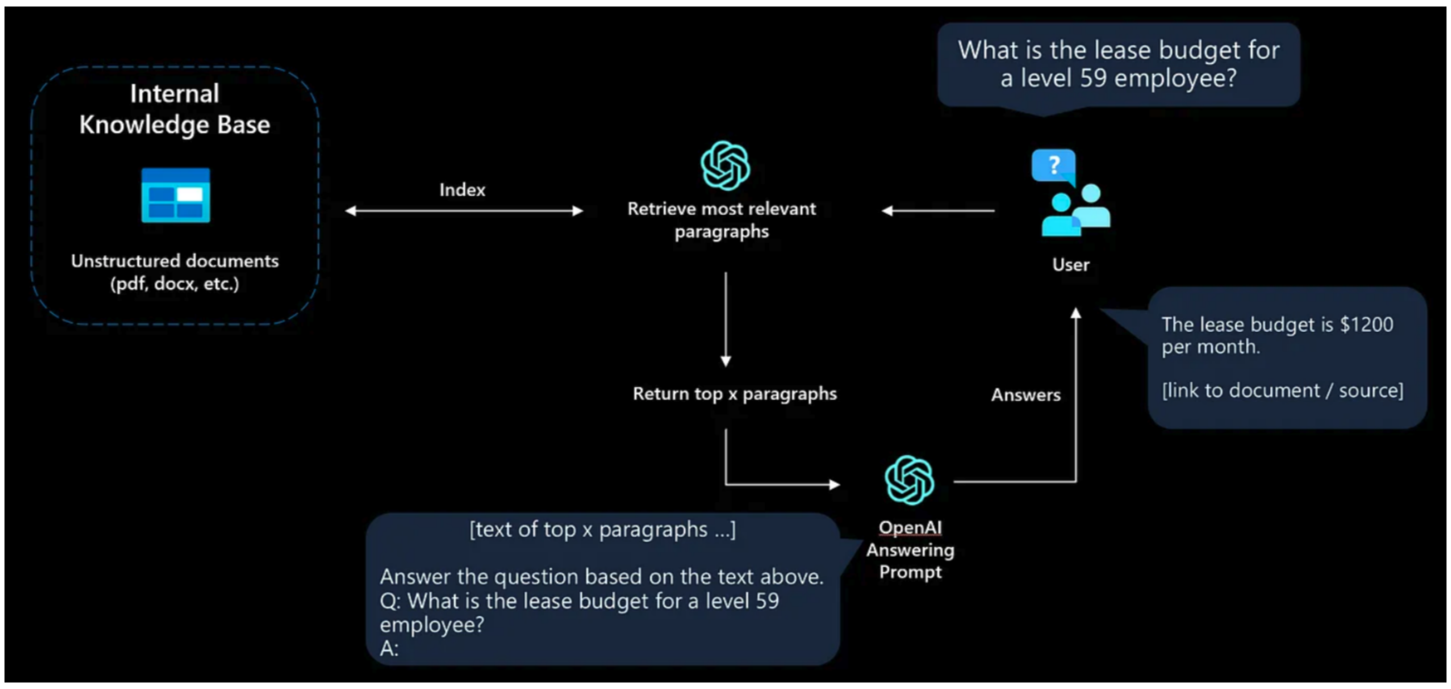
ChatGPT로 커스텀 Q&A 엔진 만들기

개요

- 원제목 : How to create a private ChatGPT with your own data
- 부제 : Learn the architecture and data requirements needed to create your own Q&A engine with ChatGPT/LLMs.
- 2023. 3

동작 요약

- 기존 자료를 수집해서
- 자료를 조각으로 쪼개서 모아놓고
- 각 자료 조각을 embedding 하고
- 질문이 들어오면, 그 질문을 embedding하고
- 질문에 가까운 자료 조각을 찾고
- 찾은 자료를 ChatGPT에 전달해서 요약해달라함



What is the lease budget for a level 59 employee?

The lease budget is $1200 per month.

[link to document / source]

[text of top x paragraphs ...]

