

SKKU SCHOLARSHIP

TEAM 스콜라, SKKULAR

강병준 | 김규진 | 박진아 | 장이준 | 주재현

CONTENTS



1. Project
Proposal 요약

2. 피드백

TEAM. 스킨라
SKKULAR



1. Project Proposal

요약

1. Project Proposal 요약 | 문제인식

공지사항

2022학년도 2학기 유당장학재단 장학생 모집 안내

최종 수정일 : 2022.08.18

학생지원팀

SNS 공유:   |  첨부파일 (1)

1. 선발인원 : 1명
2. 선발대상
 - 1) 직전 학년(2개 학기) 성적 3.5점 이상인 자
 - 2) 가정형편이 어려운 자
 - 3) 봉사정신이 투철한 자
3. 장학금액 : 1학기 200만원(등록금성)
4. 제출서류
 - 1) 장학금 지원신청서 : 재단양식(첨부파일)
 - 2) 학과장추천서 : 각 학과 사무실에 문의
 - 3) 자기소개서 : 자유양식



왜

지원하지 않을까?

1. 매일 확인하기 어려움

2. 높은 검색 피로도

3. 높아지는 무관심의 정도

즉, 자신이 받을 수 있는 장학 공고도
'몰라서' 지원하지 못한다.

내가 지원할 수 있는
장학금 공지만 **골라볼** 수 있다면?

맞춤형 장학 검색 및 알림 서비스

"SKKULAR"

(SKKU + Scholarship)

TEAM.
SKKULAR

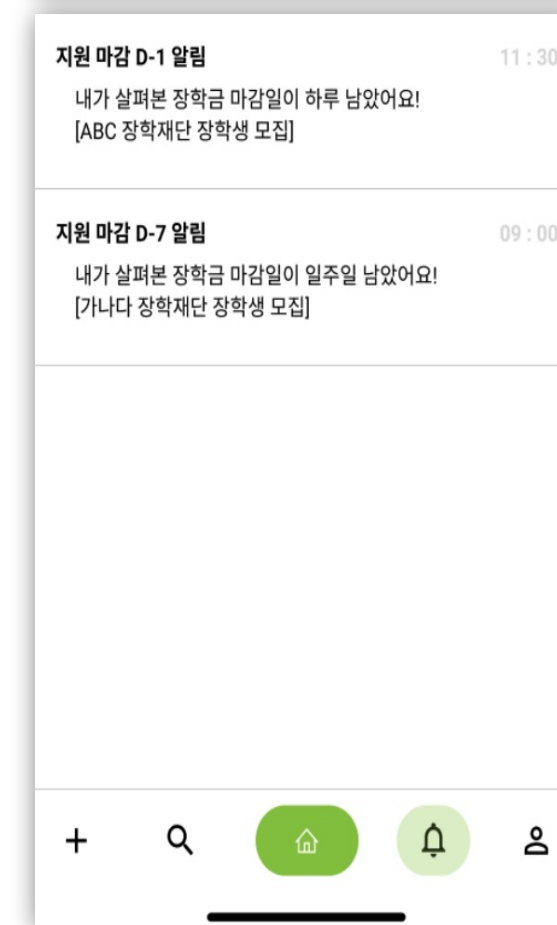
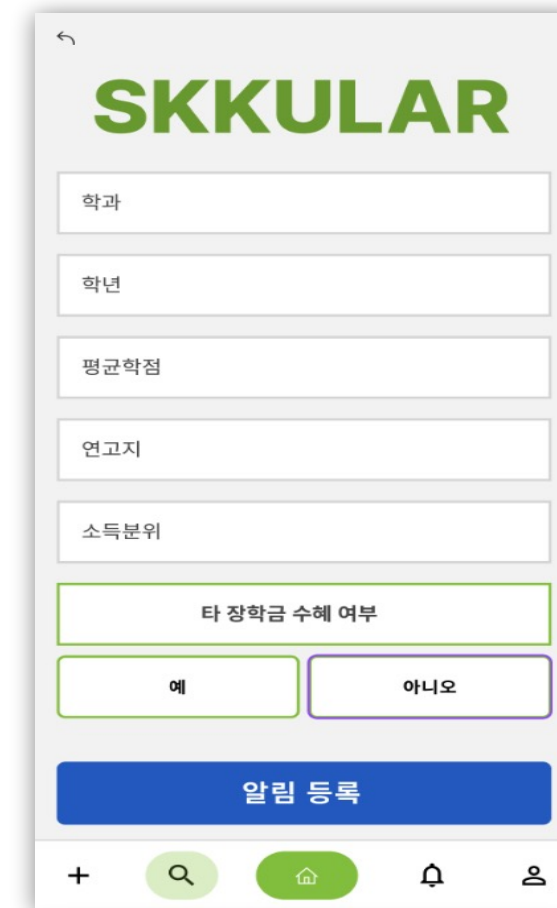
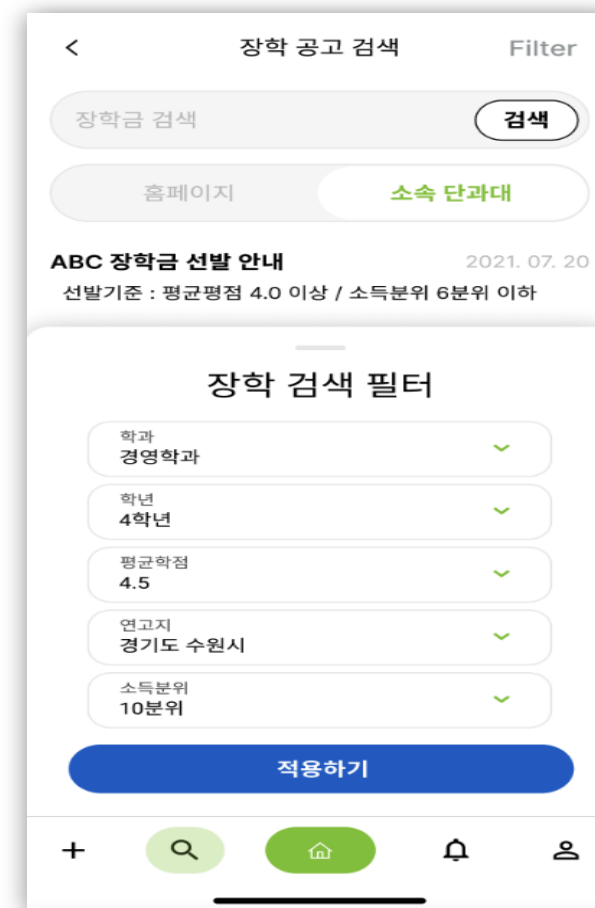
1. Project
Proposal 요약

