

YBIGTA 컨퍼런스

NoticeHub

Team: 행복하조

Contents

1. 주제 선정 동기
 - Why NoticeHub?
2. 구현 과정
 - Architecture, Data, Modeling, Frontend, Backend
3. 결과
 - 시연
4. 앞으로의 방향성
 - Limitation, Future works

팀원 및 역할

고정훈

RAG 모델링
vLLM 모델링

목종원

모델 서빙
서버 구축 및 배포
아키텍처 제작
ETL

오승옥

원천 데이터 구축

조윤영

Vector DB
AWS Lambda
ETL

정수현

프론트
프론트 엔지니어링

주제 선정 동기

Why NoticeHub?



교내 공지 사이트가
너무 많아요



장학금 관련 공지 뭐
있는지 궁금해요



저에게만 해당되는 공
지사항을 보고싶어요

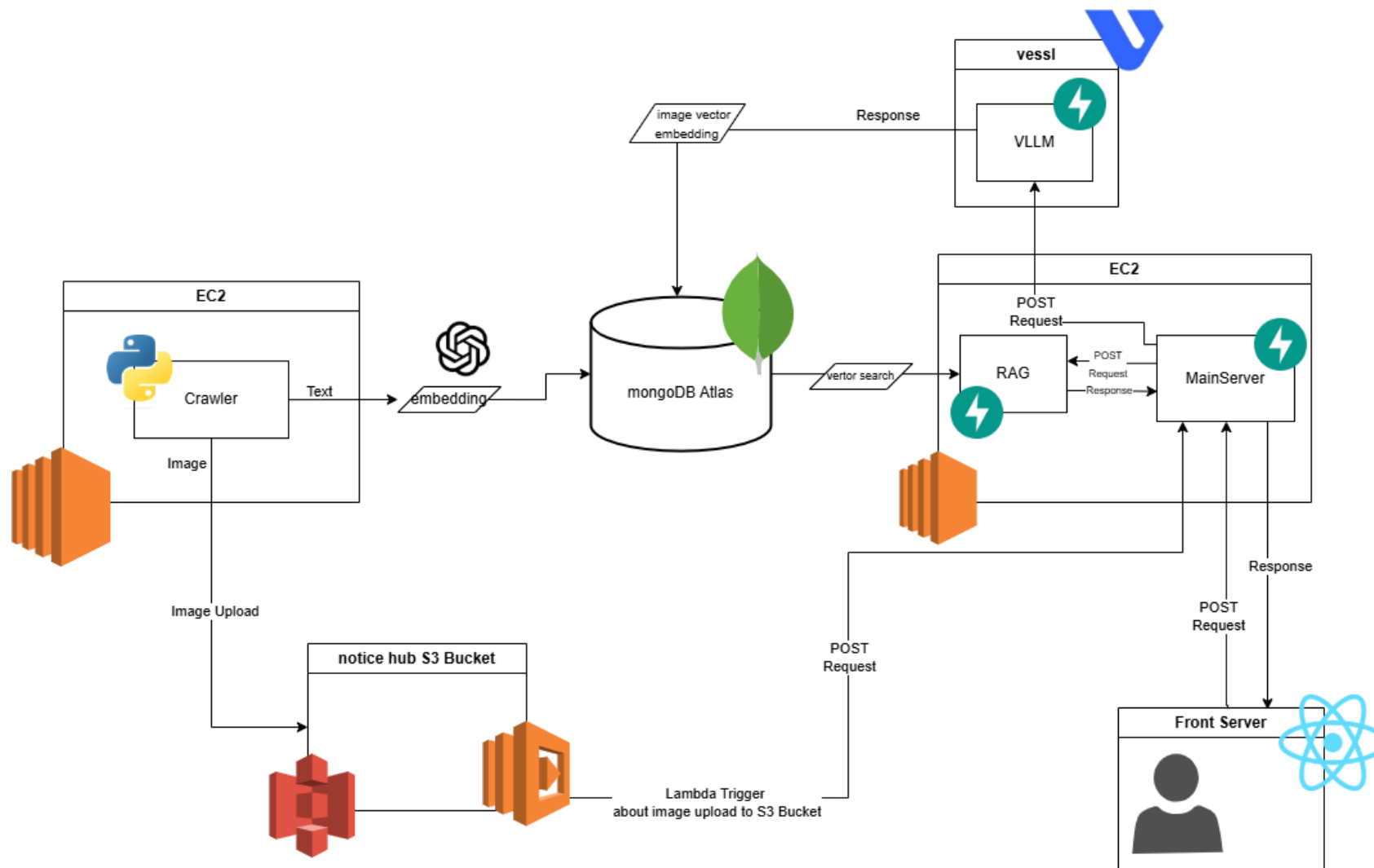


~~요새 AI 챗봇이 유행임 우리도 ㄱㄱ~~

- 1) 공지 사이트를 통합하고,
- 2) 맞춤형 공지사항 AI 챗봇을 만들어 보자...!

구현 과정

Architecture


























구현 과정

Data

각 대학 및 학과 웹사이트 공지사항

공과대학 

화공생명공학과	 	전기전자공학과	 
건축공학과	 	도시공학과	 
건설환경공학과	 	기계공학과	 
신소재공학과	 	산업공학과	 
시스템반도체공학과	 	디스플레이융합공학과	 
디지털융합엔지니어링학과		지능형데이터·최적화학과	
배터리공학과			

- 단과대학: title, date, link, content, image
- 학과: title, date, link, content
