

Chromadb Documentation

Overview

Chromadb consists of two primary databases:

1. **Correction Chromadb:** Stores doctor instructions or corrections for patient records.
2. **Books Chromadb:** Manages segmented documents from books or PDFs for medical analysis.

Correction Chromadb

Input Specifications

- **Documents:** Contains patient information and specific sections needing corrections or instructions from doctors.
- **Embeddings:** Utilizes HuggingfaceEmbeddings model "sentence-transformers/all-mpnet-base-v2" for generating document embeddings.

Output Specifications

- Retrieves the latest entered doctor instructions or corrections based on patient ID.

Process

- Documents submitted to Correction Chromadb are saved with embeddings generated using the specified Huggingface model.
- Returns the most recent instruction or correction entered.

Improvements

- **Duplicate Prevention:** Ensures no duplicate entries by checking existing page content in the collection before saving.
- **Separate Collections:** Each doctor has a dedicated collection in Chromadb identified by doctor ID, ensuring organizational clarity and data segregation.

Books Chromadb

Input Specifications

- **Documents:** Segmented chunks of PDFs or documents, with a chunk size of 1000 and overlap of 200.
- **Embeddings:** Uses HuggingfaceEmbeddings model "sentence-transformers/all-mpnet-base-v2" for embedding generation.

Output Specifications

- Determines whether a document indicates a Urinary Tract Infection (UTI) and identifies the type.
- Provides medication recommendations based on the identified UTI type.

Process

- Stores segmented PDFs/documents in Chromadb for retrieval and analysis.
- Retrieves data directly from stored books in the system.