



BASE 미니프로젝트 1

RAG 를 활용한 보아즈 Slack 봇 " BOAchat "





RAG를 활용한 보아즈 Slack 봇 "BOAchat"

팀명

나 RAG 도 RAG 이다 ~!



팀원



유하린



김이정



이혜민



전현지





#프로젝트 배경

RAG를 활용한 보아즈 Slack 봇 "BOAchat"

문제 상황 정의

- 활동별로 다양한 단체 카톡방(분석 BASE, 멘멘, 2024 활동기수, 스터디) 존재
- 필요한 정보를 단체 카톡방에서 찾아야 하는지 헷갈림 -> 비효율
- 개별 활동 점수를 즉각적으로 확인하는 데 어려움이 있음
- 공지를 읽어도 이해가 가지 않아 담당자 연결이 필요한 상황

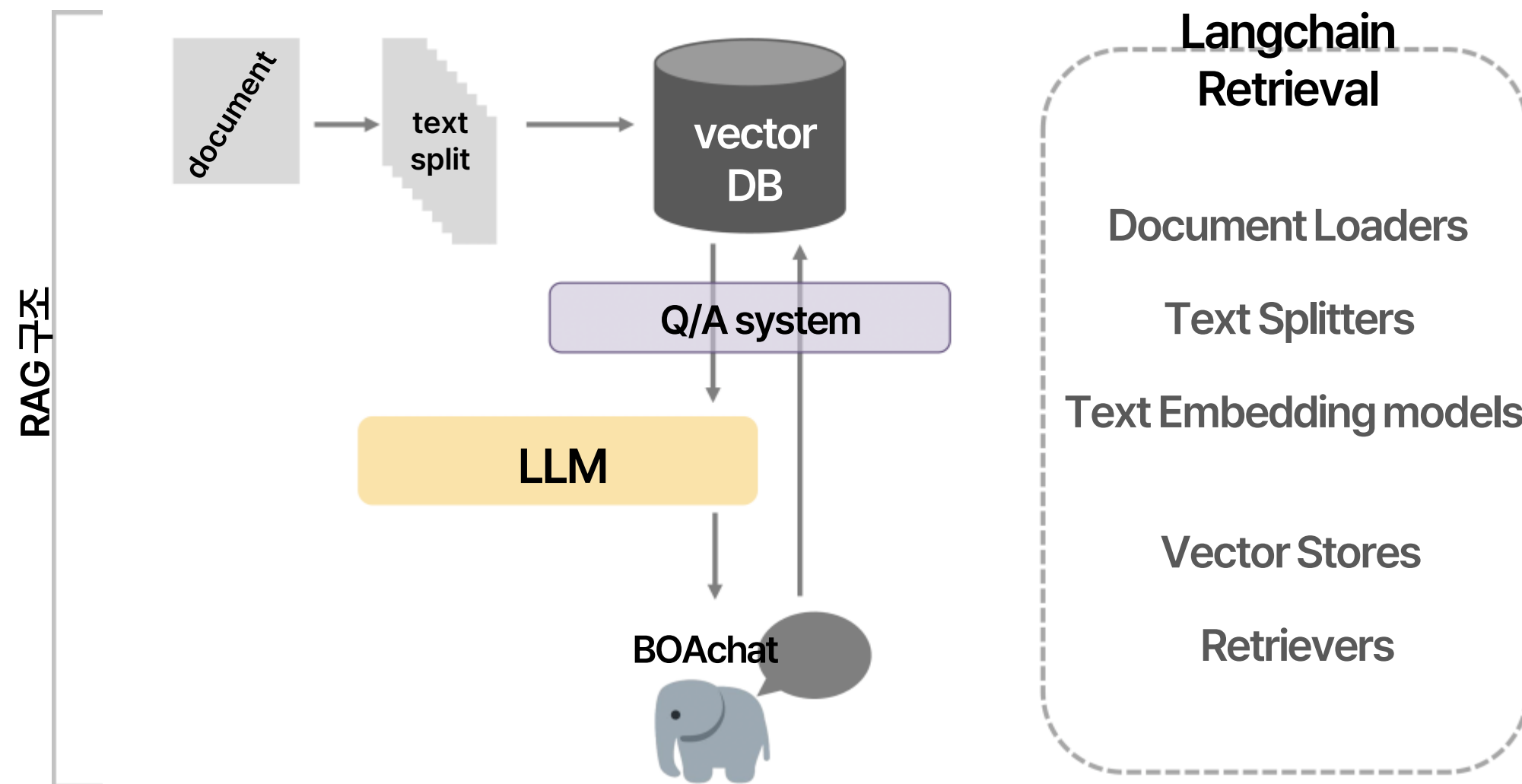


보아즈 챗봇을 만들어보면 어떨까 ?



