

SKKU SCHOLARSHIP FINAL



TEAM 스끌라, SKKULAR

강병준 | 김규진 | 박진아 | 장이준 | 주재현

CONTENTS

1. Proposal

2. AI

3. Frontend
&
Backend

TEAM. 스클라
SKKULAR



1. Proposal

Problem Recognition

NEWSIS 구독

"대학 국가장학금 받아가세요"…2차 접수 시작

입력 2013.03.01. 오전 5:00 수정 2013.03.01. 오전 5:01

장학재단 관계자는 "지난해 소득분위 기준이 충족된 대학생이 200만명 정도였는데 성적기준에 미달되거나 국가장학금 정보를 몰라 신청하지 않은 학생들이 많다"며 "등록금을 이미 냈어도 장학생으로 선정되면 개별 계좌로 입금이 되니 많은 신청 바란다"고 말했다.

건국대, 가계곤란 학생 1270명에 19억 장학금

김지수 건국대 장학사정관은 "성적 기준을 미달하거나 신청 기한을 놓쳐 국가장학금과 교내장학금 혜택을 충분히 받지 못한 가계 곤란 재학생을 위해 등록금 잔액 범위 안에서 지원해 부담을 줄여주기로 했다"고 말했다.

[2021 국감] 한국장학재단 광고비 42억 쓰는데...10명 중 1명만 학자금, 3명만 장학금 수혜

[교육플러스=지성배 기자] 한국장학재단이 광고선전비를 대폭 늘렸음에도 불구하고 오히려 학자금 및 국가장학금 신청 인원이 줄어든 것으로 나타나 개선이 요구됐다.

정경희 국회 교육위원회 소속 국민의힘 의원이 5일 공개한 자료에 따르면 학자금 대출 신청 인원은 2016년 51만667명에서 2020년 46만8500명으로 4만2167명이 줄었다. 수혜자 역시 42만4976명에서 40만7793명으로 줄었으며 수혜율은 12.1%에서 12.5%로 소폭 상승했다.

'대학생 돋는다더니'...공무원이 챙긴 코로나 장학금

[김 모씨/대학교 2학년]

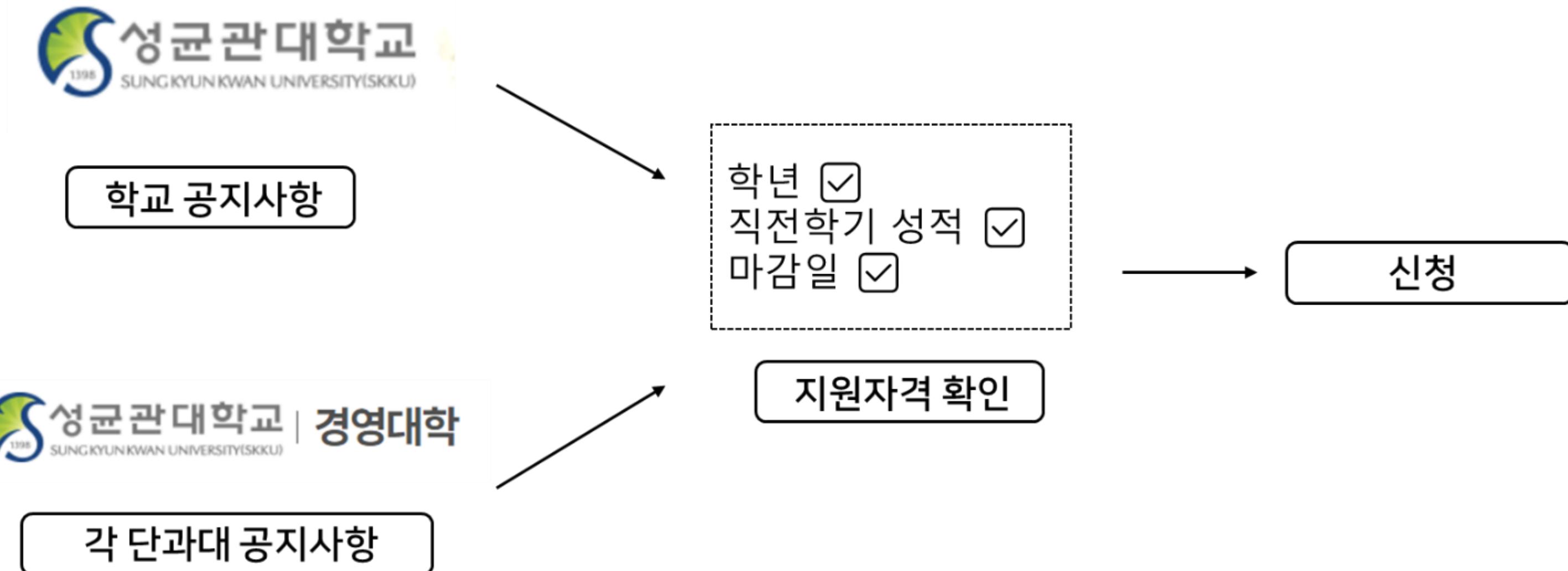
"솔직히 받을 수 있는 걸 못 받은 거니까 많이 아쉽죠."

보은군청이 지난해 11월 말 군청 홈페이지에 올려놓은 장학생 모집 공고문입니다.

12월 초부터 보름간 신청하라고 나와있습니다.

그런데 다른 공고들에 섞여 잘 눈에 띄지 않습니다.