- 웹 크롤링으로 수작업으로 dataset 구축(BeautifulSoup)
- (Title, date)에 groupby를 해서 중복 방지
- 각 사이트 별로 1-5페이지 크롤링
- 총 데이터 개수: 약 2600여개 + @ (지속적으로 늘리는 중)

engineering.yonsei.ac.kr/engineering/board/notice.do?mode=view&articleNo=214470&article.offset=20&articleLimit=10

원문링크

Yonsei Calendar 20... 내 드라이브 - Goo... DeepL 번역: 세계에... Oracle Live SQL Adobe Acrobat Replit </> Baekjoon

게시판

연구활동

학생활동

동문소식

공지사항

공학연구원

공지사항

제목 제24회 윤동주 시문학상 공모 안내

작성일 2024.11.04

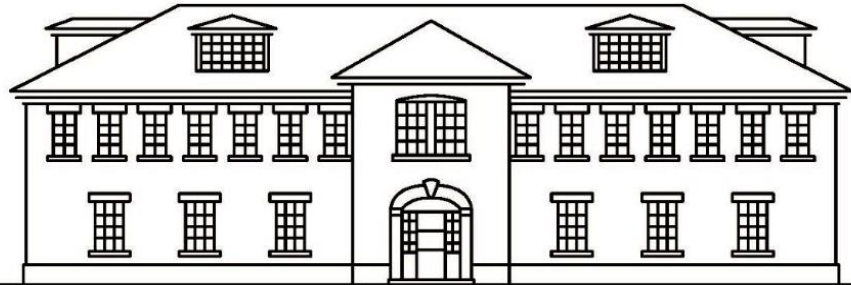
작성자 공과대학 홈페이지 관리자

제24회 윤동주 시문학상 공모

본문

연세대학교 윤동주기념사업회는 연세의 시인 윤동주의 시 정신을 기리고 그 뜻을 널리 전파하기 위해 전국의 대학(원)생을 대상으로 윤동주 시문학상을 개최합니다. 이 시대 청년들이 윤동주의 문학 정신과 삶을 만나며 자신을 되돌아보고 소중한 정신적 자산으로 가꾸는 계기로 삼길 바랍니다. 많은 분들의 적극적인 참여를 기다립니다.

이미지 (optional)



2024

제24회

윤동주 시문학상

구현 과정

Model_Text

Rag(Retrieval-Augmented Generation)?

