Codersarts Ai

Metadata Filtering & Schema-Aware Retrieval

About the Course

Course Overview

This course will teach you how to build metadata-aware retrieval systems for modern RAG applications. Metadata filtering is the practice of using structured attributes such as department, role, date, access level, document type, and source to retrieve the right information with greater precision and control.

We will start by learning why vector similarity alone is often insufficient for production RAG systems. You will explore when metadata filtering becomes necessary, how schema-aware design improves retrieval quality, and how to preserve metadata during chunking and ingestion.

Then, we will move into practical retrieval workflows, including pre-filtering, combining multiple filters, and advanced filtering patterns such as dynamic filters and query-time inference. Finally, you will apply everything in a mini project that upgrades a basic RAG pipeline into a production-ready metadata-aware system.

Course Objectives

By the end of this course, you will be able to:

- Understand why vector search alone is insufficient for many production RAG use cases
- Design metadata schemas that improve retrieval precision and business relevance
- Preserve metadata during chunking and ingestion pipelines
- Apply filtering during retrieval for role-based, time-based, and multi-constraint queries
- Build and evaluate a metadata-aware RAG workflow end to end

Prerequisites

This course is designed for developers and learners who want to build more reliable RAG systems. Basic familiarity with Python and the general idea of LLMs or vector search is helpful, but no prior experience with metadata filtering is required.

Metadata Filtering & Schema-Aware Retrieval

The ability to retrieve the right information at the right time is becoming increasingly important in AI systems. This course empowers you to move beyond basic semantic search by teaching the principles and practice of metadata filtering, schema-aware chunking, and production-ready retrieval design.

Course Syllabus

A practical module-by-module plan for learning metadata-aware retrieval and building a production-ready RAG system.

Day	Content	Project
Day 1	Introduction to metadata filtering Why vector similarity alone falls short Real-world retrieval constraints: access control, time, department Filtered vs unfiltered retrieval examples	
Day 2	Metadata-aware chunking and ingestion Designing a useful metadata schema Preserving metadata at chunk level Ingestion patterns for reliable retrieval	Design a metadata schema and ingestion plan for a chosen use case
Day 3	Filtering during retrieval Pre-filtering vs post-filtering Combining multiple constraints Query construction for precise retrieval	Implement filtered retrieval queries and compare results with baseline search
Day 4	Advanced filtering patterns Dynamic filters from user context Query-time filter inference Role, department, and time-aware retrieval	Build a dynamic filtering workflow for a realistic application scenario
Day 5	Mini project: metadata-aware RAG system Upgrade a baseline RAG pipeline Apply multiple filter types Measure improvement over baseline retrieval	Complete an end-to-end metadata-aware RAG mini project
Bonus	Extension ideas Multi-tenant retrieval Access-controlled assistants Domain-specific retrieval evaluation	Optional: adapt the mini project for your own product or dataset