



# 文章內容問答

LangChain + Data Embedding/Faiss + LLM

Python 3.13.3 + Langchain 1.2.6

Ryan Chung

20260108



# 文章內容問答

- 資料形式為文章
- 向量資料庫建立
  - 使用Embedding模型
    - all-MiniLM-L6-v2
    - Google / Gemini-Embedding-001
    - Azure OpenAI / Text-Embedding-3-Small
  - 文章內容切割
- 使用者輸入問題
  - 向量化後進行相似度搜尋
- 顯示搜尋結果

# 資料

- Text of President Joe Biden's State of the Union address

## Synopsis

Madam Speaker, Madam Vice President, our First Lady and Second Gentleman. Members of Congress and the Cabinet. Justices of the Supreme Court. My fellow Americans.



*The text of President Joe Biden's State of the Union address, as provided by the White House:*

Source : <https://economictimes.indiatimes.com/>

[https://raw.githubusercontent.com/ryanchung403/dataset/main/state\\_of\\_the\\_union.txt](https://raw.githubusercontent.com/ryanchung403/dataset/main/state_of_the_union.txt)



# requirements.txt

```
langchain_community
pandas
langchain_huggingface
faiss-cpu
sentence-transformers
ipykernel
langchain-openai
langchain_google_genai
tiktoken
chardet
```



# Embedding Model 1

- all-MiniLM-L6-v2

sentiment-transformers/all-MiniLM-L6-v2

like 2.15k

Sentence Similarity sentence-transformers PyTorch TensorFlow Rust ONNX Safetensors Transformers

21 datasets English bert feature-extraction text-embeddings-inference Inference Endpoints 5 papers

License: apache-2.0

Model card Files Community 74 Edit model card

Train Deploy Use this model

all-MiniLM-L6-v2

This is a [sentence-transformers](#) model: It maps sentences & paragraphs to a 384 dimensional dense vector space and can be used for tasks like clustering or semantic search.

Downloads last month  
47,191,134

Safetensors

Model size 22.7M params Tensor type I64 · F32

Inference API

Warm Examples Sentence Similarity

<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>



# app.py

```
from langchain.text_splitter import CharacterTextSplitter
from langchain_community.document_loaders import TextLoader
from langchain_community.vectorstores import FAISS
from langchain_core.output_parsers import StrOutputParser
from langchain_huggingface import HuggingFaceEmbeddings

embeddings = HuggingFaceEmbeddings(
    model_name="sentence-transformers/all-MiniLM-L6-v2"
)

loader = TextLoader("state_of_the_union.txt",
                    autodetect_encoding=True,
)
documents = loader.load()
text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=0)
docs = text_splitter.split_documents(documents)
db = FAISS.from_documents(docs, embeddings)

query = "What did the president say about Ketanji Brown Jackson?"
results = db.similarity_search_with_score(query, 1)
print(results[0][0].page_content)
```

<https://gist.github.com/ryanchung403/ce2bb50b55b52b2cc2f4620d4cb35f65>



# 檢視結果

query

✓ 0.0s

'What did the president say about Ketanji Brown Jackson'

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.



# Embedding Model 2

- Google Gemini / gemini-embedding-001

Property	Description	
Model code	Gemini API gemini-embedding-001	
Supported data types	Input Text	Output Text embeddings
Token limits <sup>[*]</sup>	Input token limit 2,048	Output dimension size Flexible, supports: 128 - 3072, Recommended: 768, 1536, 3072
Versions	<p>Read the <a href="#">model version patterns</a> for more details.</p> <ul style="list-style-type: none"><li>Stable: <b>gemini-embedding-001</b></li><li>Experimental: <b>gemini-embedding-exp-03-07</b> (deprecating in Oct of 2025)</li></ul>	
Latest update	June 2025	

<https://ai.google.dev/gemini-api/docs/embeddings>



# config.ini

[Gemini]

API\_KEY = XXXXXXXXXXXXXXXX



# app.py

```
from configparser import ConfigParser
# Set up config parser
config = ConfigParser()
config.read("config.ini")
from langchain_google_genai import GoogleGenerativeAIEMBEDDINGS

embeddings_gemini = GoogleGenerativeAIEMBEDDINGS(
    model="gemini-embedding-001",
    google_api_key=config["Gemini"]["API_KEY"],
)

db_gemini = FAISS.from_documents(docs, embeddings_gemini)

# Query the database
query = "What did the president say about Ketanji Brown Jackson?"
results_gemini = db_gemini.similarity_search_with_score(query, 1)
print(results_gemini[0][0].page_content)
```

