Text vectorization

Monday, 7 April 2025 10:19 PM

```
import os
from dotenv import load_dotenv
from langchain_community.document_loaders import TextLoader
from langchain_text_splitters import CharacterTextSplitter
from langchain_openai import OpenAIEmbeddings
from langchain_pinecone import PineconeVectorStore

load_dotenv()

if __name__ == '__main__':
    print("Ingesting...")
    loader = TextLoader("/Users/edenmarco/Desktop/intro-to-vector-dbs/mediumblog1.txt")
    document = loader.load()

    print("splitting...")
    text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
    texts = text_splitter.split_documents(document)
    print(f"created {len(texts)} chunks")

    embeddings = OpenAIEmbeddings(openai_api_key=os.environ.get("OPENAI_API_KEY"))

    print("ingesting...")
    PineconeVectorStore.from_documents(texts, embeddings, index_name=os.environ['INDEX_NAME'])
    print("finish")
```

```
import os
from dotenv import load_dotenv
from langchain_core.prompts import PromptTemplate
from langchain_openai import OpenAIEmbeddings, ChatOpenAI
from langchain_pinecone import PineconeVectorStore

from langchain import hub
from langchain.chains.combine_documents import create_stuff_documents_chain
from langchain.chains.retrieval import create_retrieval_chain


load_dotenv()


if __name__ == "__main__":
    print(" Retrieving...")

    embeddings = OpenAIEmbeddings()
    llm = ChatOpenAI()

    query = "what is Pinecone in machine learning?"
    chain = PromptTemplate.from_template(template=query) | llm
    # result = chain.invoke(input={})
    # print(result.content)
```

```
vectorstore = PineconeVectorStore(
    index_name=os.environ["INDEX_NAME"], embedding=embeddings
)

retrieval_qa_chat_prompt = hub.pull("langchain-ai/retrieval-qa-chat")
combine_docs_chain = create_stuff_documents_chain(llm, retrieval_qa_chat_prompt)
retrival_chain = create_retrieval_chain(
    retriever=vectorstore.as_retriever(), combine_docs_chain=combine_docs_chain
)

result = retrival_chain.invoke(input={"input": query})

print(result)
```