

LLM Apps

Advanced concepts to optimize the RAG technique

Advanced concepts to optimize the RAG technique

- Alternative techniques for indexing and storing vectors.
- Alternative techniques for measuring similarity between vectors.

Techniques for storing and indexing vectors

- Hierarchical Navigable Small World, HNSW
- Locally Sensitive Hashing, LSH
- Product Quantization, PQ
- Filtering (pre and post filtering)
- Random Projections

Techniques for measuring the similarity between vectors

- The most common are:
 - cosine similarity (angle between vectors).
 - euclidean distance (straight line between items).