

Introduction to Retrieval

Instructor

Dipanjan Sarkar

Head of Community & Principal AI Scientist at Analytics Vidhya

Google Developer Expert - ML & Cloud Champion Innovator

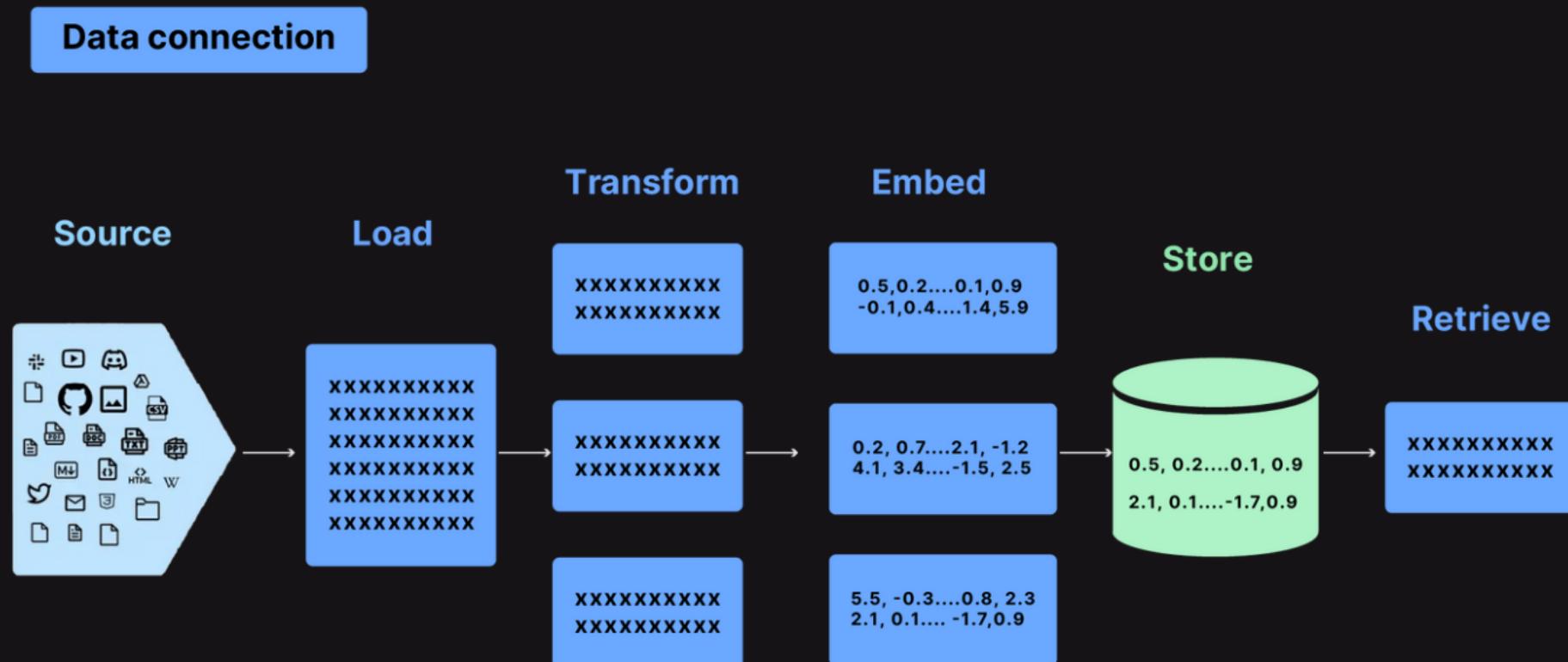
Published Author



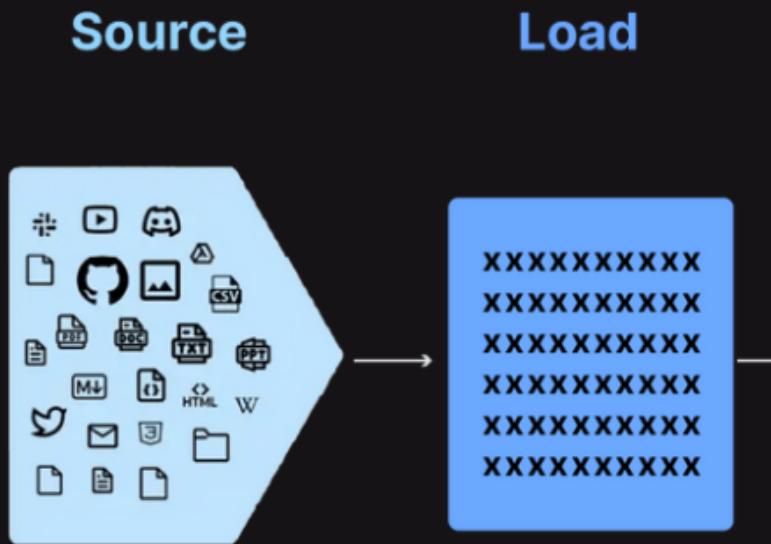
Outline

- Retrieval Workflow
- Document Loaders
- Document Splitters and Chunkers
- Embedding Models
- Vector Databases
