

# SKKU SCHOLARSHIP



| MIDTERM

TEAM 스끌라, SKKULAR

강병준 | 김규진 | 박진아 | 장이준 | 주재현

# CONTENTS

1. 프론트엔드

2. 백엔드

3. AI

TEAM. 스플라  
SKKULAR

# Proposition Remind

## 문제인식

교내 학우들의  
낮은 장학 지원율

## 원인분석

자신이 지원할 수 있는  
장학금을 직관적으로  
알기 어려움

## 해결방안

‘맞춤형’ 장학 검색  
및 알림 서비스 제안

# Proposition Remind

학교에서도 문제로 인식하고 있음

## 신청 가능 장학금 목록

더보기 >

조회된 데이터가 없습니다.



추천 장학금 시뮬레이션

# 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

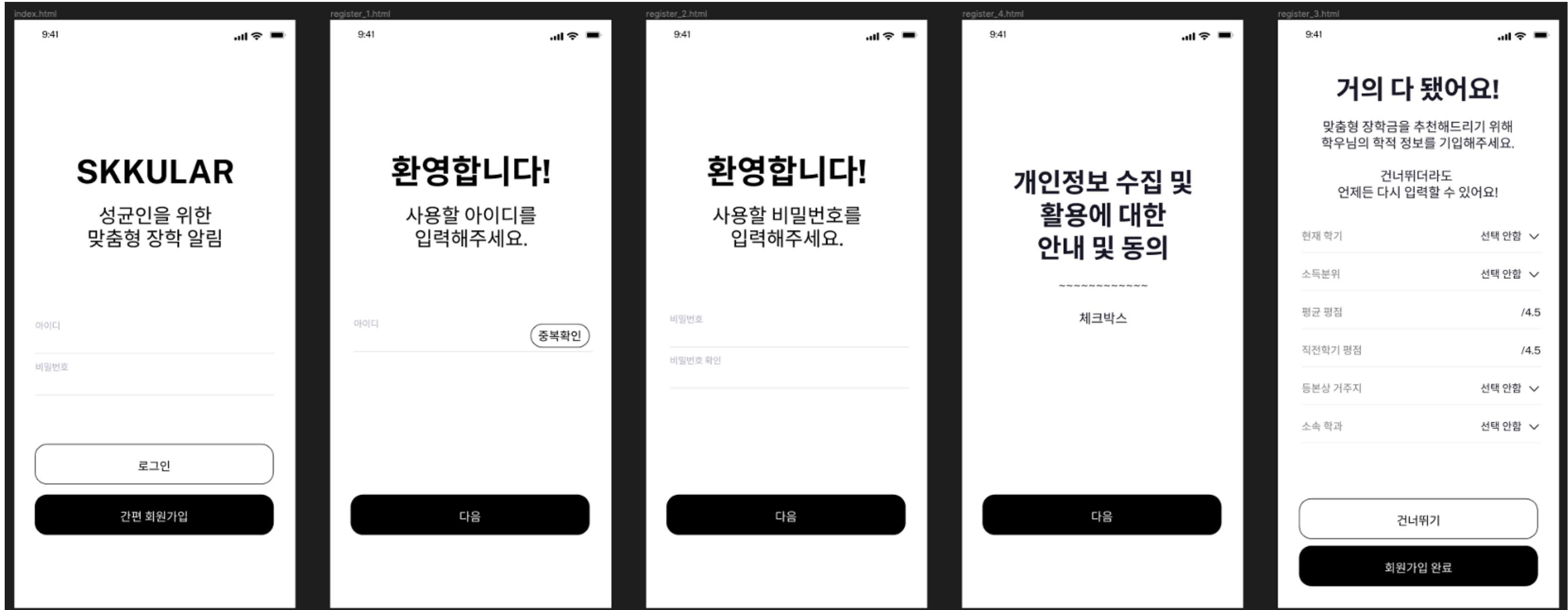
**장이준**

데이터셋 전처리  
및 모델 개발



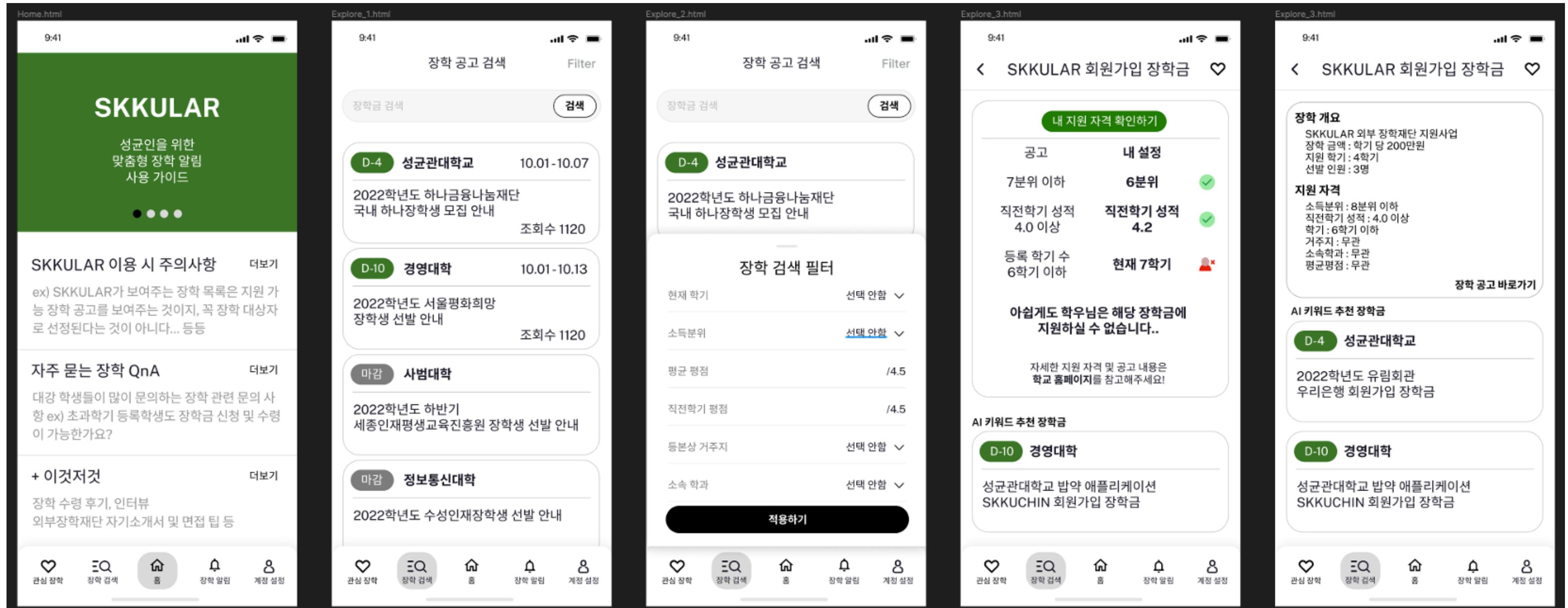
# 1. Frontend

# Front-End : UX/UI (회원가입)



개인을 특정할 수 없도록 **간단하게!**

# UX/UI (글 목록 및 상세)



# UX/UI : 핵심 기능

현재 학기	선택 안함 ▼
소득분위	선택 안함 ▼
평균 평점	/4.5
직전학기 평점	/4.5
등본상 거주지	선택 안함 ▼
소속 학과	선택 안함 ▼

< SKKULAR 회원가입 장학금 ❤

**내 지원 자격 확인하기**

공고	내 설정
7분위 이하	<b>6분위</b> ✓
직전학기 성적 4.0 이상	<b>직전학기 성적 4.2</b> ✓
등록 학기 수 6학기 이하	<b>현재 7학기</b> ✘

**아쉽게도 학우님은 해당 장학금에  
지원하실 수 없습니다..**

자세한 지원 자격 및 공고 내용은  
학교 홈페이지를 참고해주세요!