<https://gist.github.com/ryanchung403/af243430320f3c165cf948cb93f419d1>



# Embedding Model 3

- Azure OpenAI – Text-Embedding-3-Small  
Embeddings

`text-embedding-3-large` 是最新且功能最強的內嵌模型。無法在內嵌模型之間進行升級。若要從使用 `text-embedding-ada-002` 移轉至 `text-embedding-3-large`，您必須產生新的內嵌。

- `text-embedding-3-large`
- `text-embedding-3-small`
- `text-embedding-ada-002`

在測試中，OpenAI 報告顯示大型和小型第三代內嵌模型在 [MIRACL](#) 基準中提供了更好的平均多語言擷取效能，同時仍在 [MTEB](#) 基準中保持了英文工作的效能。

<https://learn.microsoft.com/zh-tw/azure/ai-services/openai/concepts/models#embeddings>



# config.ini

[AzureOpenAI]

KEY = XXXXXXXXXX

ENDPOINT = <https://XXXXXXXXX.XXXXX.azure.com/openai/v1/>

GPT4o\_DEPLOYMENT\_NAME = XXXXXXXX

Embedding\_DEPLOYMENT\_NAME = XXXXXXXXXXXX



# app.py

```
# Embedding – Azure OpenAI – TextEmbedding3-Small
```

```
from langchain_openai import OpenAIEMBEDDINGS
```

```
embeddings_azure_openai = OpenAIEMBEDDINGS(  
    model=config["AzureOpenAI"]["Embedding_DEPLOYMENT_NAME"],  
    base_url=config["AzureOpenAI"]["ENDPOINT"],  
    api_key=config["AzureOpenAI"]["KEY"],  
)
```

```
db_azure_openai = FAISS.from_documents(docs, embeddings_azure_openai)  
query = "What did the president say about Ketanji Brown Jackson?"  
results_azure_openai = db_azure_openai.similarity_search_with_score(query, 1)  
print(results_azure_openai[0][0].page_content)
```

<https://gist.github.com/ryanchung403/c51b631f926f4496995b70004e6264e4>



# LLM 問答機器人實作

RAG by LangChain + LLM + Embedding + FAISS

Ryan Chung



# 流程

- 使用者發問
- 延續前例，搜尋取得相似內容
- 將相似內容加入LLM(大型語言模型)的提示中
- 請LLM依據提示內容進行作答
- 檢視結果



# LLM 大型語言模型 1

- Google Gemini



# config.ini

[Gemini]

API\_KEY = XXXXXXXXXX



# app.py

```
from langchain_google_genai import ChatGoogleGenerativeAI

llm_gemini = ChatGoogleGenerativeAI(
    model="gemini-2.5-flash",
    google_api_key=config["Gemini"]["API_KEY"],
)

from langchain_core.prompts import ChatPromptTemplate

prompt = ChatPromptTemplate.from_template(
    """Answer the following question based only on the provided context:  

<context>
{context}
</context>
Question: {input}"""
)

output_parser = StrOutputParser()

chain = prompt | llm_gemini | output_parser
```



# app.py

```
query = "What did the president say about Ketanji Brown Jackson?"  
# query = "How much we are giving to the Ukraine?"  
# query = "How many nations in NATO Alliance?"  
results = db.similarity_search_with_score(query, 1)  
print("Retrieved related content :")  
print(results[0][0].page_content)  
print("=====  
  
llm_result = chain.invoke(  
    {  
        "input": query,  
        "context": [results[0][0]],  
    }  
)  
print("Question: ", query)  
print("LLM Answer: ", llm_result)
```

<https://gist.github.com/ryanchung403/b76df667f602dc72c319bc16989983dd>



# 測試

Retrieved related 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.

=====

Question: What did the president say about Ketanji Brown Jackson?

LLM Answer: She is one of our nation's top legal minds and will continue Justice Breyer's legacy of excellence.



# 測試

Retrieved related content :

And tonight I am announcing that we will join our allies in closing off American air space to all Russian flights – further isolating Russia – and adding an additional squeeze –on their economy. The Ruble has lost 30% of its value.

The Russian stock market has lost 40% of its value and trading remains suspended. Russia's economy is reeling and Putin alone is to blame.

