

# Azure 워크샵 (5일차-2)

<Vetor Database>

<https://colab.research.google.com/drive/1me5iRJCmrlpyUgByrRLZOjC3q1AVXmwo?usp=sharing>

- 일반적인 데이터베이스는 글자가 똑같지 않으면 검색이 안 되는데
- 벡터 데이터베이스는 오타자가 있어도 됨 (벡터 기반이기 때문!!)
- 그래서 나름 유용함!

필요한 거 싹 설치하자

```
!pip install langchain tiktoken pypdf sentence_transformers chromadb

Collecting langchain
  Downloading langchain-0.1.5-py3-none-any.whl (806 kB)
Collecting tiktoken
  Downloading tiktoken-0.5.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.0 MB)
Collecting pypdf
  Downloading pypdf-4.0.1-py3-none-any.whl (283 kB)
Collecting sentence_transformers
  Downloading sentence_transformers-2.3.1-py3-none-any.whl (132 kB)
Collecting chromadb
  Downloading chromadb-0.4.22-py3-none-any.whl (509 kB)
Requirement already satisfied: PyYAML<=5.3 in /usr/local/lib/python3.10/dist-packages (from langchain) (6.0.1)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.10/dist-packages (from langchain) (2.0.24)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.10/dist-packages (from langchain) (3.9.3)
Requirement already satisfied: async-timeout<5.0.0,>=4.0.0 in /usr/local/lib/python3.10/dist-packages (from langchain) (4.0.3)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain)
  Downloading dataclasses_json-0.6.4-py3-none-any.whl (28 kB)
Collecting jsonpatch<2.0,>=1.33 (from langchain)
  Downloading jsonpatch-1.33-py2.py3-none-any.whl (12 kB)
Collecting langchain-community<0.1,>=0.0.17 (from langchain)
  Downloading langchain_community-0.0.17-py3-none-any.whl (1.6 MB)
Collecting langchain-core<0.2,>=0.1.16 (from langchain)
  Downloading langchain_core-0.1.18-py3-none-any.whl (237 kB)
Collecting langsmith<0.1,>=0.0.83 (from langchain)
  Downloading langsmith-0.0.86-py3-none-any.whl (54 kB)
Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.10/dist-packages (from langchain) (1.23.5)
Requirement already satisfied: pydantic<3,>=1 in /usr/local/lib/python3.10/dist-packages (from langchain) (1.10.14)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.10/dist-packages (from langchain) (2.31.0)
```

- chromadb가 벡터 데이터베이스임
- 나중에 실무에선 이런 거 하나하나가 컨테이너로 되어있을 것!
- ⇒ 특정한 작업에 대한 컨테이너 숫자 탄력적으로 관리 가능. 도커 컨테이너...

```

import tiktoken

tokenizer = tiktoken.get_encoding('cl100k_base')

def tiktoken_len(text):
    tokens = tokenizer.encode(text)
    return len(tokens)

[ ] from langchain.text_splitter import RecursiveCharacterTextSplitter

text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=500,
    chunk_overlap=50,
    length_function=tiktoken_len
)

```

- 토큰 단위로 자르는 splitter로 자름
- 알아서 잘라주는 recursive splitter, 토큰 500개 단위로 자름

```

from langchain.document_loaders import PyPDFLoader

loader = PyPDFLoader('/content/[정책브리프 2021-04] 탄소중립 대응을 위한 정부 정책과 동향 (1).pdf')
pages = loader.load_and_split()

```

- pdf 파일 업로드 & 페이지 단위로 쪼개기

```
from langchain.embeddings import HuggingFaceEmbeddings

model_name='jhgan/ko-sbert-nli'
model_kwargs={'device':'cpu'}
encode_kwargs={'normalize_embeddings':True}

hf = HuggingFaceEmbeddings(
    model_name = model_name,
    model_kwargs = model_kwargs,
    encode_kwargs = encode_kwargs
)
```