1 RAG (Retrieval Augmented Generation) + LangChain

- RAG(검색 증강 생성)를 통해 외부 소스에서 검색된 추가 지식으로 입력 프롬프트를 강화
- 외부 데이터 저장소(vector DB) 자체가 knowledge source가 되고 이를 기반으로 LLM 실행



※ LangChain이란?
언어 모델(Language model)과
외부 도구를 마치 사슬(Chain)처럼
엮어 결합하는 도구(Framework)



#프로젝트 구현 개요

1 RAG (Retrieval Augmented Generation) + LangChain

- 근거 있는 생성으로 LLM의 할루시네이션 현상 방지
- 전통적인 모델 학습 과정 불필요

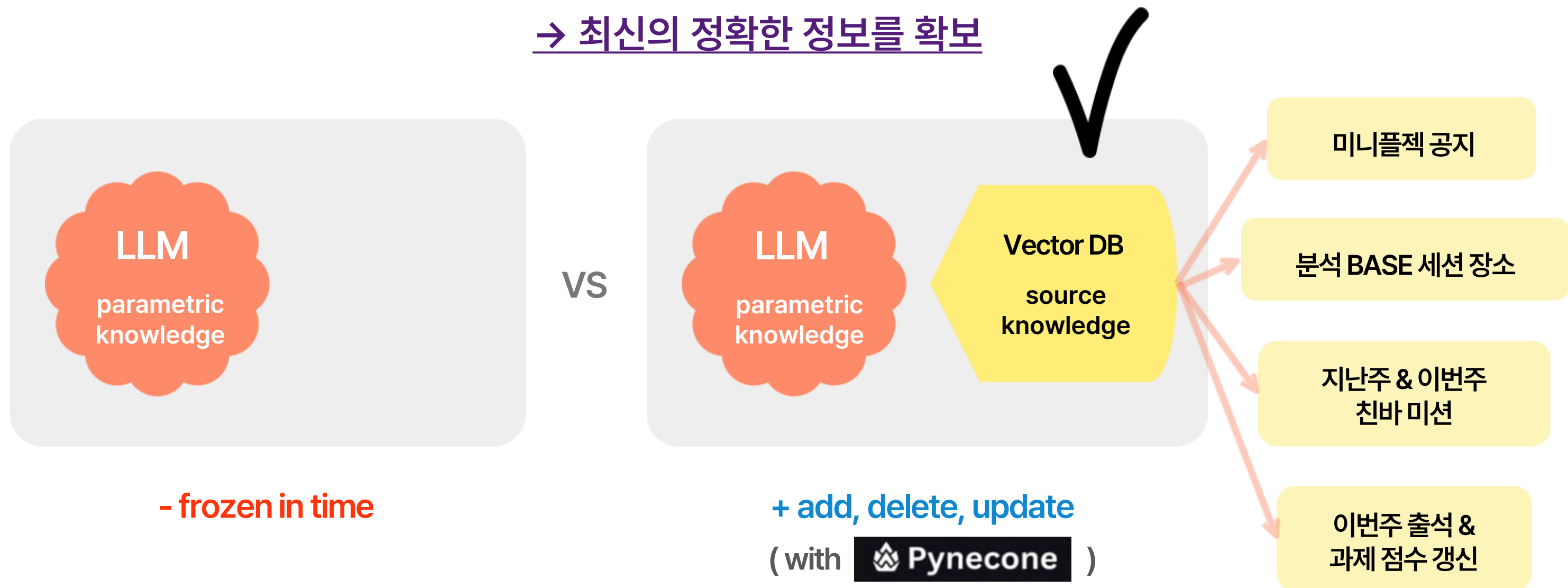
	RAG	Fine-tuning
접근 방식	DB로 기존 LLM 보강	특정 업무에 특화된 LLM
데이터 특징	소규모 데이터셋에 적합	방대한 양의 데이터 필요
컴퓨팅 리소스	적음	많음
적용 사례	QA 등	의료 문서 분석과 같은 특정 도메인전문 작업 등



2 파인콘(Pynecone) 을 이용한 Vector DB 구축

vector DB를 사용하여 보아즈의 정보 Record를 추가(add), 삭제(delete), 업데이트(update) 가능

→ 최신의 정확한 정보를 확보



2 파인콘(Pynecone) 을 이용한 Vector DB 구축

● DB 구축 데이터

보아즈 회칙_14차 개정

2024-1 활동 기수 공지방

- 친바 미션
- MT 수요조사
- 출결 관련 공지
- 스터디 모집 공지
- 회계 결산 내역
- 컨퍼런스 공지 및 안내
- 청강 및 병행 공지

22기 분석 BASE 토크방

- 멘멘 스터디 신청
- 과제 제출 관련 주의사항
- github 관련 공지
- 미니프로젝트 1 공지
- 매주차 세션 장소, 시간, 발제 주제 공지

출결/과제 제출 점수 샘플 데이터 구축



● DB 구축 데이터

24-1 세션 출석_test.pdf

BOAZ_2024_활동기수_공지방.pdf

KakaoTalk_22기분석BASE.pdf

보아즈 회칙_14차 개정_231203.pdf

[illegible][illegible]

#프로젝트 상세 설명

BOAchat 구축 과정 (with Langchain)

① 보아즈의 새로운 정보가 담긴 문서 로드(Load)

Lang Chain

Document loaders

pyPDF loader

PDF

excel loader

text loader

documents

page_content: 문서의 내용

Metadata: 문서의 위치, 제목, 페이지 넘버 등

문서에는 파일에 포함된 모든 페이지가 저장

documents

Document(page_content='대학원(원생)생원빅데이터연합동아리보아즈회칙2. 모든회원은소수1년으로, 크게2개의Term으로구분되며이는대학학사일정기준으로방학1회와동일하다. 5. Term의공식일정의시작은0이며, Term의마지막공식일정은컨퍼런스신청기간이후첫세션을의미한다. 7. ADV활동인원은2Term방학세션때반드시멘토운영진회의를통해패널티를결정해통보한다. 8. 활동과정상의모든자료는특별한사유가없는다. 단, 커리큘럼상필요할경우온라인또는녹화영상으로대체한다. 제2조. 타부문의활동으로인하여본인이가입한부문의활동에피해를주면안된다. 2. 병행

metadata={'source': 'DB/보아즈 회칙 14차 개정 231203.pdf', 'page': 2}).