Problem Recognition



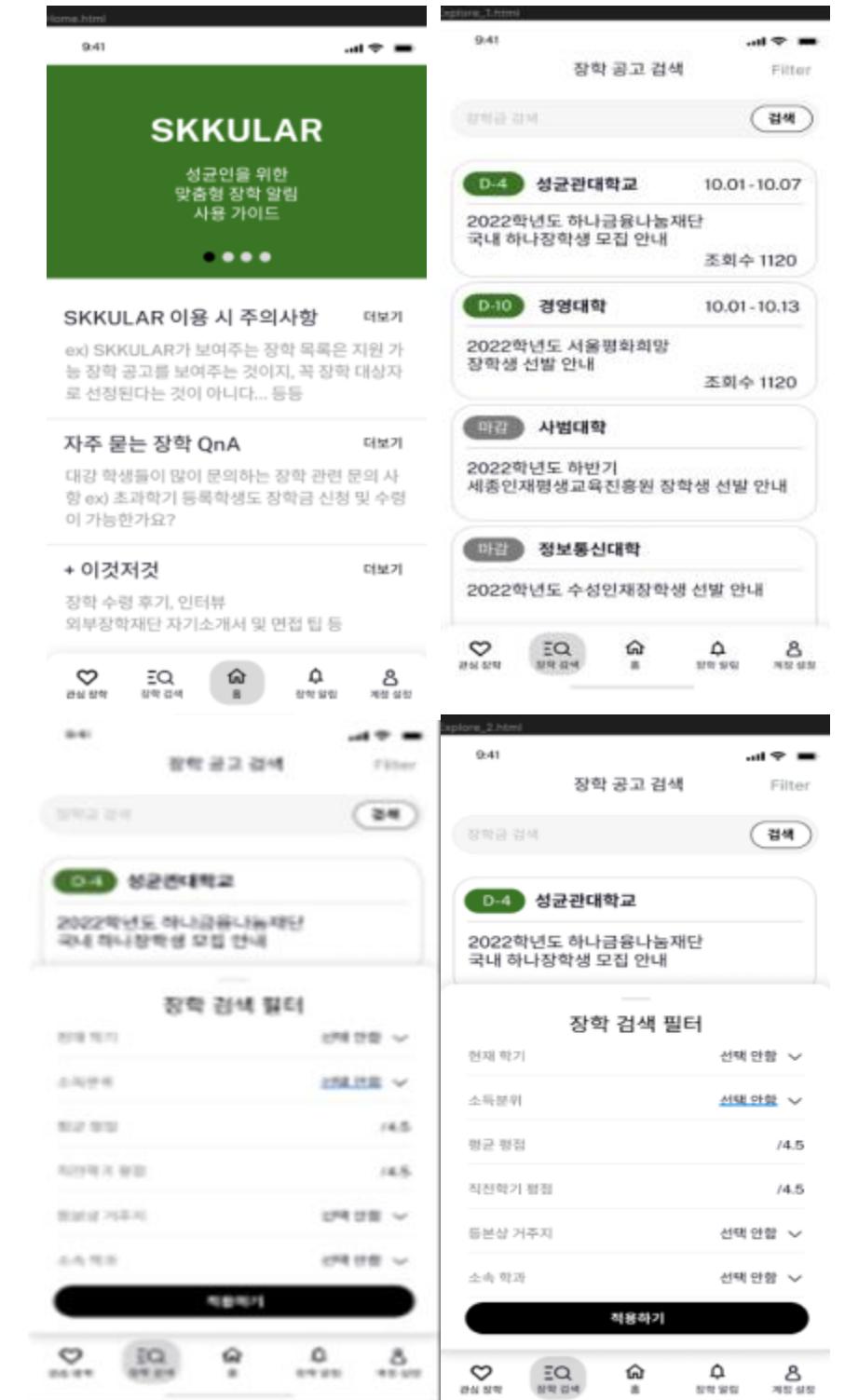
SKKULAR

1. 장학금 한눈에 보기

2. 장학금 검색

3. 맞춤 장학금 필터링

4. 맞춤 장학금 알림



SKKULAR

1. 장학금 한눈에 보기

2. 장학금 검색

3. 맞춤 장학금 필터링

4. 맞춤 장학금 알림

home.html

explore_1.html

9:41

9:41

장학 공고 검색

Filter

장학금 검색

검색

D-4 성균관대학교 10.01-10.07

2022학년도 하나금융나눔재단
국내 하나장학생 모집 안내

조회수 1120

D-10 경영대학 10.01-10.13

2022학년도 서울평화희망
장학생 선발 안내

조회수 1120

마감 사범대학

2022학년도 하반기
세종인재평생교육진흥원 장학생 선발 안내

마감 정보통신대학

2022학년도 수성인재장학생 선발 안내

더보기

더보기

더보기

더보기

관심 장학

장학 검색

장학 알림

계정 설정

관심 장학

장학 검색

장학 알림

계정 설정

SKKULAR

1. 장학금 한눈에 보기

2. 장학금 검색

3. 맞춤 장학금 필터링

4. 맞춤 장학금 알림



SKKULAR

1. 장학금 한눈에 보기

2. 장학금 검색

3. 맞춤 장학금 필터링

4. 맞춤 장학금 알림



SKKULAR

1. 장학금 한눈에 보기

2. 장학금 검색

3. 맞춤 장학금 필터링

4. 맞춤 장학금 알림



SKKULAR 활용방안



성균관대학교 학생들에게 장학금 정보 제공



추천 장학금 시뮬레이션



신청 가능 장학금 목록

[더보기 >](#)

조회된 데이터가 없습니다.

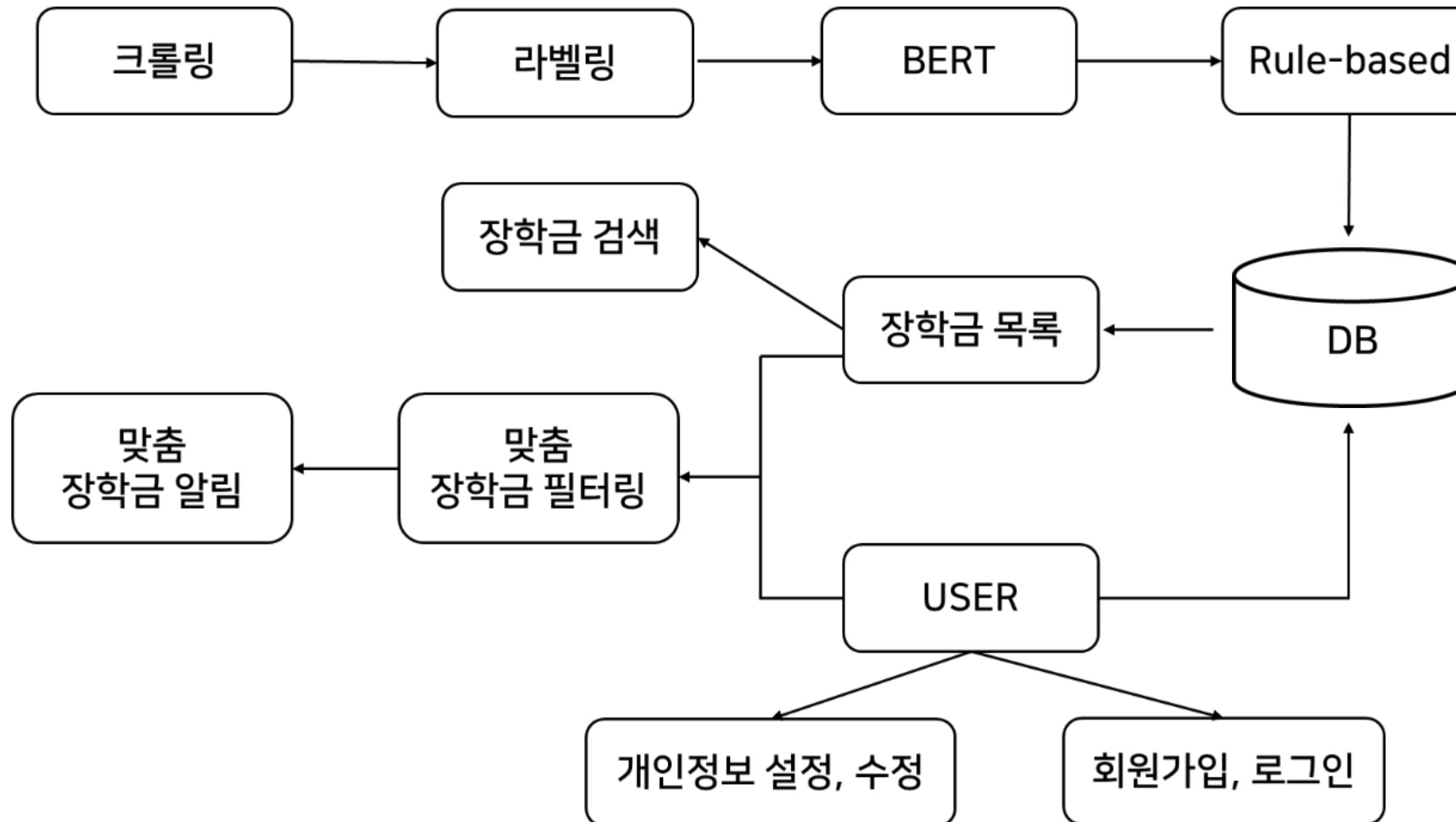
Challenge Square의 확장 프로그램으로 사용

기대효과

1. 학생들이 장학금 정보를 한눈에 보고, 자신이 받을 수 있는 장학금을 볼 수 있음
2. 장학금 신청 장려 - 장학금 신청 수가 많아짐
3. 장학금 지원 기관들에게 장학금이 학생들에게 도움이 됨을 증명
- 장학금 지원 기관들이 장학금을 더 많이 도입



Development Flow Chart





2. AI

BERT Q&A 모델

〈콩쥐팥쥐〉 원문

이에 콩쥐는 사례하고 그 물독을 기울여
두꺼비가 엉금엉금 기어 그 밑으로 들어가게 해 주었다.
그리고 독을 바로잡아 놓은 다음 물을 길어다 부으니,

...