/usr/local/lib/python3.10/dist-packages/huggingface\_hub/utils/\_token.py:88: UserWarning:  
The secret `HF\_TOKEN` does not exist in your Colab secrets.  
To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), s  
You will be able to reuse this secret in all of your notebooks.  
Please note that authentication is recommended but still optional to access public models or datasets.  
warnings.warn(  
modules.json: 100% ██████████ 229/229 [00:00<00:00, 5.73kB/s]  
config\_sentence\_transformers.json: 100% ██████████ 123/123 [00:00<00:00, 5.74kB/s]  
README.md: 100% ██████████ 4.46k/4.46k [00:00<00:00, 155kB/s]  
sentence\_bert\_config.json: 100% ██████████ 53.0/53.0 [00:00<00:00, 1.98kB/s]  
config.json: 100% ██████████ 620/620 [00:00<00:00, 24.9kB/s]  
pytorch\_model.bin: 100% ██████████ 443M/443M [00:02<00:00, 166MB/s]  
tokenizer\_config.json: 100% ██████████ 538/538 [00:00<00:00, 19.1kB/s]  
vocab.txt: 100% ██████████ 248k/248k [00:00<00:00, 7.82MB/s]  
tokenizer.json: 100% ██████████ 495k/495k [00:00<00:00, 17.0MB/s]  
special\_tokens\_map.json: 100% ██████████ 112/112 [00:00<00:00, 6.58kB/s]  
1\_Pooling/config.json: 100% ██████████ 190/190 [00:00<00:00, 10.4kB/s]

- Hugging Face에 있는 임베딩 모델 중에서 가져오겠다~
- 거대 언어 모델 말고도 작은 언어 모델에서도 사용
- microsoft에서 만든 small language model: "PHI"  
성능이 좋아지면 소프트웨어에도 탑재될 수도 있음


```
[ ] # 데이터의 Split
docs = text_splitter.split_documents(pages)

# Vector Store 저장
from langchain.vectorstores import Chroma

db = Chroma.from_documents(docs, hf)
```

```
[ ] query = '탄소중립'
docs_result = db.similarity_search(query)
```

```
print(docs_result[0].page_content)
print('-' * 100)
print(docs_result[0].metadata['page'])
```

 이미노비즈 정책브리프  
이 탄소중립과 대응 필요성  
탄소중립이란 ?  
·탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 탄소량을 같게 해 탄소 ‘순배출이 제로(0)’가 되게 하는 것으로, ‘넷-제로(Net-Zero)’라 부르기도 함  
·인간의 활동에 의한 온실가스 + 배출을 최대한 줄이고, 남은 온실가스는 흡수, 제거하여 ‘순배출이 제로(0)’가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음  
·온실가스란 ‘지구 대기를 오염시켜 온실 효과를 일으키는 가스’를 통틀어 이르는 말로, 이산화탄소 (CO2), 메탄(CH4), 아산화질소 (N2O), 수소불화탄소 (HFCs), 과불화탄소 (PFCs), 육불화황 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음  
대응 필요성  
·지구 온난화로 인한 폭염, 폭설, 태풍, 산불 등 이상기후 현상이 세계 곳곳에서 발생되고 있는 가운데 기후변화 문제에 대한 국제사회의 공동 대응 필요성이 대두됨

---

3

- 아까 설치한 chromadb를 langchain에서 랩핑해놓은 애가 Chroma
- 그냥 집어넣으면 안 되고 벡터 기반으로 임베딩해서 집어넣어야 함
- 모든 DB는 파일에 있지만 애는 메모리에 있음. 그래서 껏다 키면 날라감. 대신 속도는 엄청 빠름.
- 암튼... 지금 순식간에 파일을 임베딩해서 DB를 생성하는 작업을 함. 위에 있는 코드가...
- 정보 세 가지
  1. 페이지 콘텐츠: 갖고 온 데이터의 진짜 내용
  2. 백 개의 -
  3. 페이지 0번의 메타 데이터. 실제로 54페이지의 내용임