- Retriever

Retrieval Workflow

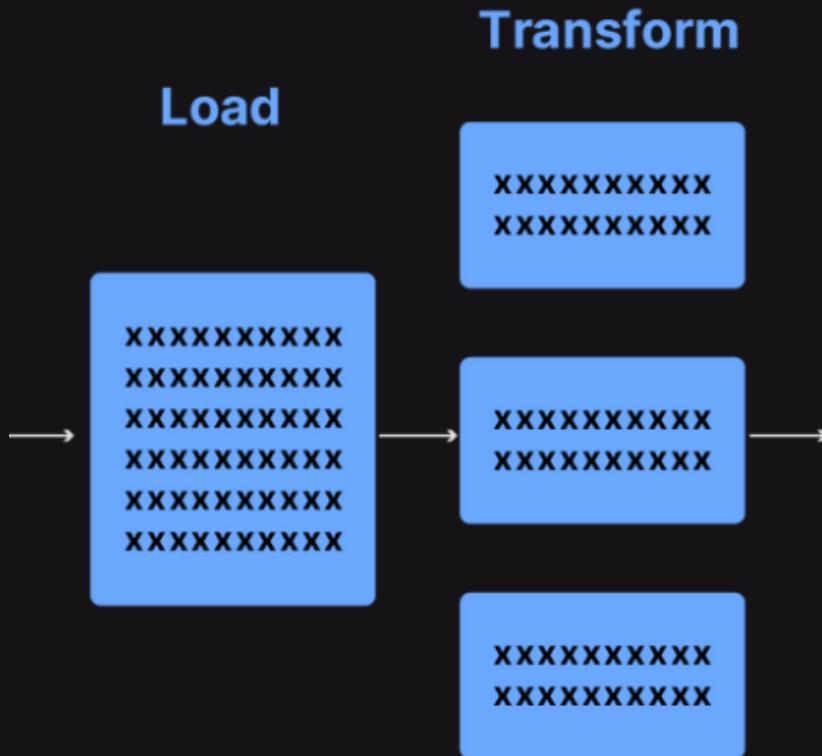


Document Loaders



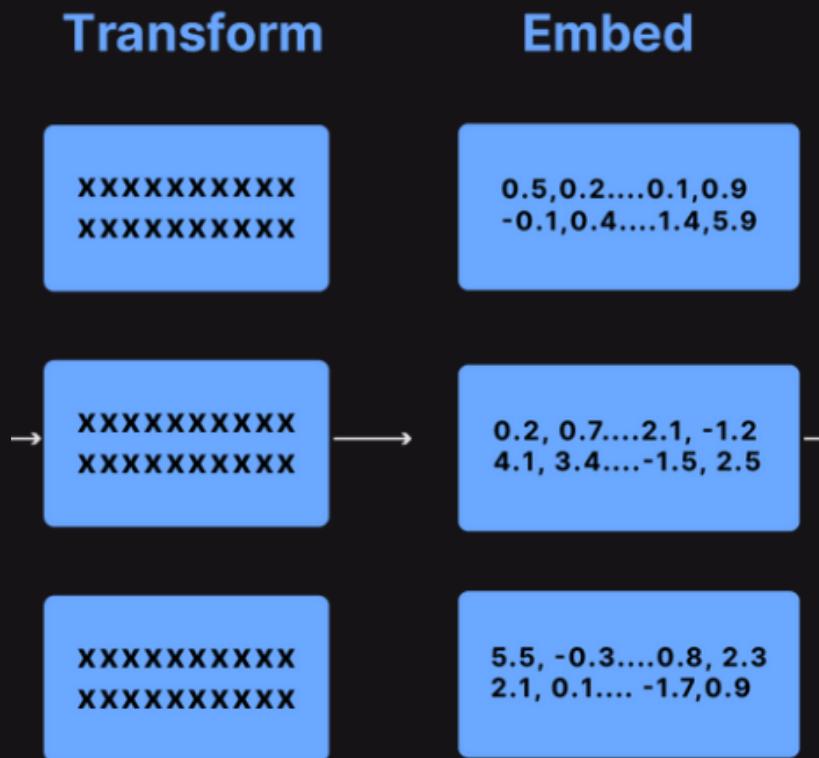
- Used to import data from various sources by LangChain as **Document** objects.
- A **Document** typically includes a piece of text along with its associated metadata.
- LangChain is integrated with over 100 document loader tools and libraries.

