semantee

CONF42 LLMS 2024

APRIL 11, 2024

# Vector databases are not new, but LLMs empowered them

Brief overview

o History dates back to 2005 (MonetDB)

o New databases have been sprouting continuously

o Big boom in 2022 thanks to ChatGPT (and others)

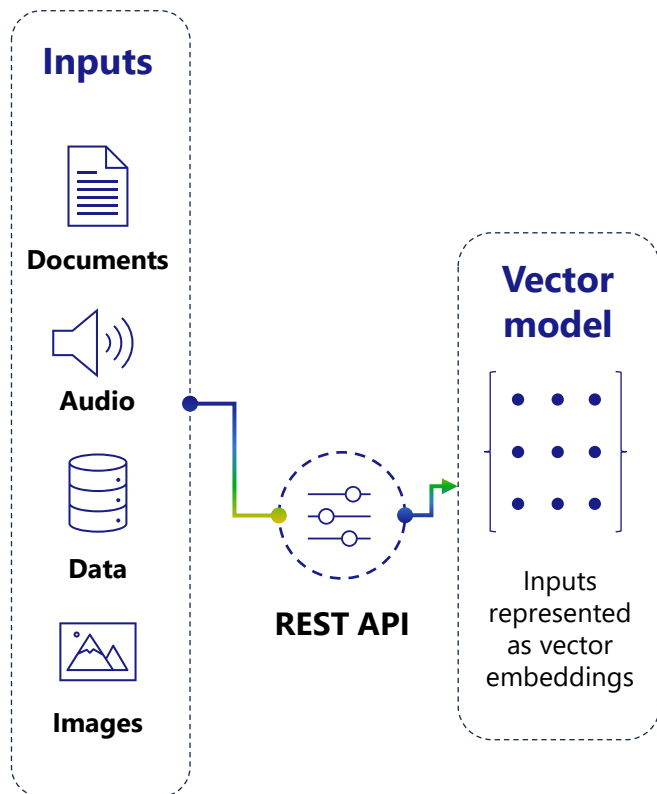o By 2026, 30% of enterprises will have adopted vector databases. (Source: Gartner)

What are vector databases?

o Store data in vector representation

o Words or sentences are stored as vectors and their similarity can be compared (cosine, Euclidian, Jaccard distance)

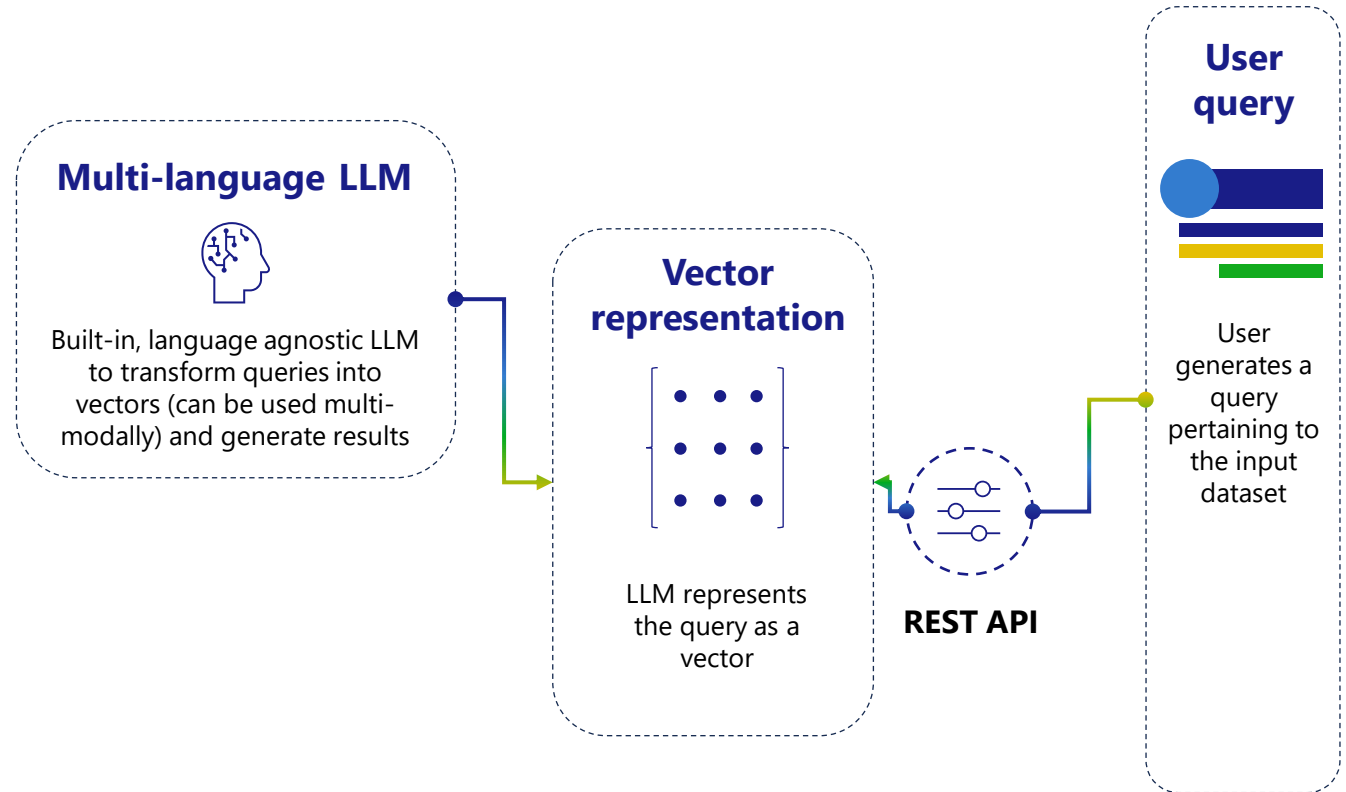o Indexing is used by ANN algorithms to search for nearest neighbor

