SKKU SCHOLARSHIP WEEKLYPROGRESS MEETING (2)

TEAM 스꿀라, SKKULAR 강병준 I 김규진 I 박진아 I 장이준 I 주재현

CONTENTS

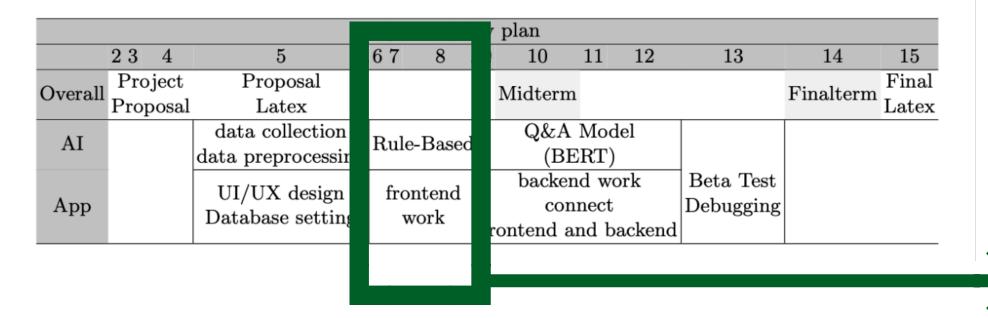


TEAM. 스골라 SKKULAR

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



(a): Al (b): Back-end (f): Front-end

week 6 1) APP: UI/UX design 보완(f) Database setting(b)

2) Al:
Data preprocessing (a)
Data labeling (a,b,f)



DB 설계 I MySQL(RDBMS)



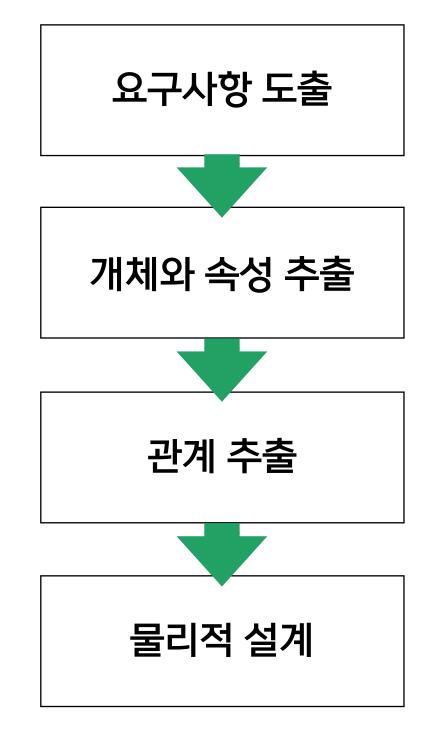


명확하게 정의 된 스키마, 데이터 무결성 보장 관계는 각 데이터를 중복없이 한번만 저장됨 변경될 여지가 없고, 명확한 스키마가 사용자와 데이터에게 중요한 경우

DB 설계 I MySQL

Schema

- □ Entity ⇒ Table
- O Attribute → Column
- ♦ Relation ⇒ PK, FK
 - Tuple → Row



DB 설계 I 요구사항 도출

유저
user_id
pw
학기
직전학기 평점은 무엇인가요?
전체학기 평점은 무엇인가요?
소득분위는 무엇인가요?
학과는 무엇인가요?
거주지는 무엇인가요?

장학금
지원가능한 학년은 무엇인가요?
직전학기 평점은 무엇인가요?
전체학기 평점은 무엇인가요?
소득분위는 무엇인가요?
관련 학과는 무엇인가요?
거주지는 무엇인가요?
마감일은 언제인가요?
특이사항은 무엇인가요?