AI 키워드 추천 장학금

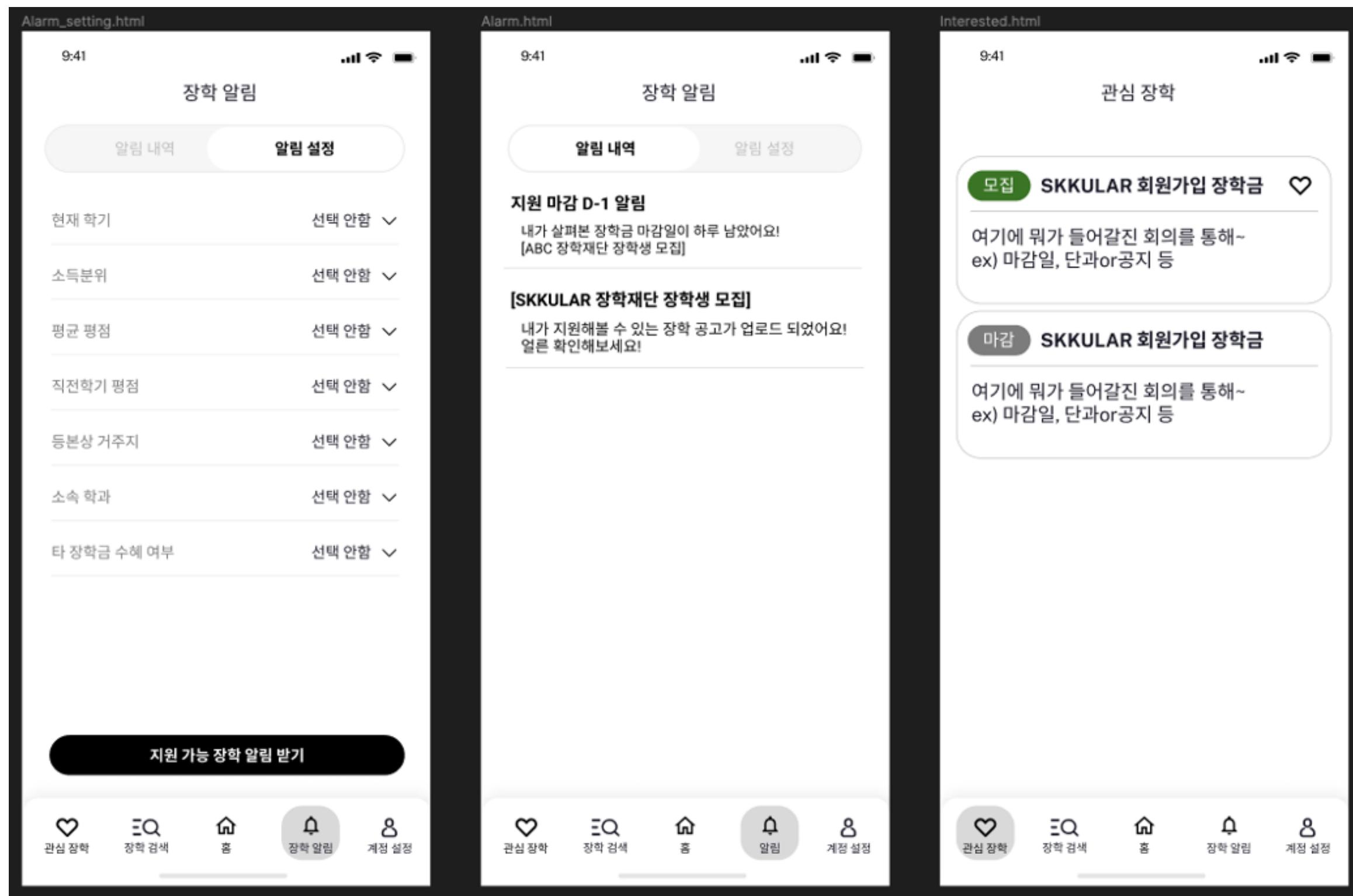
**D-4** 성균관대학교

2022학년도 유림회관  
우리은행 회원가입 장학금

**D-10** 경영대학

성균관대학교 밥약 애플리케이션  
SKKUCHIN 회원가입 장학금

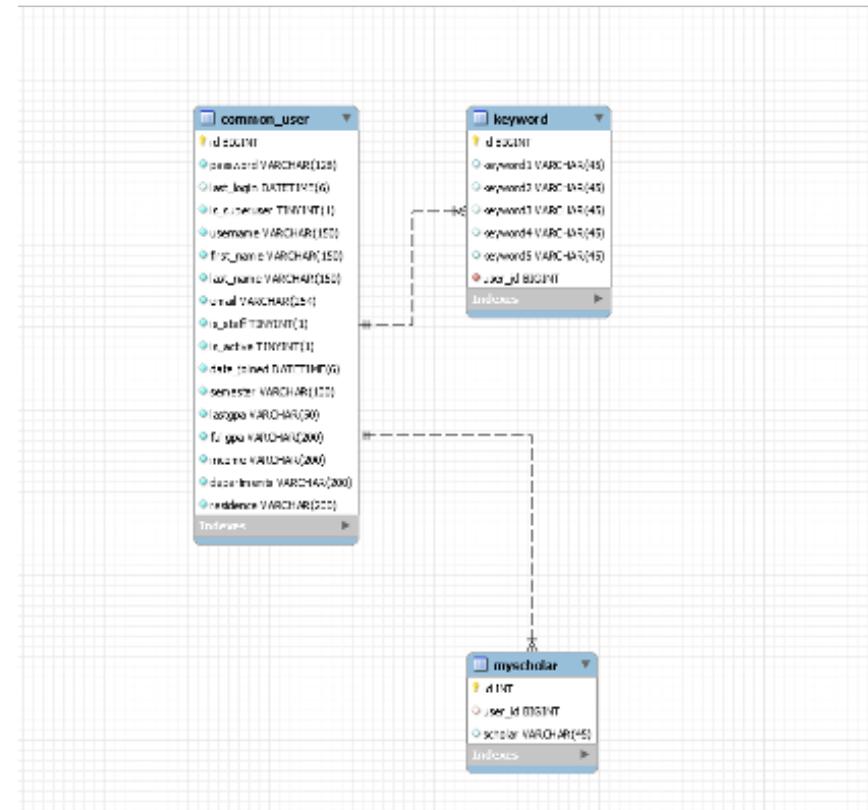
# UX/UI : 알림 및 관심 장학





## 2. Backend

# Modeling Flow Chart



## DB 설계

# MTV 패턴에 따른 장고 개발

```
Rest API Policy Rule JSON

1 {
2   "id": "1",
3   "label": "Korean Text Detection API",
4   "type": "Text Detection API",
5   "title": "Korean Text Detection API",
6   "content": "This API is designed to detect Korean text from images. It takes an image URL as input and returns the detected text along with its bounding boxes. The response includes the text content, bounding box coordinates, and confidence scores for each detected text segment. The API is currently available in English and Korean versions. The Korean version is currently in beta testing phase. Please note that the API is still under development and may contain bugs or errors. We welcome feedback and suggestions to improve the service.", 
7   "description": "The Korean version of the API is currently in beta testing phase. Please note that the API is still under development and may contain bugs or errors. We welcome feedback and suggestions to improve the service.",
8   "version": "1.0.0"
9 }
10 {
11   "id": "2",
12   "label": "Korean Text Recognition API",
13   "type": "Text Recognition API",
14   "title": "Korean Text Recognition API",
15   "content": "This API is designed to recognize Korean text from images. It takes an image URL as input and returns the recognized text along with its bounding boxes. The response includes the text content, bounding box coordinates, and confidence scores for each recognized text segment. The API is currently available in English and Korean versions. The Korean version is currently in beta testing phase. Please note that the API is still under development and may contain bugs or errors. We welcome feedback and suggestions to improve the service.", 
16   "description": "The Korean version of the API is currently in beta testing phase. Please note that the API is still under development and may contain bugs or errors. We welcome feedback and suggestions to improve the service.",
17   "version": "1.0.0"
18 }

API 명세서 작성 및 설계
```

# DB 설계

<b>유저</b>	Id pw  학기 직전학기 평점은 무엇인가요? 전체학기 평점은 무엇인가요? 소득분위는 무엇인가요? 전체학기 평점은 무엇인가요? 소득분위는 무엇인가요? 학과는 무엇인가요? 거주지는 무엇인가요?
<b>장학금</b>	지원가능한 학년은 무엇인가요?  직전학기 평점은 무엇인가요?  전체학기 평점은 무엇인가요? 소득분위는 무엇인가요?  관련 학과는 무엇인가요?  거주지는 무엇인가요?  마감일은 언제인가요?  특이사항은 무엇인가요?

## 1. 요구사항 추출

개체	속성
회원	Id, pw, 학기, 직전학기 평점, 전체 평점, 소득분위, 학과, 거주지
장학금	학기, 직전학기 평점, 전체 평점, 소득분위, 학과, 거주지, 특이사항, 마감일
키워드	Id, keyword 1~5
나의 장학금	Id, index, 장학금

## 2. 개체와 속성 추출

<b>유저와 키워드 = 1: 1</b>
<b>유저와 나의 장학금 = 1:N</b>

## 3. 관계 추출

회원(Id(pk), pw, 학기, 직전학기 평점, 전체 평점, 소득분위, 학과, 거주지)  
 장학금(학기, 직전학기 평점, 전체 평점, 소득분위, 학과, 거주지, 특이사항, 마감일)  
 키워드(Id(pk), user\_id(fk), keyword 1~5)  
 나의 장학금(Id(pk), user\_id(fk), index, 장학금 )

## 4. Relation 변환

# 장고 - 개발흐름

1. 유저가 특정 *url*로 요청을 보냄.
2. *urlConf*를 통해 해당 *url*과 매핑된 뷰를 호출.
3. 호출된 뷰는 요청에 따라 적절한 로직을 수행하며, 그 과정에서 모델에게 CRUD를 지시.
4. 모델은 ORM을 통해 DB와 소통하며 CRUD를 수행.
5. 그 후 뷰는 지정된 템플릿을 렌더링.
6. 최종 결과를 응답으로 반환.

