Vector Stores and Embeddings: The Future of Search

Learn how vector stores and embeddings transform search and recommendations.

NG by Nisaharan Genhatharan

What is a Vector Store?

Specialized Database

Designed for fast similarity searches on dense vectors.

Optimized Algorithms

Uses k-Nearest Neighbors (k-NN) for efficient retrieval.

Example Systems

- FAISS
- Pinecone
- Milvus
- Chroma

Popular Vector Store Technologies

FAISS

Facebook AI Similarity Search, high-performance library.

Pinecone

Cloud-native vector database, easy to scale and integrate.

Milvus

Open-source, supports billion-scale vector data.

Chroma

Platform focused on embeddings management and search.

What are Embeddings?

Numerical Representations

Encode text, images, or audio into high-dimensional vectors.

Semantic Meaning

Capture context and meaning beyond simple keywords.

Similarity Comparisons

Enable measuring closeness via cosine similarity.

How Embeddings are Created

1 Trained Models

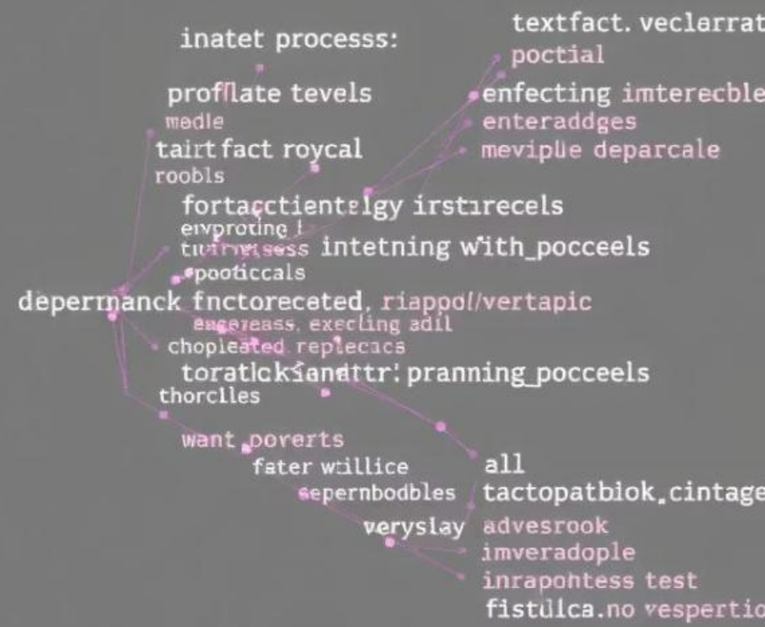
BERT, Sentence Transformers, and OpenAI create embeddings.

2 Vector Mapping

Transforms raw data into numerical vectors.

3 Contextual Relationships

Embedding space reflects semantic similarities.



Use Case: Semantic Search

Traditional Search

- Keyword-based
- Limited understanding of meaning
- Exact matches only

Semantic Search

- Understands intent and context
- Finds related and relevant results
- Boosts accuracy

Example: Netflix Movie Recommendations

Plot Embeddings

Netflix encodes movie plots into high-dimensional vectors.

FAISS Search

Instantly finds movies with similar themes using k-NN.

Personalized Suggestions

Surfaces relevant movies right after you finish watching.

Example: LinkedIn Job Matching

Embedding Profiles

Résumés and job posts are converted into vectors.

Cosine Similarity

Measures closeness to find suitable job matches.

Enhanced Matching

Improves hiring efficiency and candidate relevance.