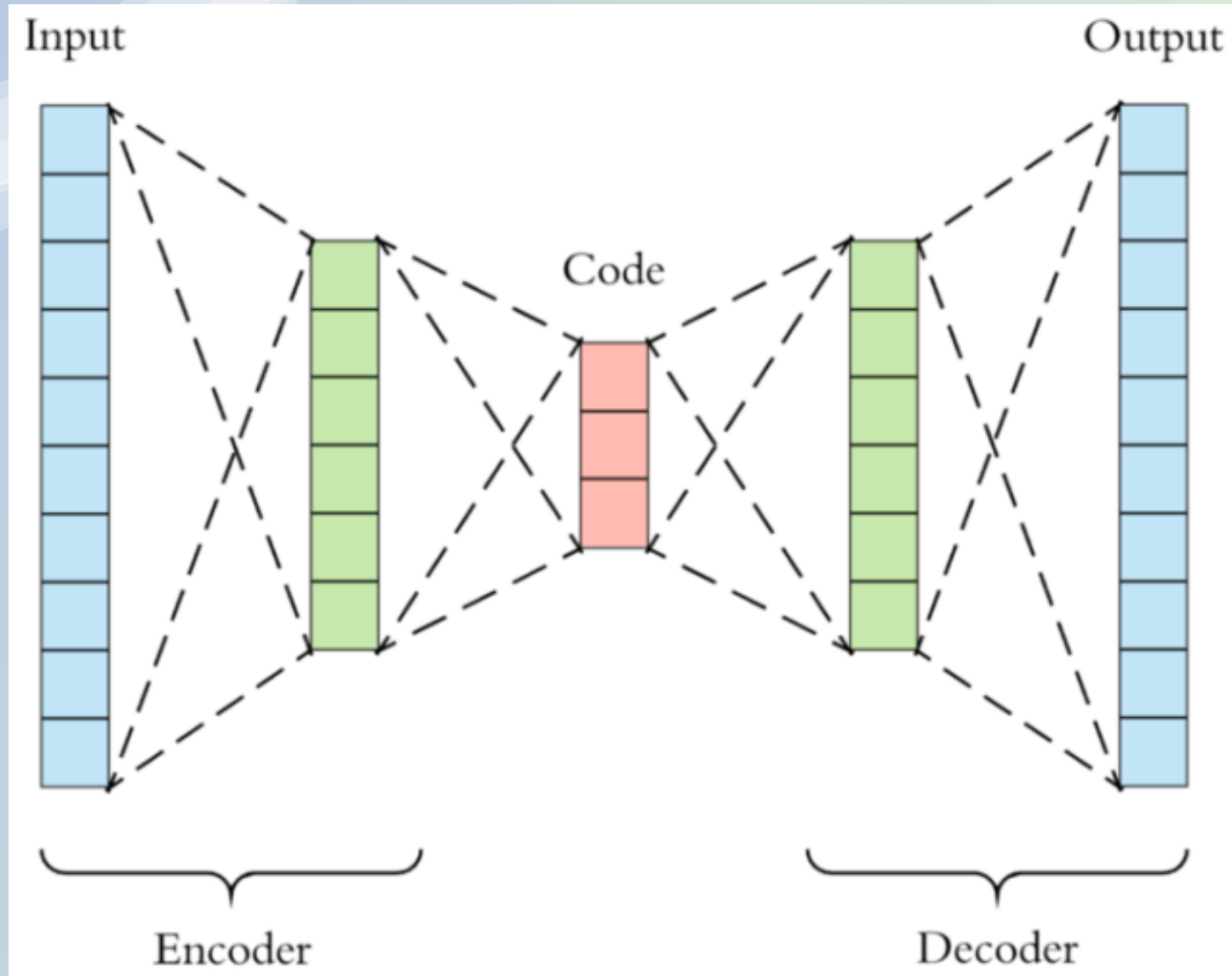
BASE DE DATOS DE VECTORES

- 12/08/2023

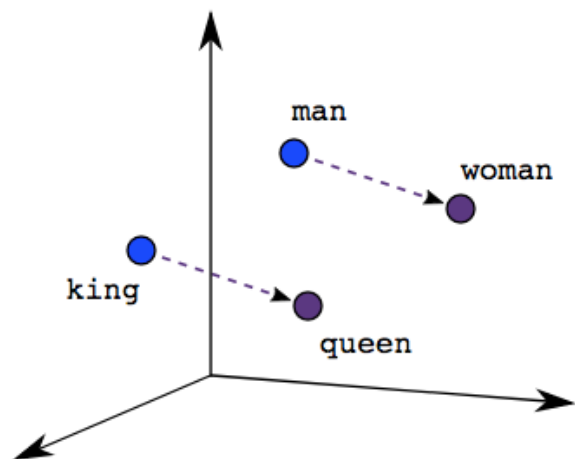
EMBEDDINGS VARIABLES LATENTES

- Un embedding es un espacio de dimensiones relativamente bajas en el que puede traducir vectores de dimensiones altas.
- Los embeddings facilitan el aprendizaje automático en entradas grandes, como vectores dispersos que representan palabras.

