

SVOP: Improving document retrieval accuracy through semantic search

17.7.2026 - | Kempelenov inštitút inteligentných technológií

Project name: SVOP semantic search.

- **Project's period:** approx 10/2024 until 12/2024, 3 months
- **Partner:** SVOP

The goal of the project was to improve full-text search to return more accurate results. Traditional word-based search approaches were to be replaced by semantic search, a method that compares text based on meaning rather than exact word matching, enabling more relevant results to be returned in response to user queries.

How we approached it

We implemented a semantic search system built around the Milvus vector database. The solution operated as:

- **Document embedding:** An embedding model from OpenAI was used to transform entire documents into vectors, which were then inserted into the Milvus database. This encoding captures the meaning of each document rather than its literal word content.
- **Query processing:** User input was transformed into vectors using the same embedding model, then compared against the stored document vectors. Documents with vectors most similar to the user input were returned as results.
- **Testing:** The solution was tested using multiple approaches to verify the quality and accuracy of the returned results.

"One of the most rewarding aspects of our work as a research institute is seeing research move beyond the research "lab" and into practice. This project is a great example of how expertise in AI models and semantic technologies can be transformed into a solution that measurably improves the performance of an operational system."

Martin Tamajka
Technology lead, KInIT

"The project was interesting because it was a prototype of semantic search, which is an approach based on the meaning of the text. The technology used, namely semantic search using vector databases, is a very powerful tool for finding duplicates or significantly similar texts, and the quality of this technology will only grow with the improvement of embedding models and vector databases."

Since the client was technically savvy and communicative, the project developed smoothly and helped the client improve their product.”

Denis Mitana
LLMs Lead, NanLab

What we delivered

The partner used our solution as a reference to implement their own semantic search system, built on the ElasticSearch database.

“Our collaboration with KInIT within the Hopero project brought us many new insights and showed us new directions that we had not dared to consider before. The consultations, combined with practical demonstrations, significantly accelerated our ongoing development and enabled its expansion in multiple directions.”

Mgr. Ján Grman, PhD.
Director, Solution architect, SVOP

<https://kinit.sk/svop-improving-document-retrieval-accuracy-through-semantic-search>