Notes on Vector Retrieval Methods and Techniques

compiled by D. Gueorguiev, 8/9/2025

# Vectors as Units of Retrieval

We often intend for the vector representation of two similar objects to be “close” to each other according to some well-defined distance function. Thus, similarity in the vector space must imply similarity between objects.

We must choose the dimensionality of the target space (a subset of ) for embedding along with the distance function . The vector space of embeddings together with the distance function defines a metric space.

# References

[1] [Foundations of Vector Retrieval, Sebastian Bruch, 2024](https://github.com/dimitarpg13/rag_architectures_and_concepts/blob/main/articles/Foundations_of_Vector_Retrieval_Bruch_2024.pdf)

[2] [Vector database management systems: Fundamental Concepts, use-cases, and current challenges, Toni Taipalus, 2024](https://github.com/dimitarpg13/rag_architectures_and_concepts/blob/main/articles/vector_db/Vector_database_management_systems-Fundamental_concepts_use-acases_and_current_challenges_Taipalus_2024.pdf)

[3] [Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, Patrick Lewis, Ethan Perez et al, 2021](https://github.com/dimitarpg13/rag_architectures_and_concepts/blob/main/articles/Retrieval-Augmented_Generation_for_Knowledge-Intensive_NLP_Tasks_Lewis_2021.pdf)