


# Day5

Hugging Face – The AI community building the future.

We're on a journey to advance and democratize artificial intelligence through open source and open science.

 <https://huggingface.co/>



**The AI community  
building the future**

언어 모델계의 깃허브 느낌

- 새로운 모델 출시할 경우 여기 오픈됨.
- AI모델도 많이 올라와있음

LangCon 2024 참고

5년 이내에 내연기관 사라질 것으로 예상

spring ⇒ spring boot ⇒ node.js

아이디에이션

- 설문조사 분류
- QR코드 발생기
  - 인식 후 자동 등록까지

마이크로소프트의 직원 평가

- 마이크로소프트는 직원이 되면 주식을 주고 매년 두 가지로 사람을 평가함
- 마이크로소프트 내에 자판기가 있는데 하루에 캔이 24만 개 소비된다는 통계도 있음

- 마이크로소프트는 직원 수가 12만 명 정도 되는데 클라이언트는 유럽과 미국에 있음

#### 모델 공개

- 새로운 모델들이 나오면 오픈임
- 공개되는 모델들은 다 올라와 있음
- 원하는 결과를 빠르게 만들어내는 게 중요함

#### 신생 분야의 장점

- 신생 분야라 경력이 필요 없고 이쪽으로 쪽 가는 게 더 좋음
- 3분의 1 정도는 못 알아듣고 다시 가서 물어보면 잘 아는 사람일수록 대답을 시원하게 잘해줌
- 학교 수업 말고도 이런 데 가서 한 번씩 잘 찾아서 들어보면 도움이 될 것임

#### 언어 모델의 한계

- 언어 모델이 처리할 수 있는 한도가 있음
- 토큰 수가 몇 개인지 확인해야 함
- 페이지를 쪼개는 방법을 해볼 것임
- 텍스트 스플리터를 사용할 때 옵션들이 있음

#### 토큰 아이디어

- 토큰 단위로 자르기 위해서는 더 복잡한 공식을 써야 됨
- PRP 인포에 오픈 소스로 나와 있는 게 있어서 그냥 그걸 사용하면 됨
- 토큰 단위로 잘라주는 걸 토큰 아이디어라고 함
- encode 방식을 써야 정확하게 잡힘

#### 저 오픈 AI의 임베딩

- 애저 오픈 AI에 있는 임베딩을 쓰겠다고 하면 애드 오픈을 가지고 와서 임베딩을 함
- 차원 수가 많았으면 더 정확하게 분류할 수 있음
- 데이터가 흩어버리면 상온의 저주가 걸림
- 데이터를 embedding 모델 만들어내는 GPU도 다 비용임

#### 허깅 페이스

- 허깅 페이스는 벡터 기반으로 바뀌어서 검색을 하면 성능에 비해서 비용이 거의 안 듬
- 허깅 페이스에 있는 모델들을 바로 다운받고 사용할 수가 있음

- 허깅 페이스에 있는 모델들은 잘 튜닝해서 CPU만으로도 돌아가게끔 튜닝해놓은 모델들이 있음

## 유사도 분석

- 유사도가 무료로 쓸 수 있는 임베딩 모델임
- 인베이딩할 때는 무료 보드를 쓰고 최적화시켜서 쓰는 방법들도 있음
- 소스 llm이기 때문에 이런 것들을 사용하는 것도 방법이 있음

## Azure for student 잘 사용할 것 200 크레딧

## 친구들것도 부탁하기

```

!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)

```

```

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
)

```

```

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()

```

```

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]

```

```

[ ] # 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 제
로(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), 수
소불화탄소 (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 가
운데 기후변화 문제에 대한 국제사회의 공동 대응 필요성이 대두됨
-----
3