정보를 탐색하고 이를 활용해 자연어 답변을 생성하는 AI 시스템

1. MongoDB 연결 및 임베딩 설정

- MongoDB Atlas를 벡터 검색 인덱스로 사용

2. 문서 검색 및 하이브리드 검색 설정

- 벡터 및 키워드 검색 조합으로 검색 정확도 향상

```
mongo_client = MongoClient(os.getenv("MONGO_PATH"))
db = mongo_client["notice-db"]
collection = db["final_notices"]
embeddings = OpenAIEmbeddings(model="text-embedding-3-small")
```

```
retriever = MongoDBAtlasHybridSearchRetriever(
    vectorstore=vector_store,
    search_index_name="default",
    top_k=10,
    fulltext_penalty=5.0,
    vector_penalty=20.0
)
```

notice-db final_notices default **search** [dynamic], "embedding"

notice-db final_notices vector_index **vectorSearch** "embedding"

구현 과정

Model_Text

3. 프롬프트 템플릿 구성

- 시스템 메시지를 통해 답변 포맷 및 조건 설정

```
PROMPT_TEMPLATE = ChatPromptTemplate.from_messages([
    ("system", """
        You are an AI assistant... Answer concisely...
        {question}
        {context}
    """),
    MessagesPlaceholder(variable_name="chat_history")
])
```

