

PHAROS Training Series - Course 9 "RAG End-to-End: Architecture, Retrieval, Generation and Evaluation"

Contribution ID: 6

Type: **not specified**

Retrieval strategies: semantic, lexical, hybrid retrieval, HyDE, RRF and reranking

Tuesday, 7 July 2026 13:05 (25 minutes)

This methodology session examines the retrieval stage, which determines what evidence the language model can use when generating an answer. It compares dense semantic retrieval, sparse keyword retrieval, structured retrieval and hybrid retrieval, explaining when each modality is useful. Participants will see why semantic search is valuable for paraphrases and user-friendly language, while keyword or BM25-style search is important for exact administrative terms, identifiers, article numbers, dates and rare phrases. The session also introduces query enhancement techniques such as rewriting, expansion and HyDE, which can bridge the gap between conversational user queries and formal source language. It then explains fusion, Reciprocal Rank Fusion and reranking as ways to combine complementary retrieval signals and improve the final Top-K evidence. Participants will understand how retrieval choices affect recall, precision, context noise, latency and final answer reliability. The session prepares participants to experiment with these choices in tutorial exercises using realistic Greek queries.

Presenter: DROSATOS, George (ATHENA RC)