# 장고 - views

```
def scholar_list(request):
    scholars = Kscholar.objects.all()
    context = {'scholars':scholars}
    return render(request,'capstone/index.html',context)

def scholar_content(request,scholars_id):
    scholars = get_object_or_404(Kscholar, pk = scholars_id)
    context = {'scholars': scholars}
    return render(request, 'capstone/content.html', context)

class Kscholarlistapi(APIView):
    def get (self, request):
        queryset = Kscholar.objects.all()
        print(queryset)
        serializer = ScholarSerializer(queryset,many = True)
        return Response(serializer.data)

    def index(request):
```

1. 애플리케이션을 동작하게 하기 위해 **CRUD(Create/Read/Update/Delete)** 등 함수를 구현하는 곳
2. 웹 애플리케이션의 "로직 부분"을 담당하여 동작

# 장고 - url

```
app_name = 'capstone'
urlpatterns = [
    path('admin/', admin.site.urls),
    path('api-auth/', include('rest_framework.urls')),
    path('capstone/', views.index, name="index"),
    path('common/', include('common.urls')),
    path('capstone/<int:question_id>/', views.detail, name='detail'),
    path('capstone/answer/create/<int:question_id>/', views.answer_create, name='answer_create'),
    path('question/create/', views.question_create, name='question_create'),
    path('', views.scholar_list, name='index1'),
    path('/<int:scholars_id>/',views.scholar_content,name ='scholar'),# '/'에 해당되는 path
    path('api/scholar',Kscholarlistapi.as_view())
```

]

1. path 함수를 통해 URL 매핑
2. url 경로를 설정하고 views.py에 있는 함수를 호출

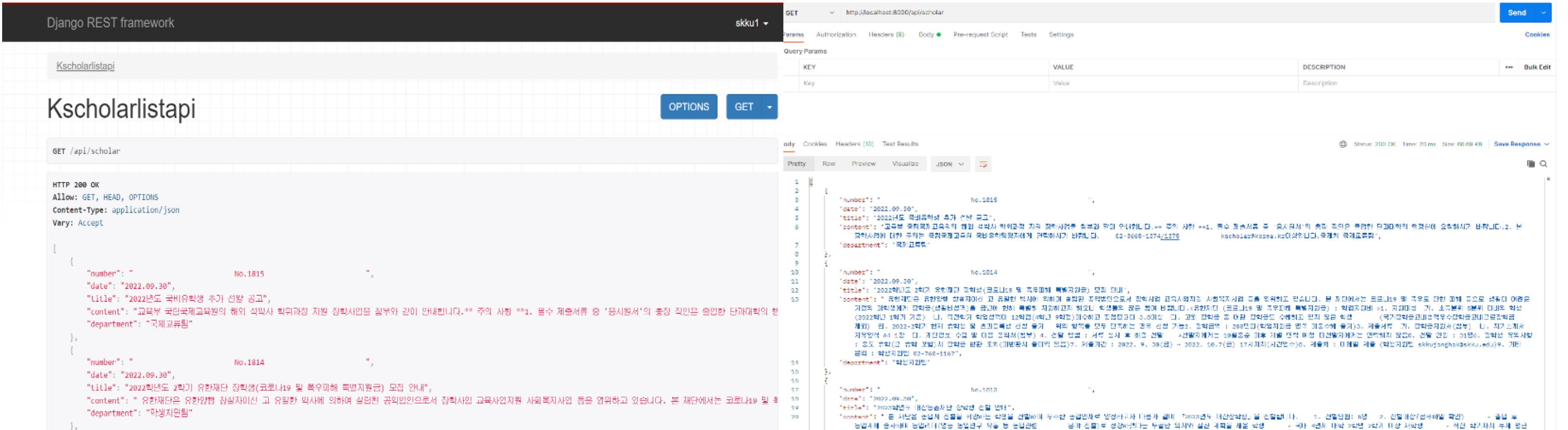
# 장고 - model

```
class Kscholar(models.Model):  
  
    number = models.TextField(blank=True, null=True)  
    date = models.TextField(blank=True, null=True)  
    title = models.TextField(blank=True, null=True)  
    content = models.TextField(blank=True, null=True)  
    attachment_url = models.TextField(blank=True, null=True)  
    attachment_content = models.TextField(blank=True, null=True)  
    current_url = models.TextField(blank=True, null=True)  
    department = models.TextField(blank=True, null=True)
```

1. ORM(SQL 쿼리 이용하지 않아도 됨)
2. Make migration 명령어와 make migrate 명령어를 이용
3. 변경사항을 DB에 적용함.

## 2. 백엔드

# API



Django REST framework

skku1 ▾

GET http://localhost:8000/api/scholar

Params Authorization Headers (8) Body Pre-request Script Tests Settings

Query Params

KEY	VALUE	DESCRIPTION
Key	Value	Description

**Kscholarlistapi**

OPTIONS GET

GET /api/scholar

HTTP 200 OK

Allow: GET, HEAD, OPTIONS

Content-Type: application/json

Vary: Accept

[

[

{

"number": "No.1815",  
"date": "2022.09.30",  
"title": "2022년도 국비유학생 추가 선발 공고",  
"content": "교육부 국립국제교육원의 해외 석박사 학위과정 지원 정책사업을 충분히 알리기 위해 주의 사항 \*\*1. 필수 제출서류 중 '응시원서'의 충장 직인은 출판한 단과대학의 행정실에 유통하시기 바랍니다. 2. 본  
정책사업에 대한 문의는 국립국제교육원 국립유학지원단에게 연락하시기 바랍니다. 02-3660-1374/1375  
kscholarlist@koaea.kz@skku.edu, 국제교류팀",  
"department": "국제교류팀"

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

{

"number": "No.1814",  
"date": "2022.09.30",  
"title": "2022년도 2학기 유학생(코로나19 및 폭우피해 특별지원금) 모집 안내",  
"content": "2022학년도 2학기 유학생(코로나19 및 폭우피해 특별지원금) 모집 안내",  
"department": "학생지원팀"

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

{

"number": "No.1813",  
"date": "2022.09.30",  
"title": "2022학년도 대선생님자선 장학생 선발 안내",  
"content": "2022학년도 대선생님자선 장학생 선발 안내",  
"department": "교직원복지부

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

Preview Visualize JSON

Status: 200 OK Time: 20 ms Size: 60.68 KB Save Response

# 1. Django Rest Framework 사용

# 2. Rest API 구축

# Demo

## SKKULAR

성균인을 위한  
맞춤형 장학 알림

아이디

비밀번호

로그인

간편 회원가입

환영합니다!

사용자 정보를  
입력해주세요.

아이디	<input type="text" value="skku1"/>	<a href="#">중복 확인</a>
비밀번호	<input type="password" value="*****"/>	
비밀번호 확인	<input type="password" value="*****"/>	
이메일	<input type="text" value="freudbj2@naver.com"/>	
학기	<input type="text" value="2-2"/>	
직전학기 평점	<input type="text" value="4.05"/>	
총 평점	<input type="text" value="4.1"/>	
소득분위	<input type="text" value="10분위"/>	
학과	<input type="text" value="소프트웨어학과"/>	
거주지	<input type="text" value="서울특별시"/>	

생성하기

# Demo

**SKKULAR**

skku1 (로그아웃) 프로필

**게시글 내용**

순번	날짜	제목	부서
1	2022.09.30	2022년도 국비유학생 추가 선발 공고	국제교류팀
2	2022.09.30	2022학년도 2학기 유한재단 장학생(...)	학생지원팀
3	2022.09.30	2023학년도 대산농촌재단 장학생 선...	학생지원팀
4	2022.09.30	2022학년도 하반기 세종인재평생교육...	학생지원팀
5	2022.09.30	2022학년도 서울평화회망 장학생 선...	학생지원팀
6	2022.09.30	2022학년도 하나금융나눔재단 국내 ...	학생지원팀
7	2022.09.27	2023학년도 1학기 공덕환 신대학원우수장학...	공덕환
8	2022.09.26	2022학년도 하반기 신안군장학재단 ...	학생지원팀
9	2022.09.26	2022학년도 수성인재장학생 선발 안내	학생지원팀

**SKKULAR**

skku1 (로그아웃) 프로필

**게시글 내용**

교육부 국립국제교육원의 해외 석박사 학위과정 지원 장학사업을 첨부와 같이 안내합니다.\*\*  
 주의 사항 \*\*1. 필수 제출서류 중 '응시원서'의 총장 직인은 졸업한 단과대학의 행정실에 요청하시기 바랍니다.2. 본 장학사업에 대한 문의는 국립국제교육원 국비유학담당자에게 연락하시기 바랍니다. 02-3668-1374/1375  
 kscholar@korea.kr입니다. 국제처 국제교류 팀

# Demo

**SKKULAR**

skku1 (로그아웃) 프로필

마이페이지

**내 정보**

이메일	freudbj2@naver.com
이름	skku1
학기	2-2
지난학기 성적	4.05
전체 성적	4.1
소득 분위	10분위
학과	소프트웨어학과
거주지	서울특별시

**정보수정**

**환영합니다!**

사용자 정보를  
입력해주세요.