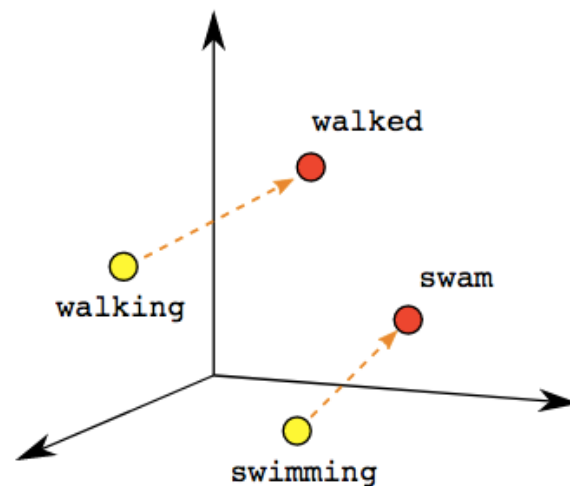
EMBEDDINGS



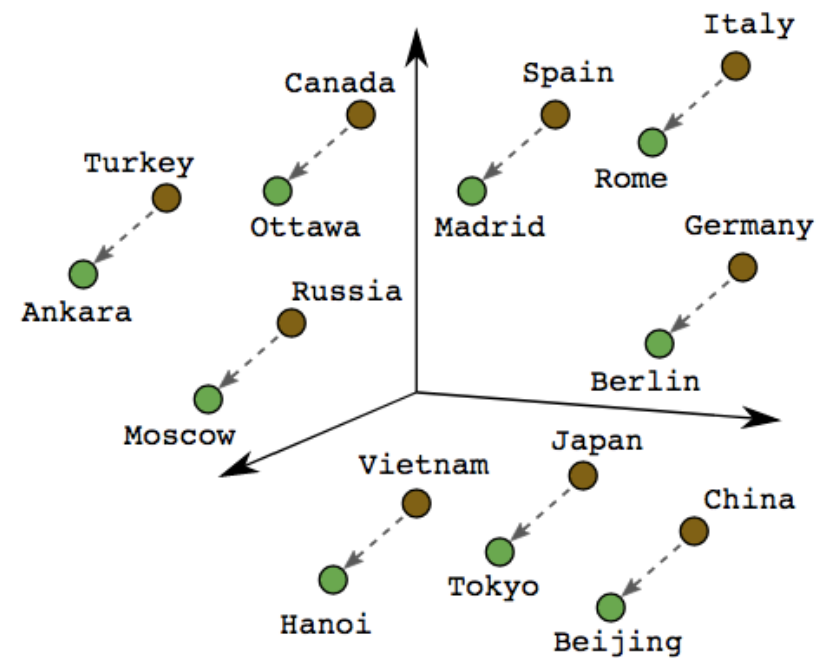
EMBEDDINGS



Male-Female



Verb Tense

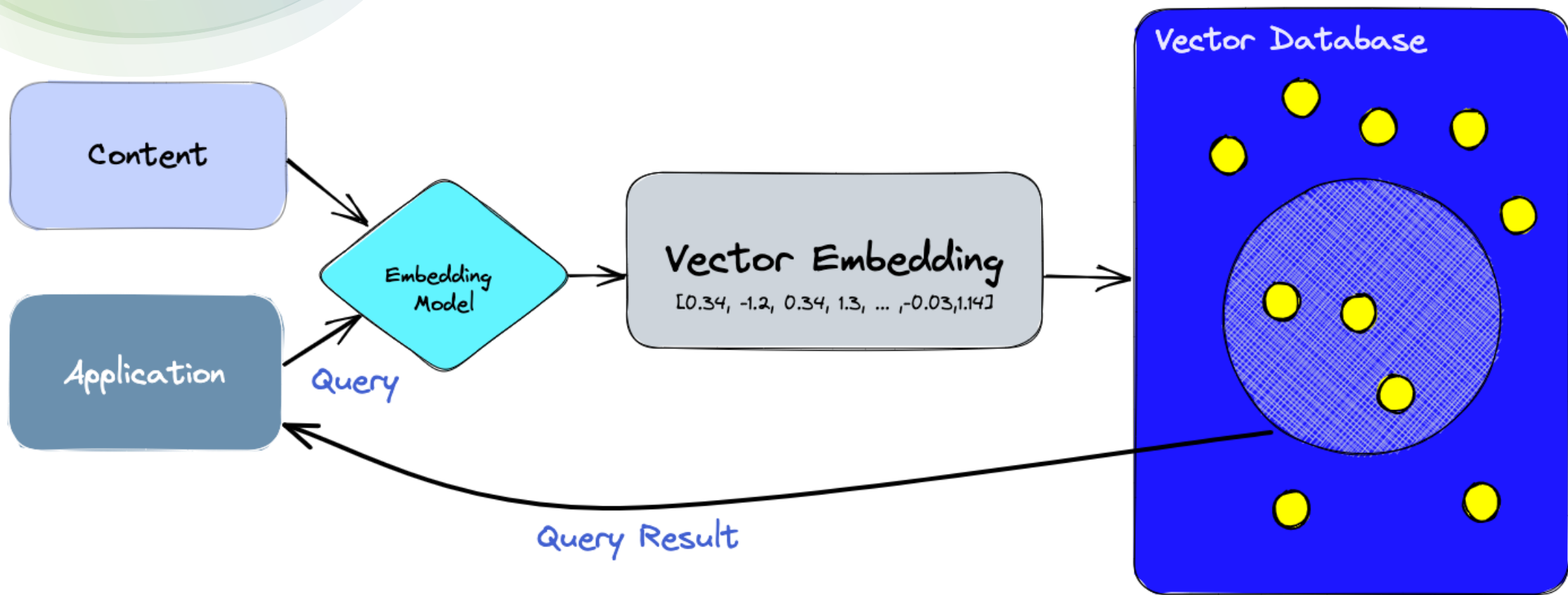


Country-Capital

BASE DE DATOS DE VECTORES

- A vector database is a type of database that stores data as high-dimensional vectors, which are mathematical representations of features or attributes. Each vector has a certain number of dimensions, which