

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

Contribution ID: 5

Type: **not specified**

Hands-on 2: chunking, metadata and vector indexing

Tuesday, 7 July 2026 12:35 (15 minutes)

This second hands-on tutorial applies the knowledge preparation concepts introduced in the methodology session. Participants will create document chunks from the loaded public-service dataset, attach useful metadata and generate embeddings for each chunk. The session demonstrates how metadata such as document title, section, source identifier and chunk position helps preserve traceability and makes later retrieval more controllable. Participants will then build a local vector index in the Colab environment and run an initial similarity search to inspect the retrieved evidence. The emphasis is on visibility and debugging: participants will read example chunks, check whether they are coherent, verify that metadata is correct and examine whether the first retrieval results are meaningful. By the end, participants will have a working indexed knowledge base and will understand how data preparation decisions shape everything that follows in the RAG pipeline. This result becomes the basis for hybrid retrieval, reranking and answer generation exercises.

Presenter: GYFTOPOULOS, Sotiris (ATHENA RC)