그 다음 할 거: 이걸 싹 긁어다가 GPT한테 묻기

그럼 GPT는 이 내용을 기반으로 해서 나한테 말을 만들어서 답할 것

"RAG" 증폭시켜서 답한다~

```
docs_result = db.similarity_search_with_relevance_scores(query, k=3)
len(docs_result)
```

 3

- 유사도가 높은 걸 세 개까지 뽑아준다
- 여기서 k=3은 k-means 알고리즘에서의 k와 비슷한 느낌임

- 정확도를 높이기 위해서 k=3를 갖고 세 개를 같이 묶어서 대답하도록 할 수 있음

```
docs_result

[(Document(page_content='01이노비즈 정책브리프01 탄소중립과 대응 필요성탄소중립이란 ?탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 탄소를 균형을 맞추고, 남은 온실가스는 흡수, 제거하여 '순배출이 제로(0)'가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음탄소중립이란 (PFCs), 육불화황 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.6235679934827092),
(Document(page_content='보제공 확대 탄소 재생원료 사용 극대화, 철강, 플라스틱을 대체하는 혁신소재 개발, 탄소발자국 + 등 친환경 제품 정보 제공 확대 산화탄소의 배출량을 의미함 탄소중립 사회로의 공정전환 (취약 산업·계층 보호) 내연기관차 완성차 및 부품업체 등 육소산업에 대한 R&D, M&A 등을 통해 적극 지원하고 직업훈련·재취업 지원 등 고용친화적 전환을 위한 정책 추진방향 마련 (지역중심의 탄소중립 실현) 지역 중심 탄소중립 실행 지원, 지역별 특성에 맞는 탄소중립 실행 방안 수립 등 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.5406716752232998),
(Document(page_content='화력발전과 같은 탄소 다배출 사업에 대한 무역 금융 중지를 제안 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.4519302212781571)]

[ ] db = Chroma.from_documents(docs, hf, persist_directory='./chroma_db')

[ ] db2 = Chroma(persist_directory='./chroma_db',
embedding_function=hf)
```

- 임베딩해서 DB를 만들면서 chroma\_db라는 파일로 저장해줌
- 그래서 이후부터는 파일만 읽어들이면 실행되도록 함
- 이걸 클라우드에 한다면 이 파일을 클라우드의 스토리지로 넣어두면 됨! (AWS EC2 같은)

```
[ ] docs_result

[(Document(page_content='01이노비즈 정책브리프01 탄소중립과 대응 필요성탄소중립이란 ?탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 탄소를 균형을 맞추고, 남은 온실가스는 흡수, 제거하여 '순배출이 제로(0)'가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음탄소중립이란 (PFCs), 육불화황 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.6235679934827092),
(Document(page_content='보제공 확대 탄소 재생원료 사용 극대화, 철강, 플라스틱을 대체하는 혁신소재 개발, 탄소발자국 + 등 친환경 제품 정보 제공 확대 산화탄소의 배출량을 의미함 탄소중립 사회로의 공정전환 (취약 산업·계층 보호) 내연기관차 완성차 및 부품업체 등 육소산업에 대한 R&D, M&A 등을 통해 적극 지원하고 직업훈련·재취업 지원 등 고용친화적 전환을 위한 정책 추진방향 마련 (지역중심의 탄소중립 실현) 지역 중심 탄소중립 실행 지원, 지역별 특성에 맞는 탄소중립 실행 방안 수립 등 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.5406716752232998),
(Document(page_content='화력발전과 같은 탄소 다배출 사업에 대한 무역 금융 중지를 제안 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.4519302212781571)]

[ ] db = Chroma.from_documents(docs, hf, persist_directory='./chroma_db')

[ ] db2 = Chroma(persist_directory='./chroma_db',
embedding_function=hf)

[ ] db2.similarity_search(query)

[Document(page_content='01이노비즈 정책브리프01 탄소중립과 대응 필요성탄소중립이란 ?탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 탄소를 균형을 맞추고, 남은 온실가스는 흡수, 제거하여 '순배출이 제로(0)'가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음탄소중립이란 (PFCs), 육불화황 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.6235679934827092),
(Document(page_content='보제공 확대 탄소 재생원료 사용 극대화, 철강, 플라스틱을 대체하는 혁신소재 개발, 탄소발자국 + 등 친환경 제품 정보 제공 확대 산화탄소의 배출량을 의미함 탄소중립 사회로의 공정전환 (취약 산업·계층 보호) 내연기관차 완성차 및 부품업체 등 육소산업에 대한 R&D, M&A 등을 통해 적극 지원하고 직업훈련·재취업 지원 등 고용친화적 전환을 위한 정책 추진방향 마련 (지역중심의 탄소중립 실현) 지역 중심 탄소중립 실행 지원, 지역별 특성에 맞는 탄소중립 실행 방안 수립 등 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.5406716752232998),
(Document(page_content='화력발전과 같은 탄소 다배출 사업에 대한 무역 금융 중지를 제안 탄소중립을 위한 정부 정책과 동향 (1).pdf'),
 0.4519302212781571)]
```