can range from tens to thousands, depending on the complexity and granularity of the data.

Workflow



METODOS DE BUSQUEDA

- ESTABLECER METRICAS Y MEDIDAS DE DISTANCIA
- COMO OBTENEMOS LOS VECTORES MÁS “PARECIDOS”



MÉTRICAS DE BUSQUEDA

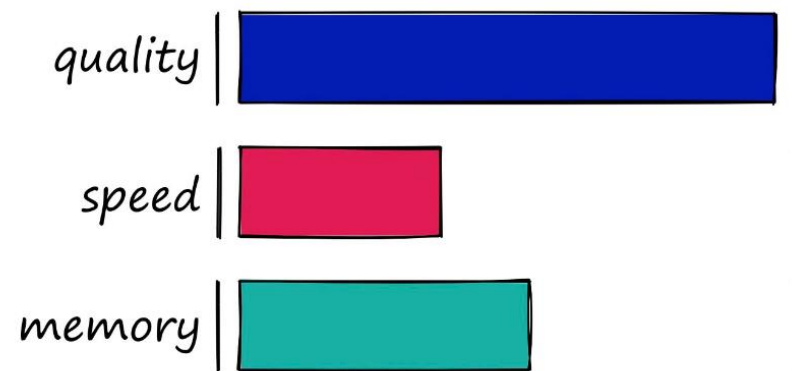
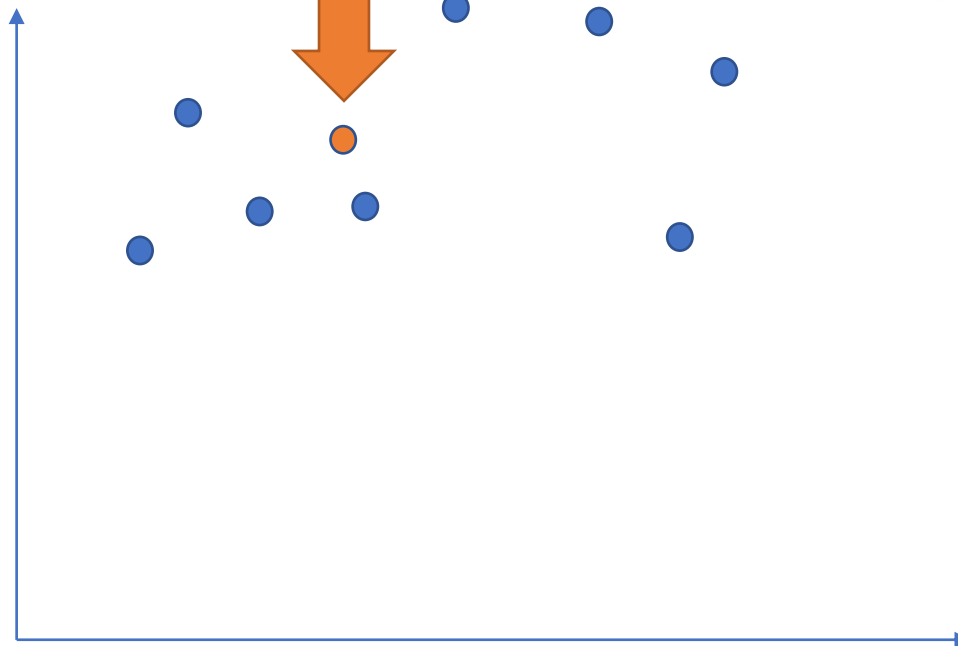
- Euclideana (**L2**)
- Producto Interno (**IP**)
- Similaridad por coseno (**normalize_L2**)