Answer the question based on the text above.
Q: What is the lease budget for a level 59 employee?
A:

환경

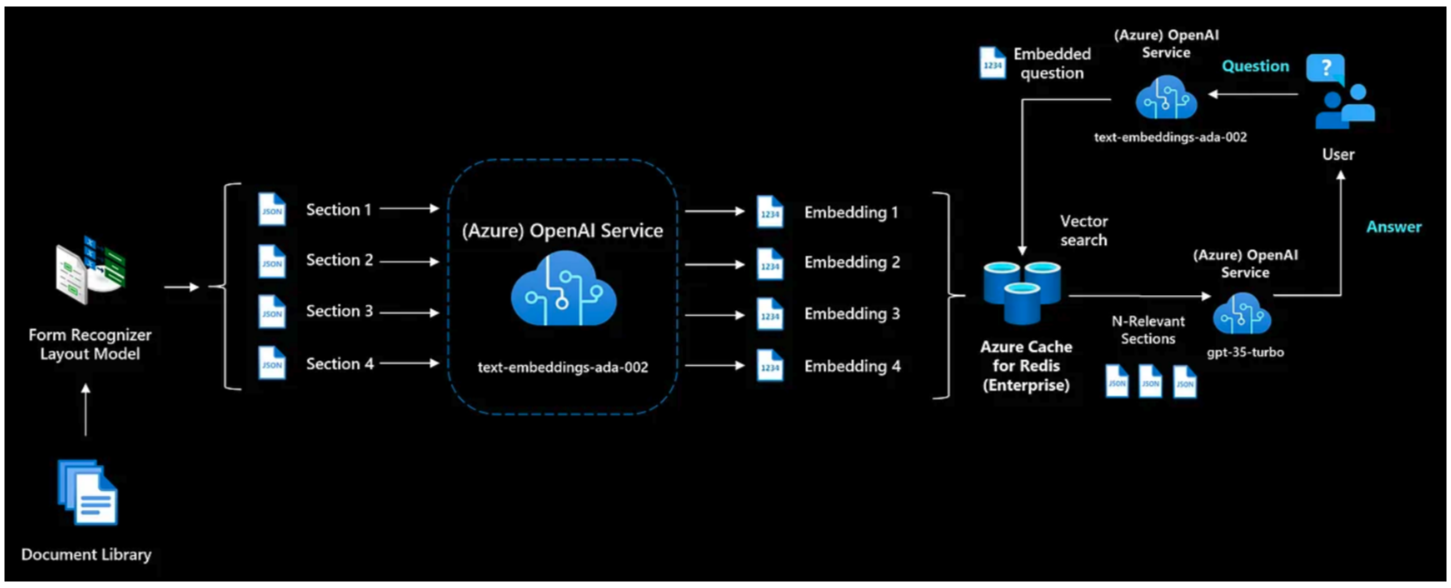
- 저자가 MS의 Solution Architect이다. 그래서 MS Azure의 서비스로 얘기하고 있다.

작업 스텝

- 문서를 파일로 모은다
- 파일을 chunk 단위로 쪼개고 검색할 수 있게 준비해 둔다.
  - option 1 : MS의 semantic ranking을 사용
  - option 2 : 별개로 쪼개서 embedding한다.
- prompt를 만들어서 입력

Split and Embedding 작업

아래 그림과 같이 MS Azure의 서비스를 사용하여 할 수 있다.



OpenAI의 ChatGPT도 Embedding도 Azure에서 서비스로 제공되고 있는 것 같다.

OpenAI의 Embedding 문서 : https://platform.openai.com/docs/guides/embeddings

Promt 예

본문에 있던 prompt

```
You are an intelligent assistant helping Contoso Inc employees with their healthcare plan questions and employee handbook questions.

Use 'you' to refer to the individual asking the questions even if they ask with 'I'.

Answer the following question using only the data provided in the sources below.

For tabular information return it as an html table. Do not return markdown format.

Each source has a name followed by colon and the actual information, always include the source name for each fact you use in the response.

If you cannot answer using the sources below, say you don't know.

###
Question: 'What is the deductible for the employee plan for a visit to Overlake in Bellevue?'
Sources:
info1.txt: deductibles depend on whether you are in-network or out-of-network. In-network deductibles are $500 for employee and $1000 for family. Out-of-network deductibles are $1000 for employee and $2000 for family.
info2.pdf: Overlake is in-network for the employee plan.
info3.pdf: Overlake is the name of the area that includes a park and ride near Bellevue.
info4.pdf: In-network institutions include Overlake, Swedish and others in the region
Answer:
In-network deductibles are $500 for employee and $1000 for family [info1.txt] and Overlake is in-network for the employee plan [info2.pdf][info4.pdf].
###
Question: '{q}'?
Sources:
{retrieved}
Answer:
"""
```