두꺼비

질문

물독의 구멍을 막아준 동물은?

BERT Q&A

Modeling Goal

직전학기 요구 성적은 얼마인가요?

The screenshot shows the homepage of Sungkyunkwan University. The main navigation bar includes links for 대학정보 (University Information), 입학 (Admission), 교육 (Education), 연구 (Research), 산학/창업 (Industry-University Cooperation), 국제화 (Internationalization), and 대학생활동 (Student Activities). A search bar is located at the top right. Below the navigation, there is a banner for the 2023 academic year regarding student loan applications. The sidebar on the left lists various university services such as 학생활동 (Student Activities), IT 서비스 (IT Services), and 성균정학 (Seonggyun Jeonghak).



장학금 게시물 크롤링

게시물에 대한 질문을 던져
BERT Q&A 모델로
핵심정보 추출

The screenshot shows a mobile application interface. At the top, there is a green button labeled "내 지원 자격 확인하기" (Check Eligibility). Below it, there are two columns of settings:

| 공고 | 내 설정 |
|----------------|---|
| 7분위 이하 | 6분위 ✓ |
| 직전학기 성적 4.0 이상 | 직전학기 성적 4.2 ✓ |
| 등록 학기 수 6학기 이하 | 현재 7학기 ✗ |

At the bottom, a message states: "아쉽게도 학우님은 해당 장학금에 지원하실 수 없습니다.." (Unfortunately, your friend cannot apply for this scholarship..)

At the very bottom, there is a note: "자세한 지원 자격 및 공고 내용은 학교 홈페이지를 참고해주세요!" (For detailed application qualifications and announcements, please refer to the university's website!).

추출된 핵심정보를 사용자의 정보와
매칭시켜 조건에 맞는 장학금 게시물 추천

Modeling Flow Chart

{

1. 지원자격 □ 국내 4년제 이상 대학의 최종 학년 재학생(휴학생 포함) □ 재학 중 성취한 탁월한 공적으로 타의 모범이 되고, 대학의 명예 선양과 발전에 크게 기여한 개인...



{

- 장학금 명: A 장학재단 희망장학금,
학위: [국내 4년제, 4학년, 재학생],
소득분위: ...
기타 조건: 교환학생

}



BERT Q&A 모델



RULE-BASED

Crawling

학교 전체 장학금 홈페이지와
단과대 장학금 게시물
1500여개 크롤링

Parsing & Labeling

데이터 전처리 진행 후
BERT 모델 학습을 위하여
학습데이터 2048쌍 생성

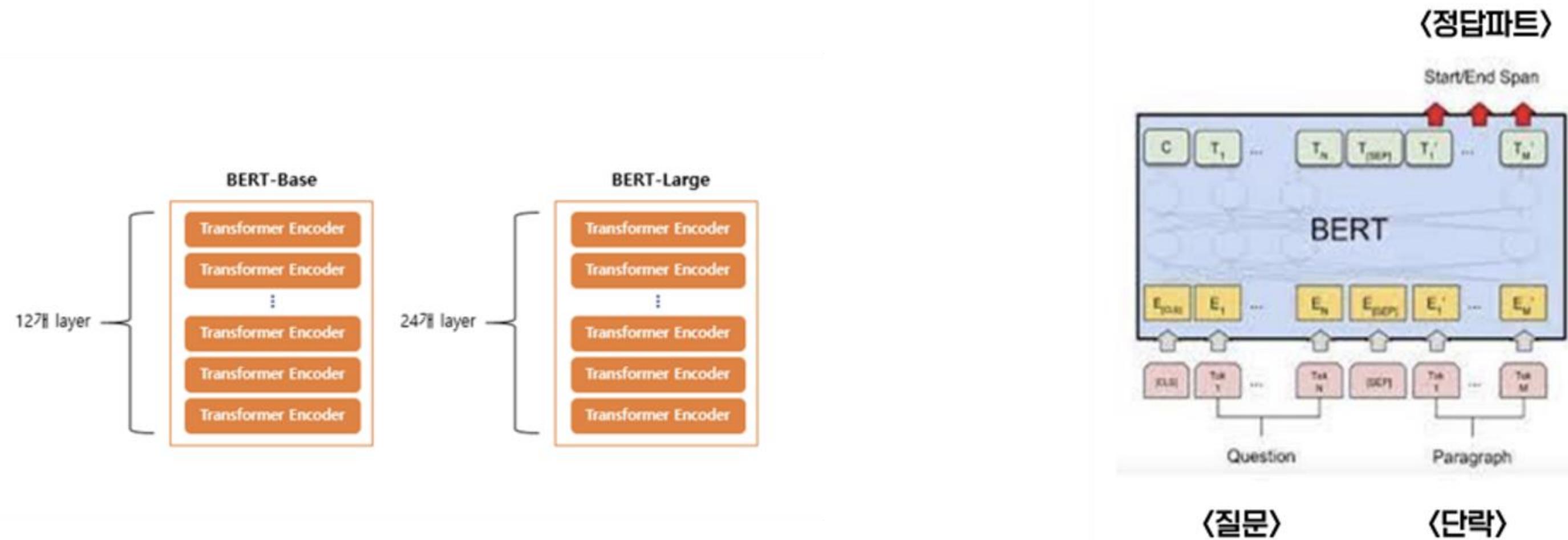
BERT Q&A 모델 training

ETRI pretrained 모델
finetuning 진행

Rule-based 전처리코드

예측된 정답을
카테고리화 진행

BERT Model: Model Description



Transformer encoder를 쌍은 구조인
BERT 모델 사용

단락과 질문을 통해 정답을 출력하는
BERT Q&A 모델 사용

BERT Model: Dataset

```
{
    "title": "2022학년도 1학기 큰사랑 이한애 장학재단 장학생 모집",
    "paragraphs": [
        {
            "context": "큰사랑이한애재단에서 2022학년도 1학기 장학생을 아래와 같이 선발하오니  
관심있는 학생들은 지원하여주시기 바랍니다. 가. 지원대상: 1학년 1학기 이상을 마친 이공계열학과 학부생 중 성적이 우수한 자 1명나.  
장학금액: 등록금 200만원/학기다. 제출서류- 추천서(날인 생략)- 장학금지급원서- 장학생신상카드- 가정환경조사서- 학습계획서- 성적증  
명서- 서약서라. 제출기한: 2022. 2. 17.(목) 오후 5시까지(시간영수)마. 제출처: skkujanghak@skku.edu문의: 02-760-1167.",
            "qas": [
                {
                    "question": "지원가능한 학년은 무엇인가요?",
                    "answers": [
                        {
                            "answer_start": "73",
                            "text": "1학년 1학기 이상"
                        }
                    ],
                    "id": "944"
                },
                {
                    "question": "직전학기 평점은 무엇인가요?",
                    "answers": [
                        {
                            "answer_start": "63",
                            "text": "."
                        }
                    ],
                    "id": "945"
                }
            ]
        }
    ]
}
```

