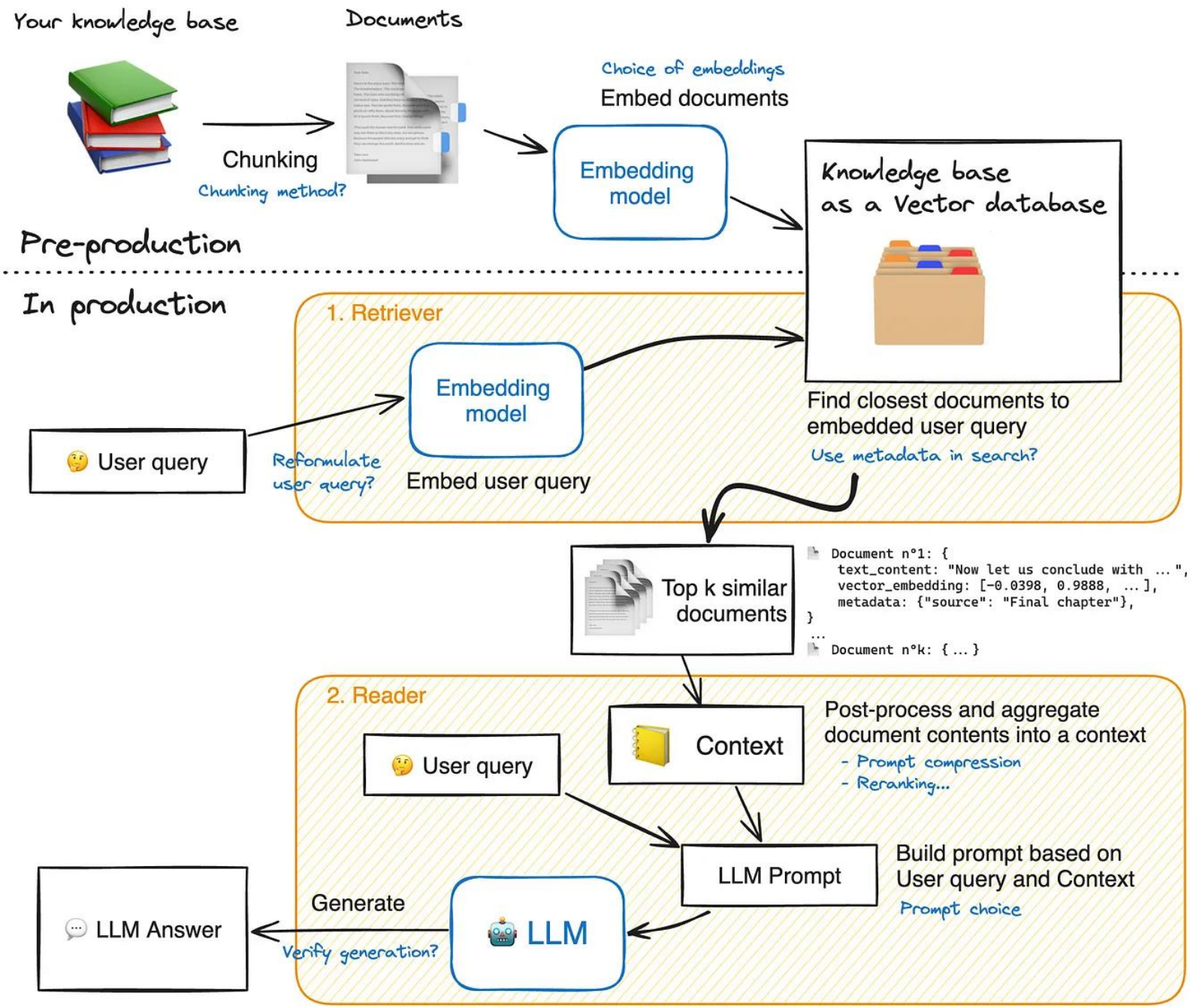