→ metadata에 따르면 '보아즈 회칙.pdf'의 '2페이지'에 해당하는 문서임을 알 수 있다
Page_content에서는 해당 문서의 정확한 내용을 확인할 수 있다

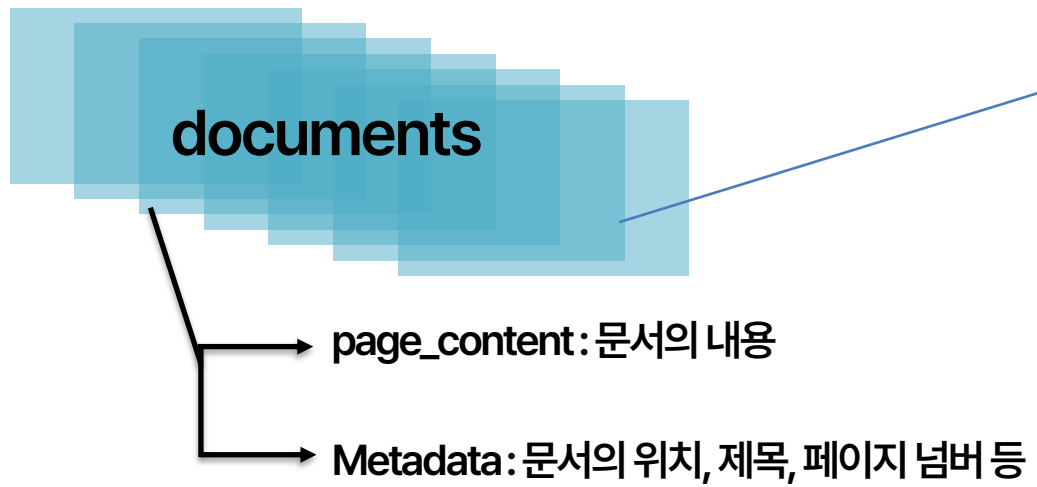
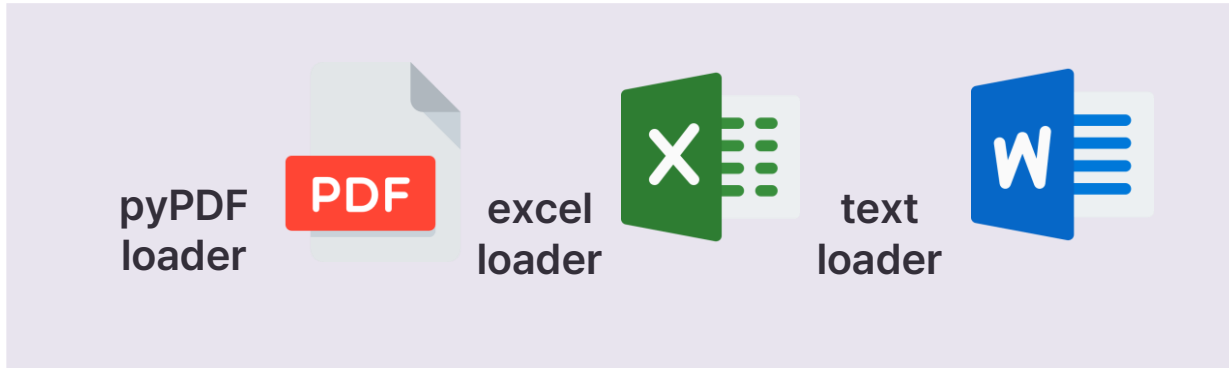
⋮

#프로젝트 상세 설명

BOAchat 구축 과정 (with Langchain)

RAG를 활용한 보아즈 Slack 봇 "BOAchat"