1. KorQuAD : 기존 기계 독해 데이터셋

context, question, answer, answer start (정답의 인덱스), answer_end (정답의 끝 인덱스)

2. 장학금이라는 특정 도메인에서만 모델을 활용할 예정이기 때문에 장학금 공고 크롤링하고, 직접 라벨링하여 JSON 파일 생성

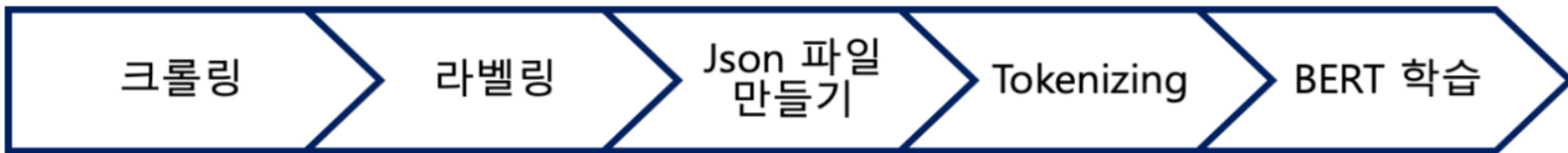
장학금 제목

장학금 게시물

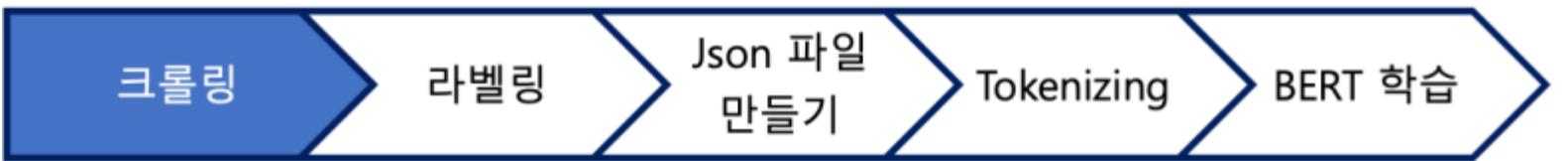
→ **장학금 게시물 관련 질문**

→ **관련 질문에 대한 답변,
정답의 인덱스**

AI 프로젝트 진행 과정



AI 프로젝트 진행 과정



```

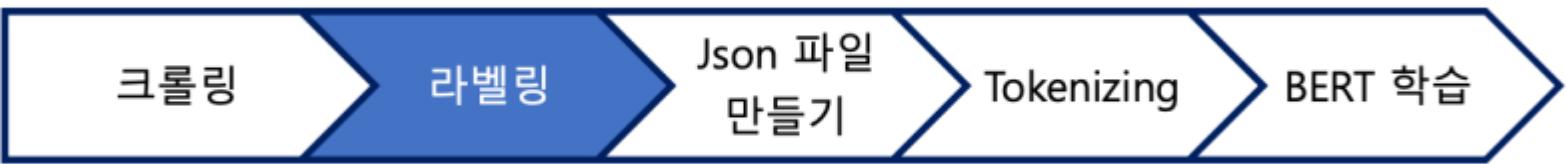
#1) 홈페이지 고유 number If using all scalar values, you must pass an index
number=post_numbers[i].text

# 세부 페이지 접속: 수집해야 하는 데이터 목록
#####
#2) date/작성날짜: 2022-09-20
date=result.select_one("#jwxe_main_content > div > div > div > div > div.board-view-title-wrap > ul > li:nth-child(2)").text.replace("최종 수정일 : ")
#3) title: 게시물 제목: 2022학년도 하나금융나눔재단 국내 하나장학생 모집 안내
title=result.select_one("#jwxe_main_content > div > div > div > div > div.board-view-title-wrap > h4").text
#4) 게시물 내용 : p 태그 전체를 감싸고 있는 상위태그를 골라주셔야 합니다!!!!!!!
content=result.select_one("#jwxe_main_content > div > div > div > div > div.board-view-content-wrap.board-view-txt").text
#5) [첨부파일 제목, 첨부파일 링크]
attachments=result.select("#jwxe_main_content > div > div > div > div > div > ul > li > a")
#####
attachment_content=""
attachment_url=""
try:
    for attach in attachments:
        # attach 제목에 attachments
        attach_title=attach.text
        if "공고문" in attach_title:
            if not os.path.exists("./content/data/"):
                os.makedirs("./content/data/")
            else:
                print("첨부파일이 이미 존재합니다.")
        else:
            print("첨부파일 제목: ", attach_title)
            print("첨부파일 링크: ", attach['href'])

```

각 공고문 내의 날짜, 게시물 제목, 게시물 내용,
첨부파일 내용(textract 이용하여 텍스트 변환), 게시물 링크, 담당부서를 크롤링

AI 프로젝트 진행 과정

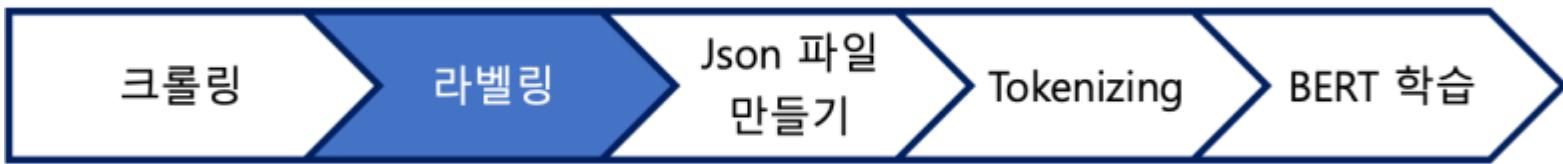


| number | date | title | content | attachment_url | attachment_conf | current_url | department | question | text | answer_start | answer_end | attach_start | attach_end | 삭제 |
|---------|------------|---|---------|--------------------------|-----------------|-------------|------------|----------|------|--------------|------------|--------------|------------|----|
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 직전학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 전체학기 평점은 무엇인가요? 3.8이상 | | | | | 287 | 292 | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 소득분위는 무엇인가요? 5분위 이내 | | | | | 224 | 230 | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 관련 학과는 무엇인가요? | | | | | | | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 거주지는 무엇인가요? | | | | | | | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 마감일은 언제인가요? | 2022. 10.7 | | | | 662 | 672 | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 특이사항은 무엇인가요? 코로나19 및 폭우피 | | | | | 180 | 197 | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 지원 가능한 학년은 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 직전학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 전체학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 소득분위는 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 관련 학과는 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 거주지는 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 마감일은 언제인가요? | 22.10.21 | | | | 577 | 589 | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 특이사항은 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 지원 가능한 학년은 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 직전학기 평점은 무엇인가요? 3 | | | | | 359 | 362 | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 전체학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 소득분위는 무엇인가요? ①기초생활수급자 | | | | | 446 | 574 | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 관련 학과는 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 거주지는 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 마감일은 언제인가요? | 2022.09.29 | | | | 1364 | 1384 | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 한나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 특이사항은 무엇인가요? | | | | | | | | | | |

질문목록:

- 1) 지원 가능한 학년은 무엇인가요? 2) 직전학기 평점은 무엇인가요?
- 3) 전체학기 평점은 무엇인가요? 4) 소득분위는 무엇인가요? 5) 관련 학과는 무엇인가요?
- 6) 거주지는 무엇인가요? 7) 마감일은 언제인가요? 8) 특이사항은 무엇인가요?

