

# SKKU SCHOLARSHIP

## | WEEKLY PROGRESS MEETING (3)

TEAM 스콜라, SKKULAR

강병준 | 김규진 | 박진아 | 장이준 | 주재현

---

# CONTENTS



1. 앱

2. AI

**TEAM. 스킨라**  
**SKKULAR**

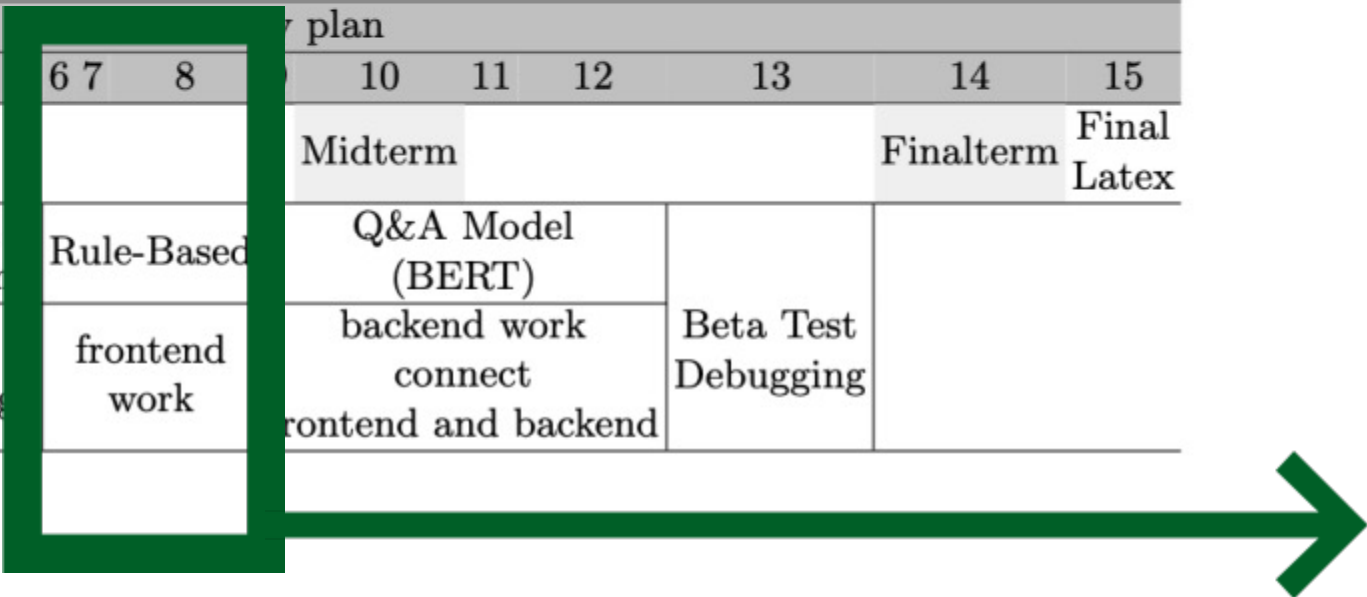
---

# 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

	2 3 4			5	6 7 8	10	11	12	13	14	15
Overall	Project Proposal			Proposal Latex		Midterm				Finalterm	Final 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		



## week 7

### 1) APP :

Frontend 개발 (f)

Database setting (b)

회원가입 기능구현 (b)

### 2) AI :

BERT 모델 스터디 (a)

Data labeling (a,b,f)