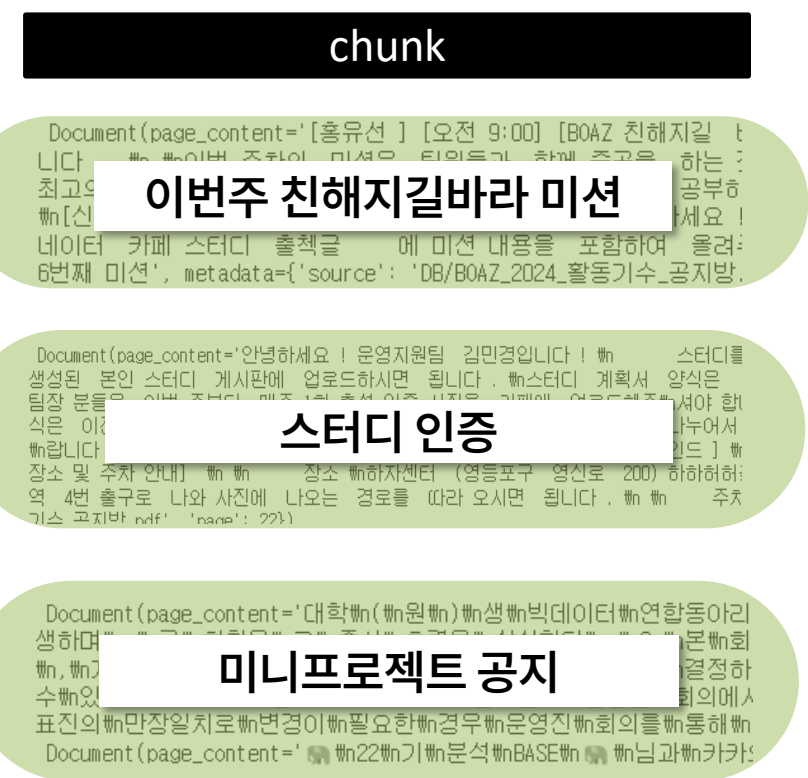
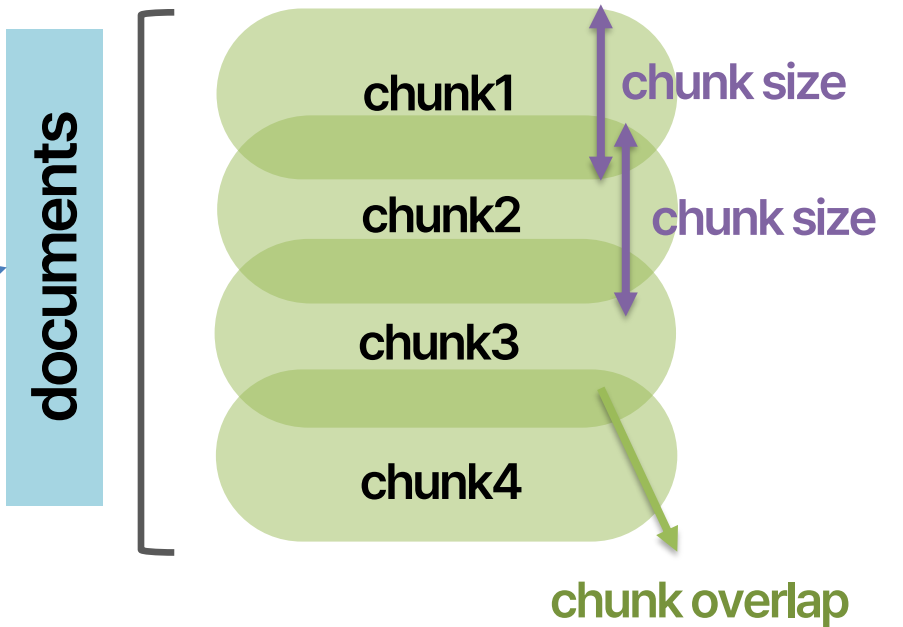
① 보아즈의 새로운 정보가 담긴 문서 로드(Load)