DB 설계 I 요구사항 도출

키워드
user_id
키워드
index

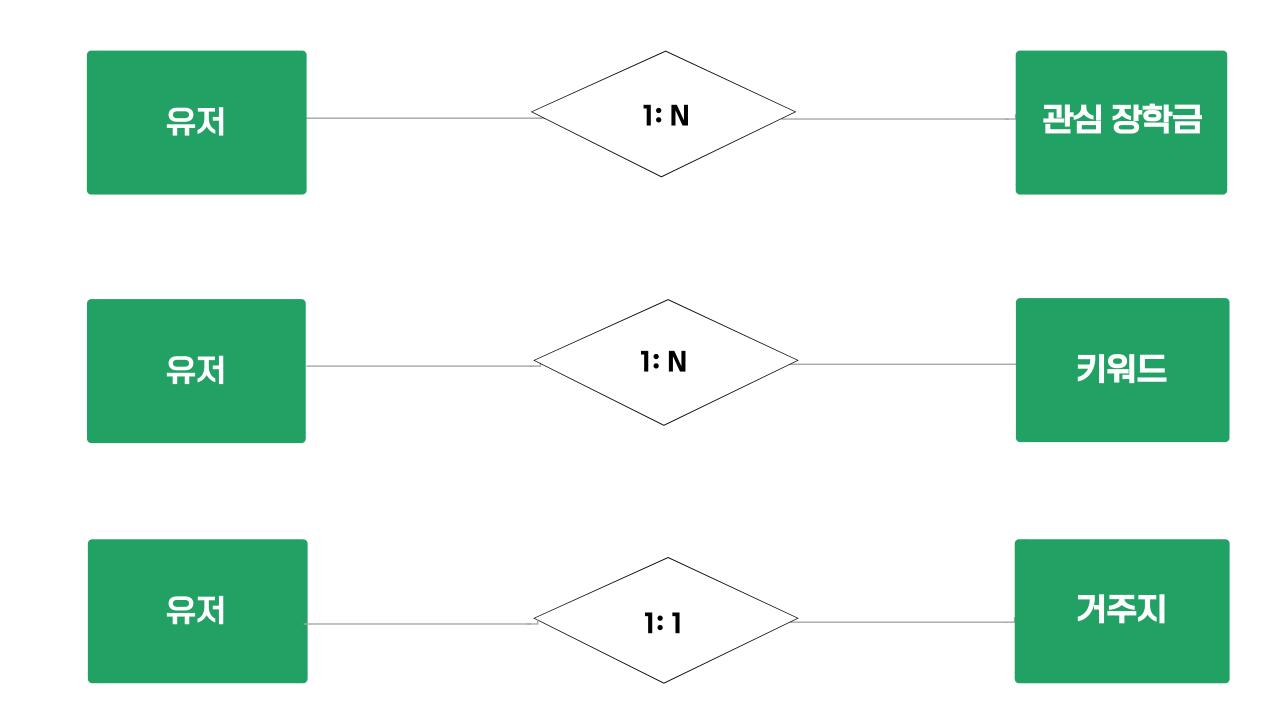
관심 장학금
user_id
장학금
index

거주지(?)
user_id
시, 군, 구
index

DB 설계 I 개체와 속성 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, user_id, keyword
관심 장학금	user_id, index, 장학금
거주지	use_id, (시,군,구), index

