

Semantic Search Overview

Semantic search overview

Semantic search uses vector embeddings to represent text meaning and retrieve relevant documents by similarity.

A typical pipeline: split documents into chunks, compute embeddings, store vectors in an index, and search with a query embedding.

Cosine similarity is common; approximate nearest neighbor search improves speed on large collections.