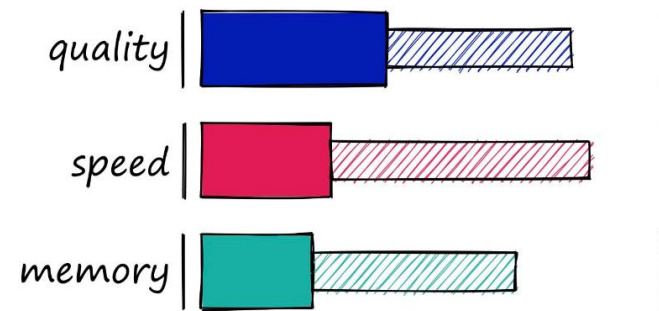
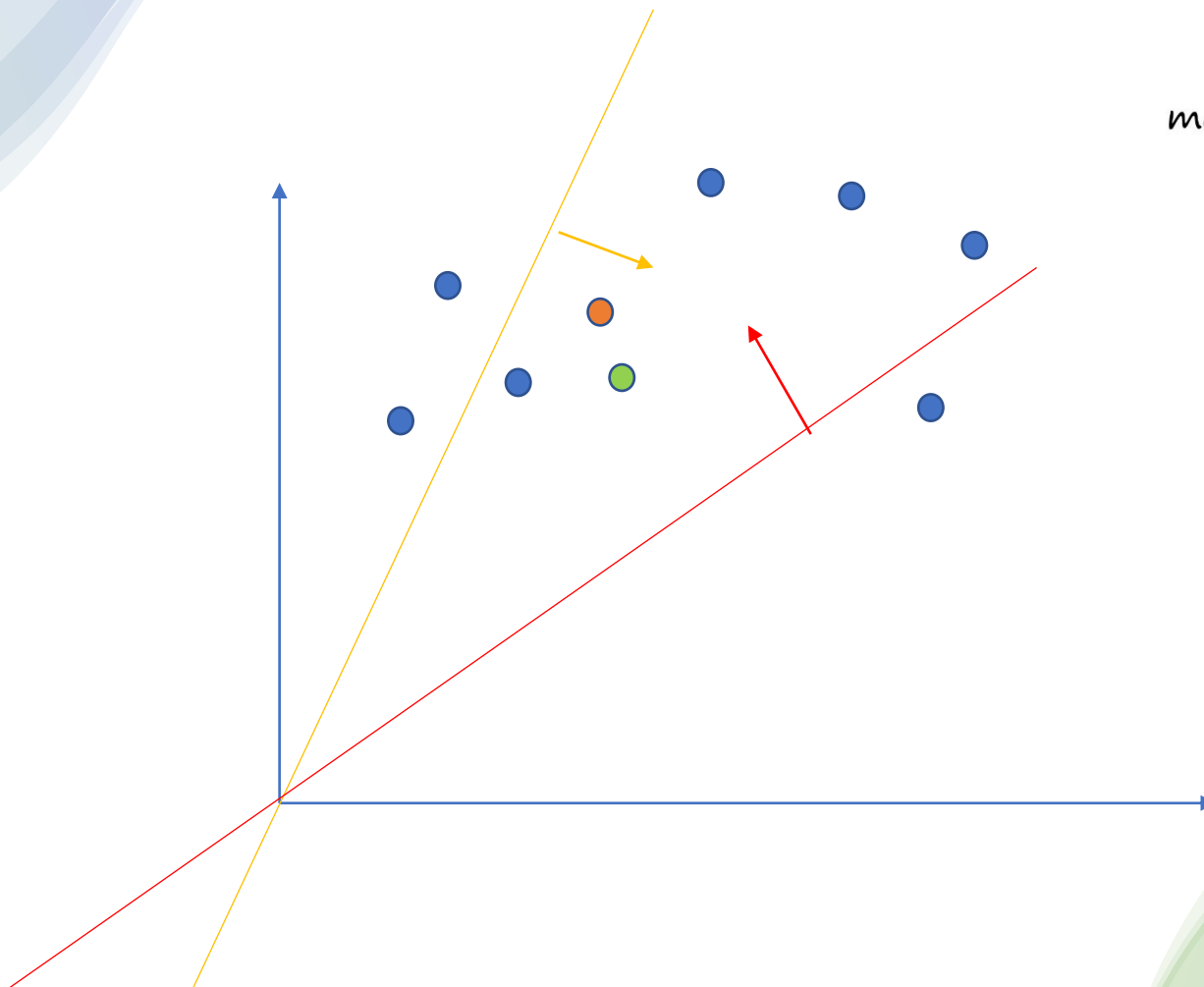
METODOS DE BUSQUEDA