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



# 1. 앱

1. 앱

1. 앱

1. 앱

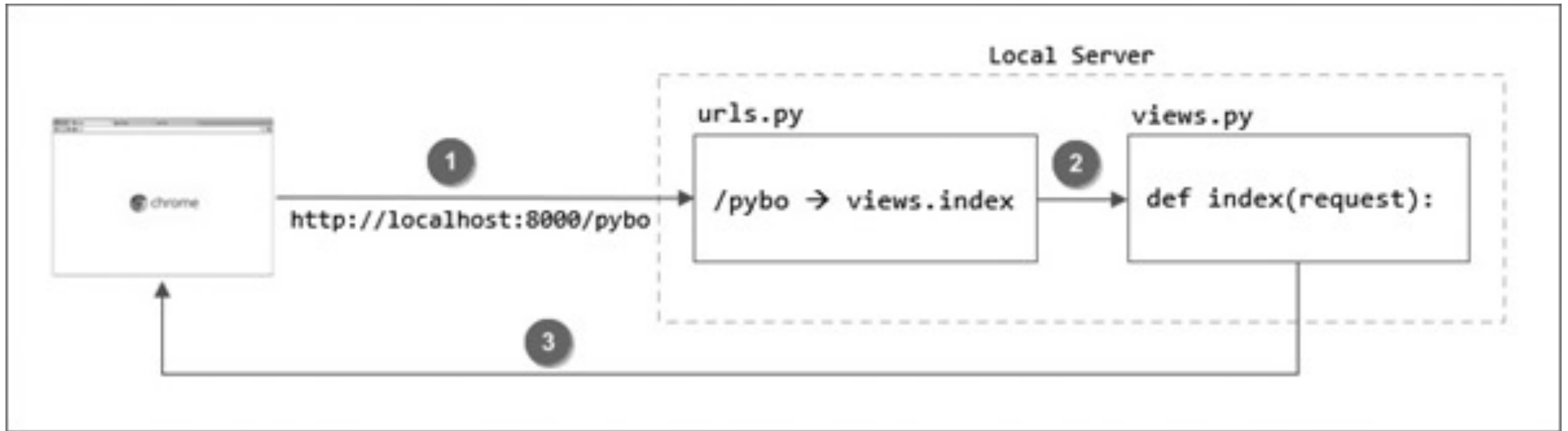
1. 앱

1. 앱

1. 앱

---

# 장고 | 개발 흐름



1. 브라우저에서 서버로 페이지 요청
2. `url.py`에서 url 매핑 확인하여 함수 호출
3. 호출한 결과를 브라우저에 반영

# 장고 | 세팅

```
INSTALLED_APPS = [  
    'capstone.apps.CapstoneConfig',  
    'django.contrib.admin',  
    'django.contrib.auth',  
    'django.contrib.contenttypes',  
    'django.contrib.sessions',  
    'django.contrib.messages',  
    'django.contrib.staticfiles',  
    'rest_framework',  
    'accounts',  
    'common',  
]
```



장고 회원 가입 기능 간편화

로그인, 로그아웃 기능 제공

관리자 권한으로 계정 관리 기능

# 장고 | urls

```
app_name = 'common'

urlpatterns = [
    path('login/', auth_views.LoginView.as_view(template_name='common/login.html'), name='login'),
    path('logout/', auth_views.LogoutView.as_view(), name='logout'),
    path('signup/', views.signup, name='signup'),
]
```

url 매핑

함수 호출

url 별칭

# 장고 | form

```
class RegisterForm(UserCreationForm):  
    class Meta:  
        model=User  
        fields= ['username', 'password1', 'password2', 'email',  
                 'semester', 'lastgpa', 'fullgpa',  
                 'income', 'departments', 'residence']
```

1. 장고의 userCreationForm 상속하는 Registerform 작성
2. 메타 클래스에 속성 추가



# 장고 | view

```
=
def signup(request):
    if request.method == "POST":
        form = RegisterForm(request.POST)
        if form.is_valid():
            form.save()
            username = form.cleaned_data.get('username')
            raw_password = form.cleaned_data.get('password1')
            user = authenticate(username=username, password=raw_password) # 사용자 인증
            login(request, user) # 로그인
            return redirect('index')
        else:
            form = RegisterForm()
    return render(request, 'common/signup.html', {'form': form})
```

1. POST 요청인 경우에는 화면에서 입력한 데이터로 사용자를 생성
2. GET 요청인 경우에는 회원가입 화면을 보여준다

# 장고 | 회원가입

← → ↻ ⓘ 127.0.0.1:8000/common/signup/

Skkular ☰

- [로그인](#)
- [회원가입](#)

사용자 이름

비밀번호

비밀번호 확인

이메일

학기

직전학기 평점

총 평점

소득분위

학과

거주지

생성하기

# 장고 | 로그인

← → ↻ ⓘ 127.0.0.1:8000/common/login/

Skkular ▾

- [로그인](#)
- [회원가입](#)