- 결과가 이렇게 나옴

<FAISS>

페이스북이 만든 벡터 DB. 근데 별로임

이런 게 있다~ 정도만 봐두기

```
~ FAISS

[ ] !pip install faiss-cpu

Collecting faiss-cpu
  Downloading faiss-cpu-1.7.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.6 MB)
    17.6/17.6 MB 45.3 MB/s eta 0:00:00

Installing collected packages: faiss-cpu
Successfully installed faiss-cpu-1.7.4

[ ] from langchain.vectorstores import FAISS

db = FAISS.from_documents(docs, hf)

> db.similarity_search(query)
```

<Retirivel 그래서 이걸 어떻게 연결하느냐~>

검색해온 결과를 토대로 말을 만들어주는 걸 볼 것

```

!pip install openai langchain pypdf tiktoken sentence_transformers chromadb

Collecting openai
  Using cached openai-1.10.0-py3-none-any.whl (225 kB)
Collecting langchain
  Using cached langchain-0.1.5-py3-none-any.whl (806 kB)
Collecting pypdf
  Using cached pypdf-4.0.1-py3-none-any.whl (283 kB)
Collecting tiktoken
  Using cached tiktoken-0.5.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.0 MB)
Collecting sentence_transformers
  Downloading sentence_transformers-2.3.1-py3-none-any.whl (132 kB)
  132.8/132.8 kB 3.3 MB/s eta 0:00:00
Collecting chromadb
  Downloading chromadb-0.4.22-py3-none-any.whl (509 kB)
  509.0/509.0 kB 22.7 MB/s eta 0:00:00
Requirement already satisfied: anyio<5,>=3.5.0 in /usr/local/lib/python3.10/dist-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in /usr/lib/python3/dist-packages (from openai) (1.7.0)
Collecting httpx<1,>=0.23.0 (from openai)
  Downloading httpx-0.26.0-py3-none-any.whl (75 kB)
  75.9/75.9 kB 9.4 MB/s eta 0:00:00
Requirement already satisfied: pydantic<3,>=1.9.0 in /usr/local/lib/python3.10/dist-packages (from openai) (1.10.14)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/dist-packages (from openai) (1.3.0)
Requirement already satisfied: tqdm<4 in /usr/local/lib/python3.10/dist-packages (from openai) (4.66.1)
Collecting typing_extensions<5,>=4.7 (from openai)
  Downloading typing_extensions-4.9.0-py3-none-any.whl (32 kB)
Requirement already satisfied: PyYAML<=5.3 in /usr/local/lib/python3.10/dist-packages (from langchain) (6.0.1)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /usr/local/lib/python3.10/dist-packages (from langchain) (2.0.24)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.10/dist-packages (from langchain) (3.9.3)
Requirement already satisfied: async-timeout<5.0.0,>=4.0.0 in /usr/local/lib/python3.10/dist-packages (from langchain) (4.0.3)
Collecting dataclasses_json<0.7,>=0.5.7 (from langchain)
  Downloading dataclasses_json-0.6.4-py3-none-any.whl (28 kB)
Collecting jsonpatch<2.0,>=1.33 (from langchain)
  Downloading jsonpatch-1.33-py2.py3-none-any.whl (12 kB)
Collecting langchain-community<0.1,>=0.0.17 (from langchain)
  Downloading langchain_community-0.0.17-py3-none-any.whl (1.6 MB)
  1.6/1.6 MB 54.8 MB/s eta 0:00:00
Collecting langchain-core<0.2,>=0.1.16 (from langchain)
  Downloading langchain_core-0.1.18-py3-none-any.whl (237 kB)
  237.0/237.0 kB 26.6 MB/s eta 0:00:00
Collecting langsmith<0.1,>=0.0.83 (from langchain)
  Downloading langsmith-0.0.86-py3-none-any.whl (54 kB)
  54.4/54.4 kB 6.8 MB/s eta 0:00:00
Requirement already satisfied: numpy<2,>=1 in /usr/local/lib/python3.10/dist-packages (from langchain) (1.23.5)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.10/dist-packages (from langchain) (2.31.0)
Requirement already satisfied: tenacity<9.0.0,>=8.1.0 in /usr/local/lib/python3.10/dist-packages (from langchain) (8.2.3)
Requirement already satisfied: regex<=2022.1.18 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2023.6.3)
Requirement already satisfied: transformers<5.0.0,>=4.32.0 in /usr/local/lib/python3.10/dist-packages (from sentence_transformers) (4.35.2)
Requirement already satisfied: torch<=1.11.0 in /usr/local/lib/python3.10/dist-packages (from sentence_transformers) (2.1.0+cu121)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from sentence_transformers) (1.2.2)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from sentence_transformers) (1.11.4)