- kNN (**FLAT**)
- Local **S**ensitivity **H**ashing (**LSH**)
- **H**ierarchical **N**avigable **S**mall **W**orld (**HNSW**)
- **P**roduct **Q**uantization (**PQ**)
- Inverted **F**ile Index (**IVF**)

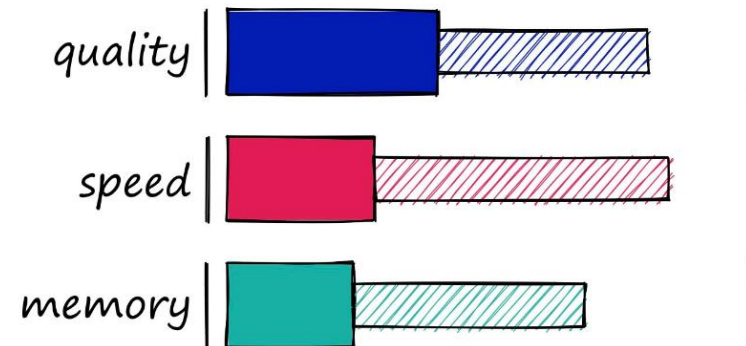
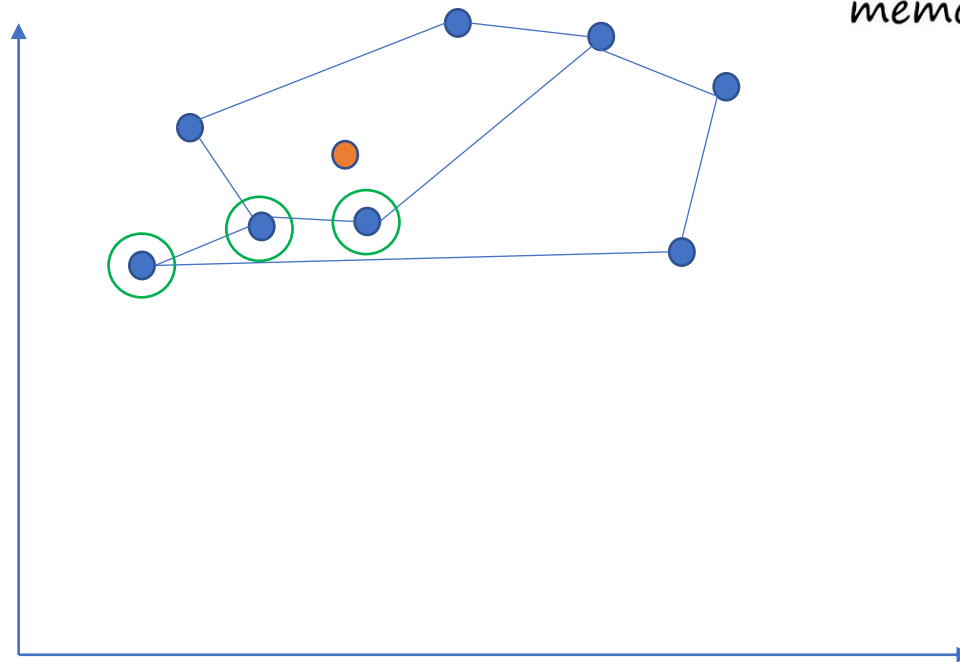
KNN