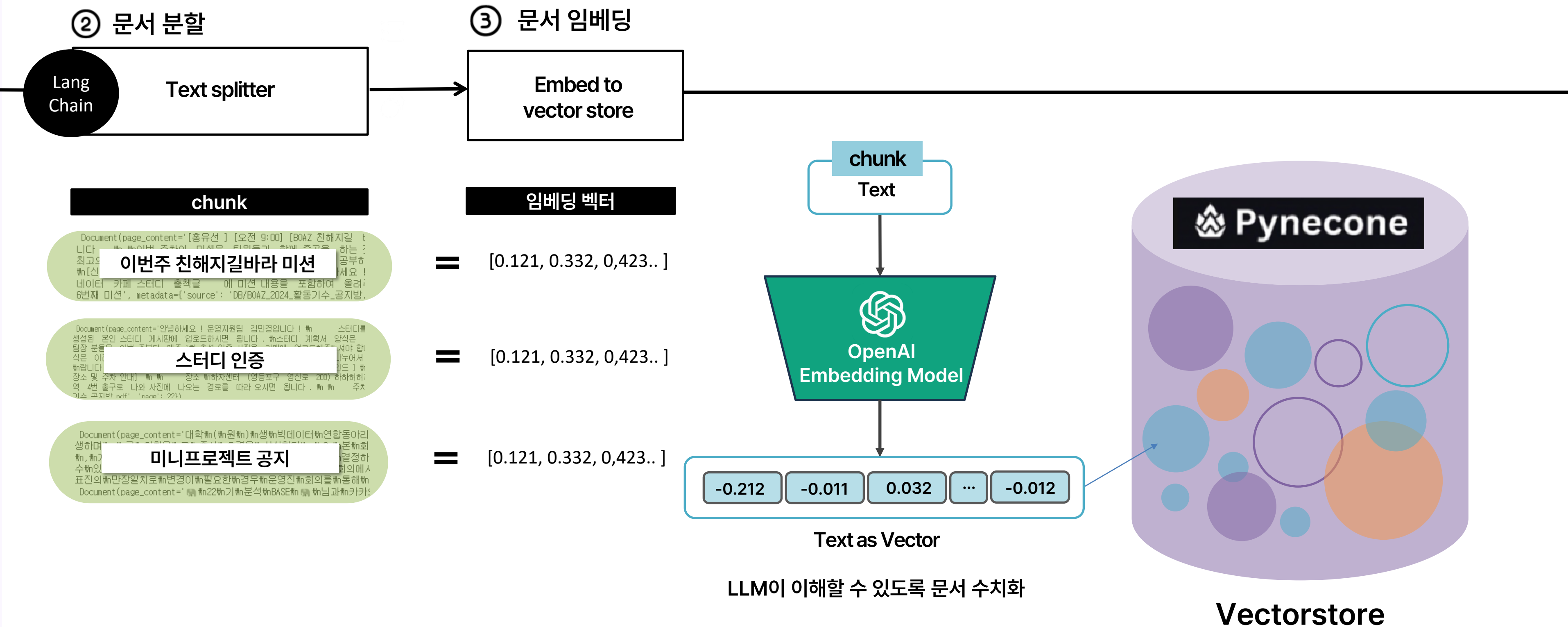
문서에는 파일에 포함된 모든 페이지가 저장

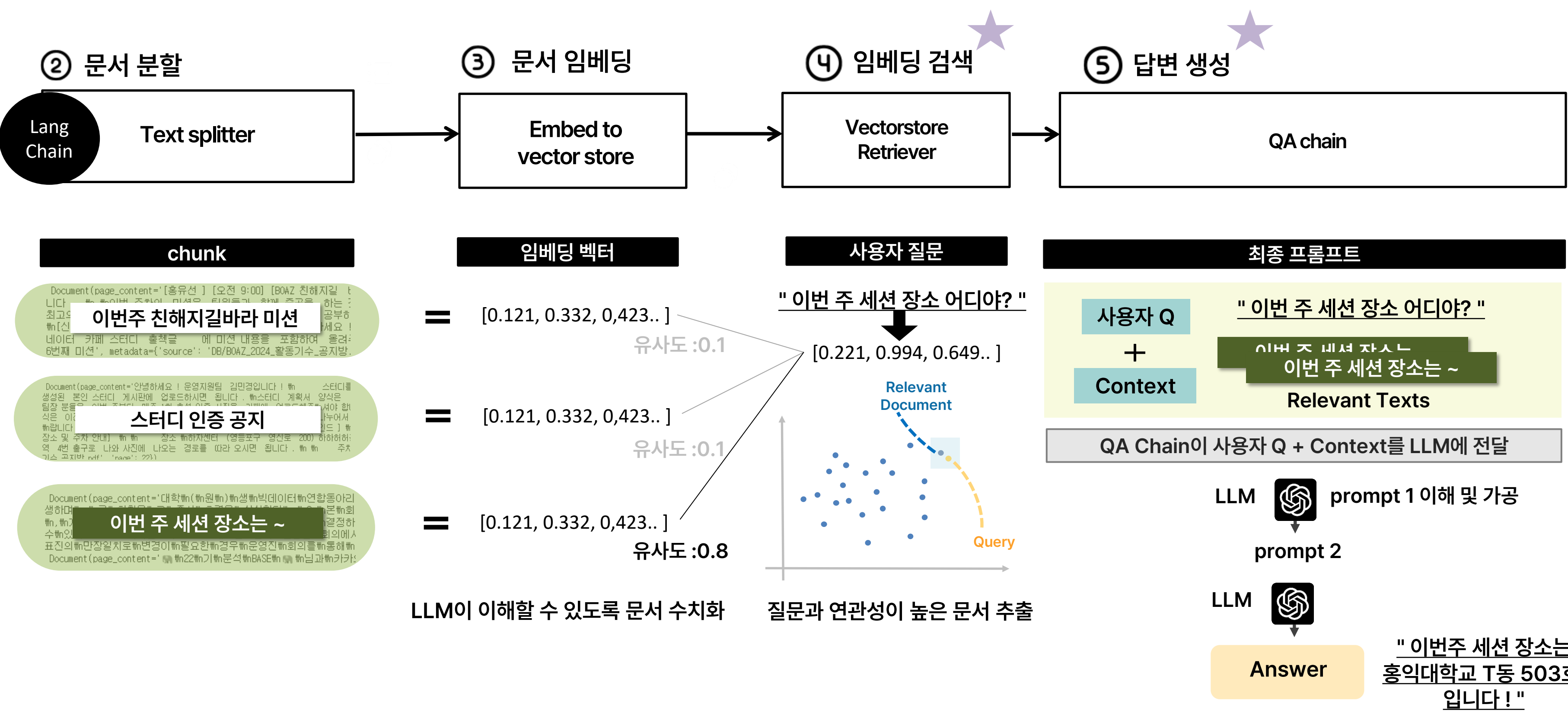
② 문서 분할

text splitter 로 문서documents를 다시 글 덩어리chunk 단위로 분할
: Documents(문서) → Chunk(글 덩어리)



```
text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=50)
texts = text_splitter.split_documents(documents)
```