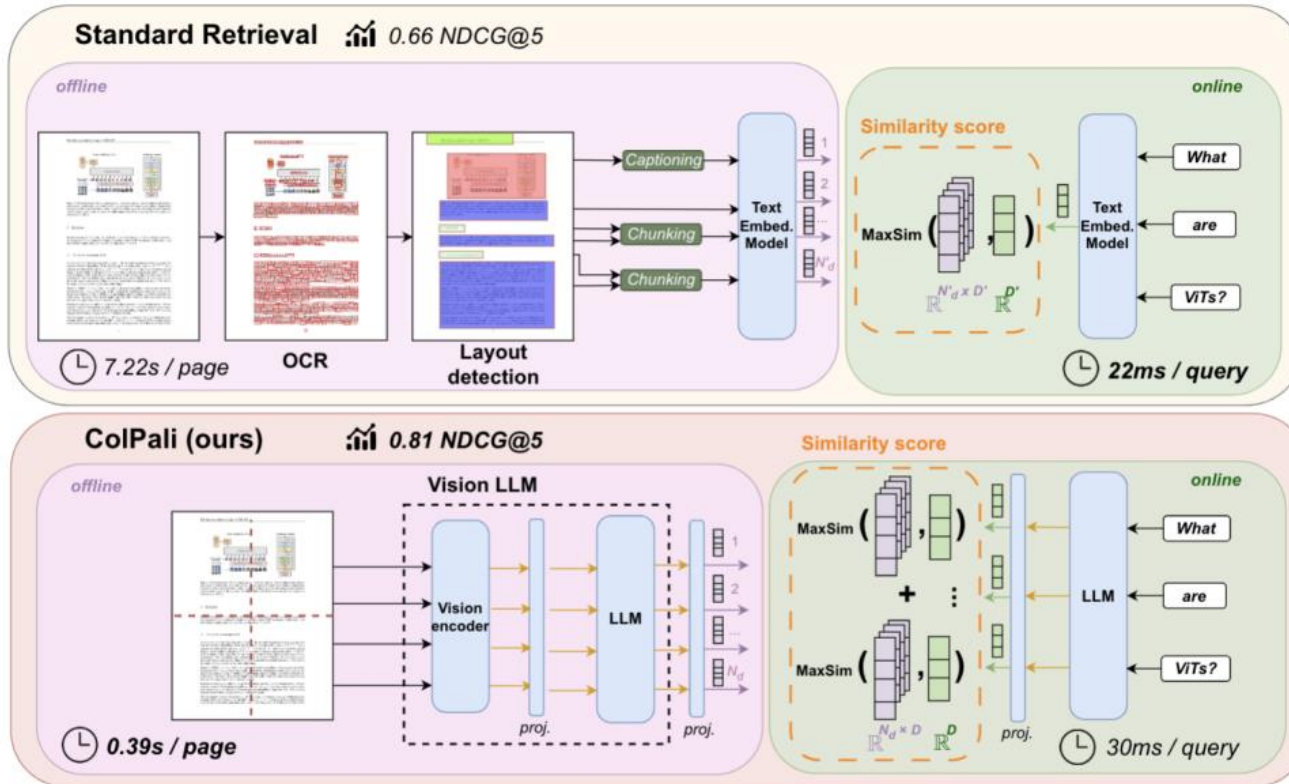
4. 세션 기록 및 답변 처리

- 대화 기록을 유지하여 문맥 기반 응답

```
store = {}
def get_session_history(session_id):
    if session_id not in store:
        store[session_id] = ChatMessageHistory()
    return store[session_id]
```


구현 과정

Model_Image



- 이미지 패치 임베딩을 언어 모델에 입력, 입력 데이터가 텍스트 입력과 동일한 잠재 공간으로 매핑
- 텍스트 토큰과 이미지 패치 간의 상호작용을 계산

```
1 RAG =  
    RAGMultiModalModel.from_pretrained(  
        "vidore/colpali-v1.2", verbose=1)  
2 RAG.model.index(  
    3     input_path=temp_dir_5,  
    4     store_collection_with_index=True,  
    5     overwrite=True)  
6 query = "공모전 정보 알려줘"  
7 results = RAG.search(query, k=5,  
    8     return_base64_results=True)
```

=== 검색 결과 ===

[원본 이미지 파일명: 2023년도 제3회 암젠한림생명공학상 추천 공고_1.png]

[원본 이미지 파일명: 2022 대학생 생물분류 연구논문 공모전_1.png]

[원본 이미지 파일명: 2024 Softeer CodeRun X현대자동차 계약학과_1.png]

[원본 이미지 파일명: 2022 대학생 생물분류 연구논문 공모전_1.png]

[원본 이미지 파일명: 2024 Softeer CodeRun X현대자동차 계약학과_1.png]

[원본 이미지 파일명: 2024 Softeer CodeRun X현대자동차 계약학과_1.png]