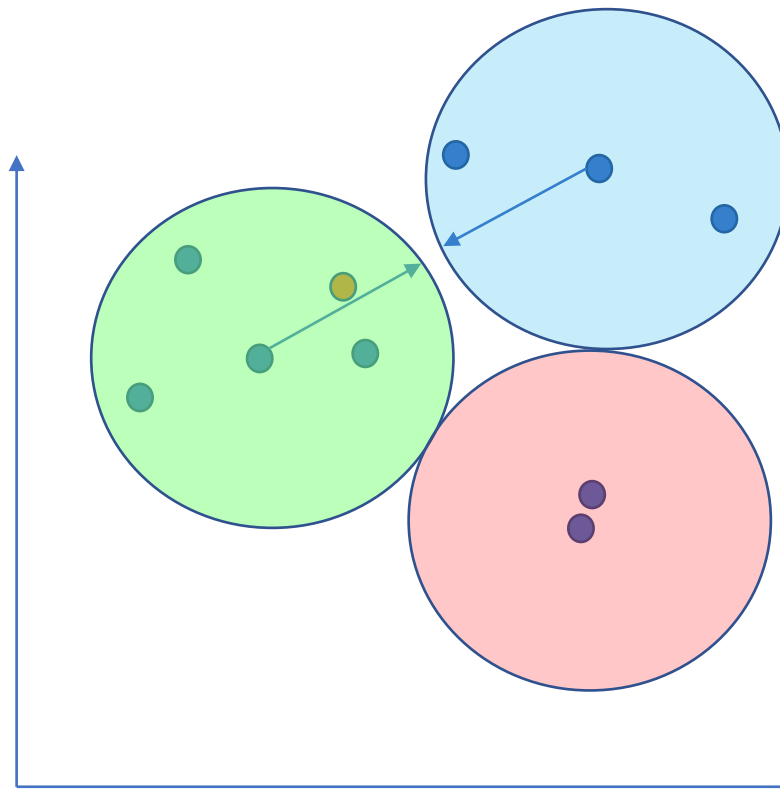
LSH



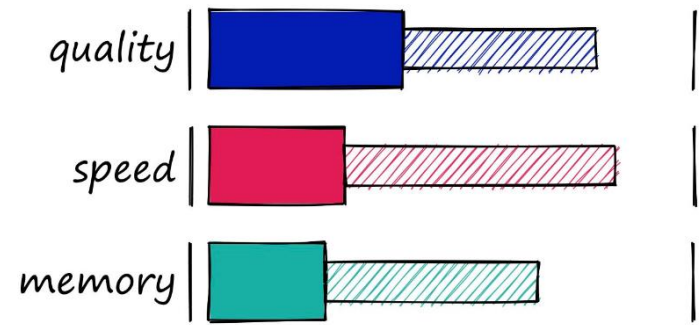
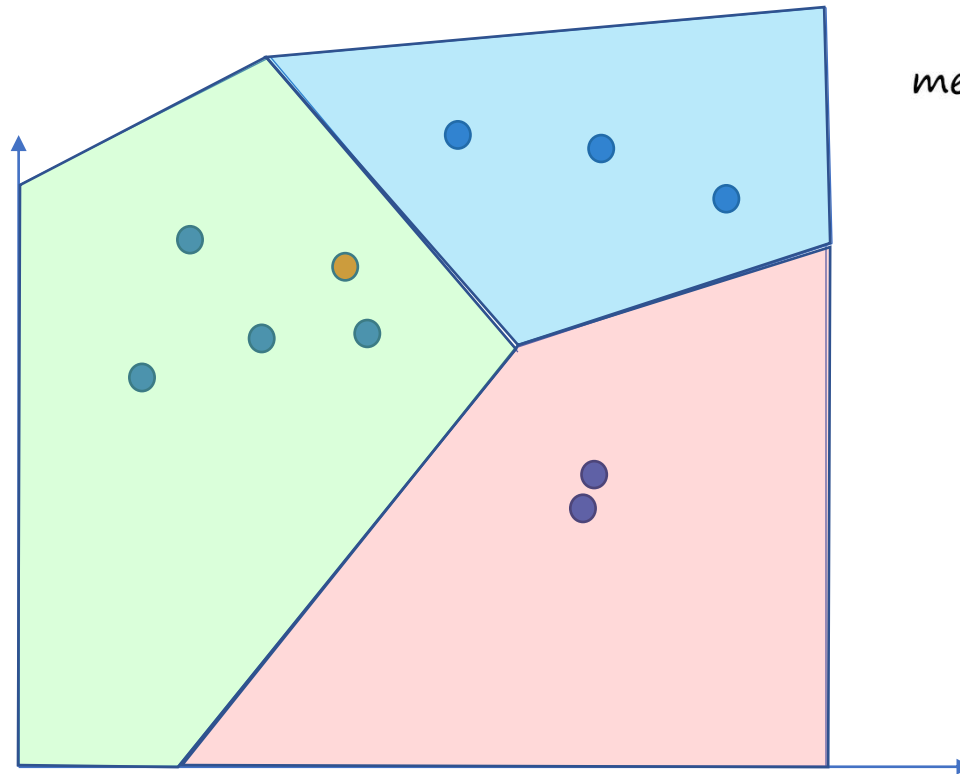
HNSW



