Requirements:

| **Library** | **Purpose** |
| --- | --- |
| langchain | Core framework for document loading, chunking, embedding, and retrieval |
| langchain-community | Community-maintained vector store (Chroma) support |
| langchain-experimental | For SemanticChunker |
| langchain-huggingface | Wrapper for Hugging Face embeddings |
| langchain-google-genai | Wrapper for Google Gemini embedding API |
| sentence-transformers | For local model-based embeddings |
| python-docx | Reading .docx files |
| PyMuPDF (fitz) | Reading .pdf files |
| chromadb | Local vector database |
| python-dotenv | Loading environment variables from .env |

Environment Variables

For **Google Generative AI Embeddings**, you must set:

GOOGLE\_API\_KEY=your\_google\_api\_key\_here

Use a .env file or set it manually with python:

os.environ["GOOGLE\_API\_KEY"] = "your\_key\_here"

How to install libraries

pip install langchain langchain-community langchain-experimental langchain-huggingface langchain-google-genai sentence-transformers python-docx pymupdf chromadb python-dotenv