```

```

[ ] import os

os.environ['OPENAI_API_KEY'] = 'bb6860ae08e147eeae418efdce560e4e'
os.environ['AZURE_OPENAI_ENDPOINT'] = 'https://helloai-openai-025.openai.azure.com/'
os.environ['OPENAI_API_TYPE'] = 'azure'
os.environ['OPENAI_API_VERSION'] = '2023-05-15'

[ ] # Data의 loading
import openai
import tiktoken

from langchain.document_loaders import PyPDFLoader
from langchain.text_splitter import RecursiveCharacterTextSplitter
from langchain.chat_models import AzureChatOpenAI
from langchain.vectorstores import Chroma
from langchain.chains import RetrievalQA

loader = PyPDFLoader('/content/[정책브리프 2021-04] 탄소중립 대응을 위한 정부 정책과 동향 (1).pdf')
pages = loader.load_and_split()

[ ] pages = loader.load_and_split()

```

- key & endpoint 값 바꿔주고 pdf 파일 업로드
- 필요한 패키지 업로드

- RetrievalQA를 통해 묶어 놓으면 질문을 던졌을 때 다 연결되어있는 채로 답함
- 경로 복사 & 로딩 & 스플릿

```
[ ] import tiktoken

tokenizer = tiktoken.get_encoding('cl100k_base')

def tiktoken_len(text):
    tokens = tokenizer.encode(text)
    return len(tokens)
```

- 토크나이저 만드는 함수

```
[ ] from langchain.text_splitter import RecursiveCharacterTextSplitter

text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=500,
    chunk_overlap=50,
    length_function=tiktoken_len
)

from langchain.embeddings import HuggingFaceEmbeddings

model_name='jlgan/ko-sbert-nli'
model_kwargs={'device':'cpu'}
encode_kwargs={'normalize_embeddings':True}

hf = HuggingFaceEmbeddings(
    model_name = model_name,
    model_kwargs = model_kwargs,
    encode_kwargs = encode_kwargs
)
```

/usr/local/lib/python3.10/dist-packages/huggingface\_hub/utils/\_token.py:88: UserWarning: The secret 'HF\_TOKEN' does not exist in your Colab secrets. To authenticate with the Hugging Face Hub, create a token in your settings tab (<https://huggingface.co/settings/tokens>), set it as secret in your Google Colab and you will be able to reuse this secret in all of your notebooks. Please note that authentication is recommended but still optional to access public models or datasets.