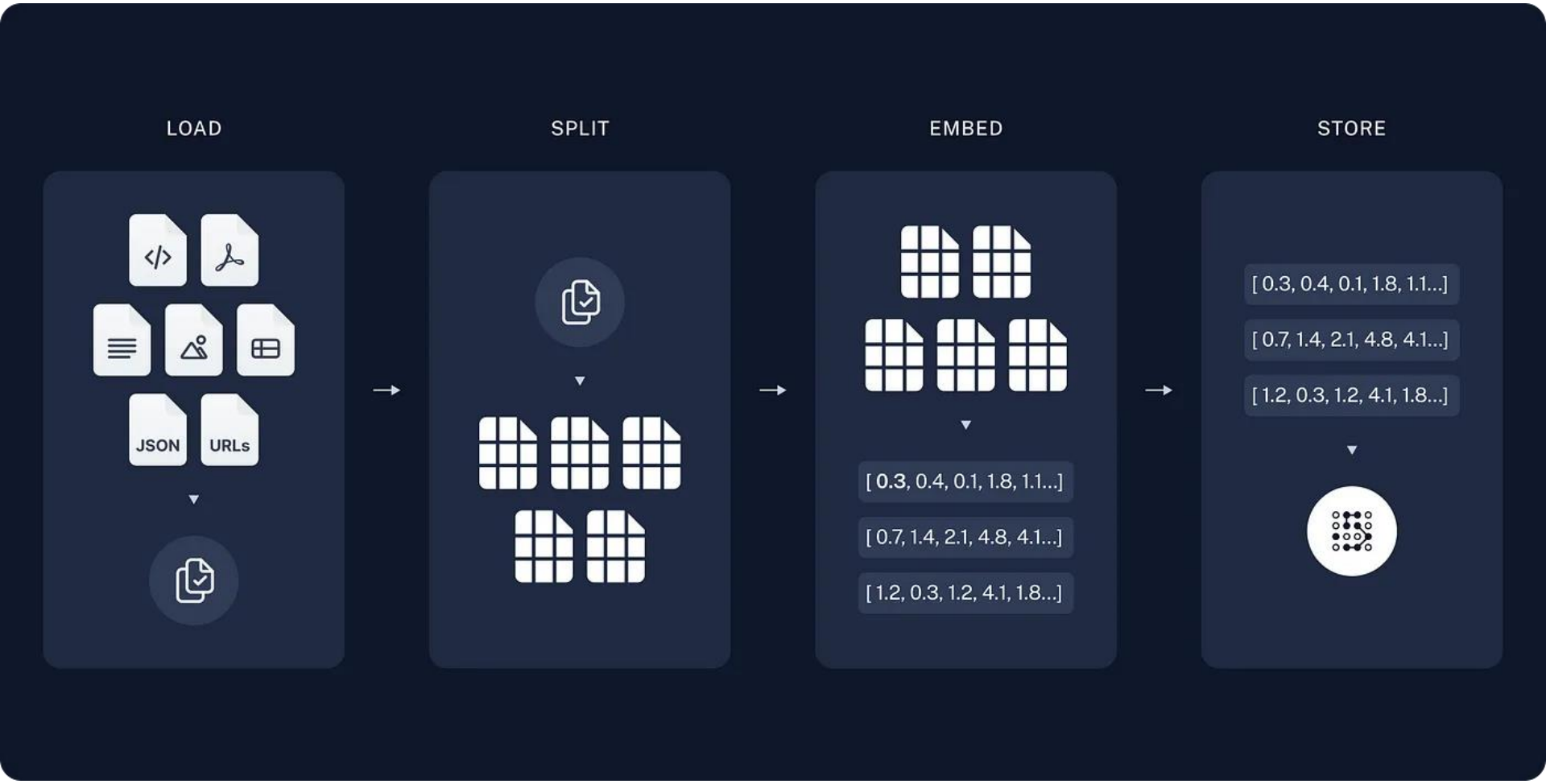
LLM+RAG實作教學
llama3 + doc
llama3 + pdf