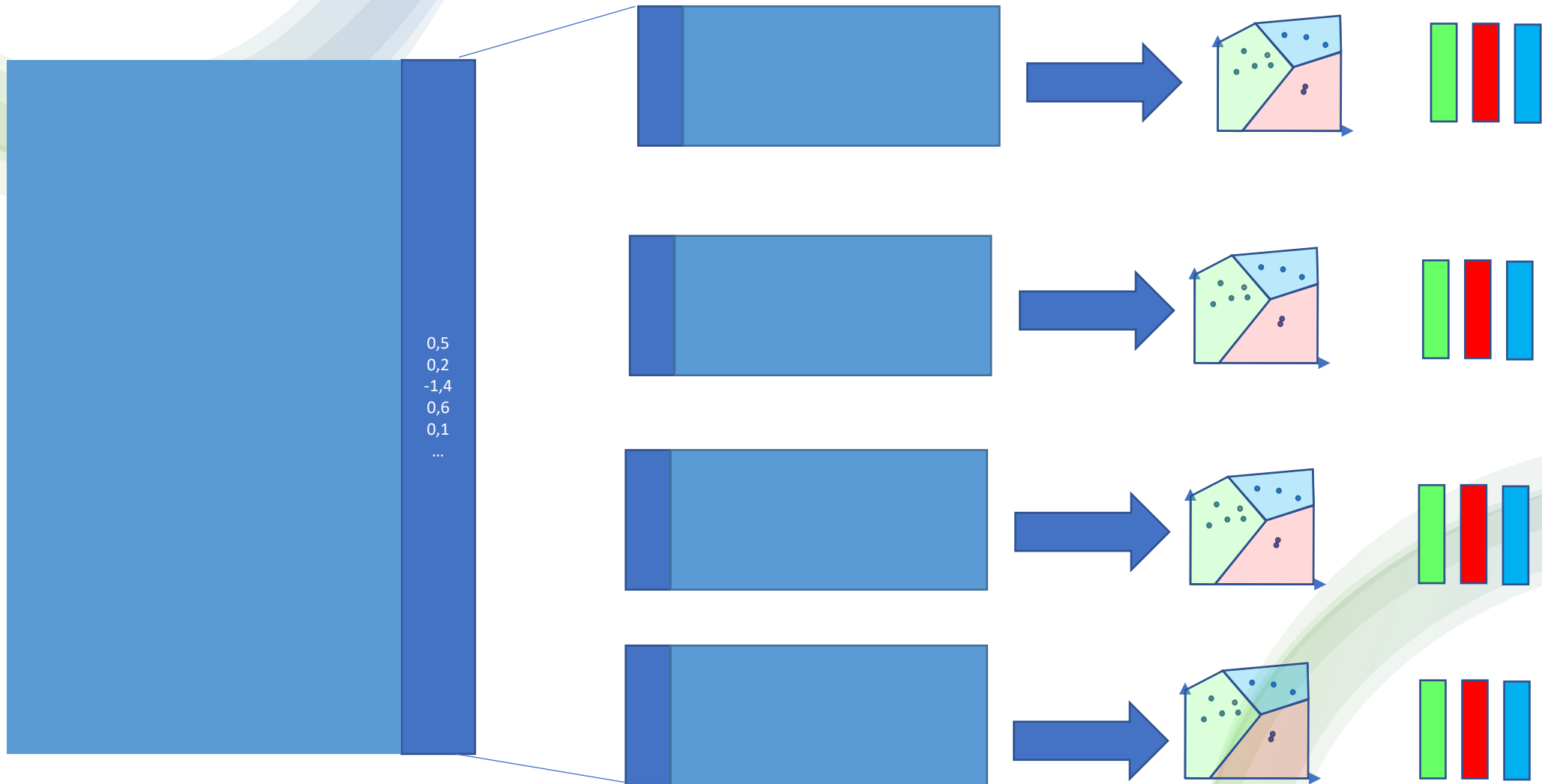
IVF



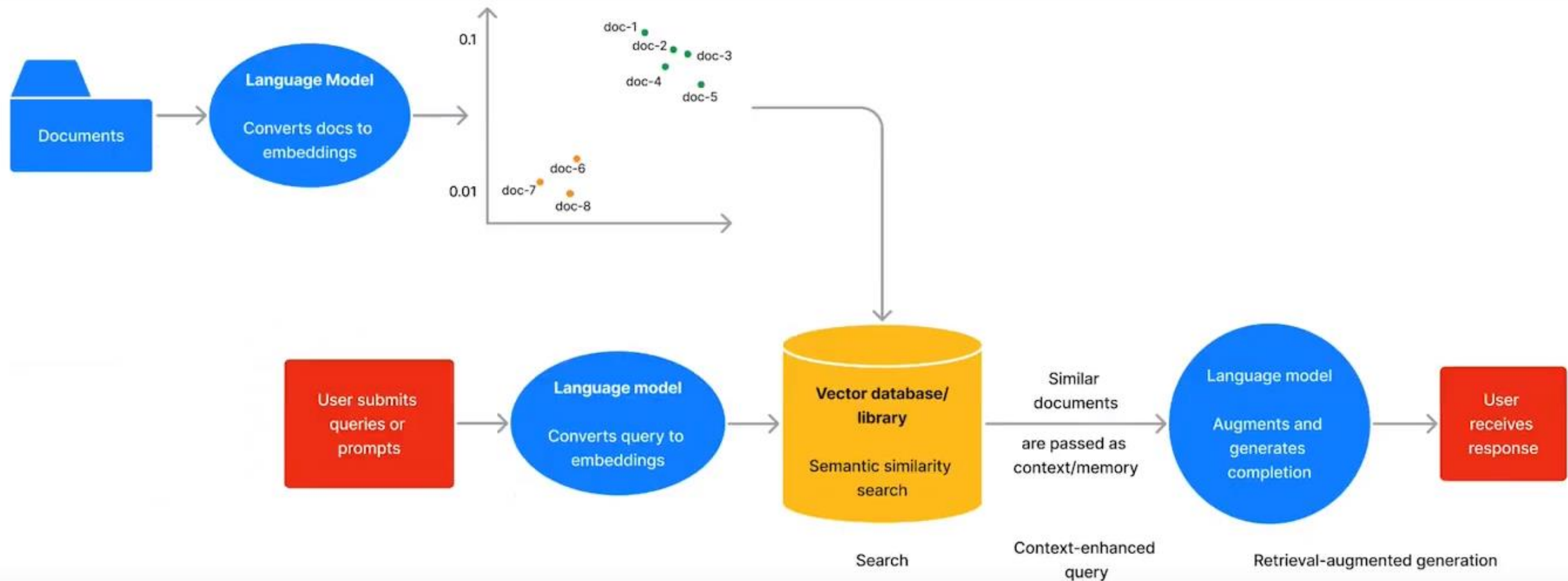
IVF



PQ Product Quantization















Workflow



Ranking

☐ include secondary database models

10 systems in ranking, July 2023

Rank			DBMS	Database Model	Score		
Jul 2023	Jun 2023	Jul 2022			Jul 2023	Jun 2023	Jul 2022
1.	1.	1.	Kdb 	Multi-model 	8.22	+0.22	-0.95
2.	2.		Chroma	Vector DBMS	2.41	+0.02	
3.	3.		Pinecone	Vector DBMS	2.27	+0.14	
4.	4.	 2.	Milvus 	Vector DBMS	1.36	+0.05	+0.99
5.	5.	 3.	Weaviate 	Vector DBMS	1.27	+0.19	+1.15
6.	6.		Vald	Vector DBMS	0.86	-0.04	
7.	 8.	 4.	Qdrant	Vector DBMS	0.61	+0.03	+0.54
8.	 9.		Deep Lake	Vector DBMS	0.55	+0.05	
9.	 7.		Vespa	Multi-model 	0.55	-0.04	
10.	10.		MyScale	Multi-model 	0.19	-0.07	

The background features a light blue-to-green gradient. On the left side, there are several overlapping, wavy, light blue shapes that curve upwards and to the right. On the right side, there are several overlapping, wavy, light green shapes that curve upwards and to the left.

PRACTICA