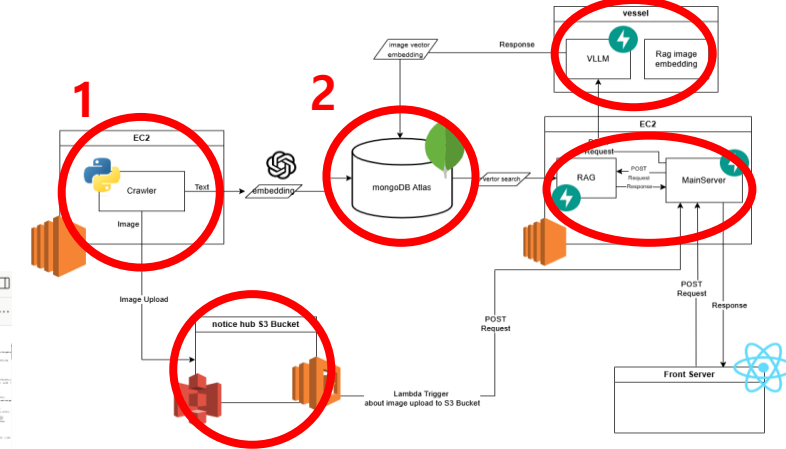
구현 과정

Backend

1. EC2 서버(Crawler)

- 텍스트 데이터 크롤링
- OpenAI를 통해 텍스트 임베딩
- MongoDB Atlas에 데이터 저장
- Amazon S3에 이미지 업로드

```
main.py
output.log
1 nohup: ignoring input
2 스케줄러가 시작되었습니다. Ctrl+C로 중지할 수 있습니다.
3 [문과대학] 공지사항 크롤링 시작
4 문과대학 1페이지 크롤링 중...
5 이미 존재하는 파일: images/2025-1 후배 사랑 함께 나눔 사업 신청 안내_1.jpg
6 이미 존재하는 파일: images/인문예술진흥사업단_ _한국문학의 저력_ 문정희 국립한국문학관장 명사 초빙 특강 신청 안내_1.png
7 이미 존재하는 파일: images/2025 인문인 동계 인턴십 참가자 모집 공고_1.png
8 문과대학 2페이지 크롤링 중...
9 이미 존재하는 파일: images/_연세대학교 창업지원단장_ 창업설명회 신청 안내_1.jpg
10 문과대학 3페이지 크롤링 중...
11 이미 존재하는 파일: images/인문_사회계열 외국인 대학원생을 위한 학술적 글쓰기 교육 프로그램 운영 안내_1.png
12 이미 존재하는 파일: images/제22회 연문인상 시상식_1.png
13 이미 존재하는 파일: images/문과대생을 위한 창업특강_타단과대생 참여 가능_1.jpg
14 이미 존재하는 파일: images/_대학원학신사업_ 연구프로세스에 따른 생성형 AI도구 활용 특강_1.png
15 이미 존재하는 파일: images/제2회 크리에이티브 공모전 안내_1.png
16 이미 존재하는 파일: images/세계 석학 초빙 강연회 안내_1.jpg
17 문과대학 4페이지 크롤링 중...
18 이미 존재하는 파일: images/2024 PwC컨설팅 신입컨설턴트 모집_1.jpg
19 이미 존재하는 파일: images/프레드릭 베르만 _오베라는 남자_ 작가 초빙 특강연 신청 안내_1.png
20 이미 존재하는 파일: images/이상업 자재회 읽기_ 이상업 선생 추모 심포지엄_1.jpg
21 문과대학 5페이지 크롤링 중...
22 이미 존재하는 파일: images/황소년 및 시인을 위한 제4회 연세 인문학 캠프 참가자 모집_1.png
23 이미 존재하는 파일: images/2024 인문_사회계열 외국인 대학원생을 위한 학술적 글쓰기 교육 프로그램 운영 안내_1.png
24 이미 존재하는 파일: images/2024-2 후배 사랑 함께 나눔 사업 신청 안내_1.jpg
```



2. MongoDB Atlas

- 텍스트 및 이미지 임베딩 벡터 저장
 - date, department, title, context
 - images, link, embedding, image_embedding
- 벡터 검색 기능 제공

notice-db.final_notices

STORAGE SIZE: 725.01MB LOGICAL DATA SIZE: 539.69MB TOTAL DOCUMENTS: 2060 INDEXES TOTAL SIZE: 104KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes

Generate queries from natural language in Compass

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

Options

```
{
  "_id": ObjectId("67504e33480794a2e954f08e"),
  "date": "2024.11.28",
  "department": "문과대학",
  "title": "인문예술진흥사업단, 『한국문학의 저력』 문정희 국립한국문학관장 명사 초빙 특강 신청 안내",
  "context": "『한국문학의 저력』 문정희 국립한국문학관장 명사 초빙 특강 신청 안내참여대상 : 연세대 구성원 및 시인 누구나일정: 2024년...",
  "embedding": Array (1536),
  "images": "https://noticehub.s3.amazonaws.com/images/인문예술진흥사업단_ _한국문학의 저력_ 문정희 국립...",
  "link": "https://libart.yonsei.ac.kr/libart/board/notice.do?mode=view&articleNo...",
  "image_embedding": Array (1030),
  "image_embedding_pca": Array (326)
}
```