warnings.warn(

modules.json: 100%	229/229	[00:00<00:00, 10.3kB/s]
config_sentence_transformers.json: 100%	123/123	[00:00<00:00, 4.97kB/s]
README.md: 100%	4.46k/4.46k	[00:00<00:00, 291kB/s]
sentence_bert_config.json: 100%	53.0/53.0	[00:00<00:00, 2.56kB/s]
config.json: 100%	620/620	[00:00<00:00, 20.6kB/s]
pytorch_model.bin: 100%	443M/443M	[00:06<00:00, 102MB/s]
tokenizer_config.json: 100%	538/538	[00:00<00:00, 27.6kB/s]
vocab.txt: 100%	248k/248k	[00:00<00:00, 10.2MB/s]
tokenizer.json: 100%	495k/495k	[00:00<00:00, 18.2MB/s]
special_tokens_map.json: 100%	112/112	[00:00<00:00, 6.18kB/s]
1_Pooling/config.json: 100%	190/190	[00:00<00:00, 5.24kB/s]

- splitter를 가져오고 아까 사용했던 HuggingFace 임베딩 모델 가져옴
- 나중에 코딩할 땐 무료라.. 오류 생길 수 있음.



```
[ ] # split
docs = text_splitter.split_documents(pages)

db = Chroma.from_documents(docs, hf)

# 언어 모델의 생성
chatgpt = AzureChatOpenAI(
    deployment_name='dev-gpt-35-turbo'
)

qa = RetrievalQA.from_chain_type(
    llm = chatgpt,
    chain_type = 'stuff',
    retriever = db.as_retriever(
        search_type = 'mmr',
        search_kwargs = {'k':3,'fetch_k':10}
    ),
    return_source_documents=True
)
```

```
[ ] query='탄소중립에 중소기업이 필요한 이유는?'
```

```
[ ] result = qa(query)
result
```

```
{'query': '탄소중립에 중소기업이 필요한 이유는?',
 'result': '중소기업의 적극적인 참여가 필요한 이유는 탄소중립을 실현하기 위해 국가 경제에서 중소기업이 차지하는 비중이 높기 때문입니다. 중소기업의 참여를 통해 탄소중립에 대한
 있는 기회를 얻을 수 있습니다.',
 'source_documents': [Document(page_content='대한 대응 필요성을 느끼지 못한다는 결과를 보여주고 있음\n\n향후 중소기업에 대상으로 도입할 정부 정책과 제도에 대한 관심과 적극적
 불가능함\n\n따라서 탄소중립에 대한 중소기업의 적극적 참여가 필요하며 , 중소기업의 참여를 유인할 수 있는 다양한 \n\n정부 정책과 제도가 마련될 예정임\n\n지원책을 적극 활용하여 업
 성장할 수 있는 동력을 마련할 수 있을 것임', metadata={'page': 11, 'source': '/content/[정책브리프 2021-04] 탄소중립 대응을 위한 정부 정책과 동향 (1).pdf'}),
 Document(page_content='속화 추진\n\n산업 분야에서는 재생에너지 , 수소, 에너지 IT 등 3대 에너지 신산업 육성에 집중\n\n(고탄소 산업구조 혁신) 철강, 석유화학 등 탄소 다배출 업
 업 등 다배출 중소기업에 대상으로 공정전단을 통해 맞춤형 공정개선 ,설비보급을 지원\n\n하는 등 산업 전과정 (연·원료→공정→제품→소비·자원 순환)에서의 저탄소 전환을 촉진\n\n(미래
 거점별 수소 충전소 구축\n\n(도시·국도 저탄소화 ) 신규 건축물 제로에너지 건축 의무화 , 국토 계획 수립 시 생태자원 활용한 탄소흡수능 강화 \n\n② 신유망 저탄소산업 생태계
 Document(page_content='이이노비즈 정책브리프\n\n01 탄소중립과 대응 필요성\n\n탄소중립이란 ?\n\n탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 탄소량을 같게 해 탄소
 줄이고 , 남은 온실가스는 흡수, 제거하여 '순배출이 제로(0)'가 되는 개념이며 , 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음\n\n온실가스란 '지구 대기를 오염시켜 온
 (PFCs), 육불화화물 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음\n\n대응 필요성\n\n지구 온난화로 인한 폭염, 폭설, 태풍, 산불 등 이상기후 현상이 세계 곳곳에서 발생되고 있는 가
 탄소중립 대응을 위한 정부 정책과 동향 (1).pdf')}]}
```

