

문서 요약 비즈니스 툴 개발

SKN 5기 프로젝트

4조 : 배운관, 신혜원, 윤상혁, 이준호

기존 문제점

- 기존 다양한 문서의 내용을 한눈에 파악하기 어려움
- 문서 별로 자체 단어 검색을 통해 해당 내용을 직접 찾아야 함
- 해당 단어 검색으로 추론 등 추가적으로 답을 찾기 위한 행동을 해야 함

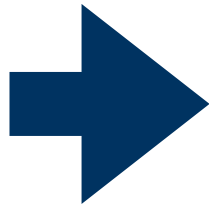
해결 방안

- LLM 모델을 활용하여 문서 검색결과 요약 모델 구축
- 업로드된 문서 관련 질문 시 학습된 문서 내용을 바탕으로 답변 도출
- 챗봇 형식의 답변
- 문서 검색 답변 보완 기능 및 출처 하단 표기
- 학습된 내용을 DB에 저장 누적된 데이터를 반영하여 답변

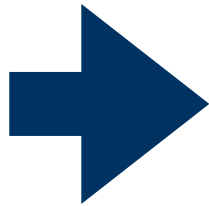
서비스



파일 업로드



학습



검색 결과 요약
출력

기술 스펙

- 엠베딩 : text-embedding-ada-002
- 검색 : pinecone
- 생성 : GPT-3.5 또는 GPT-4 (OpenAI)
- 언어지원 : 한국어, 영어 등(추후 다양한 언어 지원 예정)

Pinecone

실시간으로 데이터를 처리하고 검색할 수 있어
사용자 경험 향상에 용이

데이터 처리 속도와 저장 용량을 쉽게 확장 가능

OpenAI, Hugging Face, LangChain 등
다양한 AI 인프라와의 통합이 원활하여 개발자들이
손쉽게 활용 가능



모델 구현

Open AI

자연어 처리 능력

문맥 이해와 자연스러운 대화 생성에서 뛰어난 성능 구현

다국어 지원

한국어, 영어, 중국어 등 여러 언어에서 높은 수준의 자연어 처리를 지원

적응력과 확장성

- GPT는 다양한 도메인에 적응할 수 있는 범용 모델로, 의료, 금융, 교육 등 특정 분야의 챗봇에도 쉽게 적용 가능
- Fine-tuning 없이도 고성능을 제공하며, 필요 시 API 기반으로 확장 가능

```
pc = Pinecone(api_key=api_key)

# Pinecone 인덱스 생성 (없는 경우)
if index_name not in pc.list_indexes().names():
    pc.create_index(
        name=index_name,
        dimension=1536, # OpenAI text-embedding-ada-002의 차원
        metric="cosine",
        spec=ServerlessSpec(
            cloud="aws",
            region="us-east1"
        )
    )

# Pinecone 인덱스 로드
index = pc.Index(index_name)

# OpenAI 설정
os.environ["OPENAI_API_KEY"] = '
embedding_model = OpenAIEmbeddings(model="text-embedding-ada-002")
llm = OpenAI(temperature=0.5, max_tokens=200)

# Pinecone 벡터 저장소 초기화
vector_db = Pinecone(index, embedding_model.embed_query, "text")

# 프롬프트 템플릿 설정
prompt_template = """
다음 정보를 바탕으로 질문에 답변하세요.

{context}

질문: {question}
간단한 답변: """
PROMPT = PromptTemplate(template=prompt_template, input_variables=["context", "question"])
```

서비스 (예시 - Streamlit)

파일 업로드

파일을 업로드하세요

Drag and drop file here

Limit 200MB per file • PDF, CSV, TXT, XLSX

Browse files

문서 Q&A 시스템

질문을 입력하세요:

aws가 뭐야?

답변 생성

GPT 응답:

AWS는 Amazon Web Services의 약자로, 아마존이 제공하는 클라우드 컴퓨팅 서비스입니다. 이는 네트워크를 통해 다양한 IT 리소스와 어플리케이션을 온디맨드로 제공하는 서비스로, AWS는 이를 대표하는 퍼블릭 클라우드 플랫폼 중 하나입니다. AWS를 사용하면 사용자는 필요한 IT 자원을 클라우드에서 즉시 이용할 수 있으며, 서버, 스토리지, 네트워크 등의 IT 기반 자원을 효율적으로 관리할 수 있습니다. 또한 AWS를 통해 다양한 비즈니스적인 이점을 누릴 수 있으며, 기술적인 측면에서도 자동화, 탄력적인 확장, 개발 lifecycle 단축 등의 기능을 활용할 수 있습니다. 요약하면, AWS는 클라우드 컴퓨팅을 제공하는 아마존의 서비스이며, 다양한 이점을 제공합니다.

서비스 (앱-예시)

업로드 → 문서요약

다운로드
및 복사 ← 옵션제공