```

```

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

3

```

```
[ ] docs_result

[[{"Document(page_content='01이노비즈 정책브리핑01 탄소중립과 대응 필요성탄소중립이란 무엇 탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 양을 탄소 자산의 흡수, 제거하여 '순배출이 제로(0)'가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추스로 자리 잡고 있을수록 온실가스란 (PFCs), 육불화탄소 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있을 때 대응을 필요성 탄소 중립화로 인한 목표, 핵심, 태생, 산물 등 이상기후 현상이 세계 곳곳 탄소중립 대응을 위한 정부 정책과 동향 (1).pdf')",
  0.6235679934827092),
  {"Document(page_content='보재공 확대 탄소중립원리 사용 확대와, 열량, 플라즈마를 대체하는 혁신소재 개발, 탄소발자국 + 등 친환경 제품 정보 제공공 확대 산화탄소의 흡수량을 의미함  탄소중립 사회로의 공적전환탄 (화학 산업, 기술 보호) 내연기관차 완성차 및 부품업체 등 축산산업에 대한 R&D, M&A 등을 위한 지원사업도 직접지원·재원 지원 등 고소전환화 전환을 위한 정책 추진방향 마련탄 (지역중심의 탄소중립 실현) 지역 중심 탄소중립 실현 지원, 지역적 0.54607167322393),
  {"Document(page_content='화력발전과 같은 탄소 배출물 사안에 대한 무역 금융 중지를 제안함 내 국가에서 수입되는 모든 제품에 대하여 온실가스 감축의 적은 뒤 점차 감소할 수 있도록 목표를 수립하고 이를 실천함 교통(친환경차), 에너지 (신재생에너지) 및 탄소배출권 거래제 도입 등의 국가 단위의 정책 및 배출량 제로화를 위한 탄소중립, 수소, 자동차 등 14개의 주요 분야 목표와 실현 계획이 포함된 '그린성장전략' 발표', metadata={'page': 8, 'source': '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 정책브리핑01 탄소중립과 대응 필요성탄소중립이란 무엇 탄소중립은 개인·회사·단체 등에서 배출되는 탄소와 흡수되는 양을 탄소 자산의 흡수, 제거하여 '순배출이 제로(0)'가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 추스로 자리 잡고 있을수록 온실가스란 (PFCs), 육불화탄소 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있을 때 대응을 필요성 탄소 중립화로 인한 목표, 핵심, 태생, 산물 등 이상기후 현상이 세계 곳곳 탄소중립 대응을 위한 정부 정책과 동향 (1).pdf')",
  {"Document(page_content='보재공 확대 탄소중립원리 사용 확대와, 열량, 플라즈마를 대체하는 혁신소재 개발, 탄소발자국 + 등 친환경 제품 정보 제공공 확대 산화탄소의 흡수량을 의미함  탄소중립 사회로의 공적전환탄 (화학 산업, 기술 보호) 내연기관차 완성차 및 부품업체 등 축산산업에 대한 R&D, M&A 등을 위한 지원사업도 직접지원·재원 지원 등 고소전환화 전환을 위한 정책 추진방향 마련탄 (지역중심의 탄소중립 실현) 지역 중심 탄소중립 실현 지원, 지역적 0.54607167322393),