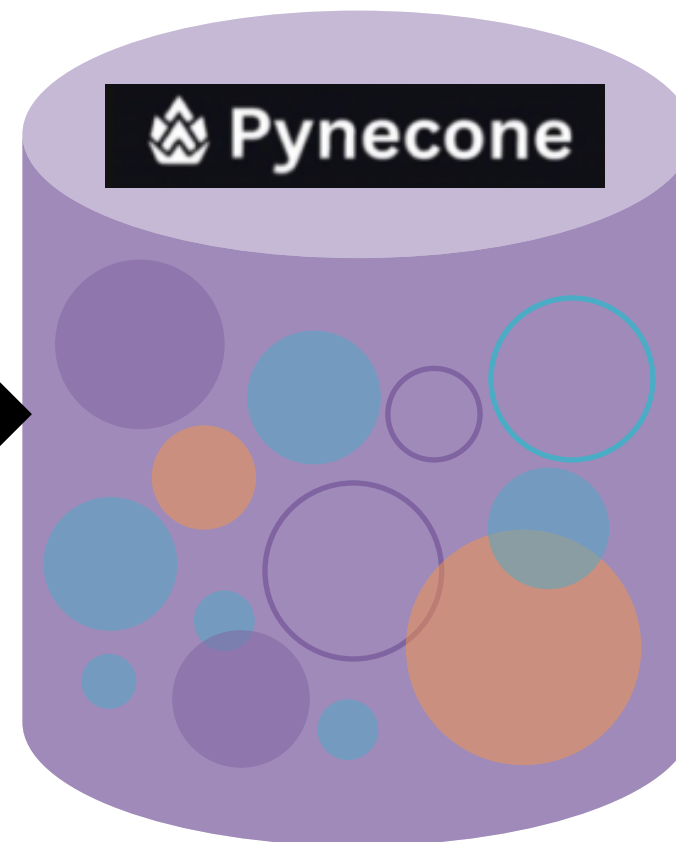


#프로젝트 상세 설명

RAG를 활용한 보아즈 Slack 봇 "BOAchat"

파인콘(Pynecone) Vector DB

Pynecone Vector DB upsert.ipynb



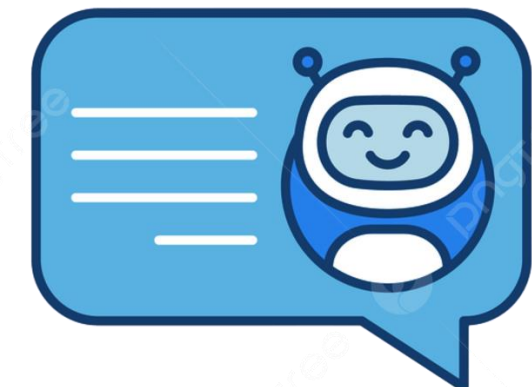
Vector DB에
새로운 Documents 임베딩해서
upsert(update + insert)



Chat Bot



Chatbot.ipynb



Pynecone에 이미 누적된
Vector Data 불러오기

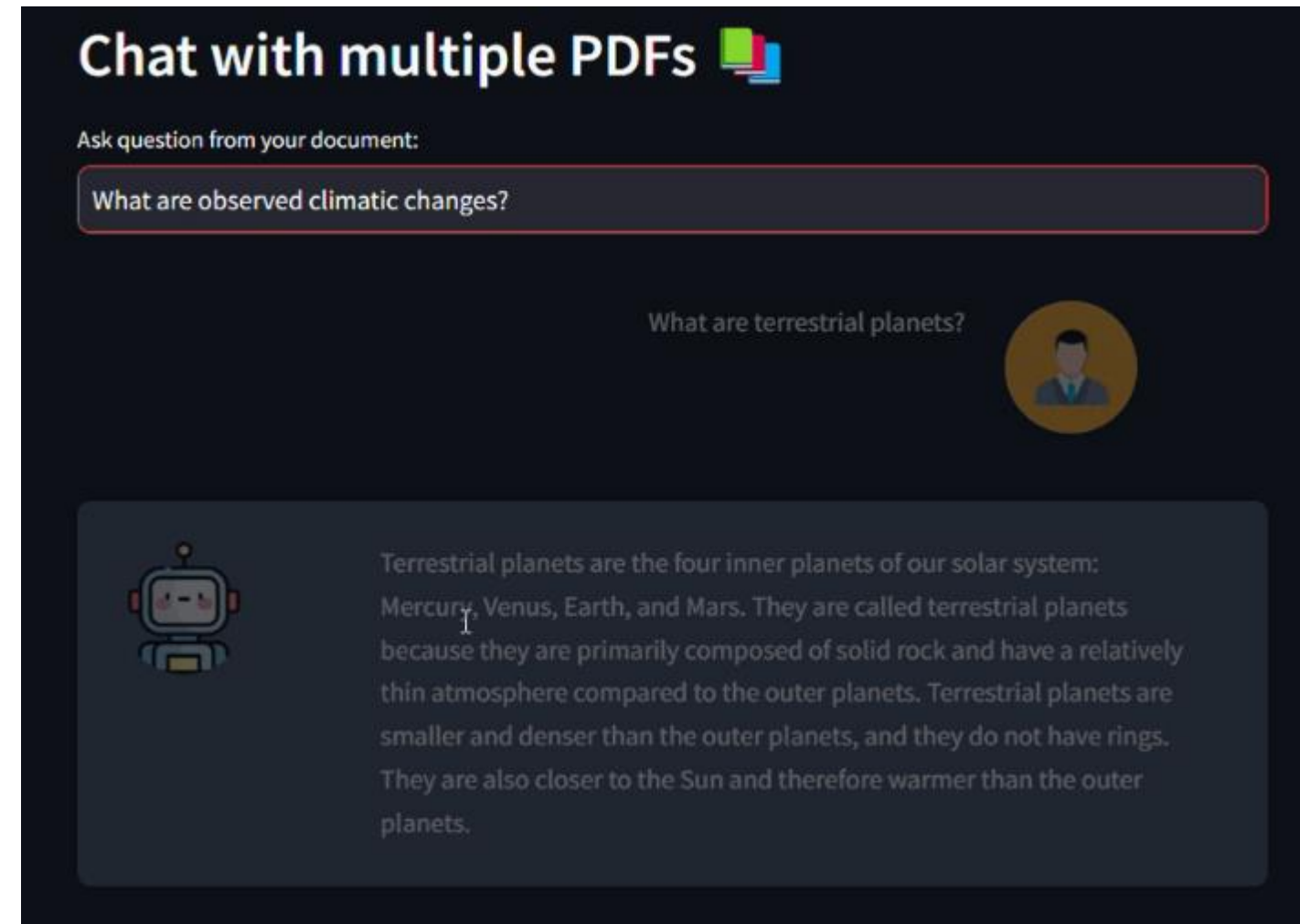
벡터 데이터 베이스를 관리하는 코드와 / Chat Bot실행 코드를 분리 ★ 유지보수에 용이 ★