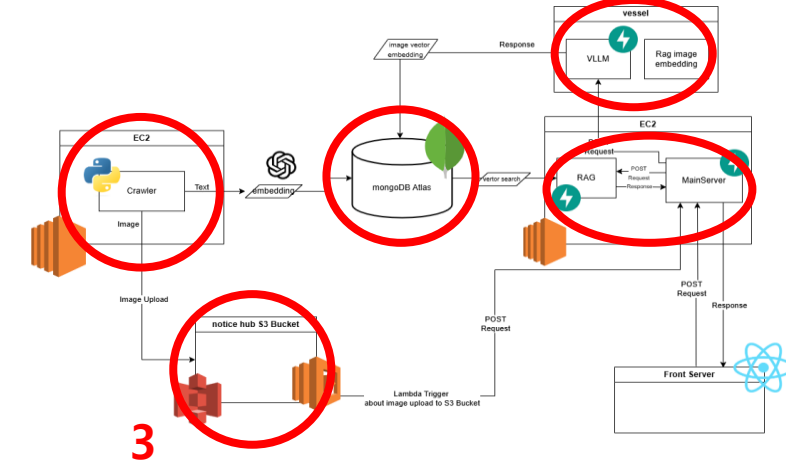
구현 과정

Backend

3. S3 버킷 & Lambda

- S3에 이미지 적재
 - Lambda를 통해 이벤트 트리거 처리
 - S3에 이미지 업로드
- > 업로드된 URL을 메인서버로 POST

s3-trigger



images/

객체 속성

S3 URI 복사

객체 (440) 정보

객체는 Amazon S3에 저장되어 있는 기본 엔티티입니다. [Amazon S3 인벤토리](#)를 사용하여 버킷에 있는 모든 객체의 목록을 얻을 수 있습니다. 다른 사용자가 객체에 액세스할 수 있게 하려면 명시적으로 권한을 부여해야 합니다. [자세히 알아보기](#)

Q. 컴퓨터로 객체 찾기

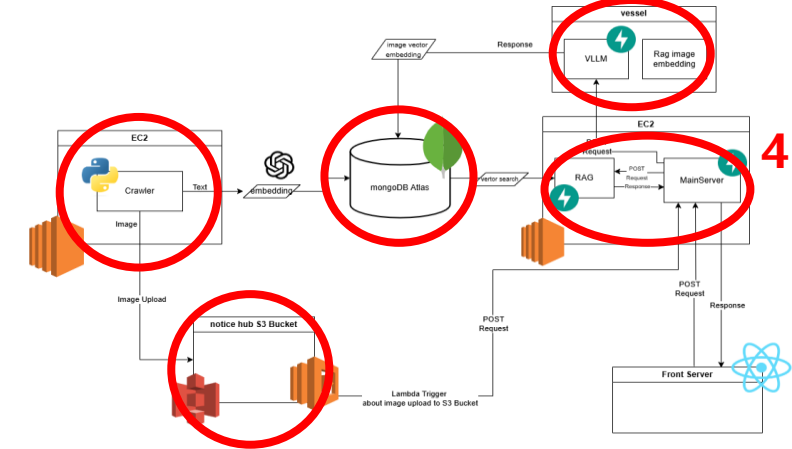
<input type="checkbox"/>	이름	유형	마지막 수정	크기	스토리지 클래스
<input type="checkbox"/>	2023-2학기_표준형 현장실습 교과목_안내_1.png	png	2024. 12. 24. am 3:27:24 AM KST	299.4KB	Standard
<input type="checkbox"/>	세미나 안내_창조산학 컨퍼런스 1_나는 창조의 하나임을 믿습니다_1.png	png	2024. 12. 24. am 3:27:12 AM KST	1.0MB	Standard
<input type="checkbox"/>	고등학생_2025 겨울 연영프럼 신청 안내_1.jpg	jpg	2024. 12. 24. am 3:25:22 AM KST	128.9KB	Standard
<input type="checkbox"/>	국제캠퍼스 건물 전체_기숙사 포함 전기시설물의 법정 정기검사 및 정비 작업으로 인한 정전통보_2024_8_17_토_14_00_18_일_02_00_12시간_1.png	png	2024. 12. 23. am 12:32:08 AM KST	55.5KB	Standard
<input type="checkbox"/>	국제캠퍼스 건물 전체_기숙사 포함 전기시설물의 법정 정기검사 및 정비 작업으로 인한 정전통보_2024_8_17_토_14_00_18_일_02_00_12시간_2.png	png	2024. 12. 23. am 12:32:08 AM KST	18.4KB	Standard
<input type="checkbox"/>	2024-2 학부대학 신입생 학사지도 안내_24_08_19_업데이트_1.png	png	2024. 12. 23. am 12:32:07 AM KST	213.5KB	Standard
<input type="checkbox"/>	교육부_2024 대한민국 인재상 선발 안내_1.jpg	jpg	2024. 12. 23. am 12:32:07 AM KST	414.7KB	Standard
<input type="checkbox"/>	2024년 12월 24일 12:32:06 AM		2024. 12. 24. am 12:32:06 AM		