DB 설계 I 관계 추출



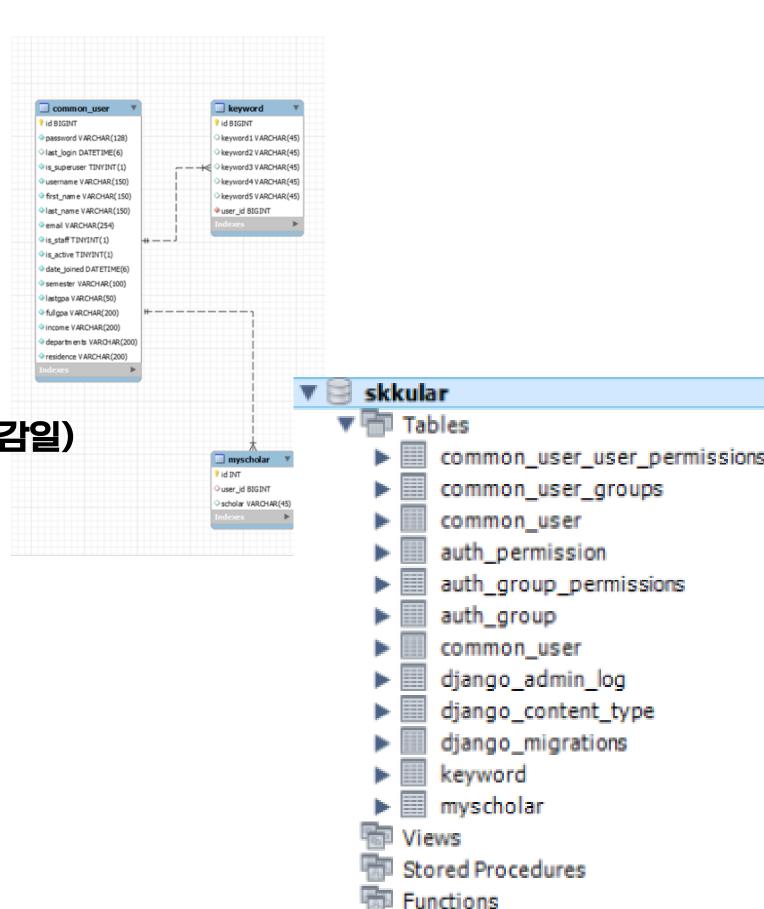
DB 설계 | Relation 변환 및 물리적 설계

회원(ld(pk), pw, 학기, 직전학기 평점, 전체 평점, 소득분위, 학과, 거주지)

장학금(학기, 직전학기 평점, 전체 평점, 소득분위, 학과, 거주지, 특이사항,마감일)

키워드(Index(pk),user_id(fk), keyword)

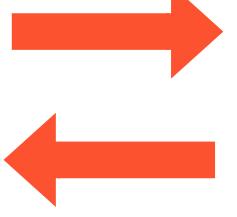
관심 장학금(index(pk), user_id(fk), 장학금)



Model View Controller Template

Model



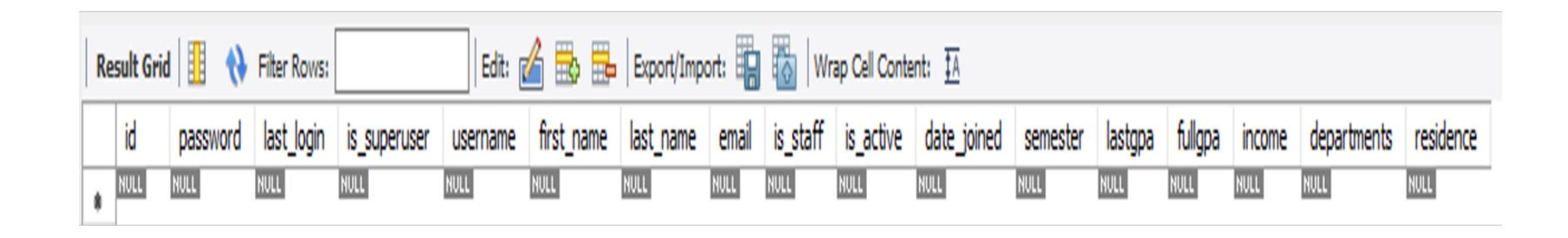




ORM

```
from django.db import models
from django.contrib.auth.models import AbstractUser
# Create your models here.
class User(AbstractUser):
    semester = models.CharField(max_length=100)
    lastgpa = models.CharField(max_length=50)
    fullgpa = models.CharField(max_length=200)
    income = models.CharField(max_length=200)
    departments = models.CharField(max_length=200)
    residence = models.CharField(max_length=200)
```