1. 客製化聊天機器人
2. 客製化專利法聊天機器人

11/24/2024

Louisa Coffee


Your knowledge base
Documents
Chunking
Chunking method?
Choice of embeddings
Embed documents
Embedding model
Knowledge base as a Vector database
Pre-production
In production
1. Retriever
Embedding model
User query
Reformulate user query?
Embed user query
Find closest documents to embedded user query
Use metadata in search?
Top k similar documents
Document n°1: {
text_content: "Now let us conclude with ...",
vector_embedding: [-0.0398, 0.9888, ...],
metadata: {"source": "Final chapter"},
}
...
Document n°k: {...}
2. Reader
User query
Context
Post-process and aggregate document contents into a context
- Prompt compression
- Reranking...
LLM Prompt
Build prompt based on User query and Context
Prompt choice
LLM
Generate
Verify generation?
LLM Answer



Chunking Embedding Vector Database

Step 1: import以下套件

- from langchain.chains.combine_documents import create_stuff_documents_chain
  from langchain.chains import create_retrieval_chain
  from langchain_core.prompts import ChatPromptTemplate
  from langchain_community.llms import Ollama
  from langchain_community.embeddings import OllamaEmbeddings
  from langchain_community.vectorstores import FAISS
  from langchain_core.documents import Document
  from langchain_community.document_loaders import PyPDFLoader

  from langchain.text_splitter import CharacterTextSplitter

- from langchain_core.documents import Document
  from langchain_community.document_loaders import PyPDFLoader

Step 2.建立模型和文件

```
# 初始化Ollama模型
llm = Ollama(model='llama3')

# 建立文件列表，每個文件包含一段文字內容
docs = [
Document(page_content='曼德珍珠奶茶草：這種植物具有強大的魔法屬性，常用於恢復被石化的受害者。'),
Document(page_content='山羊可愛蓮花石 ：是一種從山羊胃中取出的石頭，可以解百毒。在緊急情況下，它被認為是最有效的解毒劑。'),
Document(page_content='日本小可愛佐籐鱗片 ：這些鱗片具有強大的治愈能力，常用於製作治療藥水，特別是用於治療深層傷口。'),
]
```

Docs 給Llama3本來不知道的來學習

3.設定文本分割器

- # 設定文本分割器，chunk_size是分割的大小，chunk_overlap是重疊的部分
  text_splitter = CharacterTextSplitter(chunk_size=20, chunk_overlap=5)

- # 將文件分割成更小的部分
  documents = text_splitter.split_documents(docs)

chunk_size (塊大小)

定義: 每個分割塊的大小，以字符數量為單位。
作用: 決定每個文本塊包含多少字符。

chunk_overlap (塊重疊)

定義: 相鄰文本塊之間重疊的字符數量。
作用: 確保每個分割後的文本塊之間有一些重疊部分，以保證連貫性和上下文不丟失。

為什麼需要 chunk_overlap?

在自然語言處理和其他文本分析任務中，連貫性和上下文信息非常重要。
通過設置塊重疊部分，我們可以確保每個分割後的文本塊仍然包含足夠的上下文信息，避免因切割造成的信息丟失或語義斷裂。

4.建置embeddings和向量資料庫

- # 初始化嵌入模型
  embeddings = OllamaEmbeddings()

  # 使用FAISS建立向量資料庫
  vectordb = FAISS.from_documents(docs, embeddings)
-

  # 將向量資料庫設為檢索器
  retriever = vectordb.as_retriever()

5.設定提示模板

- # 設定提示模板，將系統和使用者的提示組合
  prompt = ChatPromptTemplate.from_messages([
  ('system', 'Answer the user\'s questions in Chinese, based on the context provided below:\n\n{context}'),
  ('user', 'Question: {input}'),
  ])

6.llm和提示模板結合

- # 創建文件鏈，將llm和提示模板結合
  document_chain = create_stuff_documents_chain(llm, prompt)