#프로젝트 시연

RAG를 활용한 보아즈 Slack 봇 "BOAchat"

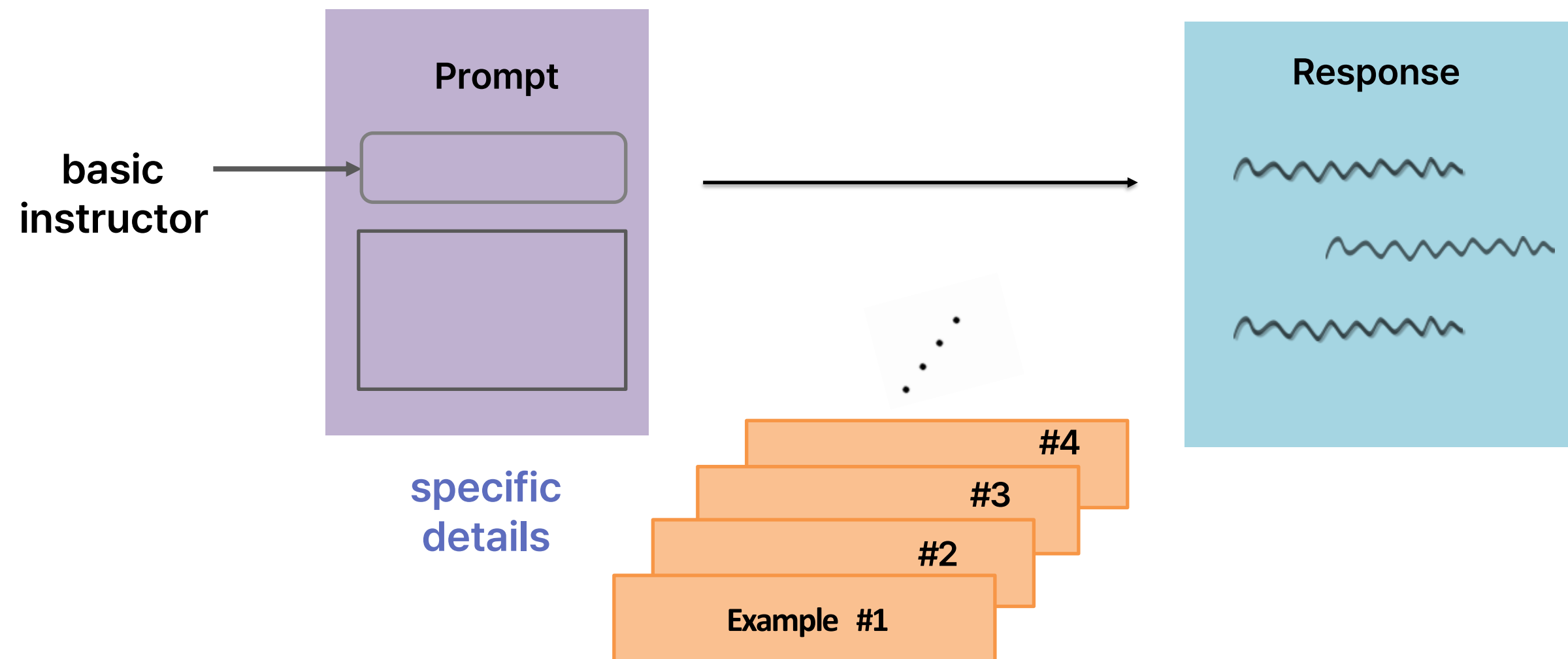


BOAchat TEST



✓ Trial & Error - Prompt Template

- Prompt Template은 관련성 있고 일관된 언어 기반 출력을 생성하는 데 도움





#프로젝트 상세 설명 # TRIAL & ERROR

RAG를 활용한 보아즈 Slack 봇 "BOAchat"



Trial & Error - 시각화



slack 봇으로 챗봇 구성 시
연결 오류 발생



동아리의 회칙은 언제부터 효력이 발생해?

회칙의 효력 발생 시점

회칙은 전 회원을 대상으로 커뮤니티에 공표된 날로부터 효력이 발생합니다.