구현 과정

Backend

4. Main Server(Rag, 데이터 처리)

- Rag 모델 요청 및 응답 처리
- VLLM 및 Rag 이미지 임베딩 처리와 통신

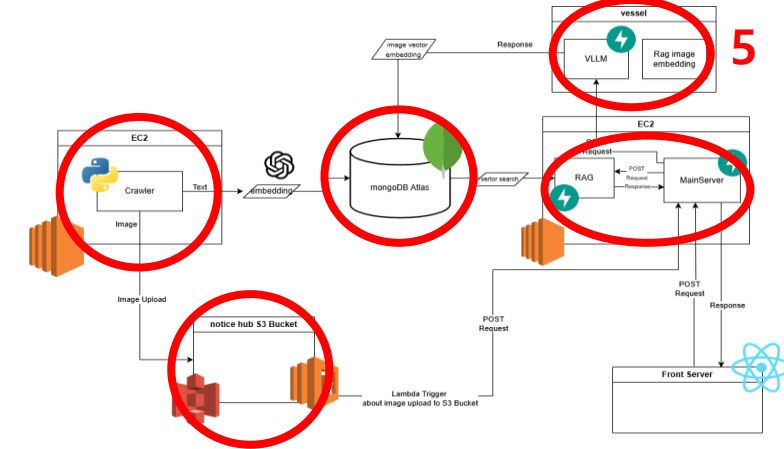


구현 과정

Backend

5. Vessel 서버

- VLLM 모델 실행
- 이미지 벡터 임베딩 생성



결과

결과 시연

기다려주세요...!

앞으로의 방향성

Limitation & Future works

1. 교내 기관 웹사이트 공지사항 데이터 추가

이름	홈페이지
연세대 공지사항	https://www.yonsei.ac.kr/sc/support/notice.jsp
연세대 외부기관공고	https://www.yonsei.ac.kr/sc/support/etc_notice.jsp
연세대 장학금 안내	https://www.yonsei.ac.kr/sc/support/scholarship.jsp
신촌생활관	https://dorm.yonsei.ac.kr/
송도학사 (Songdo Dormitory)	https://yicdorm.yonsei.ac.kr
스포츠센터	https://www.facebook.com/ywspots/
연세대학교 교무처 교무실	https://aad.yonsei.ac.kr/aad/index.do
도서관	https://library.yonsei.ac.kr/
국제처 (Office of International Affairs)	http://oia.yonsei.ac.kr
정보통신처	http://yis.yonsei.ac.kr
학술문화처	https://lamp.yonsei.ac.kr
커리어연세	https://career.yonsei.ac.kr/

2. 이미지 데이터 처리의 한계

Index Name

vector_index_

```
1  {
2    "fields": [
3      {
4        "type": "vector",
5        "path": "<field-name>",
6        "numDimensions": 0,
7        "similarity": "<similarity-function>"
8      }
9    ]
10 }
```

- MongoDB Vector Index 설정 시 차원의 한계
(128x1030 vs 4096)