AI 프로젝트 진행 과정

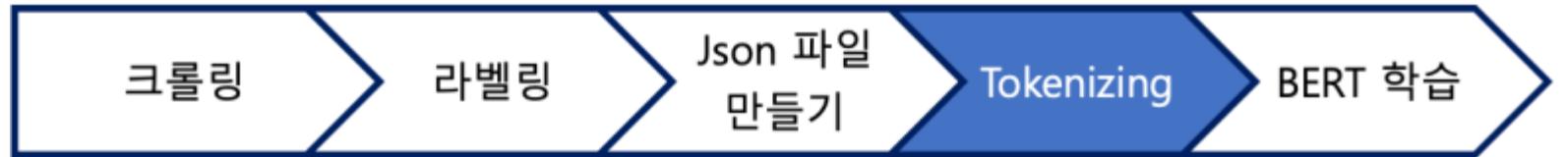


| number | date | title | content | attachment_url | attachment_conf | current_url | department | question | text | answer_start | answer_end | attach_start | attach_end | 삭제 |
|---------|------------|---|---------|-------------------|-----------------|-------------|------------|----------|------|--------------|------------|--------------|------------|----|
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 직전학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 전체학기 평점은 무엇인가요? | 13.8이상 | | | | 287 | 292 | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 소득분위는 무엇인가요? | 5분위 이내 | | | | 224 | 230 | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 관련 학과는 무엇인가요? | | | | | | | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 거주지는 무엇인가요? | | | | | | | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 마감일은 언제인가요? | 2022. 10.7 | | | | 662 | 672 | | | | |
| No.1814 | 2022.09.30 | 2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 https://www.skku | 학생지원팀 | 특이사항은 무엇인가요? | 코로나19 및 폭우피 | | | | 180 | 197 | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 지원 가능한 학년은 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 직전학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 전체학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 소득분위는 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 관련 학과는 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 거주지는 무엇인가요? | | | | | | | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 마감일은 언제인가요? | 22.10.21 | | | | 577 | 589 | | | | |
| No.1811 | 2022.09.30 | 2022학년도 서울장학재단은 https://www.skku 1111 2022년 서울 https://www.skku | 학생지원팀 | 특이사항은 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 지원 가능한 학년은 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 직전학기 평점은 무엇인가요? | 13 | | | | 359 | 362 | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 전체학기 평점은 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 소득분위는 무엇인가요? | ①기초생활수급자 | | | | 446 | 574 | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 관련 학과는 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 거주지는 무엇인가요? | | | | | | | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 마감일은 언제인가요? | 2022.09.29 | | | | 1364 | 1384 | | | | |
| No.1810 | 2022.09.30 | 2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장 https://www.skku | 학생지원팀 | 특이사항은 무엇인가요? | | | | | | | | | | |

각 게시물에서 질문에 대한 답을 찾아 정답값 태깅(start_index & end_index)
 → 총 256개(256개 게시물*8쌍 질문=2048쌍)의 공고문 라벨링 완료

train : valid : test = 8:1:1

AI 프로젝트 진행 과정



```

def parser_add_argument(parser):
    ## Required parameters
    parser.add_argument("--openapi_key", default="79dd5e46-8977-4355-aef0-ceedb2ed32c7", type=str, required=False, help="OpenAPI key information for morphology analysis")
    ## Other parameters
    parser.add_argument("--input_file", default="total.json", type=str, help="all file that include train file and test file")
    parser.add_argument("--output_file", type=str, default="tokenizing_utf8.json", help="Tokenizing file")

    return parser

def do_lang(openapi_key, text):
    openApiURL = "http://aiopen.etri.re.kr:8000/wisentu"
    requestJson = {"access_key": openapi_key, "argument": {"text": text, "analysis_code": "morp"}}

    http = urllib3.PoolManager()
    response = http.request("POST", openApiURL, headers={"Content-Type": "application/json; charset=UTF-8"}, body=json.dumps(requestJson))

    return response.data.decode()
  
```

→

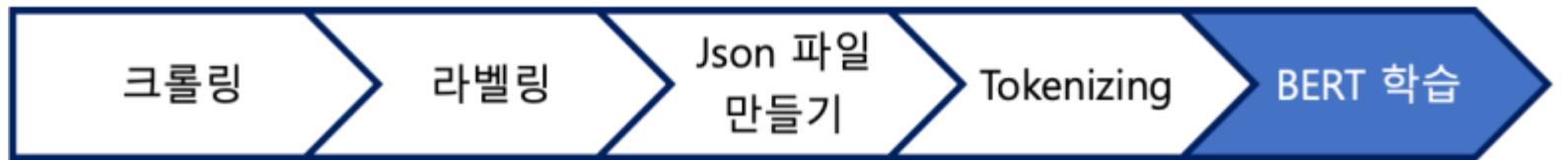
```

': 
'eight\":0.0,\\"title\":
":[],\\"sentence\":
인가요?\",\\"morp\":[{\\"id\":0.0,\\"lemma\":\"관련
\\id\":1.0,\\"lemma\":\"학과
\\id\":2.0,\\"lemma\":\"는
\\id\":3.0,\\"lemma\":\"무엇
\\id\":4.0,\\"lemma\":\"이
\\id\":5.0,\\"lemma\":\"ㄴ가
\\id\":6.0,\\"lemma\":\"요
\\id\":7.0,\\"lemma\":\"?
":[],\\"word\":[{\\"id\":0.0,\\"text\":\"관련
\\text\":\"학과는
\\text\":\"무엇인가요?
Link\":[],\\"dependency\":[],\\"SRL\":[[]]],\\"entity\":
  
```

[TTA 표준 형태소 태그셋 \(TTAK.KO-11.0010/R1\) 기반](#)

ETRI의 tokenizing API를 request하여 입력될 문장들을 형태소 단위로 tokenize 진행 후 임베딩 진행

AI 프로젝트 진행 과정



```
!python3 /content/drive/Shareddrives/skkular/code/QA_BERT/001_bert_morp_pytorch/src_examples/run_squad_ETRI.py \
--openapi_key "79dd5e46-8977-4355-ae10-ceedb2ed32c7" \
--bert_model '/content/drive/Shareddrives/skkular/code/QA_BERT/001_bert_morp_pytorch/' \
--bert_model_file '/content/drive/Shareddrives/skkular/code/QA_BERT/001_bert_morp_pytorch/pytorch_model.bin' \
--train_file '/content/drive/Shareddrives/skkular/data/장학공고/train.json' \
--predict_file '/content/drive/Shareddrives/skkular/data/장학공고/valid.json' \
--output_dir '/content/drive/Shareddrives/skkular/code/QA_BERT/' \
--tokenized_file '/content/drive/Shareddrives/skkular/code/QA_BERT/001_bert_morp_pytorch/tokenizing_utf8.json' \
--vocab_file '/content/drive/Shareddrives/skkular/code/QA_BERT/001_bert_morp_pytorch/vocab.korean_morp.list' \
--do_train \
--do_predict \
--train_batch_size 4 \
--learning_rate 3e-5 \
--num_train_epochs 5.0
```