One-Shot Learning을 해야 한다고 한다. prompt안에서 Question/Answer 예를 들어 주는 것을 의미한다.

기타 - OpenAI Platform의 fine tuning 문서

- https://platform.openai.com/docs/guides/fine-tuning
- 4가지 섹션이 있다.
    - Preparing your dataset
    - Advanced usage
    - Weights & Biases
    - Example notebooks
- 실제 코드 예까지 있는 것 같다. 파악 하자

기타 - MS Azure의 OpenAI 서비스

- https://azure.microsoft.com/ko-kr/pricing/details/cognitive-services/openai-service/
- 사용 가능한 언어 모델과 가격
    - Text-Ada : $0.0004
    - Text-Babbage : $0.0005
    - Text-Curie : $0.002
    - Test-Davinci : $0.02
    - Code-Cushman : $0.024
    - Code-Davinci : $0.10
    - ChatGPT : $0.002
    - GPT-4 8K : $0.03(prompt), $0.06(답변)
    - GPT-4 32K : $0.06(prompt), $0.12(답변)
- 서비스 카테고리
    - 언어 모델
    - 이미지 모델
    - 미세 조정된 모델 ← 파인튜닝인것 같고
    - 컴퓨팅 학습 시간당 ← 학습 시킬수 있나 보다
    - 호스팅 시간당 ← 별개로 호스팅도 가능하고
    - 모델 포함

기타 - Q&A notebook

- 노트북 제목 : Question answering using embeddings-based search
- https://github.com/openai/openai-cookbook/blob/main/examples/Question_answering_using_embeddings.ipynb
- 본문과 같은 내용인데 1개의 노트북으로 보여주고 있다.
- 전체 코드 구조
    1. Prepare search data (once)

    a. Collect: download Wikipedia about the 2022 Olympics
    b. Chunk
    c. Embed: with OpenAI API
    d. Store
2. Search (once per query)
    a. generate an embedding for query
    b. rank the text sections
3. Ask (once per query)
    a. Insert the question and the most relevant sections into a message to GPT
    b. Return GPT's answer

기타 - Your Own Data

- 노트북 제목 : Power your products with ChatGPT and your own data
- https://github.com/openai/openai-cookbook/blob/main/apps/chatbot-kickstarter/powering_your_products_with_chatgpt_and_your_data.ipynb
-

Reference

- 리뷰 포스트. How to create a private ChatGPT with your own data : https://medium.com/@imicknl/how-to-create-a-private-chatgpt-with-your-own-data-15754e6378a1
- 리뷰 포스트를 구현한 코드 : https://github.com/Azure-Samples/azure-search-openai-demo
- 본문에 언급된 Embeding 소개 문서. Neural Network Embeddings Explained : https://towardsdatascience.com/neural-network-embeddings-explained-4d028e6f0526
- OpenAI Platform docs : https://platform.openai.com/docs
- Prompt Engineering Guide : https://github.com/dair-ai/Prompt-Engineering-Guide
- 비슷한 제목의. How To Build Your Own Custom ChatGPT With Custom Knowledge Base : https://betterprogramming.pub/how-to-build-your-own-custom-chatgpt-with-custom-knowledge-base-4e61ad82427e
- 비슷한 제목의. How To Build Your Own Custom ChatGPT Bot : https://medium.com/gitconnected/how-to-build-your-own-custom-chatgpt-bot-cf4af959adcc
- LangChain : https://python.langchain.com/en/latest/index.html
- MS Semantic Ranking : https://learn.microsoft.com/en-us/azure/search/semantic-ranking