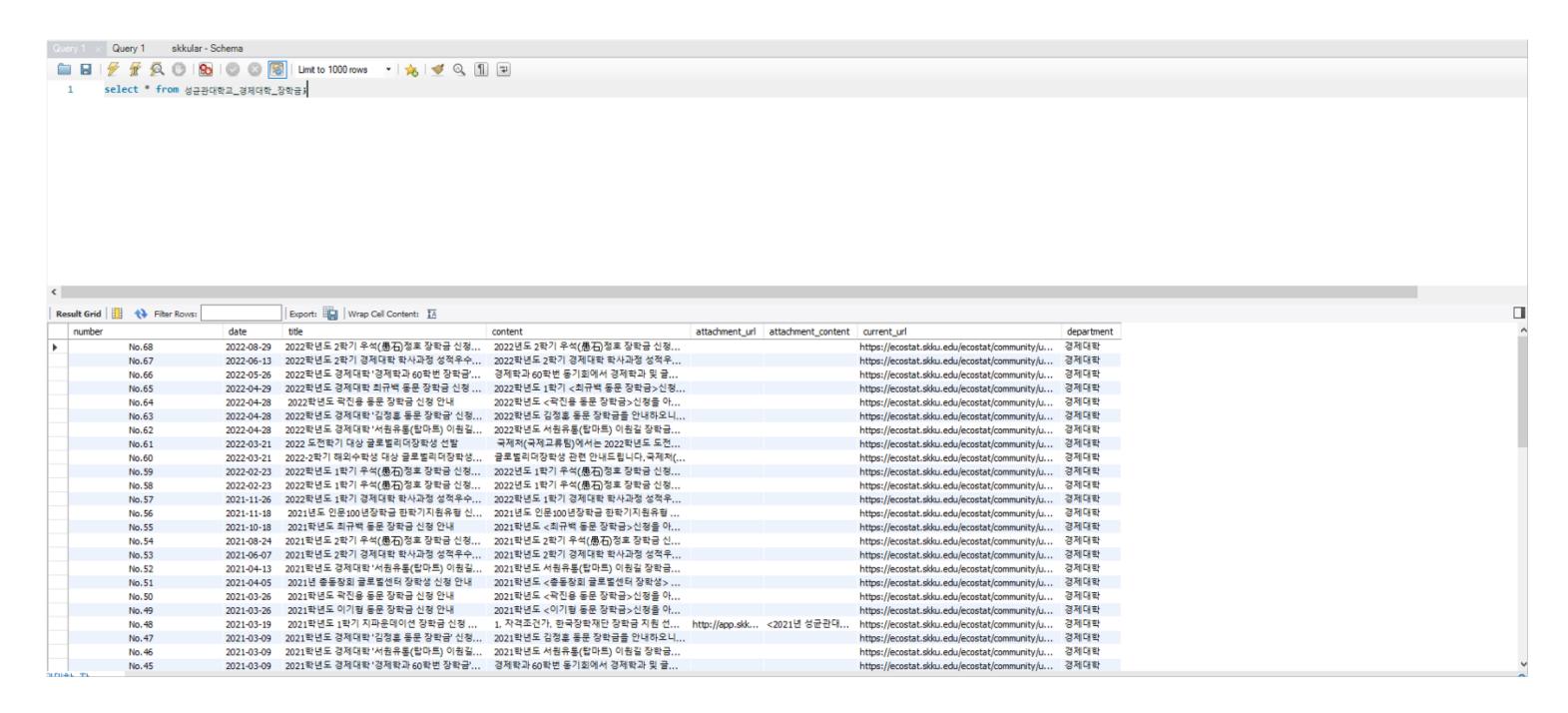
AbstractUser를 이용하여 Django의 로그인 기능을 사용하면서 User Model의 칼럼(데이터)들을 수정