Parameter

Epochs: 5

Learning_rate: 3e-5

Batch_size: 4

BERT finetuning 진행. pre-trained 된 모델로는 ETRI의 사전학습 모델 사용

BERT Modeling: Result

Evaluation)

1. **F1 Score**: 예측한 정답과 실제 정답 간의 정밀도와 재현율의 조화 평균 값
2. **EM (Exact Match)** : 한국어 기준 정답의 어절 단위가 정확하게 일치하는 정도

```
{  
  "exact": 76.4423076923077,  
  "f1": 78.92628205128204,  
  "total": 208  
}
```

Valid set 26개(단락-질문 208개쌍)으로 test를 진행한 결과, EM은 76%, F1-score는 78%

BERT Modeling: Good example

```
"paragraphs": [
    {
        "context": "1. 선발인원 : 1명 2. 선발대상 1) 직전 학년(2개 학기)  
200만원(등록금성) 4. 제출서류 1) 장학금 지원신청서 : 재단양식(첨부파일) 2) 학과장추천서  
1년도 세목별과세증명서(보호자) 7) 2021년도 건강보험료 납부확인서(보호자) 8) 통장사본(학  
및 제출처 1)지원마감 : 8/22(월) 15시(시간정수) 2)제출처 -이메일 : skkujanghak@skku.  

    "qas": [
        {
            "question": "지원가능한 학년은 무엇인가요?",  

            "answers": [
                {
                    "answer_start": "1",
                    "text": "."
                }
            ],
            "id": "248"
        },
        {
            "question": "직전학기 평점은 무엇인가요?",  

            "answers": [
                {
                    "answer_start": "39",
                    "text": "3.5점 이상"
                }
            ],
            "id": "249"
        },
        {
            "question": "전체학기 평점은 무엇인가요?",  

            "answers": [
                {
                    "answer_start": "1",
                    "text": "."
                }
            ],
            "id": "250"
        }
    ]
}
```

248: .,
249: 3.5점 이상,
250: .,
251: .,
252: .,
253: .,
254: 8/22,

AFTER BERT: Rule-Based

BERT 학습 결과

"1288": "123학년 재학생",

"904": "3학년 1학기",



Rule-based 후

[1, 6]

[7, 7]



BERT 학습 결과는 게시물에서 text를 추출
Rule-based로 서로 다른 결과값을 일정하게 통일

AFTER BERT: Rule-Based

| number | title | date | content | con_age | con_bef_score | con_total_score | con_income | con_major | con_where | con_end_date |
|----------|---------------------------------------|---------------------|----------------|---------|---------------|-----------------|------------|------------|-----------|--------------|
| No. 1844 | 2022학년도 상반기 익산시 대학(원)생 학자금 ... | 2022-11-23 00:00:00 | ▣ 지원대상 : 공 | None | None | None | [0, 6] | None | None | 2022-05-13 |
| No. 1843 | 2023년도 1학기 1차 국가근로장학금 학생신청... | 2022-11-23 00:00:00 | 2023년도 1학기 | None | None | None | None | None | None | 2022-11-15 |
| No. 1842 | 한국고등교육재단 제 2기 인재림 장학생 선발 ... | 2022-11-22 00:00:00 | 한국고등교육재단 | None | None | None | None | [글로벌경영학... | None | None |
| No. 1841 | 2023-2024 헐가리 정부초청 장학생(Stipendium ... | 2022-11-22 00:00:00 | 2023-2024 헐가리 | [5, 8] | 3.6 | None | None | None | None | 2022-11-10 |
| No. 1840 | 2023-2024 이스라엘 정부초청 장학생 선발 안내 | 2022-11-22 00:00:00 | 2023-2024 이스라엘 | None | 3.5 | None | [0, 7] | None | None | 2022-10-19 |
| No. 1839 | 2022학년도 2학기 한국선원복지고용센터 장학... | 2022-11-17 00:00:00 | 우리센터는 선원 | None | 2.0 | 2.0 | None | None | None | 2022-10-30 |
| No. 1838 | 취업후상환 학자금대출 재무자신고 및 해외이... | 2022-11-17 00:00:00 | 안녕하세요, 학자금 | None | 3.5 | 2.0 | [0, 7] | [글로벌경영학... | None | 2022-10-19 |
| No. 1837 | 2023학년도 덕신하우징 건축안전/환경개선 장학... | 2022-11-17 00:00:00 | (주)덕신하우징 | None | None | None | None | [글로벌경영학... | None | None |
| No. 1836 | 2022학년도 한국산림복지진흥원 녹색장학사업... | 2022-11-14 00:00:00 | 가. 사업명: | [3, 6] | None | None | None | [반도체시스템... | None | 2022-12-31 |
| No. 1835 | 2023학년도 KC미래장학금 제6기 장학생 모집 ... | 2022-11-14 00:00:00 | 가. 선발 대상 : | None | None | None | None | [나노공학과,... | None | 2022-08-15 |

〈DB: BERT 핵심 정보 추출 후〉

Trial 1. unanswered question

```
'answer': {'text': '피막 흄',
            'html_answer_start': 23296,
            'html_answer_text': '피막 흄',
            'answer_start': 5909},
    'question': '섬유모세포의 막에는 몇 퍼센트의 피막 흄이 있을까?',
    'id': '66612'}]}
```



1. 장학금 공고에 지원 학년 제한이 없으면 모델에서는 ‘답이 없음’을 추출해야 함
2. KorQUAD2.0와 SQuad2.0에서는 unanswered question이 있었지만 임의의 답을 추출하고 answer에 넣었음 : 모델에서 “**답이 없음**”을 추출하는 것을 학습하기 원함

Trial 1. unanswered question

```
{  
  "data": [  
    {  
      "title": "2021년 상반기 군산시 대학생 학자금이자 지원사업",  
      "paragraphs": [  
        {  
          "context": "군산시에서는 군산지역 대학생들의 고등교육기회를 보장하고 재학기간동안 학업에 전념할 수 있도록 교육비  
부담을 경감하기 위하여 「군산시 대학생 학자금 이자지원 사업」을 시행중에 있습니다. 해당 학생들은 많은 지원 바랍니다. 가. 신청기간 : 2021. 5. 3.(월) ~ 6. 4.  
(금) 나. 대상 : 주민등록상 군산시에 주소를 두고 있는 전국대학의 재(휴)학생 다. 지원범위 : 2016년 1월 이후 한국장학재단에서 대출 받은 학자금의2021년 상반기  
('21. 1. 1. ~ 6. 30.) 발생이자 라. 신청방법 : 온라인 또는 방문신청 - 온라인신청 : 군산시청 홈페이지(https://www.gunsan.go.kr) - 방문신청 : 군산시청 교  
육지원과(7층) 마. 제출서류 : 신청서(홈페이지 다운로드) 재(휴)학증명서*자세한 내용은 불임 참조.",  
        "qas": [  
          {  
            "question": "지원가능한 학년은 무엇인가요?",  
            "answers": [  
              {  
                "answer_start": "100",  
                "text": "."  
              }  
            ],  
            "id": "2064"  
          },  
        ]  
      ]  
    }  
  ]  
}
```