**[1] 보아즈 회칙_14차 개정_231203.pdf(12) [2] 보아즈 회칙_14차 개정_231203.pdf(9)

"챗봇 결과를 보다 직관적으로 이해할 수 있도록
Gradio 활용해 시각화 진행"



- UI 를 프로토타입으로 제작하는 프레임워크
- 블록을 쌓듯이 UI 를 구현해 간단히 시각화 가능

"추후 초기 연결 오류 해결 후
slack 봇을 통한 챗봇 구현에 다시 도전할 계획"



기대효과



보아즈 회원들의 정보 접근성 향상

보아즈의 정보가 담긴 챗봇에
동아리회원들은 언제든지 챗봇에 접속해
회칙과 상벌점 규정을 검색할 수 있음

LLM 검색 기능 활용 가능

일반적인 텍스트 검색보다 LLM 챗봇은
질문에 대한 정확한 답변을 제공하는 데 특화
→ 원하는 정보를 더 빨리 찾을 수 있게 해줌

운영진들의 업무 효율성 향상

챗봇을 통해 질문&검색
→ 운영진에 대한 단순 문의량을 줄여
시간을 절약/업무 효율성 향상

업데이트 및 유지보수 용이성

회칙/공지사항 변경 시
파인콘의 Vector DB를 업데이트
→ 자동으로 챗봇의 Answer에 업데이트 반영

기능적 측면

기술적 측면





#향후 계획

What's Next?

1. 데이터

- 분석 뿐만아니라 시각화, 엔지니어링 카톡 채팅방으로 확장
- 운영진 톡방으로 확장

2. 챗봇 성능 향상

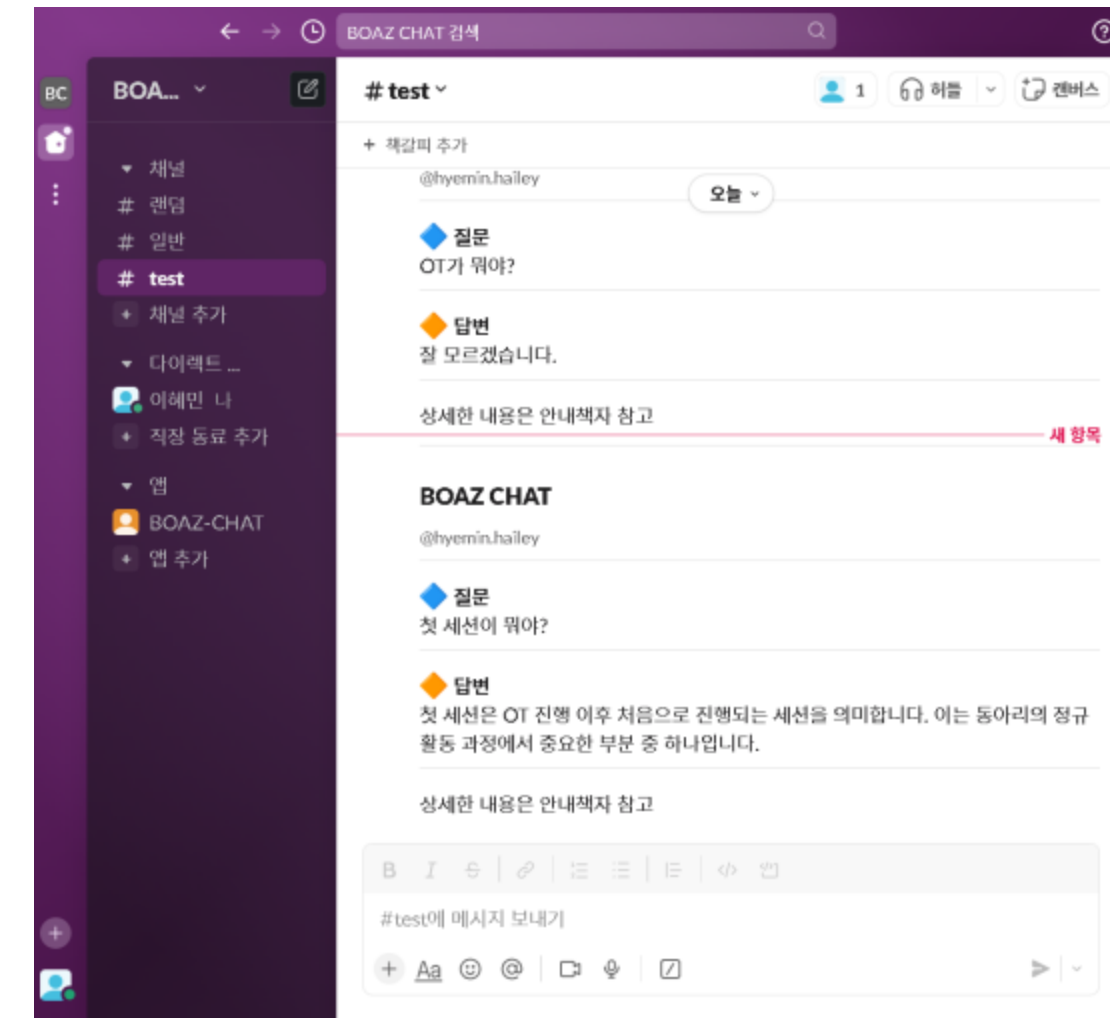
- Prompt Template 사용
- 임베딩 모델 TEST(hugginh face 한국어 임베딩 모델)

3. Slack 봇 구현

4. 기능 추가

- 일정 리마인드 기능
- 로그인 등 개인 식별 기능 추가 -> 개별 활동 점수 안내 기능

RAG를 활용한 보아즈 Slack 봇 "BOAchat"



BOAchat





#나RAG도 RAG이다

1

Q & A

RAG를 활용한 보아즈 Slack봇 "BOAchat"





#나RAG도 RAG이다

발표를 들어주셔서 감사합니다

1



17:00 pm

RAG를 활용한 보아즈 Slack봇 "BOAchat"