Make migrations 명령어 이용하여 모델의 변경사항 기록

Migrate 명령어 이용하여 기록한 변경사항을 DB 적용을 위해 migrate 명령어 사용

DB 설계 I csv 파일 저장



Csv 파일 mysql 에 저장



크롤링 결과

CSV 파일로 저장

number	date	title	content	attachment_url	attachment_content	current_url	department
No.1805	2022.09.20	2022학년도 2학기 DE	당 재단은 DB그룹(구	동부그룹) 설립자 김준	기 회장이 기금 출연하여	https://www.skku.edu/	학생지원팀
No.1804	2022.09.16	2022학년도 (재)광산경	(재)광산장학회에서는	우리 지역의 우수한 인	재를 발굴·육성하고 학·	https://www.sku.edu/	학생지원팀
No.1803	2022.09.16	2022학년도 2학기 안	안산인재육성재단에서	서는 교육비 부담 경감을	통한 지역 인재육성 및	https://www.skku.edu/	학생지원팀
No.1802	2022.09.16	2022학년도 2학기 고	고양시에서는 대학생	https://www.skku.edu/	본문내용 입력 (2022년	https://www.skku.edu/	학생지원팀
No.1801	2022.09.15	2022학년도 2학기 63	2022년도 2학기 63점	학금 장학생은 선발하	고자 하오니 아래를 참고	https://www.skku.edu/	학생지원팀
No.1800	2022.09.14	2022학년도 하반기 (기	1. 재단 안내 ㅇ명 🤞	https://www.skku.edu/	1111 재단법인 은평구	https://www.skku.edu/	학생지원팀
No.1799	2022.09.14	2022학년도 하반기 (기	가. 선발인원: 26명(성	https://www.skku.edu/	skku/campus/skk_com	https://www.skku.edu/	학생지원팀
No.1798	2022.09.06	2022학년도 수림재단	수림재단에서 설립자	인 동교 김희수 전 이사	장의 교육이념은 받들어	https://www.skku.edu/	학생지원팀
No.1797	2022.09.06	2022학년도 2학기 지	본 재단은 국내 디자인	민 예술 산업 발전에 믿기	H름이 될 인재육성을 위	https://www.skku.edu/	학생지원팀
No.1796	2022.09.06	2022학년도 일문과학	1.선발분야 및 규모 :	0명내외 가.전공제	한 없이 지원 가능 다	https://www.sklu.edu/	학생지원팀
No.1795	2022.09.06	2022학년도 서울장학	서울장학재단에서는	https://www.skku.edu/	2022년 <서울희망 디	https://www.sku.edu/	학생지원팀
No.1794	2022.09.06	2022학년도 광주광역	1. 지원자격 : 공고일	https://www.skku.edu/	재 (재)광주광역시 서-	https://www.skku.edu/	학생지원팀
o.1793	2022.09.05	2022학년도 하반기 &	■ 신청기간: 2022. 9.	26.(월) ~ 9. 30.(금) (5	일간)■ 접 수 처: 애향장	https://www.skku.edu/	학생지원팀
No.1792	2022.09.02	2022학년도 제 25기	「두울장학재단」은 삼	성그룹의 창업주 故이병	철 선대회장을 내조하셨	https://www.skku.edu/	학생지원팀
No.1791	2022.09.02	2022학년도 경주시장	1. 선발 인원 및 신청기	자격 🗆 선발인원 - 총 45	0명(일반전형: 423명 =	https://www.sku.edu/	학생지원팀
No.1790	2022.09.02	2022학년도 포항시장	1. 신청기간 : 2022. 0	08. 24(수) 09:00 ~ 202	2. 09. 20(화) 18:002.	https://www.sku.edu/	학생지원팀
No.1789	2022.09.02	2022학년도 인송문화	'仁松文化財團'이 학양	업에 뜻이 있고 바른 사회	J관을 가진 대학생으로	https://www.sku.edu/	학생지원팀
No 1788	2022 08 31	2022한녀드 2한기 도	도사작하레느 1995년	[선린되어 매년 결제점	이거우이 이느 하색트(https://www.ekki.edu/	하세기의티

- Number : 공지사항 번호

- Date : 공지사항 작성 날짜

- Title : 공지사항 제목

- Content : 공지사항 내용

- Attachment_url : 첨부파일 url

- Attachment_content : 첨부파일 내용

- Current_url : 공지사항 url

- Department : 공지사항 부서

학교 공식 홈페이지 공지사항은 약 1,000개, 단과대 홈페이지 공지사항은 대략 30개씩 크롤링 진행

라벨링 방식

STX장학재단에서 2023년도 신규 장학생을 선발합니다. 당 재단은 창의와 도전정신이 투철한 인재들의 지원 및 육성을 목적으로 2006년에 설립된 공익 법인입니다.