모든 content에 "."을 추가하고 unanswered question에 대한 answer을 그러한 '.'으로 지정하고 태깅

Limitation

| | |
|-------|---|
| 크롤링 | <ol style="list-style-type: none"> 1. 장학 선발결과 공고 제외 2. 학부생 대상 장학금이 아닌 경우 제외 (ex. 석,박사 대상) |
| 라벨링 | <ol style="list-style-type: none"> 1. 소득 산정 기준 중 건강보험료 제외 <ul style="list-style-type: none"> -> 주로 건강보험료 지급명세서를 제출하라는 방식. 기준으로 보기 어려움. -> 또한, 보통 차상위계층 등 다른 기준과 함께 중복 조건으로 제시 2. 직전학기 이수학점 제외 <ul style="list-style-type: none"> -> 일반적으로 12학점, 기재 안 하는 장학금 다수 3. 타 장학금 수혜 여부 제외 <ul style="list-style-type: none"> -> 교내장학금의 경우 기준이 비교적 명확하나, 교외장학금의 경우 기준이 없거나 특수한 경우가 많음 |
| 학습 제외 | <ol style="list-style-type: none"> 1. 한 장학금 공고에 여러 장학금이 있는 경우 제외 2. 첨부파일이 이미지 형식이거나 hwp 파일인 경우 제외 <ul style="list-style-type: none"> (*다만, 최근 1년간 장학금 공고 중 첨부파일에만 내용이 있는 경우는 없음) |
| 목록 제외 | <ol style="list-style-type: none"> 1. 신청할 필요가 없는 장학금 제외 (ex. 문행장학금) |



3. Backend & Frontend

Development Tool_Backend



+



+



+



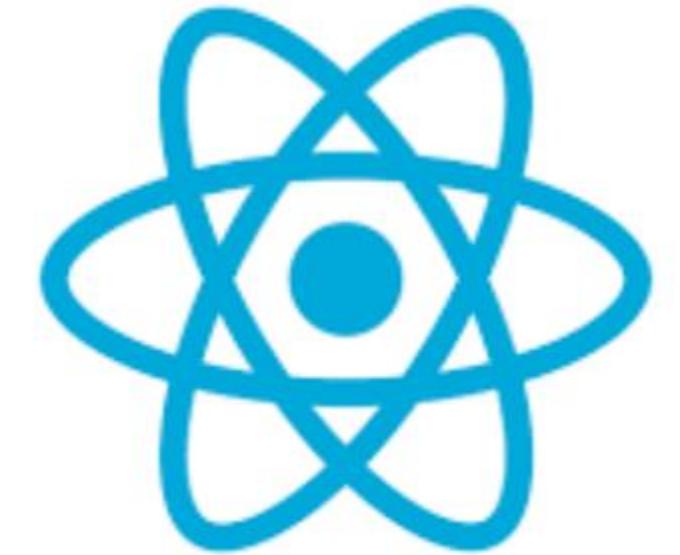
웹 프레임워크

REST API
서버 구축

서버

REST API
TEST

Development Tool_Frontend



React Native

expo

1. React Native CLI

어떠한 툴을 사용하지 않고 React Native만을 이용해서 개발하는 방식
– 자신이 원하는 기능이나 모든 요소들을 제어할 수 있으며, 많은 라이브러리를 사용할 수 있음

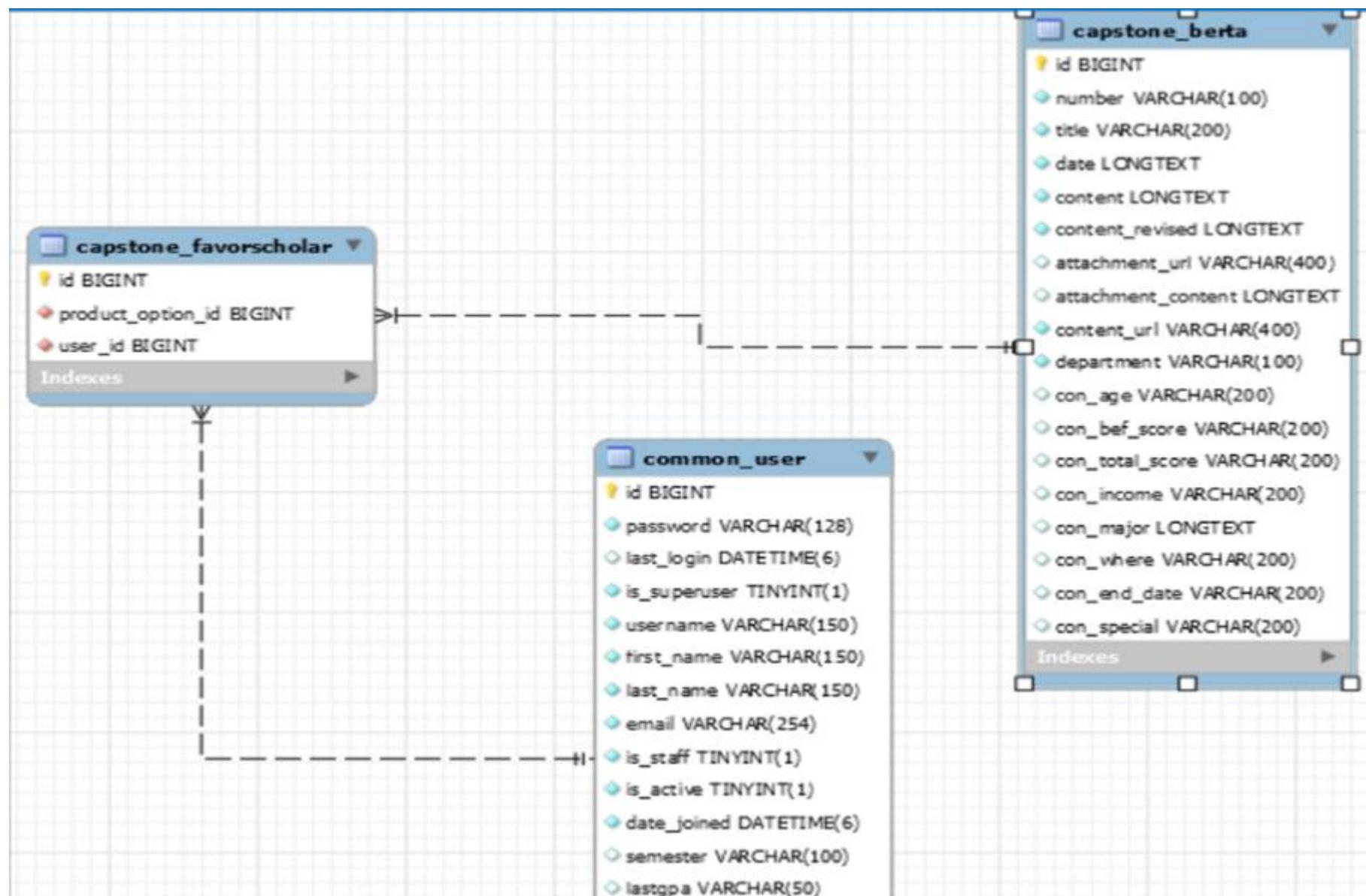
=> 자유도가 높지만, 초기 설정이 까다롭고 개발 과정이 비교적 복잡.