총 평점  
4.5

**수정하기**

**SKKULAR**

skku1 (로그아웃) 프로필

마이페이지

**내 정보**

이메일	freudbj2@naver.com
이름	skku1
학기	2-2
지난학기 성적	4.05
전체 성적	4.5
소득 분위	10분위
학과	소프트웨어학과
거주지	서울특별시

**정보수정**

# Demo

The image displays four screenshots of a user profile management application, likely a mobile application, showing different states of the user profile screen.

**Screenshot 1 (Left): Profile Overview**

Header: SKKULAR  
Top Bar: skku1 (로그아웃) 프로필  
Section: 게시글 내용  
Content: 교육부 국립국제교육원의 해외 석박사 학위과정 지원 장학사업을 첨부와 같이 안내합니다.\*\* 주의 사항 \*\*1. 필수 제출서류 중 '응시원서'의 총장 직인은 졸업한 단과대학의 행정실에 요청 하시기 바랍니다.2. 본 장학사업에 대한 문의는 국립국제교육원 국비유학담당자에게 연락하시 기 바랍니다. 02-3668-1374/1375  
kscholar@korea.kr

**Screenshot 2 (Second from Left): My Page**

Header: SKKULAR  
Top Bar: skku1 (로그아웃) 프로필  
Section: 마이페이지  
Section: 내 정보  
Content: 이메일: freudbj2@naver.com, 이름: skku1, 학기: 2-2, 지난학기 성적: 4.05, 전체 성적: 4.1, 소득 분위: 10분위, 학과: 소프트웨어학과, 거주지: 서울특별시

**Screenshot 3 (Center): Edit Profile**

Header: SKKULAR  
Top Bar: skku1 (로그아웃) 프로필  
Section: 환영합니다!  
Text: 사용자 정보를 입력해주세요.  
Section: 수정하기  
Content: 총 평점: 4.5

**Screenshot 4 (Right): Updated Profile**

Header: SKKULAR  
Top Bar: skku1 (로그아웃) 프로필  
Section: 마이페이지  
Section: 내 정보  
Content: 이메일: freudbj2@naver.com, 이름: skku1, 학기: 2-2, 지난학기 성적: 4.05, 전체 성적: 4.5, 소득 분위: 10분위, 학과: 소프트웨어학과, 거주지: 서울특별시

# Demo

# 정보 수정 전

skku1	freudbj2@naver.com	0	1	2022-10-31 15:27:21.168996	2-2	4.05	4.1	10분위	소프트웨어학과	서울!
-------	--------------------	---	---	----------------------------	-----	------	-----	------	---------	-----

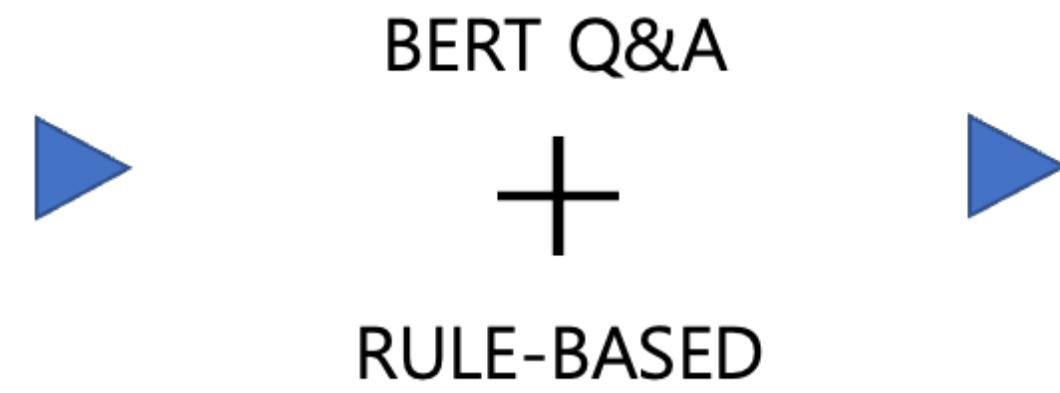
# 정보 수정 후



### 3. AI

# Modeling Goal

{
   
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 게시물 크롤링

자연어 모델과 rule based  
를 통하여 게시물 내 핵심  
정보 추출 및 카테고리화

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

}

## Crawling

학교 전체 장학금 홈페이지와 단과대 장학금 게시물 1500 여개 크롤링

## Parsing & Labeling

데이터 전처리 진행 후 BERT 모델 학습을 위하여 학습데이터 2048쌍 생성  
(Squad 형태. 질문-단락 쌍)

BERT Q&A

**BERT Q&A 모델  
training**

**진행 중**

RULE-BASED

**Rule-based  
전처리코드**

# Limitation 1 - 한계 인식 후 수정

## 라벨링 후 작업

1. 학점 (xx/4.5 외에도 A,B,C or 90점 등 기준 다양)  
→ 추출 이후 4.5점 기준으로 환산
2. 학과 (ex, 자연과학계열, 공학계열, 상경계열 등)  
→ 추출 이후 각 학과로 환산