Together with our allies we are providing support to the Ukrainians in their fight for freedom. Military assistance. Economic assistance. Humanitarian assistance.

We are giving more than \$1 Billion in direct assistance to Ukraine.

And we will continue to aid the Ukrainian people as they defend their country and to help ease their suffering.

Let me be clear, our forces are not engaged and will not engage in conflict with Russian forces in Ukraine.

Our forces are not going to Europe to fight in Ukraine, but to defend our NATO Allies – in the event that Putin decides to keep moving west.

=====

Question: How much we are giving to the Ukraine?

LLM Answer: The context states that the US is giving \*\*more than \$1 Billion in direct assistance to Ukraine\*\*.



# 測試

Retrieved related content :

Groups of citizens blocking tanks with their bodies. Everyone from students to retirees teachers turned soldiers defending their homeland.

In this struggle as President Zelenskyy said in his speech to the European Parliament “Light will win over darkness.” The Ukrainian Ambassador to the United States is here tonight.

Let each of us here tonight in this Chamber send an unmistakable signal to Ukraine and to the world.

Please rise if you are able and show that, Yes, we the United States of America stand with the Ukrainian people.

Throughout our history we've learned this lesson when dictators do not pay a price for their aggression they cause more chaos.

They keep moving.

And the costs and the threats to America and the world keep rising.

That's why **the NATO Alliance** was created to secure peace and stability in Europe after World War 2.

**The United States is a member along with 29 other nations.**

It matters. American diplomacy matters. American resolve matters.

---

Question: How many nations in NATO Alliance?

LLM Answer: **According to the provided context, there are \*\*30\*\* nations in the NATO Alliance.**



# LLM 大型語言模型 2

- Azure OpenAI – GPT-4o-mini



# config.ini

[AzureOpenAI]

KEY = XXXXXXXXXX

ENDPOINT = https://XXXXXXXXX.XXXXX.azure.com/openai/v1/

GPT4o\_DEPLOYMENT\_NAME = XXXXXXXXXX

Embedding\_DEPLOYMENT\_NAME = XXXXXXXXXX



# app.py

```
# LLM - Azure OpenAI - GPT4o-mini
from langchain_openai import ChatOpenAI

llm_gpt4o = ChatOpenAI(
    model=config["AzureOpenAI"]["GPT4o_DEPLOYMENT_NAME"],
    api_key=config["AzureOpenAI"]["KEY"],
    base_url=config["AzureOpenAI"]["ENDPOINT"],
)
from langchain_core.prompts import ChatPromptTemplate

prompt = ChatPromptTemplate.from_template(
    """Answer the following question based only on the provided context:
<context>
{context}
</context>
Question: {input}"""
)

output_parser = StrOutputParser()
```



# app.py

```
chain_gpt4o = prompt | llm_gpt4o | output_parser

query = "What did the president say about Ketanji Brown Jackson?"
# query = "How much we are giving to the Ukraine?"
# query = "How many nations in NATO Alliance?"
results = db.similarity_search_with_score(query, 1)
print("Retrieved related content :")
print(results[0][0].page_content)
print("=====")

llm_result_gpt4o = chain_gpt4o.invoke(
    {
        "input": query,
        "context": [results[0][0]],
    }
)
print("Question: ", query)
print("LLM Answer: ", llm_result_gpt4o)
```

<https://gist.github.com/ryanchung403/4796c46ac694de65cf9f1c9efa149c27>



# 測試

Retrieved related 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.

---

Question: What did the president say about Ketanji Brown Jackson?

LLM Answer: The president said that he nominated Circuit Court of Appeals Judge Ketanji Brown Jackson to serve on the United States Supreme Court and described her as one of the nation's top legal minds who will continue Justice Breyer's legacy of excellence.



# 測試

Retrieved related content :

And tonight I am announcing that we will join our allies in closing off American air space to all Russian flights – further isolating Russia – and adding an additional squeeze –on their economy. The Ruble has lost 30% of its value.

The Russian stock market has lost 40% of its value and trading remains suspended. Russia's economy is reeling and Putin alone is to blame.

Together with our allies we are providing support to the Ukrainians in their fight for freedom. Military assistance. Economic assistance. Humanitarian assistance.

We are giving more than \$1 Billion in direct assistance to Ukraine.