2. Expo CLI

React Native 개발을 쉽게 할 수 있도록 도와주는 프레임워크/플랫폼
– 애플리케이션 개발을 단순화하는 많은 편의 유틸리티 기능 존재
– 기본적인 세팅이 되어 있으며 native 파일을 숨겨놓고 자동으로 관리
=> 자유도가 낮지만, 초기 설정이 간단하고 비교적 편리하게 개발 가능.

*리액트 네이티브는 페이스북에서 만든 오픈소스 모바일 애플리케이션 프레임워크
JavaScript로 개발하며, IOS나 Andriod, Web, UWP에 대한 모바일앱을 동시에 개발을 할 수 있는 크로스 플랫폼

MySQL 스키마



Capstone_favorscholar : 관심 장학

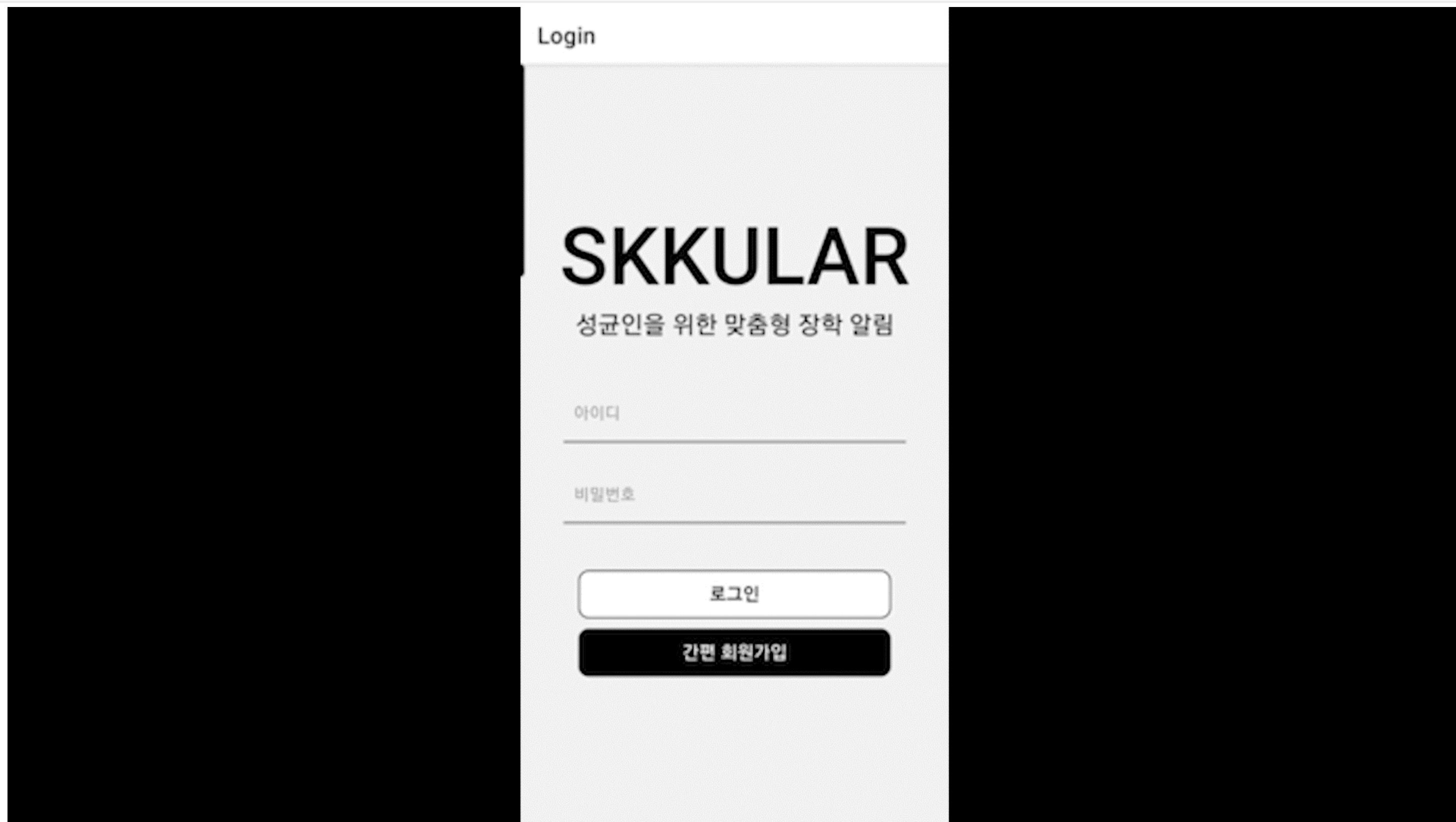
Capstone_Berts : 크롤링 + 버트 학습 데이터

Common_user: 회원 정보 데이터

사용한 API

| | |
|-------------|-----------------|
| 회원가입 | 회원정보 보여주기 |
| 로그인/로그아웃 | 관심장학 추가/삭제/보여주기 |
| JWT 토큰 리프레시 | 맞춤형 장학 선별/보여주기 |
| 회원정보 수정 | 전체 게시물 보여주기 |
| | 특정 게시물 보여주기 |

Demo



Trial & Error

* setInterval & useInterval

실시간으로 업데이트 되는 장학 정보를 받아와야 함
 => 기존에 useState만 사용할 때에는 앱을 다시 로딩해야만,
 새로운 장학 정보 업데이트가 가능했음.

주기적으로 DB에 접근하여 장학 정보를 받아와야 함
 => setInterval과 useInterval을 사용하여 일정 주기마다
 api를 fetch 하도록 하여 실시간으로 업데이트된
 장학 정보를 반영할 수 있도록 하였음
 (request가 많아져 앱 자체의 속도는 살짝 느려졌지만)

=> 반복하여 정보를 받아옴으로써 이전 상태와 새로 받아온 게시글 개수를 비교하여
 나에게 맞는 맞춤 장학 공고가 업데이트 되었음을 확인할 수 있음 => 맞춤 알림 기능 제공!



Limitation

* background 맞춤 알림

앱이 실행되고 있는 active 상태에서는 정보가 업데이트 되었을 때 맞춤 알림이 정상적으로 작동하지만, 사용자가 앱을 실행하고 있지 않은 때에는 맞춤형 알림 기능이 제공되지 않음

대부분 Firebase의 FCM(Firebase Cloud Messaging) 사용->
백그라운드 푸시알림을 관리

이미 프로젝트를 진행할 때에 장고를 활용하여 DB를 관리하고 있어,
expo가 푸시 알림 용도로 제공하는
'expo-notifications'과 Expo SDK를 사용하여 알림 기능을 구현



=> 다시 앱으로 돌아가면 (background->active)
그동안 업데이트 된 맞춤 장학 알림을 받아볼 수 있도록 하였음 (제한적 알람)

Role of each member

역할분담 및 진행일정

강병준

데이터 수집
서버, 백엔드 개발

박진아

데이터 수집
프론트엔드 개발

주재현

데이터 수집
프론트엔드 개발

김규진

데이터셋 전처리
및 모델 개발

장이준

데이터셋 전처리
및 모델 개발

Q & A

지금까지 TEAM. SKKULAR 였습니다.
감사합니다.