# Limitation 2 - 한계 인식 후 제외

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



# BERT 모델 설명

# 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"
                }
            ]
        }
    ]
}
```

**장학금 제목**

**장학금 게시물**

**장학금 게시물 관련 질문**

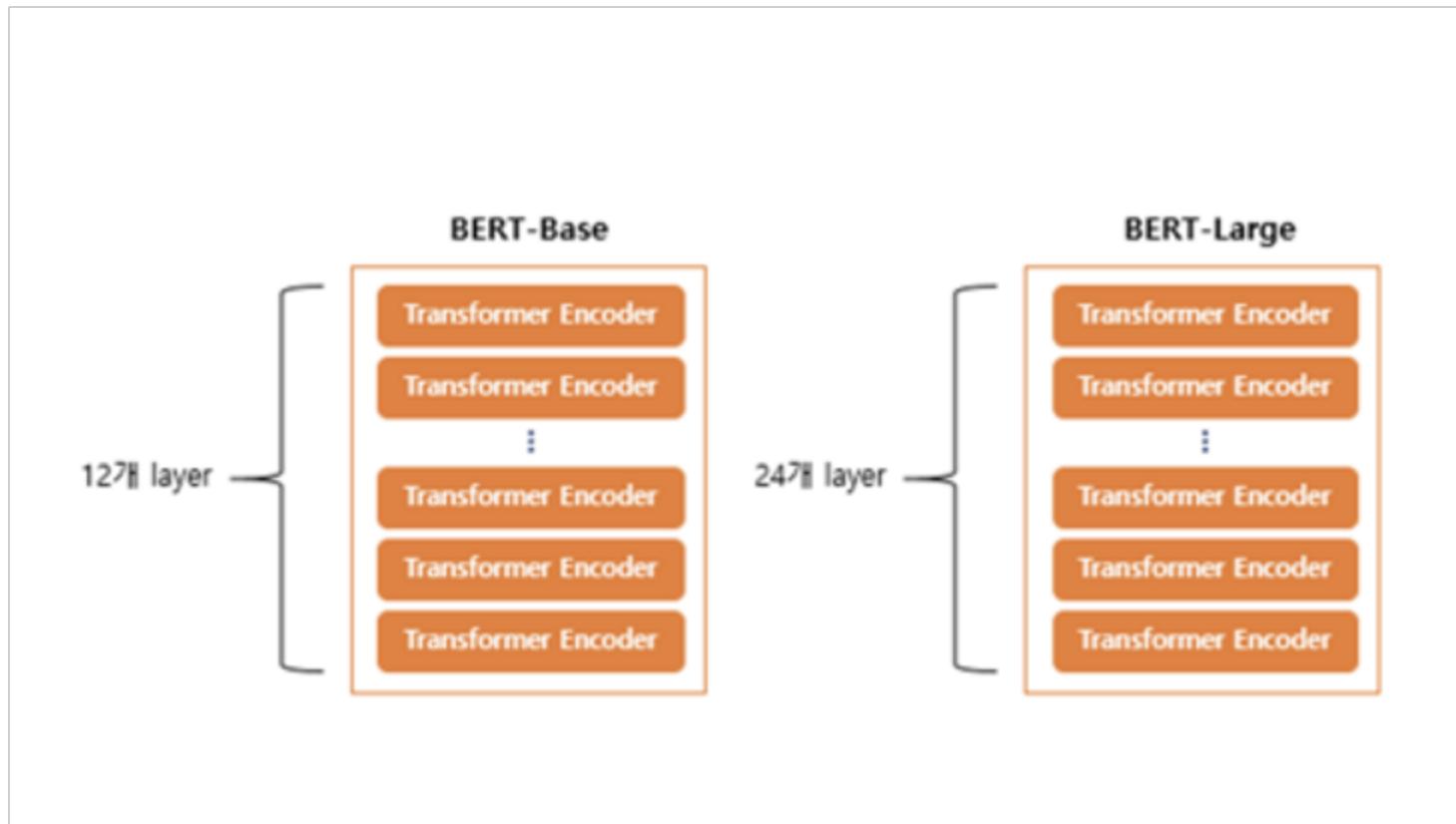
**관련 질문에 대한 답변,  
정답의 인덱스**

## 1. KorQuAD : 기존 기계 독해 데이터셋

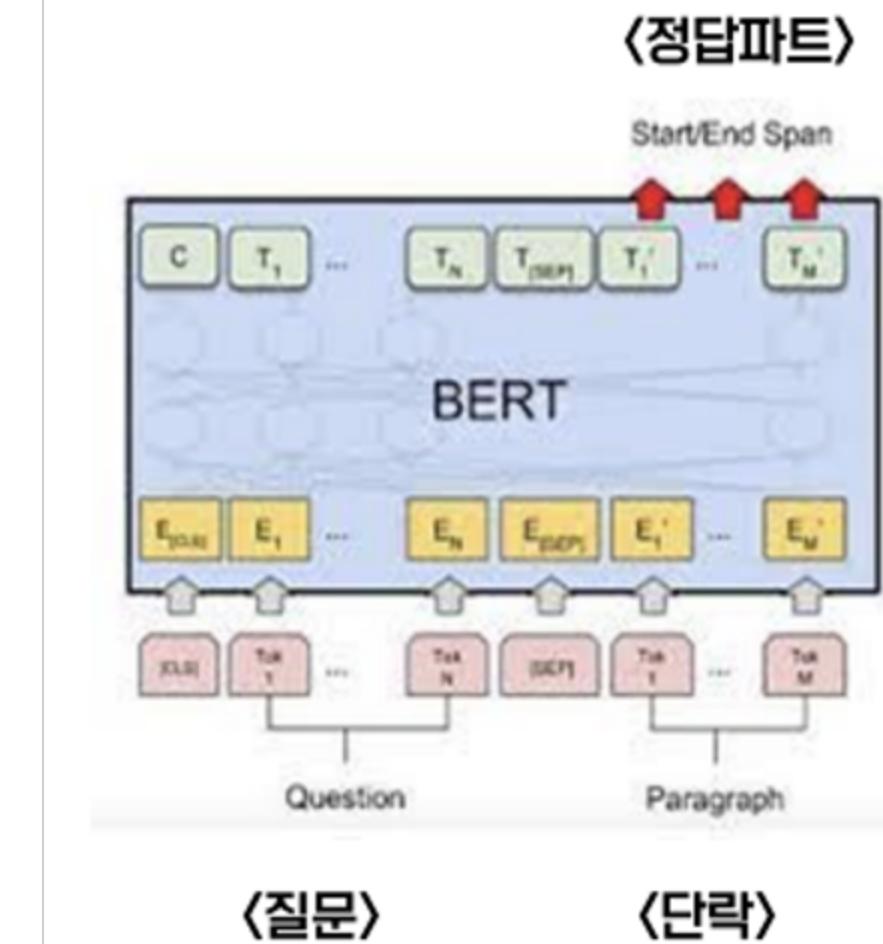
-> 데이터는 context, question, answer, answer start (정답의 인덱스), answer\_end (정답의 끝 인덱스)

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

# BERT Model \_ Model Description



Transformer encoder를 쌓은 구조인  
BERT 모델 사용



단락과 질문을 통해 정답을 출력하는  
BERT Q&A 모델 사용

# BERT \_ Evaluation, hyperparameter

## The F-score

The F-score is the Harmonic mean of Precision and Recall.

$$F = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

Alternatively

$$F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

## Evaluation)

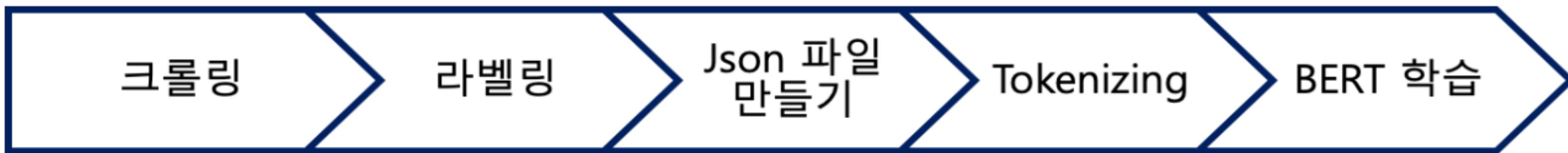
1. **F1 Score**: 예측한 정답과 실제 정답 간의 정밀도와 재현율의 조화 평균 값
2. **EM (Exact Match)** : 한국어 기준 정답의 어절 단위가 정확하게 일치하는 정도

## Hyper parameter)

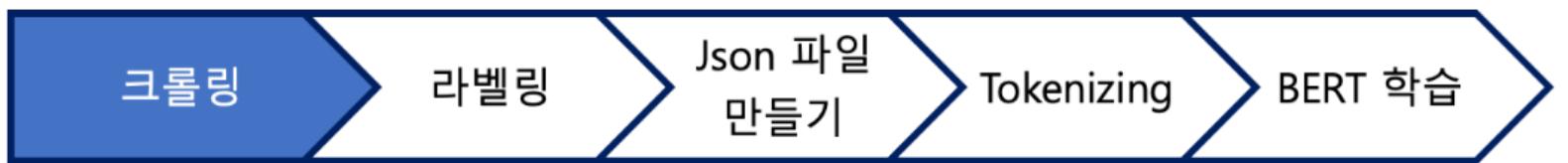
추후에 학습하면서 정할 예정



# AI 프로젝트 진행 과정



# AI 프로젝트 진행 과정



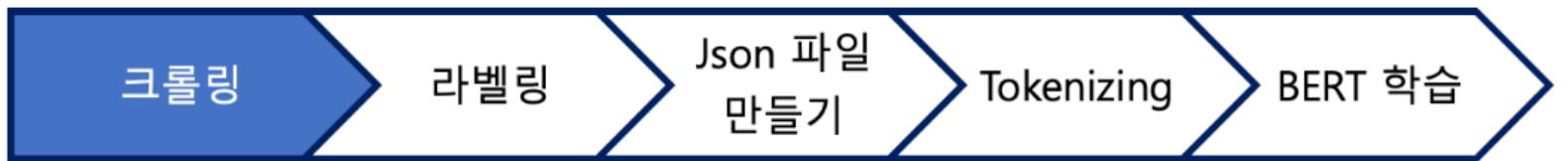
```
for page_num in crawling_number:
    #####수정#####
    url="https://ice.skku.edu/ice/community/notice.do?mode=list&&articleLimit=10&srCategoryId=816&article.offset="+str(page_num)
    #####
    headers = {'User-Agent':'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/83.0.4103.61 Safari/537.36'}
    html = requests.get(url, headers = headers)
    result = BeautifulSoup(html.content, "html.parser")

    post_numbers=[31,32,33,34,35]
    # 수집해야 하는 데이터 목록:
    # 1) number/고유번호: No.1806(중복 거르기용)

    #####
    post_urls=result.select("#jwxe_main_content > div > div > div > ul > li > dl > dt > a")
    post_numbers=result.select("#jwxe_main_content > div > div > div > ul > li > dl > dd > ul > li:nth-child(1)")
    #####
```

장학금 목록에서 장학금 공고의 url과 number(장학금 고유번호) 얻기

# AI 프로젝트 진행 과정



```

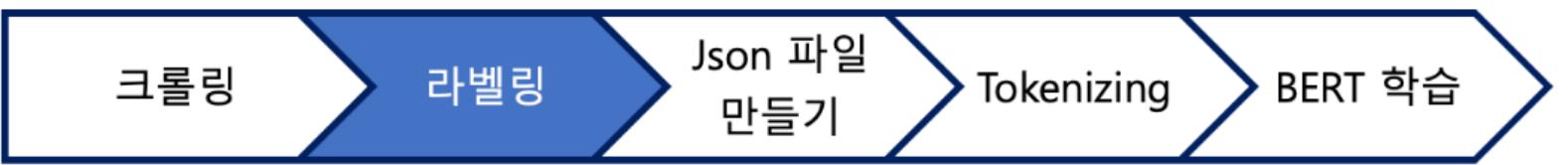
#1) 홈페이지 고유 number If using all scalar values, you must pass an index
number=post_numbers[i].text