- 1. 선발대상 및 인원
- □ 선발대상 : 국내 국.공립, 사립 종합 4년제 대학교 3학년 1학기(5학기) 진급예정자(4학기 재학생 혹은 4학기 이수한 복학예정자
- □ 선발인원 : 총 20명 이내
- □ 지원자격
- 1) 성적: 학점평균 B+이상 (3.3/4.3 이상 또는 3.5/4.5 이상)
- 2학년 1학기까지 평점평균(3학년 1학기 복학예정자는 2학년 2학기까지)
- 2) 선발계열
- 전계열
- * 예체능계, 의대, 약대, 간호대, 사범대 등 특수.전문계열 및 전공 제외
- 2. 장학생 지원사항
- □ 금액: 학기당 등록금 300만원 한도 + 학업보조금 월 50만원(1인 연간 최대 1,200만원 지원)
- □ 기간 : 최대 2년까지(조기 졸업 시 지원종료)
- * 매학기 성적 및 결격사유 등을 심사하여 계속 지원여부 결정
- 3. 선발일정
- □ 접수기간: 2022년 11월 7일(월) 오전 10시~11월 21일(월) 오후 5시까지 (온라인 접수에 한하며, 접수기간 내에만 지원서 양식 확인 가능)

Content 및 Attach_content를

- 지원가능한 학년
- 요구되는 직전학기 평점
- 요구되는 전체학기 평점
- 소득분위
- 관련 학과
- 거주지
- 마감일
- * 특이사항 ex) 교환학생, 탈북민

각 팀원별 400문제씩 담당하기로 함 -> 약 1500개의 학습데이터 생성

BERT Q&A 학습 데이터

Q. 지원 가능학 학년은 무엇인가요?

A. 5개~8개 학기 - answer_start = 147, answer_end = 155

Q. 직전학기 평점은 무엇인가요?

A. 4.0 이상 (첨부파일 내용에 있음) - attach_start = 1437, attach_end = 1442

답의 위치를 파악하기 위한 코드

```
● 1 import pandas as pd 2 import numpy as np

[] 1 def find_start_end_index(content, keyword):
2 try:
3 start_index=content.index(keyword)
4 end_index=start_index+len(keyword)
5 print("찾는 문자열: "+content[start_index:end_index])
6 except:
7 print("해당하는 문자열이 없습니다! 다시 입력해주세요(띄어쓰기, 기호 등의 오류가 없는지 다시 확인해보자구요(/④9④)/*: (*◇)")
8 return 0,0
9 return start_index, end_index

[] 1 temp_str = "□ 지원대상 : 공고일(2018.09.17.) 기준 현재 부모 또는 학생이 김제시에 주민등록이 되어 있는 대학 재학생□ 지원내용 : '16.1월 ~ '18.9월에 한국장학재단에서 함

[] 1 find_start_end_index(temp_str,"학자금")

찾는 문자열: 학자금 (105, 108)
```

- 1. CSV에서 전체 장학금 공고 내용 전처리,
- 2. Content: 장학금 공고 내용, Keyword: 찾으려는 답
- 3. Str의 index 함수를 이용하여 start 위치 찾음
- 4. End위치는 start 위치 + keyword의 길이

라벨링 결과

1	date	title	content		attachmer	nt_url	attachment	_cont current_url	department	question	text	answer_	st answer_	e attach_	s attach_
450	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한호	학기지원유형	신규 <u>https://scos.skku</u>	유학대학	지원가능한 학년은 무엇인가요?	5개~8개 학기	150	0 15	8	
451	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한호	막기지원유형	신규 https://scos.skku	유학대학	직전학기 평점은 무엇인가요?	3.6 이상	34	1 34	7	
452	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한호	막기지원유형	신규 https://scos.skku	유학대학	전체학기 평점은 무엇인가요?					
453	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한호	학기지원유형	신규 <u>https://scos.skku</u>	유학대학	소득분위는 무엇인가요?					
454	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한학	학기지원유형	신규 <u>https://scos.skku</u>	유학대학	관련 학과는 무엇인가요?	유학대학	16	5 16	9	
455	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한호	탁기지원유형	신규 <u>https://scos.skku</u>	유학대학	거주지는 무엇인가요?					
456	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한학	학기지원유형	신규 <u>https://scos.skku</u>	유학대학	마감일은 언제인가요?	2021년 11월 29일	74	4 76	0	
457	2021-11-23	(유학대학) 2021	제목없음	2021년도 인	민문100년장학	금 한호	학기지원유형	신규 <u>https://scos.skku</u>	유학대학	특이사항은 무엇인가요?					
458	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 증) https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	지원가능한 학년은 무엇인가요?					
459	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 장) https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	직전학기 평점은 무엇인가요?	4.0이상			1437	144
460	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 증) https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	전체학기 평점은 무엇인가요?					
461	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 장) https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	소득분위는 무엇인가요?	중위소득 65% 이	하		1543	155
462	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 장) https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	관련 학과는 무엇인가요?					
463	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 장) https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	거주지는 무엇인가요?	광주광역시 서구		3 1	1	
464	2020.10.05	(재)광주광역시	(재)광주광	}역시 서구 장	남학 [;] <u>https://ww</u>	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	마감일은 언제인가요?	2020. 10. 16(금)			745	76
465	2020.10.05	(재)광주광역시	(재)광주광	·역시 서구 장	남학 https://ww	w.skkı	재 (재)광주	광역/ <u>https://www.skku</u>	학생지원팀	특이사항은 무엇인가요?					

