Updated on 27/06/2025

Sign up

ChromaDB Training

2 days (14 hours)

Presentation

Master ChromaDB to build powerful semantic search engines and vector databases ready to integrate your AI applications. This step-by-step course guides you through the design, indexing and querying of vector data.

You'll learn how to generate and manipulate text-based embeddings, index them efficiently in Chroma, and optimize your collections for fast, accurate, metadata-filtered searches.

You'll know how to create persistent collections, query your data with advanced criteria, visualize the semantic distances between your content, and maintain a vector base suitable for production.

You'll also discover how to integrate Chroma into a complete RAG system via LangChain, combining semantic retrieval and contextual response generation with an LLM like GPT. This will enable you to create your own intelligent assistants connected to your knowledge base.

As with all our training courses, it will take place on my latest version of the tool: ChromaDB.

Objectives

- Understand the principles of vector search, the role of embeddings, and the foundations of semantic similarity in modern AI systems.
- Install, configure and administer a ChromaDB vector database in memory or persistent mode, with fine-tuned management of collections, documents and metadata.
- Efficiently generate, index and query text embeddings using pre-trained templates (OpenAI, SentenceTransformers), with filtered and precise queries.
- Design complete workflows for ingesting, slicing, indexing and querying textual data textual data in a Chroma database for intelligent search.

- Exploit ChromaDB in a RAG architecture by integrating it with LangChain to build conversational assistants connected to a vectorized document base.
- Supervise, secure and optimize vectorized databases in production, addressing performance performance, governance, persistence and scaling.

Target audience

- Developers
- Data Scientists
- AI researchers

Prerequisites

- Basic knowledge of REST APIs
- Knowledge of a modern Python development language

ChromaDB training program

Introduction to vector search

- Semantic text representation
- Common embeddings : OpenAI, HuggingFace, BERT
- Cosine similarity, Euclidean, Manhattan
- Notions of top-k, score thresholds

ChromaDB architecture and operation

- Collections, Documents, Embeddings, Metadata
- Internal index (FAISS-like, local storage)
- Memory only (:memory:)
- Persistent (local folder)
- Text documents, key/value pairs, optional metadata

Setting up ChromaDB

- pip install chromadb
- Useful dependencies (tqdm, langchain for testing)
- chromadb.Client()
- Storage folder (persist_directory)
- create_collection(), naming, duplication, management
- Adding documents and embeddings

Data indexing and querying

- Embeddings via OpenAI or sentence-transformers
- add() and upsert(): when to use what
- query(query_embeddings=..., n_results=...)
- Metadata filtering
- delete(), modify(), reset(): manage your data
- Export / import (JSON, local backup)

View and analyze results

- ids, documents, metadatas, distances
- Use pandas or prettytable for rapid inspection
- t-SNE, PCA for 2D representation

Practical use cases

- Embedding transcript sentences
- Recovering relevant passages
- PDFs converted to indexed chunks
- Displaying results with context
- Use of metadata

Integration with LangChain

- Embedding ? Vector Store ? Retriever ? LLM
- Chroma.from_documents(documents, embedding)
- as_retriever(), similarity_search()
- Using RetrievalQA
- Injecting context into the prompt

Best practices & Resources

- Official Chroma documentation
- Optimal chunking, persistent storage, lightweight embeddings
- In-depth projects
  - Legal chatbot
  - HR search engine
  - customer support assistant

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new, advanced IT technology, or to acquire specific business knowledge or modern methods.

acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the forthcoming course, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Certification

A certificate will be awarded to each trainee who has completed the entire course.