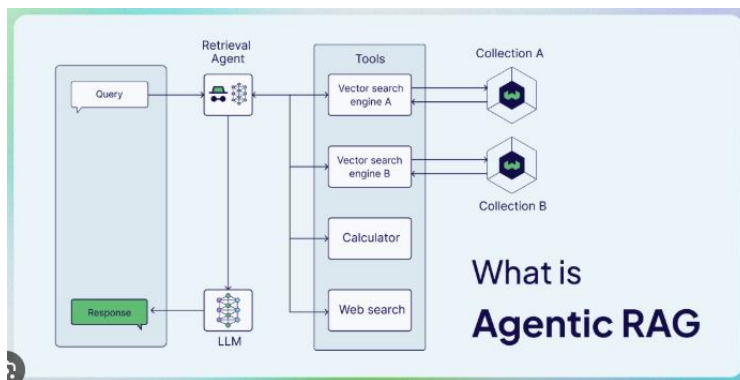
앞으로의 방향성

Limitation & Future works

3. 평가 지표 / Agent

	question	contexts	ground_truth	evolution_type	metadata	episode_done
0	What are the key considerations regarding AI f...	[3개국은 기판모델 전반에 대한 규제가 기술 중립적이고 위험 기반의 AI 규제 ...	Key considerations regarding AI foundation mod...	simple	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
1	What are the key features and improvements of ...	[1. 정책/법제 2. 기업/산업 3. 기술/연구 4. 인력/교육]인공지능의 미래와 클라우...	The key features and improvements of the LLM '...	simple	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
2	What is the purpose of the AAAI Conference?	[인공지능 분야의 주요 행사인 AAAI Conference의 목적과 중요성에 대해 설명하십시오.]	The purpose of the AAAI Conference is to advan...	simple	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
3	What role does content moderation play in the ...	[콘텐츠 관리, 특히 인공지능 생성 콘텐츠에 대한 규제 요청에도 대응 계획]인...	Content moderation plays a crucial role in AI ...	simple	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
4	What FTC concerns exist on AI comp. & consumer...	[1. 정책/법제 2. 기업/산업 3. 기술/연구 4. 인력/교육]미국 연방거래위원회...	The FTC has raised concerns regarding AI-relat...	reasoning	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
5	What key algorithms, architectures, and functi...	[구글 딥마인드, 오픈AI 모델의 기능과 동작에 대한 분류 체계 발표]인...	The context does not provide specific details ...	reasoning	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
6	What strategies ensure safe and ethical AI adv...	[문제를 방지하는 조치를 확대]인 - 형사사법 시스템에서 AI 사용 모범 사례를 개발하...	The context discusses various strategies for e...	multi_context	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
7	What key metrics gauge AI app effectiveness an...	[1. 정책/법제 2. 기업/산업 3. 기술/연구 4. 인력/교육]인공지능의 효과 측정...	The key metrics that gauge AI app effectiveness...	multi_context	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
8	What does the complementarity principle mean f...	[영국 옥스포드 인터넷 연구소 AI 기술자의 임금에 평균 21% 높아]인...	The complementarity principle in AI suggests t...	conditional	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True
9	How might new copyright laws affect FTC's AI r...	[1. 정책/법제 2. 기업/산업 3. 기술/연구 4. 인력/교육]미국 연방거래위...	The context does not provide specific informat...	conditional	['source': 'data/SPRI_AI_Brief_2023년12월호_F.pd...	True

합성 테스트 데이터 생성 (RAGAS 파이프라인)



Agentic RAG

- 장학금 검색기
- 학칙 검색기

4. 서비스 기능 추가

- 피드백을 통한 지속적인 모델 업데이트
- 추가 서비스 계획
 - 원하는 학과 및 부서의 공지사항만 선택
 - 중요한 공지사항 따로 스크랩
 - 이미지 제공

5. AI 챗봇 판매 / IT 아웃소싱

- 공지사항 AI 챗봇 솔루션 판매
- 타 대학 및 기업으로의 확장 가능성 검토
- ~~경제적 자유(?)~~

감사합니다.