2. 피드백

1. Project Proposal 요약 | 핵심 기능

SKKULAR 핵심기능

1. 맞춤형 장학 검색
2. 맞춤형 장학 알림
3. 유사 키워드 장학 안내
4. 지원 마감일 알림



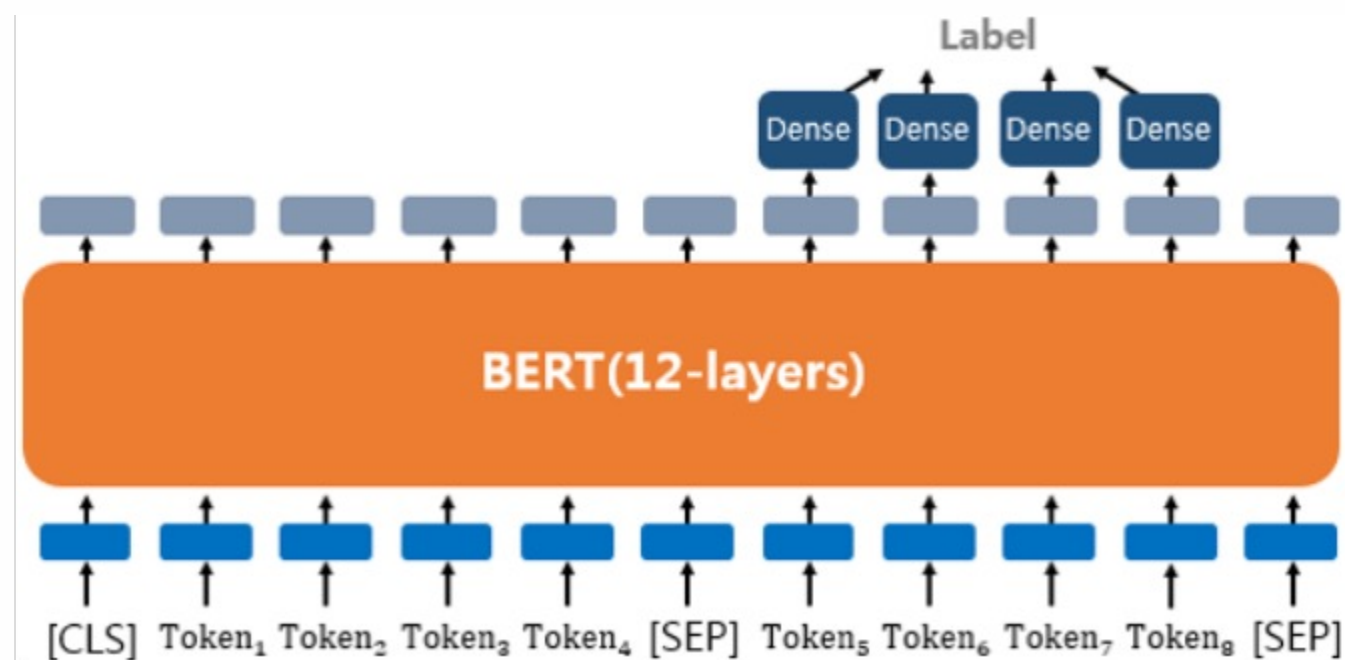
TEAM.
SKKULAR

1. Project Proposal 요약

2. 피드백

1. Project Proposal 요약 | 기술 배경 설명

AI



BERT

BERT Q&A 모델

$$w_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

TF-IDF

Term x within document y

$tf_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

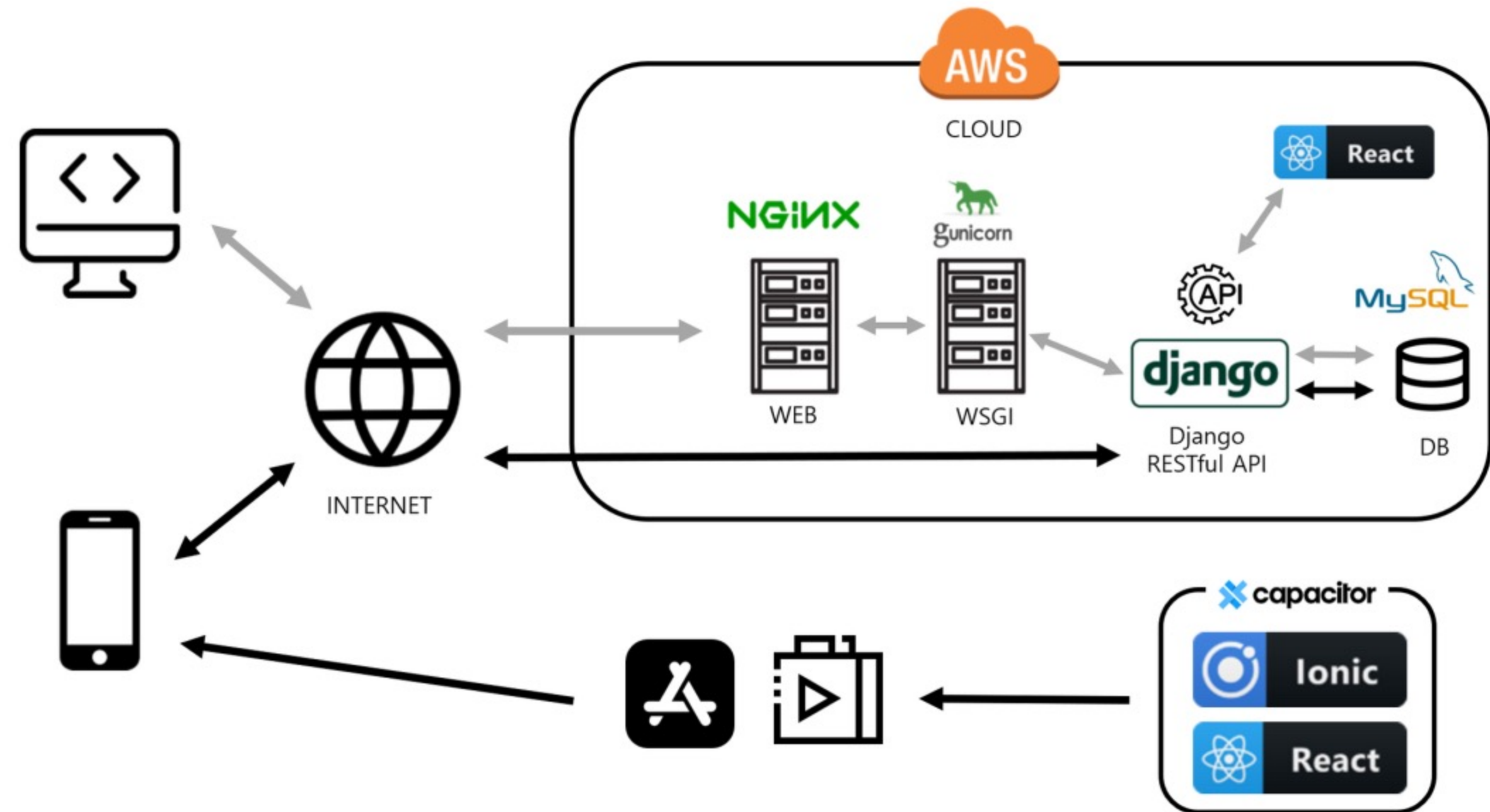
TF-IDF

TEAM.
SKKULAR

1. Project
Proposal 요약

2. 피드백

웹앱 | 아키텍처





2. 피드백



Q1. Project Range : (1) 장학금 범위

1. 학교 공식 홈페이지 공지사항

공지사항		
전체	검색어를 입력하세요	검색
2022학년도 (재)광산장학회 장학생 선발 안내		
No.1804	학생지원팀 2022-09-16 조회수333	NEW 
2022학년도 2학기 만산시 대학생 본연부담 등록금 반값지원 안내		
No.1803	학생지원팀 2022-09-16 조회수132	NEW 
2022학년도 2학기 고양시 대학생 본연부담 등록금 지원사업 안내		
No.1802	학생지원팀 2022-09-16 조회수141	NEW 
2022학년도 2학기 63장학재단 장학생 선발 안내		
No.1801	학생지원팀 2022-09-15 조회수2850	NEW 

2. 단과대 홈페이지 공지사항

학부	
제목 +...	검색어를 입력해 주세요
전체 학사 입학 취업 채용/모집 장학 행사/세미나 일반 학생회 소식	
[장학] 2022-2학기 경영대학 성적우수장학생 선발결과 공지	
No.56	bizskk 2022-08-18 조회수 1858