  {"Document(page_content='화력발전과 같은 탄소 배출물 사안에 대한 무역 금융 중지를 제안함 내 국가에서 수입되는 모든 제품에 대하여 온실가스 감축의 적은 뒤 점차 감소할 수 있도록 목표를 수립하고 이를 실천함 교통(친환경차), 에너지 (신재생에너지) 및 탄소배출권 거래제 도입 등의 국가 단위의 정책 및 배출량 제로화를 위한 탄소중립, 수소, 자동차 등 14개의 주요 분야 목표와 실현 계획이 포함된 '그린성장전략' 발표', metadata={'page': 8, 'source': '0.4519302212781571}),
  {"Document(page_content='국가 계획과 연계한 지역맞춤형 탄소중립 계획 수립 등을 지원하여 지역 주도의 탄소중립 실현이 가능하도록 추진함 (탄소중립, 융합기술(가칭)' 신규조성, 세제·부담금·비영리거래제 등 탄소가치 체계 제구축, 탄소인지에 산책제도 도입 검토함 (녹색금융) 정책금융지원의 녹색분야 확대와, 태양전지 등 탄소중립을 위한 핵심기술 개발 집중 지원', metadata={'page': 6, 'source': '/content/[정책브리핑 2021-04] 탄소중립 대응을 위한 정책'
  ]}
```

```
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)

[ ] [Document(page_content='010노비즈 정책브리핑'이 탄소중립과 대응 필요성탄소중립이란? 탄소중립은 개인, 회사, 단체 등에서 배출되는 탄소와 흡수되는 탄소를 같게 해 탄소중립이, 남은 온실가스는 흡수, 제거하여 순배출이 제로(0)가 되는 개념이며, 기후변화에 대응하기 위한 글로벌 주제로 자리 잡고 있습니다.온실가스는 지구 대기를 온실효과(FE), 즉 온실효과 (GE)으로 인해 지구의 대기가 더워지고 있을 때를 필요성탄소 지구 온난화로 인한 폭염, 폭설, 태풍, 산불 등 이상기후 현상이 세계 곳곳에서 발생되고 있는 것을 위한 정부 정책과 방안 (1) page : 3)

Document(page_content='보통 정책 확대 탄소 재정원으로 사용 극대화, 일감, 불타소식을 대체하는 혁신산책 개발, 탄소발자국 등 친환경 제품 생산 제정공 확대 등 순환 사회로의 전환 탄소중립 탄소를 의미한다 탄소중립 사회로의 공적정확화(원자 산업, 기술 보호) 내연기관차 완성차 및 부품업체 등 축소산업을 대한 R&D, M&A 등을 통해 친환경, 유망분야로 적극 지원한다' 직접본문 제4면 지면 등 공통된 정책의 전환을 위한 정책 추진방안 마련함 (지역중립의 탄소중립 실현) 지역 중심 탄소중립 실현 지원, 지역별 맞춤형 정책 개발, 지역별 특장점 활용 등 추진방안과 같은 내용 다배출 시대에 대한 무의 관할 중기적 제안함) 한국에서 추진되는 모든 계층에 대해서 온실가스 감축 의무를 충족시켜 계획한 중점적 및 전략적 추진소 수 있도록 목표를 설정하고 이를 실천함( 교통(친환경차), 에너지 (신재생에너지), 및 탄소배출권 거래제 도입 등의 국가 단위의 정책, 발전 정책과 일반민(2020) 배출량 자료를 위한 해상정책, 수도, 자식자 등 14개의 주요 분야 목표와 실행 계획이 포함됨 '그린성장전략' 발표, metadata={ 'source': '/content/정책브리핑2021-04 탄소배출량 자료' } ) 국가 계획과 연계한 지역맞춤형 탄소중립 계획 수립 등을 지원하여 지역 주도의 탄소중립 실현이 가능함탄소중립 추진전략(탄소배출 사회에 대한 국민인식 제고) (가정) : 친환경적 제품 구매, 배출권거래제 등 탄소가격 체계 구축, 탄소인자(배출량)에 도입 검토(탄소) (복합성) : 정책결정기관의 복합분야 저감지침 내용 확대, 저탄소화, 다중정책 등 탄소중립을 위한 핵심기술 개발, 정책 지원, metadata={ 'source': '/content/정책브리핑2021-04 탄소중립 대응을 위한 정부 정책과 방안 (1), page : 3)