KING [1, 1]
QUEEN [2, 2]
MAN [1, 1.5]
WOMAN [2, 2.5]
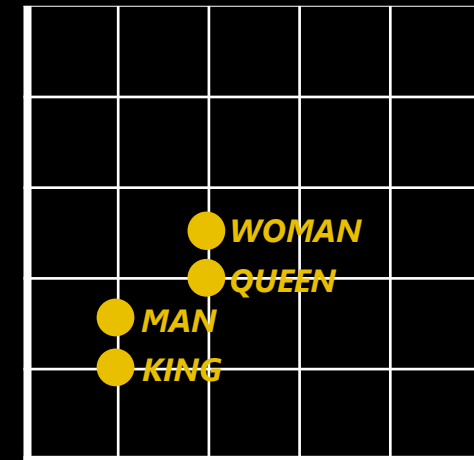
semantee

To set up semantic search, several technologies must be assembled.
Some offer just the base vector model...

...others will give you the LLM...

Multi-language LLM

Built-in, language agnostic LLM to transform queries into vectors (can be used multi-modally) and generate results

Vector representation

LLM represents the query as a vector

REST API

User query

User generates a query pertaining to the input dataset

semantee

...so we decided to take the dirty work out of the game.

semantee

Inputs
- Documents
- Audio
- Data
- Images

REST API

Vector model
Inputs represented as vector embeddings

Multi-language LLM
Built-in, language agnostic LLM to transform queries into vectors (can be used multi-modally) and generate results

Similarity Search
Nearest neighbor represented via HNSW

Vector representation
LLM represents the query as a vector

REST API

User query
User generates a query pertaining to the input dataset
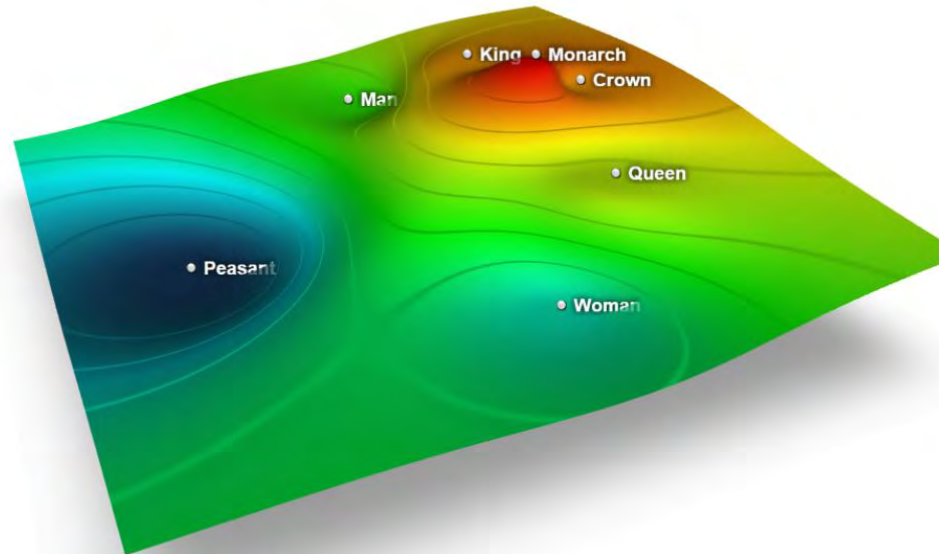
Results

semantee

And we eliminated numerous showstoppers on the way

o   Which LLM to use?

o   Which similarity search algorithm works best?

o   What to do about large datasets?

o   How to curb the costs?

o   What to do about re-indexing the database?

o   How to implement semantic search in practice?

o   When to fall back to keyword search?

o   How to semantically search a large SQL database?

o   **And we could go on...**

semantee

We added a nice visualization layer to top it all off

3D visualization of the semantic proximity of words contained in a dataset

Almost no relevance   Very low relevance   Moderate relevance   Very high relevance   Extremely high relevance

The closer the objects appear on the map, the higher their semantic similarity.

1 **Visualize semantic search results** both for your internal use or as a service to the end user

2 Click through the map to **reveal nested relationships** between datapoints

3 **Uncover surprising connections** between datapoints in an easily digestible manner

4 Depict the **position of a datapoint relative to the whole** embedding space

semantee

The real world use cases for this technology supercede search

Semantic search and categorization open a whole host of new horizons for interesting use cases:

Semantic Text Search
Semantee search doesn't rely on specific words; it **understands the meaning** behind them. It works great with many long texts and descriptive, contextual queries. You can easily set up a smart search for your data in just a few minutes.

Recommendations
Think of how people act as a special pattern that can be represented mathematically. Semantee helps turn this into **real-time suggestions for your customers**.

Similar Image Search
Find pictures that look alike, discover duplicates, or locate a picture using text descriptions. Use pre-trained models and tweak them to make them work even better based on your data.

Anomaly Detection
Semantee can even **help find things that are out of the ordinary**. It can detect fraud, security issues or bugs much more efficiently and scalably than traditional database models.

Chatbots
Semantee can **supercharge your chatbots**. It understands what people are asking and can help answer questions automatically drawing from your internal data. Combine it with other tools to build a smart FAQ system.

Matching Engines
When you want to find things that are similar but have a lot of **different conditions**, Semantee is the perfect tool.

semantee

Talk is cheap: let's see a demo

Click the thumbnail above or [THIS LINK]

Thanks for paying attention!

Ben Fistein

Co-Founder
Semantee.ai

✉ **ben@semantee.ai**

semantee