[장학] 2022년 중봉 국외 유학 장학생 선발 안내	
No.55	bizskk 2022-07-15 조회수 443 
[장학] 2022 한국고등교육재단 유학후보장학생 모집 안내	
No.54	bizskk 2022-06-16 조회수 340 

학교 및 단과대 공지사항 이외의 공지사항은
범위가 넓고 정확하지 않은 정보가 많기 때문에 수집하지 않음

Q2. Project Range : (2) polling 방식

대부분의 장학금 신청기간이

1주~2주



=> 매일 크롤링

대상 :

1. 새로 올라온 공지사항
2. 모집기간이 끝나지 않은 공지사항

특징 :

- 데이터 크기가 적어서 크롤링에 크게 문제 X
- 장학공고 내용이 수정되어도 데이터 수집 가능

Q3. Limitation



- PDF 일 경우 : python에서 textract를 이용하여 pdf를 text로 변형한 후 정보를 얻음
- PDF가 이미지 형식이거나 공지사항이 이미지 형식이면 정보수집 하지 않음.

Q3. Limitation



Textract :
워드, 파워포인트 PDF 파일 등의 텍스트를
추출하는 모듈



장학금 관련 text



Rule-based 기반 데이터 전처리
Bert Q&A 모델 사용

Q4. TF-IDF 활용방안

$$w_{x,y} = \text{tf}_{x,y} \times \log \left(\frac{N}{\text{df}_x} \right)$$

TF-IDF

Term x within document y

$\text{tf}_{x,y}$ = frequency of x in y

df_x = number of documents containing x

N = total number of documents

TF-IDF 를 통해 문서의 **키워드**를 추출한 후,
해당 키워드를 기준으로 특정 장학 게시물과
유사한 장학금들을 추천해줄 예정

Q5. BERT Q&A 모델

1. BERT 기계독해 모델

Pretrained된 한국어 BERT 모델에 기계독해 task를 해결할 수 있도록 **finetuning**을 진행할 예정
(처음부터 학습하기에는 관련 데이터셋도 적고 오랜 학습 시간이 요구된다.)