사용자ID

비밀번호

Skkular ▾

- [freudbj\\_\(로그아웃\)](#)
-

## 장고 | DB

id	password	last_login	is_superuser	username	first_name	last_name
3	pbkdf2_sha256\$390000\$3l8SDnBMGwrdrvrohEy...	2022-10-17 12:37:50.477942	0	freudbj		
4	pbkdf2_sha256\$390000\$teGWEIQVwhoITDkPjR...	2022-10-17 11:19:24.498783	0	freudbj1		
5	pbkdf2_sha256\$390000\$j6r2hUi9EfBLOfK2sera...	2022-10-17 11:35:33.222890	1	freudbj2		
NULL	NULL	NULL	NULL	NULL	NULL	NULL

[illegible]



## 2. AI



AI

---

# 현 라벨링 상황

장학공고 ☆ 📁 📄

파일 수정 보기 삽입 서식 데이터 도구 확장 프로그램 도움말 김규진님이 1시간 전에 마지막으로 수정했습니다.

100% 🔍 % .0 .00 123 기본값 (Arl... 10 B I S A 🔍 📄 📑 📊 📈 📉 📉 📉 ...

A1	E	F	G	H	I	J	K	L	M	N	O
1	attachment_url	attachment_cont	current_url	department	question	text	answer_start	answer_end	attach_start	attach_end	삭제
426	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	지원가능한 학년은 무엇인가요?							
427	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	직전학기 평점은 무엇인가요? 2.0 이상			168	174			
428	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	전체학기 평점은 무엇인가요?							
429	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	소득분위는 무엇인가요?							
430	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	관련 학과는 무엇인가요?							
431	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	거주지는 무엇인가요?							
432	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	마감일은 언제인가요?	2022년 7월 25일		470	482			
433	경우 제출 시 메일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	특이사항은 무엇인가요?	국가유공자		69	74			
434	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	지원가능한 학년은 무엇인가요?							
435	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	직전학기 평점은 무엇인가요? 3.0 이상			1191	1197			
436	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	전체학기 평점은 무엇인가요?							
437	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	소득분위는 무엇인가요?							
438	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	관련 학과는 무엇인가요?							
439	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	거주지는 무엇인가요?							
440	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	마감일은 언제인가요?	2022년 7월 25일		1397	1409			
441	일 본문에 "학번(수험번호) + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	특이사항은 무엇인가요?	성균가족장학금		783	790			
442	일 본문에 "학번 + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	지원가능한 학년은 무엇인가요?							O (지원자)
443	일 본문에 "학번 + 이름 + 연락처"가 포함되어 있는지 확인합니다.	https://www.skku.ac.kr/	학생지원팀	직전학기 평점은 무엇인가요?							O

현재 장학공고 라벨링 중. 1500개쌍 중 80% 라벨링된 상황. 금일 중으로 완료될 예정.



# 라벨링 RULES

0. 최근에 있는 장학금 위주로 라벨링 → 날짜별 정렬함
1. 한 장학금에 여러 장학금이 있는 경우는 학습데이터에 추가하지 않음
2. 장학금 내 중요한 단어 위주로 start와 end 정하기  
(ex. 군산시가 아무데나 있는 경우 & 지원자격에 군산시 있는 경우 지원자격 뽑기)
3. 지원가능 학기(남은 정규학기가 n개 이상), 학과 이름 (ex상경/농식품계열학과), 지원 학년  
(2학년 1학기 진학예정자 : 2학년 1학기) 등 특정되지 않더라도 텍스트에서 그대로 추출
4. 소득분위 : 건보료는 무시, 차상위계층, 기초생활수급자는 그대로 추출하기
5. 마감일 기록 시 요일은 기입 X

## 이번주 개발 계획

- 1) 라벨링 완료 후 KorQuard 데이터셋 형태로 json 변환  
- train, valid(필요시), test set 분리
- 2) ETRI의 korBERT 모델로 학습을 진행해볼 예정



# KorBERT 세부사항

## 1. BERT 모델 파라미터

- 12 layer / 768 hidden / 12 heads

## 2. BERT 모델 유형

1) 001\_bert\_morp\_pytorch, 002\_bert\_morp\_tensorflow .

- 형태소분석 결과 기반 BERT 학습 모델
- 입력 예: ETRI/SL 에서/JKB 한국어/NNP BERT/SL 언어/NNG 모델/NNG 을/JKO 배포/NNG 하/XSV 었/EP 다/EF ./SF

2) 003\_bert\_eojeol\_pytorch, 004\_bert\_eojeol\_tensorflow

- 어절 기반 BERT 학습 모델 (형태소분석 미수행)
- 입력 예: ETRI에서 한국어 BERT 언어 모델을 배포하였다.

