

Vector stores

One of the most common ways to store and search over unstructured data is to embed it and store the resulting embedding vectors, and then at query time to embed the unstructured query and retrieve the embedding vectors that are 'most similar' to the embedded query. A vector store takes care of storing embedded data and performing vector search for you.

Get started

This walkthrough showcases basic functionality related to VectorStores. A key part of working with vector stores is creating the vector to put in them, which is usually created via embeddings. Therefore, it is recommended that you familiarize yourself with the [text embedding model](#) interfaces before diving into this.

This walkthrough uses the `FAISS` vector database, which makes use of the Facebook AI Similarity Search (FAISS) library.

```
pip install faiss-cpu
```

We want to use OpenAIEmbeddings so we have to get the OpenAI API Key.

```
import os
import getpass

os.environ['OPENAI_API_KEY'] = getpass.getpass('OpenAI API Key:')
```

```
from langchain.document_loaders import TextLoader
from langchain.embeddings.openai import OpenAIEmbeddings
from langchain.text_splitter import CharacterTextSplitter
from langchain.vectorstores import FAISS

raw_documents = TextLoader('../.../state_of_the_union.txt').load()
text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
documents = text_splitter.split_documents(raw_documents)

embeddings = OpenAIEmbeddings()
db = FAISS.from_documents(documents, embeddings)
```

Similarity search

```
query = "What did the president say about Ketanji Brown Jackson"
docs = db.similarity_search(query)
print(docs[0].page_content)
```

Tonight. I call on the Senate to: Pass the Freedom to Vote Act. Pass the John Lewis Voting Rights Act. And while you're at it, pass the Disclose Act so Americans can know who is funding our elections.

Tonight, I'd like to honor someone who has dedicated his life to serve this country: Justice Stephen Breyer—an Army veteran, Constitutional scholar, and retiring Justice of the United States Supreme Court. Justice Breyer, thank you for your service.

One of the most serious constitutional responsibilities a President has is nominating someone to serve on the United States Supreme Court.

And I did that 4 days ago, when I nominated Circuit Court of Appeals Judge Ketanji Brown Jackson. One of our nation's top legal minds, who will continue Justice Breyer's legacy of excellence.

Similarity search by vector

It is also possible to do a search for documents similar to a given embedding vector using `similarity_search_by_vector` which accepts an embedding vector as a parameter instead of a string.

```
embedding_vector = embeddings.embed_query(query)
docs = db.similarity_search_by_vector(embedding_vector)
```