- ESG 평가... 탄소 얼마나 사용하는지 다 따짐. 준수 X→ 펀드도 안 들어옴
- 답을 만드는 모델을 좋은 걸 쓰면 또 더 좋으니까... 우리는 GPT 3.5 turbo를 사용할 것
- chain\_type: 결과가 나오면 묶어줌. 오늘은 가장 기본적인 방법을 쓸 것
- retriever: db는 아까 위에서 만든 chromadb... chromadb의 유사도 검색을 할 것  
k=3 유사도 3개... fetch\_k=10: 내가 질문을 던지면 chromadb에서 10개를 뽑아내고, 그  
중에 가장 괜찮을 걸 뽑겠다는 거
- return\_source\_documents=True: 어느 pdf 파일의 몇 번째 페이지... 이런 정보 돌려줌
- // 암튼 이렇게 qna chain을 연결함. chain한테 일을 시키면 거기 묶여있는 애들까지 다~

```
[ ] result = qa(query)
result

{'query': '탄소중립에 중소기업에 필요한 이유는?',
 'result': '중소기업의 적극적인 참여가 필요한 이유는 탄소중립을 실현하기 위해 국가 경제에서 중소기업이 차지하는 비중이 높기 때문입니다. 중소기업의 참여는 기회를 얻을 수 있습니다.',
 'source_documents': [{'Document(page_content='대한 대응 필요성을 느끼지 못한다는 결과를 보여주었고 있음w>한중 중소기업을 대상으로 도입될 정부 정책과 제도 불가합w> 따라서 탄소중립에 대한 중소기업의 적극적 참여가 필요하w>, 중소기업의 참여를 유인할 수 있는 다양한 w>한정부 정책과 제도가 마련될 예정임w> 구성장w> 수 있는 동력을 마련할 수 있을 것임', metadata={'page': 11, 'source': '/content/정책브리프 2021-04/ 탄소중립 대응을 위한 정부 정책과 동향 (1).p', 'document_id': '국회 차원에서는 재설계w>지, w>수, 에너지 IT 등 3대 에너지 산업 육성에 집중w> (고탄소 산업구조 혁신) 철강, 석업 등 다배출 중소기업을 대상으로 공정전단을 통해 맞춤형 공정개선, 설비보급을 지원w>하는 등 산업 전과정 (연·원료→공정→제품→소비·자원 순환)에서의 저탄거점별 w>수 출전소 구축w> (도시·국도 저탄소화) 신규 건축물 제로에너지 건축 의무화, 국도 계획 수립 시 생태자원 활용한 탄소흡w>수 기능 강화 w> 등 신Document(page_content='01이노비즈 정책브리프w>01 탄소중립과 대응 필요성w>탄소중립이란 ?w>탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 w>출이고, 남은 온실가스는 흡수, 제거하여 '순배출이 제로(w>0)'가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음w>수온실가스란 ' (PFCS), 육불화탄 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음 w>한대응 필요성w>지구 온난화로 인한 폭염, 폭설, 태풍, 산불 등 이상기후 현상이 세계 곳곳탄소중립 대응을 위한 정부 정책과 동향 (1).pdf)}}}]
```

- qa 체인한테 질문을 던지고 결과를 가져옴
- 결과 잘 나옴~
- 그냥 하면 조각 데이터만 나왔을 것. 근데 여긴 gpt가 말을 만들어서 답해줌
- 소스 document... page 60에서 가져왔다고까지 알려줌

// 끝...