# 기존 KorBert 모델의 입력

## 〈BPE〉

바이그램 쌍	빈도	바이그램 쌍	빈도	바이그램 쌍	빈도
b, u	4	b, u	4	b, un	4
g, s	5	h, <u>u</u> a	15	h, <u>u</u> a	15
h, u	15	p, u	12	p, <u>u</u> a	5
p, u	17	p, <u>u</u> a	5	p, <u>u</u> n	12
u, g	20	u, n	16	<u>u</u> a, s	5
u, n	16	<u>u</u> a, s	5		

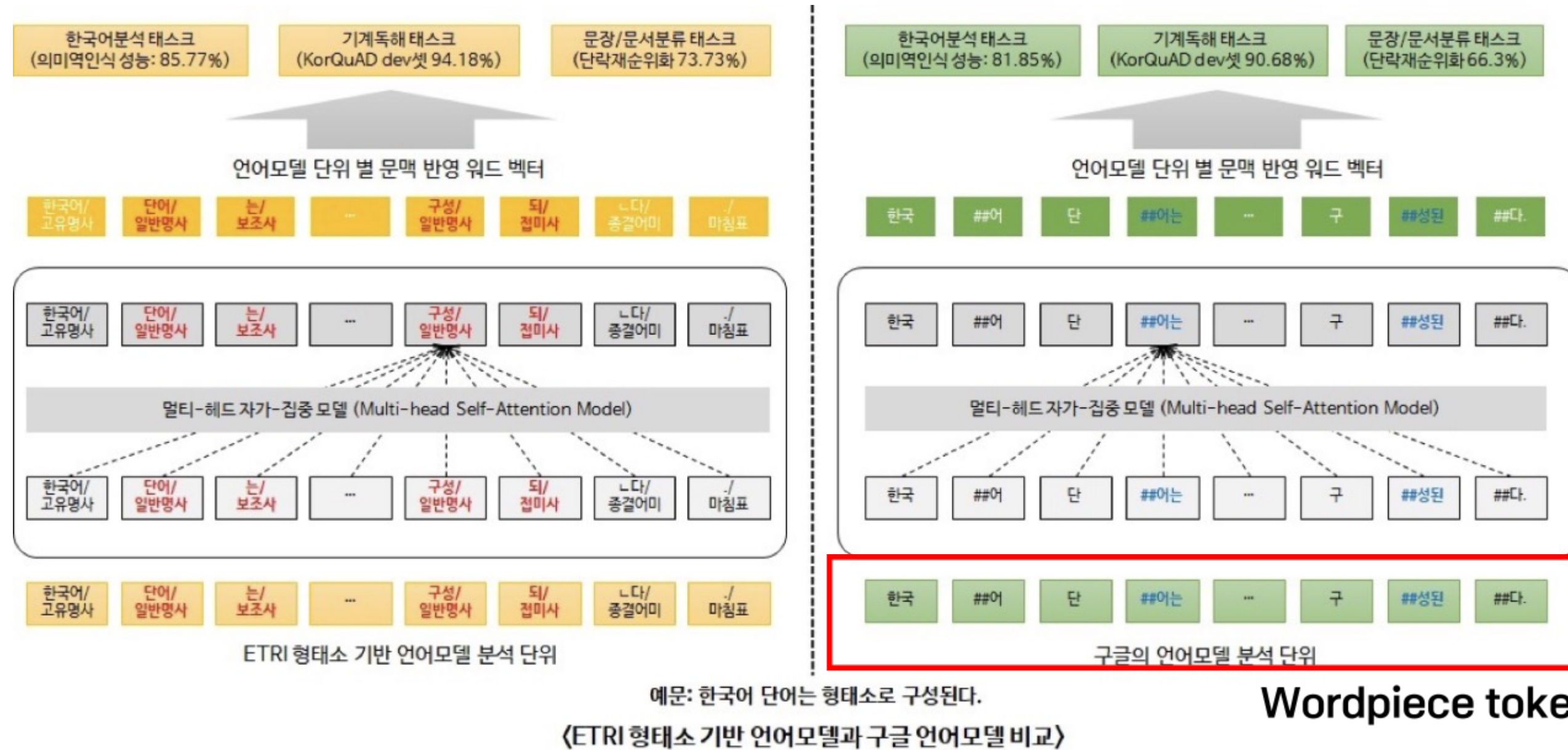
## 〈Wordpiece tokenizer〉

워드피스 tokenizer는 BPE처럼 단순히 빈도를 기준으로 병합하는 것이 아니라, 병합했을 때 말뭉치의 우도(likelihood)를 가장 높이는 쌍을 병합. 모든 언어에 적용가능하다는 장점이 존재.

$$\frac{\frac{\#ab}{n}}{\frac{\#a}{n} \times \frac{\#b}{n}}$$

문장 -> tokenizer를 통해  
토큰으로 쪼개어지고 모델에 입력되게 됨.