Document Splitters and Chunkers



- After loading documents, you might need to transform them for optimal use with LLMs
- One common transformation is splitting a long document into smaller segments or chunks to fit the LLM's context window
- LangChain provides several built-in document transformers to facilitate splitting and chunking

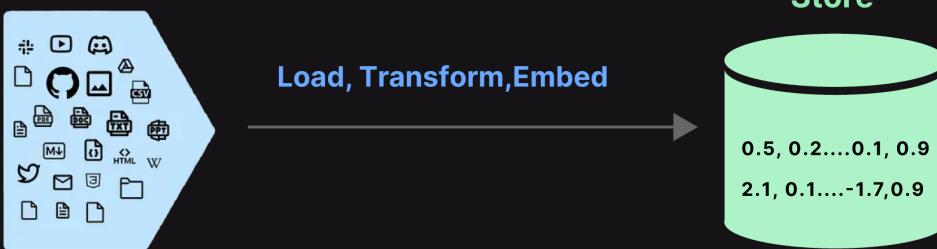
Embedding Models



- Create a vector representation for any document which is called an embedding
- The base `Embeddings` class in LangChain provides two methods:
 - one for embedding documents
 - one for embedding a query
- LangChain has integrations with all popular Embedding model providers like:
 - Open AI
 - HuggingFace

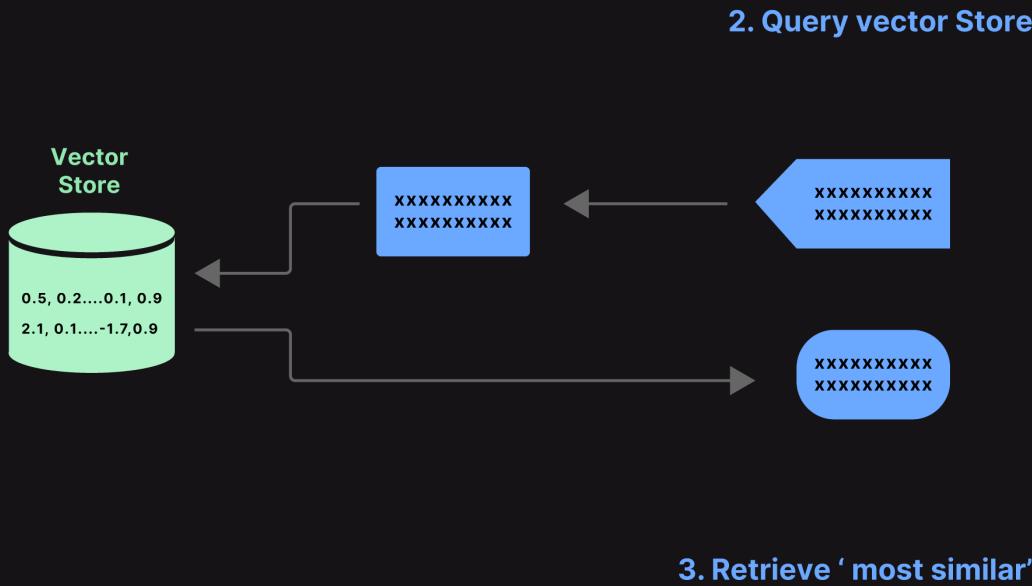
Vector Databases

1. Load Source Data



- Manages document storage, embeddings, and performs semantic vector search.
- Supports common operations including create, update, delete and querying
- LangChain has integrations with all popular vector DB providers like:
 - Chroma
 - Pinecone
 - Weaviate
 - Milvus and more

Retrievers



- An interface which accepts a string **query** as input and returns a list of **Documents** as output
- Retrievers can be created from vector databases or other search APIs
- LangChain enables us to use retrievers with a variety of retrieval strategies like:
 - cosine similarity
 - maximal marginal relevance
 - multi-query retrieval and more

Thank You