```

```

!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<=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'] = 'bb6860ae08e147eeae418efdc560e4e'
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()

```

```

[ ] 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
)

[ ] from langchain.embeddings import HuggingFaceEmbeddings

model_name='jhsrn/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 use it as 'HF_TOKEN' in your code.
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]
```

```
[ ] # 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 관심을 높일 수 있습니다.',
 'source_documents': '[Document(page_content='대한 대응 필요성을 느끼지 못한다는 결과를 보여주고 있음'한후 중소기업을 대상으로 도입할 정부 정책과 제도'에 대한 관심과 적극적인 참가를 유도'에 대한 중소기업의 적극적인 참여가 필요하며 , 중소기업의 참여를 유도할 수 있는 다양한 '행정' 정책과 제도가 마련될 예정임' 지원책을 적극 활용하여 '탄소중립'을 달성할 수 있는 동력을 마련할 수 있을 것임 , metadata={'page': 11, 'source': '/content/('정책브리프 2021-04) 탄소중립 대응을 위한 정부 정책과 방안 (1).pdf' )],
 Document(page_content='속화 추진' 산업 분야에서는 재생에너지 , 수소, 에너지 IT 등 3대 에너지 산업 육성에 집중' (고탄소 산업구조 혁신) 일환, 석유화학 등 탄소 대비율 50% 인 등 다배출 중소기업을 대상으로 공정전환을 통해 맞춤형 공정개선 '설비'보급을 지원'하는 등 산업 전과정 (연.원료→공정→제품→수배.지원.순환)에서의 저탄소 전환을 촉진' (이대.기업별 수소 중점소 구축' (도시.국도 저탄소화) 신규 건축물 제로에너지 건축 의무화 , 국토 계획 수립 시 생태자랑 활용'한 탄소중립'수 기반 강화 'n O' 신유망 저탄소산업 생태계'
Document(page_content='이)노비즈 정책브리프'이 탄소중립과 대응 필요성'탄소중립이란 '탄소중립'은 개인, 회사, 단체 등에서 배출되는 탄소와 흡수되는 탄소를 함께 해 탄소중립이고 , 남은 온실가스는 흡수, 제거하여 '순배출'이 제로'인)가 되는 개념이며 , 기후변화에 대응하기 위한 글로벌 추세로 자리 잡고 있음'온실가스는 '지구 대기'를 오염시켜 온(PFCs), 육불화화 (SF6) 등은 배출 규제 및 거래의 대상이 되고 있음 '탄소'를 필요성' 지구 온난화로 인한 폭염, 폭설, 태풍, 산불 등 이상기후 현상이 세계 곳곳에서 발생되고 있는 기'탄소중립' 대응을 위한 정부 정책과 방안 (1).pdf' )]]
```

```
[ ] 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  

   하는 기회를 얻을 수 있습니다.'},
 {'source_documents': '[Document(page_content='\"대한 대응 필요성을 느끼지 못한다면 결과를 보여주고 있을텐~한후 중소기업들 대상으로 도입할 정부 정책과 제도  

   불가능하냐~따라서 탄소를 줄이는 데 대한 중소기업의 적극적 참여가 필요해\", page_index=0), Document(page_content='\"중소기업의 참여를 유인할 수 있는 다양한 인센티브 정책과 제도가 마련될 예정인데~거  

   성상할 수 있는 동력을 마련할 수 있을 것임\", metadata={'page': 11, 'source': '/content/정책브리핑-자료 2021-04}) 탄소를 줄이는 대응을 위한 정부 정책과 통찰(1).  

   Document(page_content='\"확화 추진원~산업 분야에서는 재정지원(1), 수소, 에너지 IT 등 3대 에너지 신산업 육성에 집중원~(고탄소 산업구조 혁신) 포함, 석,  

   연 등 대체용 중소기업들 대상으로 공정전환을 통해 맞춤형 공정전~설비보급을 지원하자는 등 산업 권조장(연·연료→제품→치환·수환)에서의 저탄  

   거점별 수소 충전소 구축원~(도시·국도 주변소재) 및 신규 용역 제공원~다만 중소기업은 조달·공급·제조 원·소재·에너지 신·생태계와 관련한 탄소가중수소 사용 증가와 관련  

   Document(page_content='\"이제부터는 정책브리핑에 중소기업과 대응 가능한중소기업들이 많은~한 중소기업들의 개인·단체 등에서 발굴되는 탄소와 중소기업  

   줄이고, 남은 온실가스는 흡수, 제거한다.\" 순배출이 목표함(0) 가 되는 개념이다. 기후변화에 대응하기 위한 글로벌 추세로 인해 감소되고 있는온실가스(온  

   (PFCs), 과불화합물(SF6) 등은 배플 구멍 및 기어의 대상이 되고 있음 해당등 필요성강~지구 온난화로 인한 폭염, 폭설, 대풍, 산불 등 이상기후 현상이 세계 곳곳  

   탄소중립 미래상을 위한 정부 정책과 통찰(1.러더)~}}]
```