* Pretrained bert 모델 종류

1) ETRI 한국어 bert 언어모델

[공공 인공지능 오픈 API·DATA 서비스 포털 \(etri.re.kr\)](http://etri.re.kr)

- 신문기사와 백과사전 등 23GB의 대용량 텍스트를 대상으로 47억개의 형태소를 사용하여 학습
- encoder only model: encoder layer 12개
- ETRI에서 배포하는 한국어 tokenizer 사용
- KorQuad 데이터 기반 기계독해 성능:
[기계 독해] KorQuAD 1.0 EM 80.70%, F1 91.94%

2) SKTbrain

[GitHub - SKT-AI/KoBART: Korean BART](https://github.com/SKT-AI/KoBART)

- 40GB 이상의 한국어 텍스트에 대해서 학습한 한국어 bert 모델
- encoder-decoder model: encoder 6개, decoder 6개
- tokenizers 패키지의 character BPE tokenizer 사용
- KorQuad 데이터 기반 기계독해 성능: 테스트해봐야!

Q5. BERT Q&A 모델

2. 데이터셋

기존 기계 독해 모델의 데이터셋: KorQUAD

Input으로 context(단락), question(단락에 대한 질문),
output은 answer(질문에 대한 답)으로 구성됨.

=> Context가 게시물 형태, 그리고 장학금이라는 분야에만 적용
된다는 점에서 장학금 domain에 특화된 데이터셋을 생성할 필요
존재. 장학금 게시물을 크롤링한 후 데이터를 SQUAD 데이터셋 형
식에 맞게 labeling할 예정. (데이터 증강까지 가정하여 대략 200
0-3000쌍 데이터 마련 예정. 경험적 판단)

3. Challenges

Data augmentation(데이터 증강), 혹은 tokenizin
g 방법을 다양하게 사용해볼 예정

```
{
  "creator": "KorQuAD",
  "data": [
    {
      "title": "존재와 무",
      "paragraphs": [
        {
          "context": "《존재와 무》(l'Être et le Néant: essai d'ontologie phénoménologique)는 프랑스의 철학자 사르트르가 1943년 출판한 책이다. 사르트르의 주된 목적은 개인의 존재에 앞서 개인의 실존을 주장하는 것이다. 이 책을 작성하면서 최우선적으로 염두에 둔 것은 자유가 존재한다는 것이다. 1940년과 1941년에 전쟁 포로로 억류되어있던 시절 사르트르는 마르틴 하이데거의 《존재와 시간》을 읽고 자기 자신만의 자유를 전개하였다. 하이데거의 영향을 받았음에도 사르트르는 하이데거의 존재와의 가설적인 재조우와 비교하여 인간성(humanity)이 개인적인 성취의 상태를 달성할 수 있다는 방법에 회의적이었다.",
          "qas": [
            {
              "id": "9_f9_wiki_4511-1",
              "answers": [
                {
                  "answer_start": 69,
                  "text": "사르트르/nnp"
                }
              ],
              "question": "책 존재와 무의 저자가 누구야"
            },
            {
              "id": "9_f9_wiki_4512-1",
              "answers": [
                {
                  "answer_start": 60,
                  "text": "프랑스"
                }
              ],
              "question": "사르트르는 어느나라 철학자야"
            }
          ]
        }
      ]
    }
  ]
}
```

Q6. 홍보방안

1) '꾸'페셜 장학금 프로모션

:스꼴라만의 특별 장학 이벤트 진행

ex. '스꼴라 생일 기념 장학'

지원자격

1. 성균관대학교 소속 학생
2. 스포라 회원가입한 자

cf. A+ 응원 커피 장학
해피뉴이어 2023장학

2) 학교 및 다양한 교내 단체와 협업

컨택예정

성균3사, 성균웹진
성균관대학교 공식 인스타/유튜브
학지팀 협업, 킹고엠 알림

cf. 에브리타임/총학생회 홍보는
현실적 어려움 존재

야, 너두 장학금 받을 수 있어

“스콜라로 나한테 딱 맞는 장학금 확인하자!”

Q&A

지금까지 TEAM. SKKULAR 었습니다.
감사합니다.