첨부파일이 없는 경우는 content에서 내용을 추출하였으며, 첨부파일이 있는 경우는 첨부파일에서도 내용 추출

Trial and error _ 라벨링 Question

- 1. 선발대상 및 인원
- □ 선발대상: 국내 국.공립, 사립 종합 4년제 대학교 3학년 1학기(5학기) 진급예정자(4학기 재학생 혹은 4학
- 기 이수한 복학예정자)
- □ 선발인원 : 총 20명 이내
- □ 지원자격
- 1) 성적: 학점평균 B+이상 (3.3/4.3 이상 또는 3.5/4.5 이상)
- 2학년 1학기까지 평점평균(3학년 1학기 복학예정자는 2학년 2학기까지)
- 2) 선발계열
- 전계열
- * 예체능계, 의대, 약대, 간호대, 사범대 등 특수.전문계열 및 전공 제외

△ 이수학점 X / 휴학생, 초과등록생 가능

<유한재단 (코로나19 및 폭우피해 특별지원금) : 학업지원비 >

- 1. 지원대상
- 가. 소득분위 5분위 이내의 학생(2022학년 1학기 기준)
- 나. 직전학기 학업성적이 12학점(4학년 9학점)이수하고 평점평균이 3.8이상
- 다. 교외 장학금 중 어떤 장학금도 수혜하고 있지 않은 학생 (국가장학금,교내성적우수장학금,교내근로장학금 제외)
- 라. 2022-2학기 현재 휴학생 및 초과등록생 신청 불가

△ 이수학점 O / 휴학생, 초과등록생 불가능

원래는 이수학점을 일괄적으로 12학점으로 제한하려고 했으나, 그렇지 않은 장학금도 많았고, 4학년은 9학점, 휴학생은 이전 학기 등 고려해야 할 부분이 많았음. 심지어 휴학생과 초과등록생은 신청이 불가능한 장학금도 많음.

Trial and error _ 라벨링 Question

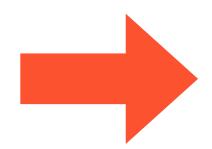
- 1. 선발대상 및 인원
- □ 선발대상: 국내 국.공립, 사립 종합 4년제 대학교 3학년 1학기(5학기) 진급예정자(4학기 재학생 혹은 4학
- 기 이수한 복학예정자)
- □ 선발인원 : 총 20명 이내
- □ 지원자격
- 1) 성적: 학점평균 B+이상 (3.3/4.3 이상 또는 3.5/4.5 이상)
- 2학년 1학기까지 평점평균(3학년 1학기 복학예정자는 2학년 2학기까지)
- 2) 선발계열
- 전계열
- * 예체능계, 의대, 약대, 간호대, 사범대 등 특수.전문계열 및 전공 제외