And we will continue to aid the Ukrainian people as they defend their country and to help ease their suffering. Let me be clear, our forces are not engaged and will not engage in conflict with Russian forces in Ukraine. Our forces are not going to Europe to fight in Ukraine, but to defend our NATO Allies – in the event that Putin decides to keep moving west.

---

Question: How much we are giving to the Ukraine?

LLM Answer: We are giving more than \$1 billion in direct assistance to Ukraine.



# 測試

Retrieved related content :

Groups of citizens blocking tanks with their bodies. Everyone from students to retirees teachers turned soldiers defending their homeland. In this struggle as President Zelenskyy said in his speech to the European Parliament “Light will win over darkness.” The Ukrainian Ambassador to the United States is here tonight.

Let each of us here tonight in this Chamber send an unmistakable signal to Ukraine and to the world. Please rise if you are able and show that, Yes, we the United States of America stand with the Ukrainian people. Throughout our history we've learned this lesson when dictators do not pay a price for their aggression they cause more chaos. They keep moving.

And the costs and the threats to America and the world keep rising.

That's why **the NATO Alliance** was created to secure peace and stability in Europe after World War 2.

**The United States is a member along with 29 other nations.**

It matters. American diplomacy matters. American resolve matters.

---

---

Question: How many nations in NATO Alliance?

LLM Answer: **The NATO Alliance consists of 30 nations, including the United States.**



# LLM 大型語言模型 3

- Ollama 地端模型



# requirements.txt

```
langchain_community
pandas
langchain_huggingface
faiss-cpu
sentence-transformers
ipykernel
langchain-openai
langchain_google_genai
tiktoken
chardet
langchain_ollama
```



# app.py

```
from langchain_ollama.llms import OllamaLLM
ollama_llm = OllamaLLM(model="gemma3:1b")
query = "What did the president say about Ketanji Brown Jackson?"
# query = "How much we are giving to the Ukraine?"
# query = "How many nations in NATO Alliance?"
results = db.similarity_search_with_score(query, 1)
print("Retrieved related content :")
print(results[0][0].page_content)
print("=====")

chain_ollama = prompt | ollama_llm | output_parser
ollama_result = chain_ollama.invoke(
    {
        "input": query,
        "context": [results[0][0]],
    }
)
print("Question: ", query)
print("LLM Answer from Ollama: ", ollama_result)
```

<https://gist.github.com/ryanchung403/9774055374d1c2cf5376089d0182f8d7>



# 測試

Retrieved related 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.

---

=====  
Question: What did the president say about Ketanji Brown Jackson?

LLM Answer from Ollama: The president nominated Circuit Court of Appeals Judge Ketanji Brown Jackson.



# 測試

Retrieved related content :

And tonight I am announcing that we will join our allies in closing off American air space to all Russian flights – further isolating Russia – and adding an additional squeeze –on their economy. The Ruble has lost 30% of its value.

The Russian stock market has lost 40% of its value and trading remains suspended. Russia's economy is reeling and Putin alone is to blame.

Together with our allies we are providing support to the Ukrainians in their fight for freedom. Military assistance. Economic assistance. Humanitarian assistance.

We are giving more than \$1 Billion in direct assistance to Ukraine.

And we will continue to aid the Ukrainian people as they defend their country and to help ease their suffering. Let me be clear, our forces are not engaged and will not engage in conflict with Russian forces in Ukraine. Our forces are not going to Europe to fight in Ukraine, but to defend our NATO Allies – in the event that Putin decides to keep moving west.

---

Question: How much we are giving to the Ukraine?

LLM Answer from Ollama: \$1 Billion.



# 測試

Retrieved related content :

Groups of citizens blocking tanks with their bodies. Everyone from students to retirees teachers turned soldiers defending their homeland. In this struggle as President Zelenskyy said in his speech to the European Parliament "Light will win over darkness." The Ukrainian Ambassador to the United States is here tonight. Let each of us here tonight in this Chamber send an unmistakable signal to Ukraine and to the world. Please rise if you are able and show that, Yes, we the United States of America stand with the Ukrainian people. Throughout our history we've learned this lesson when dictators do not pay a price for their aggression they cause more chaos. They keep moving. And the costs and the threats to America and the world keep rising.

That's why the **NATO Alliance was created to secure peace and stability in Europe after World War 2**. The United States is a member along with 29 other nations. It matters. American diplomacy matters. American resolve matters.

---

Question: How many nations in NATO Alliance?

LLM Answer from Ollama: **29**