# 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.get("href"))

```

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

# AI 프로젝트 진행 과정

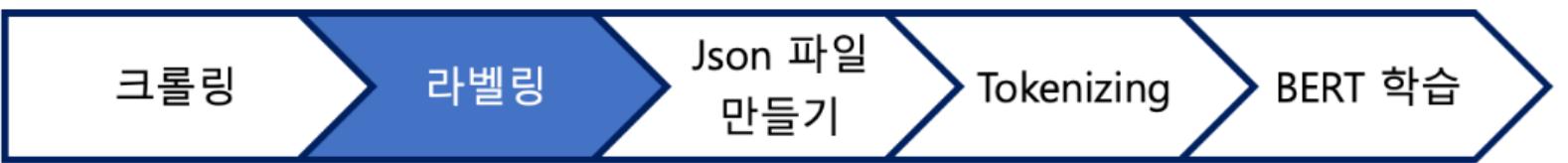


number	date	title	content	attachment_url	attachment_conf	current_url	department	question	text	answer_start	answer_end	attach_start	attach_end	삭제
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	직전학기 평점은 무엇인가요?										
No.1814	2022.09.30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

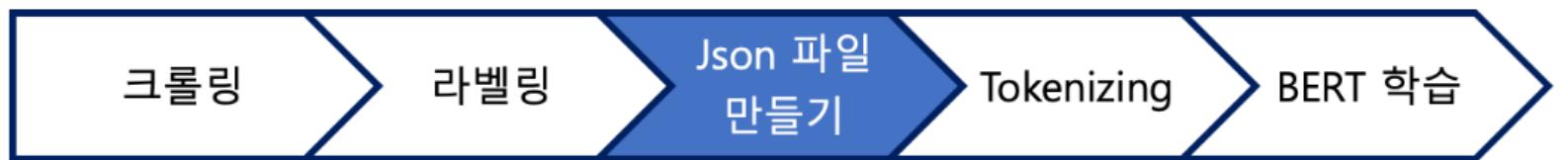
# AI 프로젝트 진행 과정



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 <a href="https://www.skku">https://www.skku</a>	학생지원팀	직전 학기 평점은 무엇인가요?										
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	전체 학기 평점은 무엇인가요?	13.8이상			287	292					
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	소득분위는 무엇인가요?	5분위 이내			224	230					
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	관련 학과는 무엇인가요?										
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	거주지는 무엇인가요?										
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	마감일은 언제인가요?	2022. 10.7			662	672					
No.1814	2022.09.30	2022학년도 2학: 유한재단은 유한양행 창설자이신 고 유일한 박사에 <a href="https://www.skku">https://www.skku</a>	학생지원팀	특이사항은 무엇인가요?	코로나19 및 폭우피			180	197					
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	지원 가능한 학년은 무엇인가요?										
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	직전 학기 평점은 무엇인가요?										
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	전체 학기 평점은 무엇인가요?										
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	소득분위는 무엇인가요?										
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	관련 학과는 무엇인가요?										
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	거주지는 무엇인가요?										
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	마감일은 언제인가요?	22.10.21			577	589					
No.1811	2022.09.30	2022학년도 서울장학재단은 <a href="https://www.skku">https://www.skku</a> 1111 2022년 서울 <a href="https://www.skku">https://www.skku</a>	학생지원팀	특이사항은 무엇인가요?										
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	지원 가능한 학년은 무엇인가요?										
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	직전 학기 평점은 무엇인가요?	13			359	362					
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	전체 학기 평점은 무엇인가요?										
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	소득분위는 무엇인가요?	①기초생활수급자			446	574					
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	관련 학과는 무엇인가요?										
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	거주지는 무엇인가요?										
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	마감일은 언제인가요?	2022.09.29			1364	1384					
No.1810	2022.09.30	2022학년도 하니 하나장학생은 하나금융나눔재단에서 실시하는 장학 <a href="https://www.skku">https://www.skku</a>	학생지원팀	특이사항은 무엇인가요?										

각 content에서 질문에 대한 답을 골라 start\_index와 end\_index를 고르기  
 → 총 256개(256개 게시물\*8쌍 질문=2048쌍)의 공고문 라벨링 완료

# AI 프로젝트 진행 과정



장학공고.CSV



Json\_parsing.py

```

article = []
answer = []

for row in range(int(len(data_rm)/5)) :
    nodata=False

    # article.append(len(ws.cell_value(row,2)))

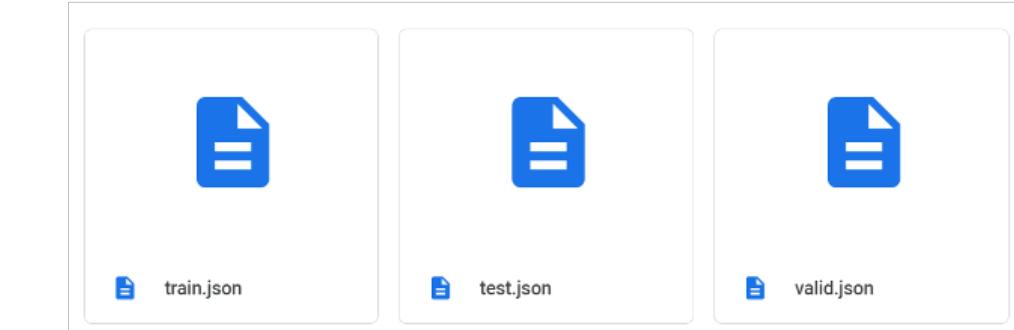
    for question in range(0,5) :
        row_num=row*5+question

        print(row_num)
        text = str(data_rm.loc[row_num,"text"])
        startNum = data_rm.loc[row_num,"answer_start"]
        endNum = data_rm.loc[row_num,"answer_end"]
        print(text)

        if (text=="nan") :
            nodata=True
            text=""
            start_index=str(data_rm.loc[row_num,"content"]+".").index(".")
            startNum=start_index
            print(start_index)
            endIndex=str(data_rm.loc[row_num,"content"]+".").index(text)+1
            endNum=end_index

            ansMap['answer_start'] = str(int(startNum))
            #ansMap['answer_end'] = endNum-1
            ansMap['text'] = text
            ansList.append(ansMap)
            ansMap = {}