△ 이수학점 X / 휴학생, 초과등록생 가능

<유한재단 (코로나19 및 폭우피해 특별지원금): 학업지원비 >

- 1. 지원대상
- 가. 소득분위 5분위 이내의 학생(2022학년 1학기 기준)
- 나. 직전학기 학업성적이 12학점(4학년 9학점)이수하고 평점평균이 3.8이상
- 다. 교외 장학금 중 어떤 장학금도 수혜하고 있지 않은 학생 (국가장학금,교내성적우수장학금,교내근로장학금 제외)
- 라. 2022-2학기 현재 휴학생 및 초과등록생 신청 불가

△ 이수학점 O / 휴학생, 초과등록생 불가능



이를 모두 고려하기는 어렵다고 판단. 개인별로 추천 장학금에 자신이 적합한지 파악해야 할 필요가 있음

-> UI/UX 보완 필요

Trial and error _ 크롤링 대상

[장학] 2018년 2학기 풀뿌리학회 지원 사업 공모

최종 수정일: 2018.09.03

SNS 공유: **f** 🔰 📗 🖉 첨부파일 (1)

CORE

2018년 2학기 풀뿌리학회 지원사업 공고

성균관대학교 코어사업단에서는 2018년 2학기 풀뿌리학회 지원사업을 아래와 같이 시행합니다. 많은 신청 바랍니다.

1. 지원대상

- 아래의 학생들이 전공 학습력 강화를 위하여 자율적으로 결성한 5~15인 규모의 소모임 및 스터디.
- 유학동양학과, 국어국문학과, 영어영문학과, 중어중문학과, 사학과, 철학과, 한문학과, 프랑스어문학과, 독어독문학과, 러시아어문학과 등 10개 학 과에 재학 중이거나 복수전공하고 있는 학부생
- 글로컬문화콘텐츠, 미래인문학, 융합언어학, 비교문화, 일본학, 고전학 등 연계전공을 복수전공하고 있는 학부생

장학금 대상이 개인이 아닌 학회와 소모임인 장학금이 있음

-> 이러한 장학금은 크롤링하지 않기로 판단!

Trial and error _ 첨부III일

□ 신청방법

○ (온라인신청): 우수인재 / 학교 밖 청소년 제외 전체 장학금 분야

※ 보완사항 발생 시 '마이페이지 → 장학금 신청내역'에서 수정 가능

→ 신청 장학금 클릭 → 서류 업로드 및 신청서 제출 → 처리완료 대기

- 진흥원 누리집(www.sjhle.or.kr) 로그인 → 상단 '인재육성사업' 탭 → 장학금 신청

장학 공고에 내용이 충분하지 않았음에도, 첨부파일이 크롤링되지 않은 경우가 많았음.



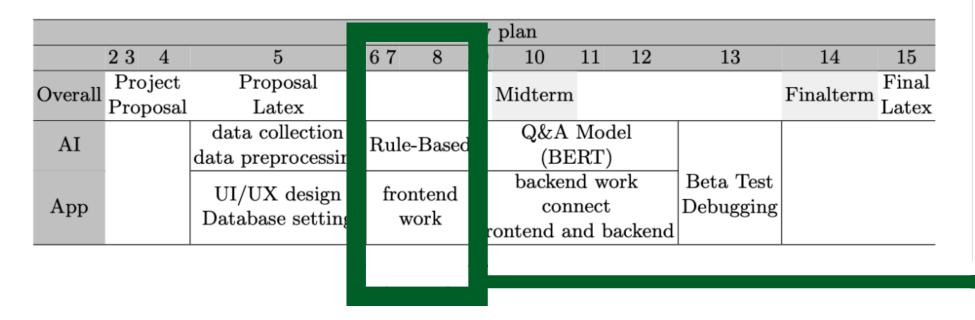
BERT Q&A 학습을 위한 학습 데이터에는 추가하지 않기로 함.

사용자 추천을 위한 장학금 크롤링에는 '공고' 등 더 다양한 단어를 추가하여 첨부파일 크롤링 예정

Brief Schedule [next week]

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



(a): AI (b): Back-end (f): Front-end

week 7
1) APP:
Database setting(b)
Frontend 개발(f)

2) AI: Data labeling (a,b,f) Rule-Based 구현(a)

지금까지 TEAM. SKKULAR 였습니다. 감사합니다.