# 기존 KorBert 모델과의 차이점



형태소분석 기반의 언어모델은 교착어인 한국어의 특성을 반영한 모델. 명사/동사에 조사/접미사가 결합된 어절을 의미의 최소단위인 형태소로 토큰이 구성됨. 한국어의 특성을 더 잘 살릴 수 있음.

## 추가 계획

- 1) ETRI의 korBERT\_형태소 버전의 모델 학습이 성공적으로 이루어진 경우, 어절 버전의 모델로도 학습을 진행할 예정.

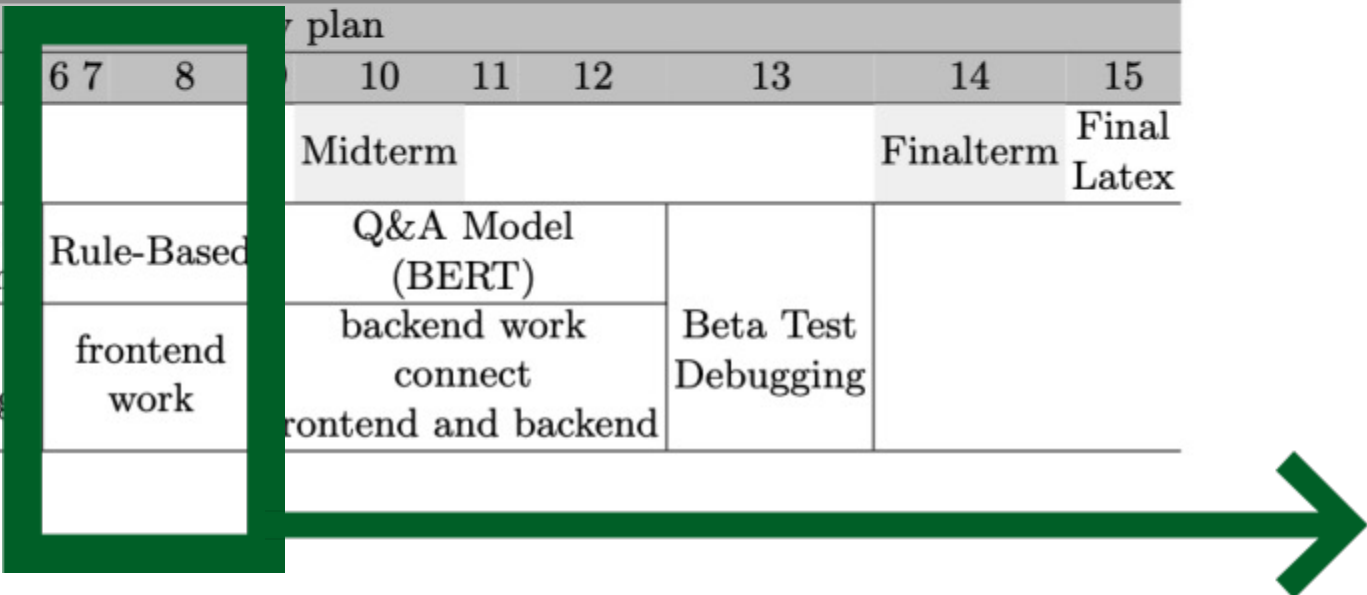


# Brief Schedule [next week]

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

	2 3 4			5	6 7 8	9	10	11	12	13	14	15
Overall	Project Proposal			Proposal Latex			Midterm				Finalterm	Final 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			



**week 8**  
1) APP :  
Backend-frontend 연결 (b,f)  
Database setting (b)  
Frontend 개발 (f)

2) AI :  
BERT 학습 (f)

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

---

# Q&A

지금까지 TEAM. SKKULAR 었습니다.  
감사합니다.