  # 創建檢索鏈，將檢索器和文件鏈結合
  retrieval_chain = create_retrieval_chain(retriever, document_chain)

7.從用戶輸入中獲取問題，並用retrieval_chain來回答

- ```
  context = []
  input_text = input('>>> ')
  while input_text.lower() != 'bye':
  response = retrieval_chain.invoke({
  'input': input_text,
  'context': context
  })
  print(response['answer'])
  context = response['context']
  input_text = input('>>> ')
  ```
  ```

Q&A

- >>> 什麼是中華文化?
- >>> 請告訴我珍珠奶茶是？
- >>>曼德珍珠奶茶草有什麼用?

LLM only

```
(metaverse) c:\Python3\my_project\project_llm>ollama run llama3
pulling manifest
pulling 6a0746a1ec1a... 100% ▕▏ 4.7 GB
pulling 4fa551d4f938... 100% ▕▏  12 KB
pulling 8ab4849b038c... 100% ▕▏  254 B
pulling 577073ffcc6c... 100% ▕▏  110 B
pulling 3f8eb4da87fa... 100% ▕▏  485 B
verifying sha256 digest
writing manifest
success
>>> Send a message (/? for help)
```

```
>>> 請回達中文，奶茶如何製作？
Let me translate that for you! 😊

**怎麼做奶茶？**

如果您想喝一杯美味的奶茶，以下是製作方法：

**材料：**

* 2茶匙黑茶葉（或1個黑茶袋）
* 1碗熱水
* 1/2碗牛奶（或任何非牛奶替代品，如 almond milk 或 soy milk）
* 2茶匙糖（選擇）
* 冰塊
*攪拌器或奶沫機（選擇）

**步驟：**

1. **煮茶**: 將熱水傾入茶壺中，然後添加黑茶葉（或茶袋）。讓它浸泡3-5分鐘，直到您想要的濃度。
2. **準備牛奶**: 在另一個容器中，用微波爐或 stove 溫暖牛奶，直到它變得沸騰hot。這樣可以創造奶茶時混和的 creamy 文本。
3. **添加糖（選擇）：** 如果您喜歡您的奶茶甜，可以加2茶匙糖，然後攪拌直到溶解。
4. **結合茶和牛奶**: 將浸泡好的茶傾入大碗或咖啡杯中，然後添加溫暖的牛奶。攪拌直到混合好。您也可以使用奶沫機創造奶沫。
5. **添加冰（選擇）：** 如果您喜歡您的奶茶涼快，可以添加一些冰塊，然後攪拌一下。
6. **享用**: 現在，您的美味奶茶準備好了！您可以根據自己的喜好調整甜度和creaminess。

注意：

* 使用高質量的茶葉以獲得更好的口感。
* 調整茶和牛奶的比例以符合您的喜好。
*嘗試使用不同的牛奶替代品或添加一些 vanilla_extract 或鹽以增強口感。
```

LLM + RAG

```
>>>  什麼是中華文化
中華文化是一種含有繁多內涵和精神的文化，由中國、台灣、香港、澳門等地共同構成。它包含了古老的哲學思想、美術、音樂、飲食、文學、舞蹈以及宗教信仰，形成了一個獨特而豐富的文化體系。
>>> 請告訴我珍珠奶茶是？
珍珠奶茶（Mandarin Pearl Tea）是一種從小可愛佐籐植物的根部提取的藥物，有助於消除毒素和加速傷口癒合。它被用來治療多種疾病，包括石化、瘀血和感染。
>>> 曼德珍珠奶茶草有什麼用
曼德珍珠奶茶草（Mandrake）是一種具有特殊物理和魔法特性的植物，可以用來治療多種疾病。它可以被用於：

1.恢復被石化的受害者。
2.解毒：曼德珍珠奶茶草被認為是最有效的解毒劑，尤其是在緊急情況下。
3.治療深層傷口：日本小可愛佐籐鱗片常與曼德珍珠奶茶草組合使用，以製作治療藥水。