        ansMap['answer_start'] = str(int(startNum))
        #ansMap['answer_end'] = endNum-1
        ansMap['text'] = text
        ansList.append(ansMap)
        ansMap = {}
  
```

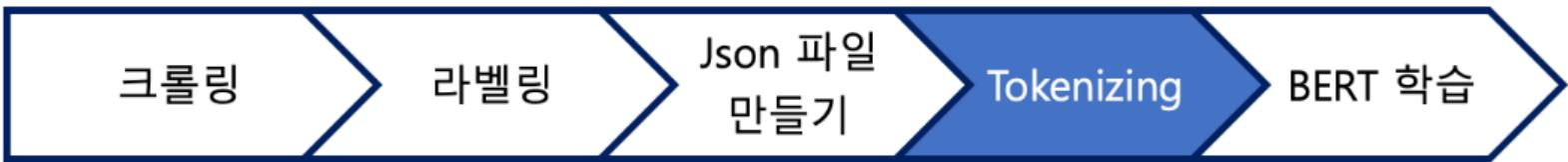


(train: 212개, valid : 26개, test : 27개)  
Train: valid: test = 8: 1: 1

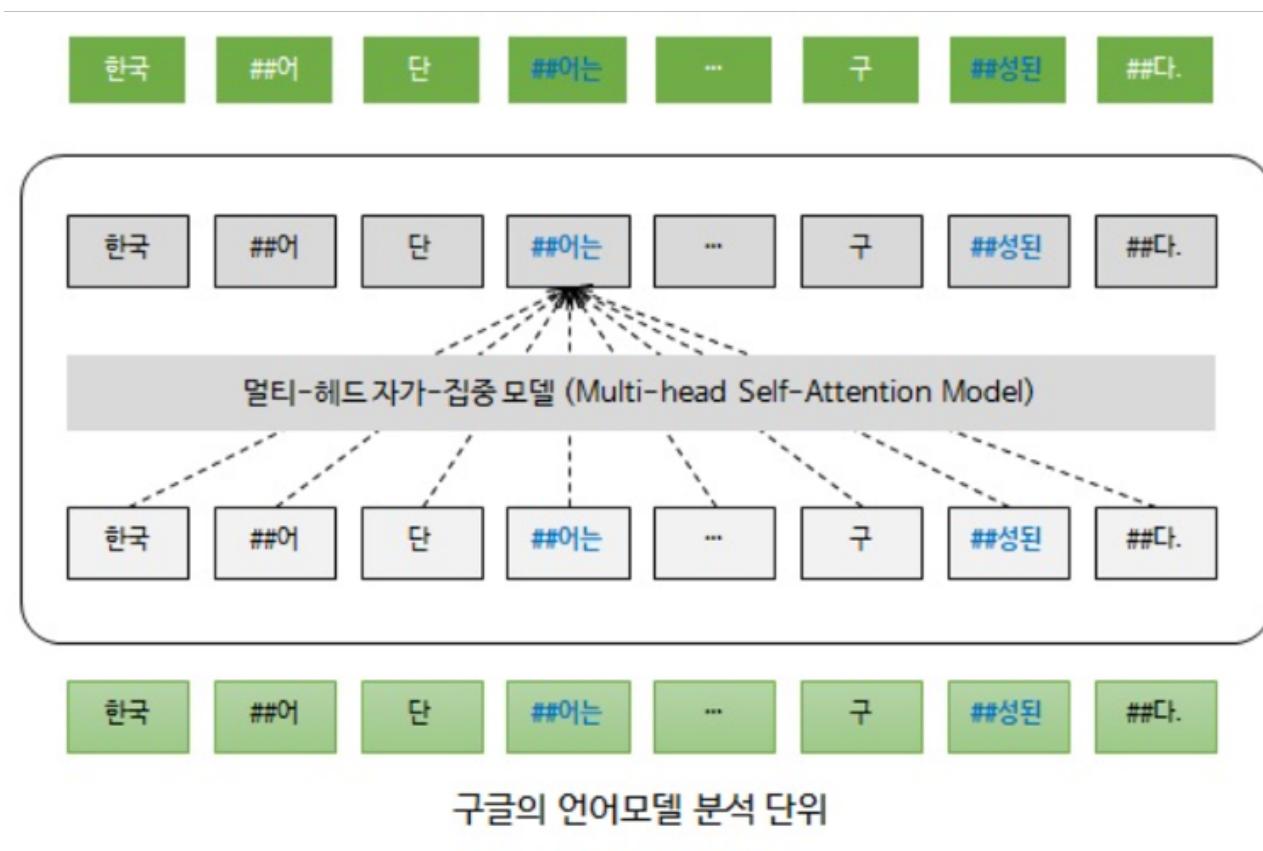
Limitation에 따라 학습데이터에 맞지 않는 데이터는 삭제하기  
위 json 파일 형식에 맞게 parsing 한 후, dataset split 진행

→ train.json, test.json, valid.json, total.json 만들어서 저장하기

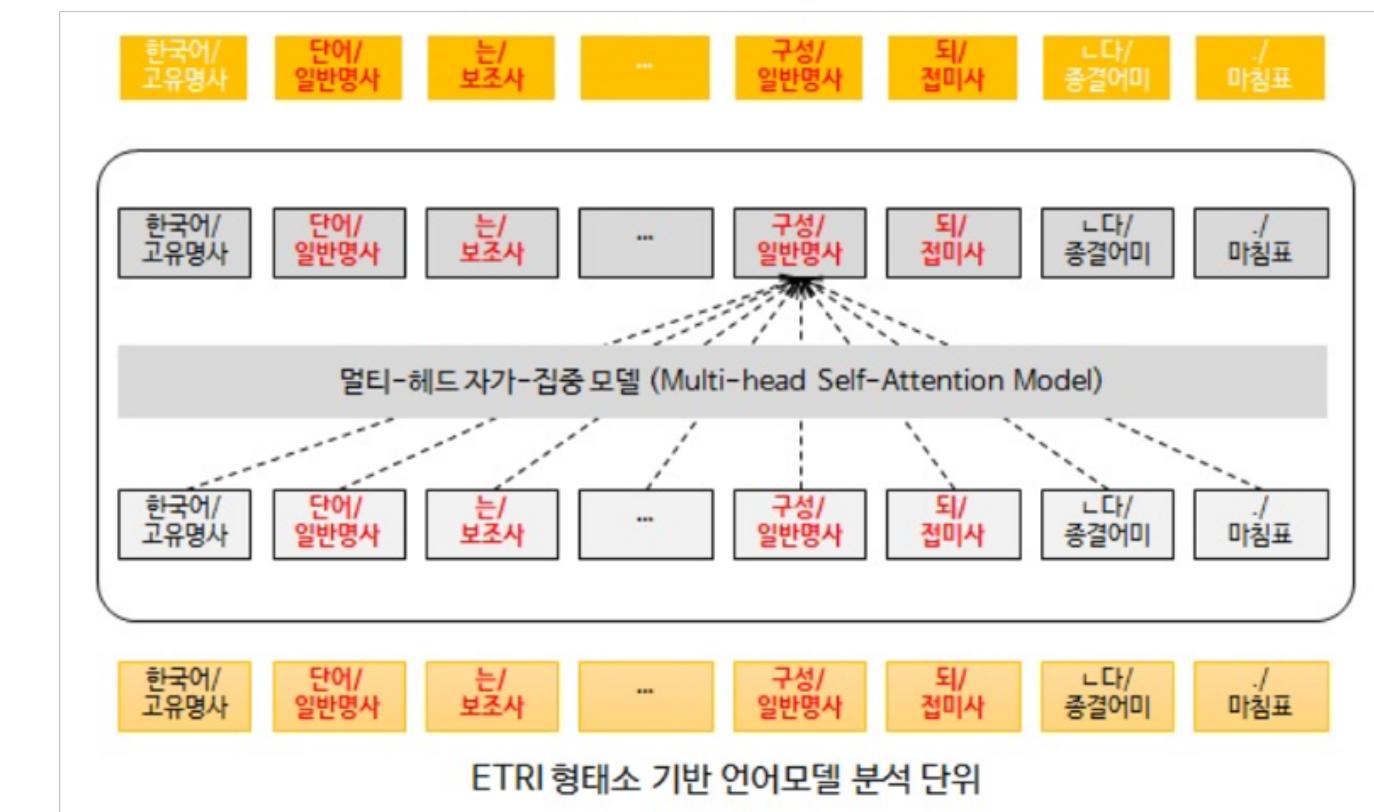
# AI 프로젝트 진행 과정



## 기존 BERT → wordpiece tokenizer

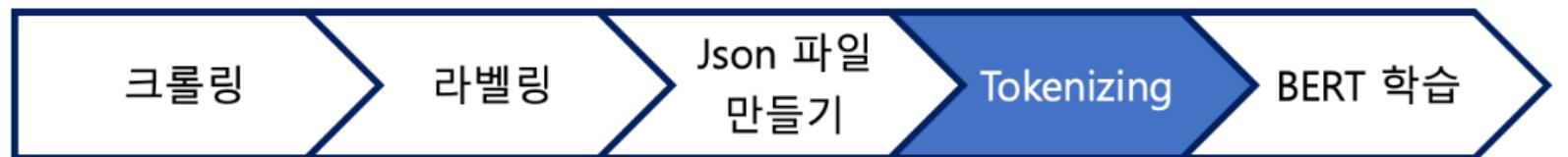


## ETRI koBert → ETRI 고유 tokenizer



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

# AI 프로젝트 진행 과정



```

def parser_add_argument(parser):
    ## Required parameters
    parser.add_argument("--openapi_key", default="79dd5e46-8977-4355-ae10-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/wiseNLU"

    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()
  
```

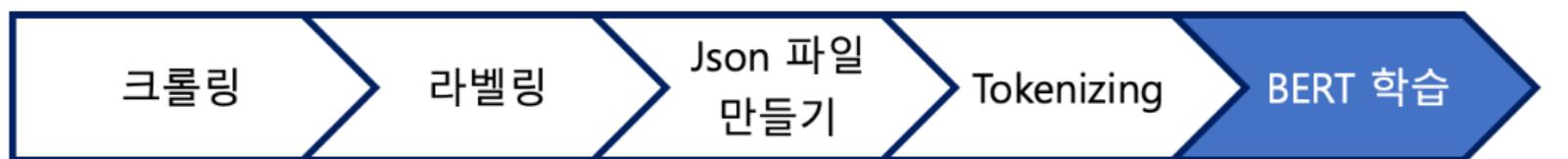
→

```

':
  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\":\\\"ㄴ 가
  \"id\":6.0,\"lemma\":\\\"요
  \\"id\":7.0,\"lemma\":\\\"?
  \":[],\"word\":[{\\"id\":0.0,\"text\":\\\"관련
  xt\":\\\"학과는
  xt\":\\\"무엇인가요?
  Link\":[],\"dependency\":[],\"SRL\":[[]]],\"entity\":
  
```

ETRI의 tokenizing API를 request하여 입력될 문장들의 형태소를 미리 tokenize 진행.  
**Argument : api key, input file : total.json, output file : 형태소 분석 완료한 json 파일**

# 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

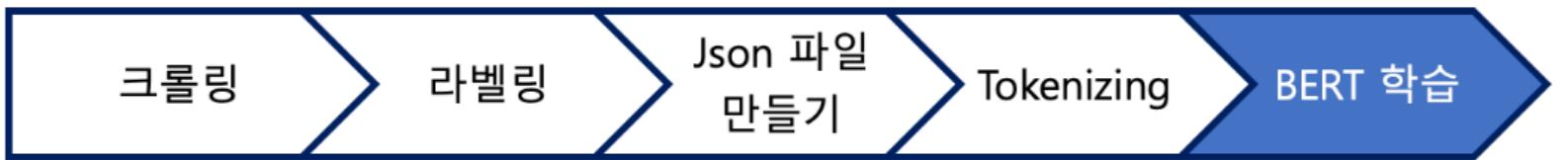
Epochs: 5

Learning\_rate: 3e-5

Batch\_size: 4

BERT train 진행. (Hugging face의 QA 모델 finetuning 코드를 참고)

# AI 프로젝트 진행 과정



## [학습 시 Embedding 방법]

```
["id":0.0,"lemma":"관련\n\"학과\n":\n:"\n는\n:\n무엇\n:\n\"이\n:\n\"ㄴ가\n:\n\"요\n:\n?\n"]
```

문장 토큰

이/JKS_	8938636
,/SP_	8614178
는/JX_	8594108
를/JKO_	8310256
어/EC_	7491203
은/JX_	7238132
는/ETM_	6614467
고/EC_	6544461
가/JKS_	5607159
/SS_	5290740
'/SS_	5184635

Vocab.korean\_morp.list