但是，由於曼德珍珠奶茶草的魔法屬性非常強大，所以它需要在適當的條件下使用，並且需要對其使用的方法進行深入研究。
>>> 山羊石是什麼？
山羊蓮花石是一種從山羊胃中取出的石頭，可以解百毒。在緊急情況下，它被認為是最有效的解毒劑。
>>> 奶茶如何製作？
奶茶的製作與使用了曼德珍珠奶茶草。先將曼德珍珠奶茶草放入熱水中，讓其充分釋放其魔法屬性。然後，在茶水中加入適量的甜味和奶油，調整風味即可。
>>> bye
```

{'input': '有哪些工具可以用？',

'context': [Document(metadata={}, page_content='山羊可愛蓮花石 ：是一種從山羊胃中取出的 石頭，可以解百毒。在緊急情況下，它被認為是最有效的解毒劑。'),

Document(metadata={}, page_content='日本小可愛佐籐鱗片 ：這些鱗片具有強大的治愈能力，常用於製作治療藥水，特別是用於治療深層傷口。'),

Document(metadata={}, page_content='曼德珍珠奶茶草：這種植物具有強大的魔法屬性，常用於恢復被石化的受害者。')],

'answer': '可愛蓮花石、佐籐鱗片和奶茶草都是具有特殊功效的工具，可以用來治療不同類型的傷害或病況。'}

LLM+RAG實作教學
llama3 + pdf

客製化專利法聊天機器人

Step 2.建立模型和文件

- # 初始化Ollama模型
  llm = Ollama(model='llama3')

  # 載入並分割PDF文件
  loader = PyPDFLoader("文件名稱.pdf")
  docs = loader.load_and_split()

LLM + RAG

```
(metaverse) c:\Python3\my_project\project_llm>python my_rag1.py
>>> 專利權期間幾年？
專利權期限為十五年。
>>> 申請日起幾年？
申請專利之發明，經審查確定後，給予專利權，並發證書。
>>> 專利權期間申請日起幾年？
專利權期限為十五年，自公告之日起算。
>>> 專利權期限為幾年由申請日起？
專利權期限為十五年，由公告之日起算，但自申請之日起不得逾十八年。
>>> 那些款不予發明專利？
那些不予發明專利的是：

一、飲食品及嗜好品，但其製造方法不在此限。

二、動、植物及微生物新品種，但植物新品種及微生物新菌種育成方法不在此限。

三、人體或動物疾病之診斷、治療或手術方法。

四、科學原理或數學方法。

五、遊戲及運動之規則或方法。

六、其他必須藉助於人類推理力、記憶力始能實施之方法或計畫。

七、物品新用途之發現，但化學品及醫藥品不在此限。
>>>
```

```
>>> 本法所稱專利，分為幾種？
根據本法的規定，本法所稱專利，分為三種：

一、發明專利權
二、新型創造人、新型專利權
三、設計專利人之報酬
>>> 本法主管機關是誰？
根據本法的第1條，「本法主管機關」為「中華民國專利局」。
>>> 專利申請權人，除本法另有規定或契約另有約定外，指誰？
C N

專利申請權人，除本法另有規定或契約另有約定外，指專利權所有者或被授權人。

(Translation: The patent applicant, except for the provisions or agreements otherwise specified in this law, refers to the patent owner or authorized party.
)
>>> 專利申請權及專利權，如何繼承？
根据本法第6條和第7條，專利申請權及專利權可以繼承。除非契約另有約定外，即發明人、新型創作人、設計人或其受讓人或繼承人的專利申請權屬於雇用人，雇用人應支付適當報酬。但是，如果契約另有約定從該約定。

因此，在本法下述的情況下，專利申請權及專利權不可以繼承：

一、受到讓的專利申請權及專利權之所有權書；
二、因法律效力性強制授權。
三、對方有良心願意繼承的聲明書。

總之，本法規定在特定的情況下繼承，但在不iolated 掛款或費用。

Translation:

According to Article 6 and 7 of the Patent Law, patent application rights and patents can be inherited. Unless otherwise provided by a contract, the inventi
on person's rights belong to the employer, and the employee should pay reasonable compensation. However, if the contract is not, then it will follow.

Therefore, in cases where one or more of the following conditions:

One, a written authority;
Two, due to changes the facts are changed, no longer necessary to grant patent license.

Three, the authorized person does not intend to inherit.

In general, the patent law, but under certain circumstances, and there is no provision.
>>>
```

https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=J0070007