```
*** Example ***
unique_id: 1000000008
example_index: 8
doc_span_index: 0
tokens: [CLS] 지원/NNG_ 가능/NNG_ 하/XSA_ ㄴ/ETM_ ㅎ
token_to_orig_map: 13:0 14:0 15:1 16:2 17:3 18:4 19
token_is_max_context: 13:True 14:True 15:True 16:True
input_ids: 2 415 226 42 10 2348 21 1381 15 1657 461
input_mask: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
segment_ids: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
```

임베딩

문장의 각 token들을 vocab.Korean\_morp.list에 따라 임베딩 진행.



# Modeling Result



# BERT Modeling\_result

```
{  
    "exact": 73.5576923076923, "f1":  
        74.99313186813187, "total": 208  
}
```

```
"422": "7월 19일",  
"423": ".,"  
"1984": ".,"  
"1985": "2.75점 이상",  
"1986": ".,"  
"1987": ".,"  
"1988": ".,"  
"1989": ".,"  
"1990": "06월 17일",  
"1991": ".,"  
"544": ".,"  
"545": "2.75점 이상",  
"546": ".,"  
"547": "기초생활수급자/차상위계층: 1.88점 이상",  
"548": ".,"  
"549": ".,"  
"550": "2022년 06월 23일",  
"551": ".,"  
"920": "3학년 1학기",  
"921": ".,"  
"922": "4.0",  
"923": ".,"  
"924": ".,"  
"925": ".,"  
"926": "02. 25",
```

Valid set 26개(단락-질문 208개쌍)으로 test를 진행한 결과, **EM은 73%, F1-score는 74%**

# 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(시간영수) 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",

# BERT Modeling\_Bad example

```
"768": "3.00",
"769": "3.00점 이상",
"770": "3.00",
"771": "3.00",
"772": "3.00",
"773": "3.00",
"774": "3.00",
"775": "3.00",
"1864": ".",
"1865": ".",
"1866": ".",
"1867": ".",
"1868": ".",
"1869": ".",
"1870": ".",
"1871": ".",
"1872": "
```

실제 정답이 없는 것들은 문장 내의 “.”을 찾아 정답 태깅을 하고 학습을 진행하였는데, 그 과정에서 “평점 3.00”과 같이 유의미한 정보를 담고 있는 곳에서 “.”이 뽑혀서 학습된 경우 존재.

정확한 원인은 파악하지 못하였지만, prediction 시 그러한 “.”을 뽑아내려다가 양방향 학습인 bert 특성상 유의미한 정보인 3.00 도 통째로 뽑히지 않았나 싶음.

→ “.” 말고 다른 기호로 “정답 없음”을 태깅할 필요 존재.



AI  
Trial & Error

---

# 크롤링\_공지사항 구성

### 3. AI - Trial & Error

```
> <thead>...</thead>
<tbody>
  <tr>
    <td>
      <div class="boardView_txtWrap">...</div>
      <dl class="board-write-box board-write-box-v03">
        <dt class="hide replyNone">게시글 내용</dt>
        <dd>
          <pre class="pre"> == $0
            "교류부 국립국제교육원의 해외 석박사 학위과정 지원 장학
            을을 협부와 같이 안내합니다. ** 주의 사항 ** 1. 필수 제
            서류 중 '응시원서'의 출장 직인은 졸업한 단과대학의 謹정
            에 요청하시기 바랍니다. 2. 본 장학사업에 대한 문의는 국
            국제교육원 국비유학금 담당자에게 연락하시기 바랍니다. 02-
            3668-1374/1375 kscholar@korea.kr 이상입니다. 국제처 국
            교류팀"
        </pre>
      </dd>
    </td>
  </tr>
</tbody>
```

```
    <span style="font-family: Dotum;">
        <span style="font-family: Dotum;">
            <p>
                <span style="font-family: 'Dotum'; font-size: 10pt;">경
                    학과 60학번 졸기회에서 경제학과 및 글로벌경제학과 학생들
                    학습의욕 고취와 </span>
            </p>
        <p>
            <span style="font-family: 'Dotum'; font-size: 10pt;">원
                한 학업수행률 위하여 다음과 같이 장학생을 선발하고자 합니
            </span>
        </p>
        <p>
            <span style="font-family: 'Dotum'; font-size: 10pt;">여
                분들의 많은 신청 바랍니다.</span>
        </p>
        <p>
            <span style="font-family: 'Dotum'; font-size: 10pt;">&nbsp;</span>
        </p>
```

## <학교 전체 공지>

<경제대 공지>

<경영대 공지>

전체 공지사항과 단과대 공지사항의  
Html 구조가 다름  
→ <dd>에 있는 text와 <p>에 있는 text 따로 추출

경영대학 등 일부 대학은 JS에서 selector 값이  
보이는 것과 다르기 때문에 html에서  
직접 요소를 select 하는 방식을 취함

# Json 파일 \_ 라벨링 수정

```

    "title": "2022학년도 우덕재단 대학원 장학생 선발",
    "paragraphs": [
        {
            "context": "1. 장학금 내용 - 당 재단 장학금은 '등록금'에 한하여 지원함 - 2022학년도 1년간 우덕장학생 지원 유
- 최대 2명 예정2. 지원대상자(1) 건축 재료공학 건설환경 학과(2) 가정환경이 어렵지만 학업 성적 등이 우수한 학생(3) 2학기 이상 남은 자(2022-1
기 포함)3. 제출서류-장학생 추천 및 추천서 (당 재단 소정 양식 충장날인 제외하고 모두 작성해서 제출)-자기소개서 (당 재단 소정 양식)-개인정보 제공 및 활용 동
서-보호자(주 소득자) 건강보험료 납입증명서-주민등록등본-재학증명서-성적증명서 (직전학기 교환학생인 경우 P/F 증명 포함)-등록금 납입증명서4. 지원마감 및 제출
1) 지원마감 : 3/29(화) 15시 (시간엄수) 2) 제출처 -이메일 : skkujanghak@skku.edu -직접제출 : 인사캠 600주년기념관 1층 학생지원팀 3) 문의 : 02-760-
67.",
            "qas": [
                {
                    "question": "지원가능한 학년은 무엇인가요?",
                    "answers": [
                        {
                            "answer_start": "138",
                            "text": "2학기 이상 남은 자"
                        }
                    ],
                    "id": "704"
                },
                {
                    "question": "직전학기 평점은 무엇인가요?",
                    "answers": [
                        {
                            "answer_start": "1",
                            "text": "."
                        }
                    ],
                    "id": "705"
                }
            ]
        }
    ]
}

```

## 1. 통일성 부여

ex) 직전학기 성적은 무엇인가요?

A1. 직전 학기 3.0이상

A2. 3.0이상

→ 모두 “3.0 이상” 으로 바꿔줌

## 2. 라벨링 오류 수정

index 라벨링이 잘못 기입된 경우 존재,
학습에 오류가 생기기 때문에 수정

# Json 파일 \_ 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

# Json 파일 \_ 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 학자금의 2021년 상반기 ('21. 1. 1. ~ 6. 30.) 발생이자 라. 신청방법 : 온라인 또는 방문신청 - 온라인신청 : 군산시청 홈페이지 (https://www.gunsan.go.kr) - 방문신청 : 군산시청 교육지원과(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"
            },
            ...
          ]
        }
      ]
    }
  ]
}
```

1. Unanswered question에 대한 answer을 '.'으로 지정하고 start\_index에는 '.'의 임의의 위치를 찾아서 저장
2. 공고에 '.'이 없는 paragraph에는 마지막에 임의로 .을 넣어서 답을 지정하고 index 수집

# Tokenizing \_ API Connection 에러

```
WARNING:urllib3.connectionpool:Retrying (Retry(total=2, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError('<urllib3.connectio
WARNING:urllib3.connectionpool:Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError('<urllib3.connectio
WARNING:urllib3.connectionpool:Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'NewConnectionError('<urllib3.connectio
-----
TimeoutError
Traceback (most recent call last)
/usr/local/lib/python3.7/dist-packages/urllib3/connection.py in _new_conn(self)
    158         conn = connection.create_connection(
--> 159             (self._dns_host, self.port), self.timeout, **extra_kw)
    160
-----  
 19 frames -----  
TimeoutError: [Errno 110] Connection timed out  
  
During handling of the above exception, another exception occurred:  
  
NewConnectionError
Traceback (most recent call last)
NewConnectionError: <urllib3.connection.HTTPConnection object at 0x7f6c87befcd0>: Failed to establish a new connection: [Errno 110] Connection timed out
```

1. API connection 에러 원인 분석 : 방화벽, json 파일 오류 등등 > local로 돌리면서 해결
2. 1번으로 해결이 되지 않을 경우를 대비해서 Tokenizing 코드 분석도 진행.

# Brief Schedule

---

Based on the roles that were largely divided into two categories, we planned the project schedule by dividing it into AI and app parts. The schedule may vary depending on the situation for more efficient project progress.

**Table 2.** weekly project schedule

weekly plan														
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Overall	Project Proposal	Proposal Latex							Midterm			Final Proposal	Finalterm Latex	
AI			data collection data preprocessing	Rule-Based		Q&A Model (BERT)								
App			UI/UX design Database setting	frontend work		backend work connect frontend and backend		Beta Test Debugging						

---

# Q & A

지금까지 TEAM. SKKULAR 였습니다.  
감사합니다.