

아직도 혼자 공부해?

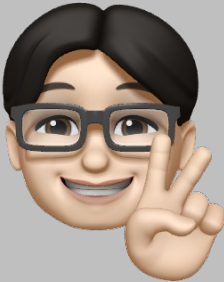
한국어 자연어 처리 기술에 기반한 맞춤형 학습 도움 서비스



박의성

🔧 백엔드 개발과
데이터 분석을 맡고 있습니다.

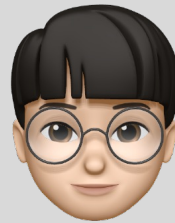
소프트웨어학과
스마트정보통신공학과



김정혁

🔧 프론트엔드 개발과
데이터 분석을 맡고 있습니다.

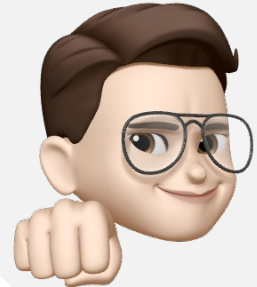
스마트정보통신공학과
그린화학공학과



이동재

🔧 프론트엔드 개발과
데이터 분석을 맡고 있습니다.

스마트정보통신공학과



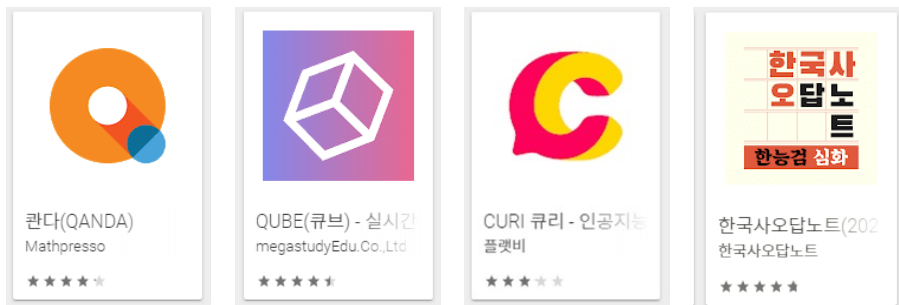
서범규

🔧 데이터 분석을 맡고 있습
니다.

스마트정보통신공학과



1	2	3	4	5	6	7	8
주제	시스템 흐름도	모델 비교	모델 학습	문제 생성	앱 화면	데이터 베이스	참고문헌



<기존의 학습 보조 어플리케이션>

	새로운 문제생성	오답 노트	기출 문제
혼공 머신	o	o	x
한국사 오답노트	x	o	o
큐리	x	o	x
관다	x	o	x

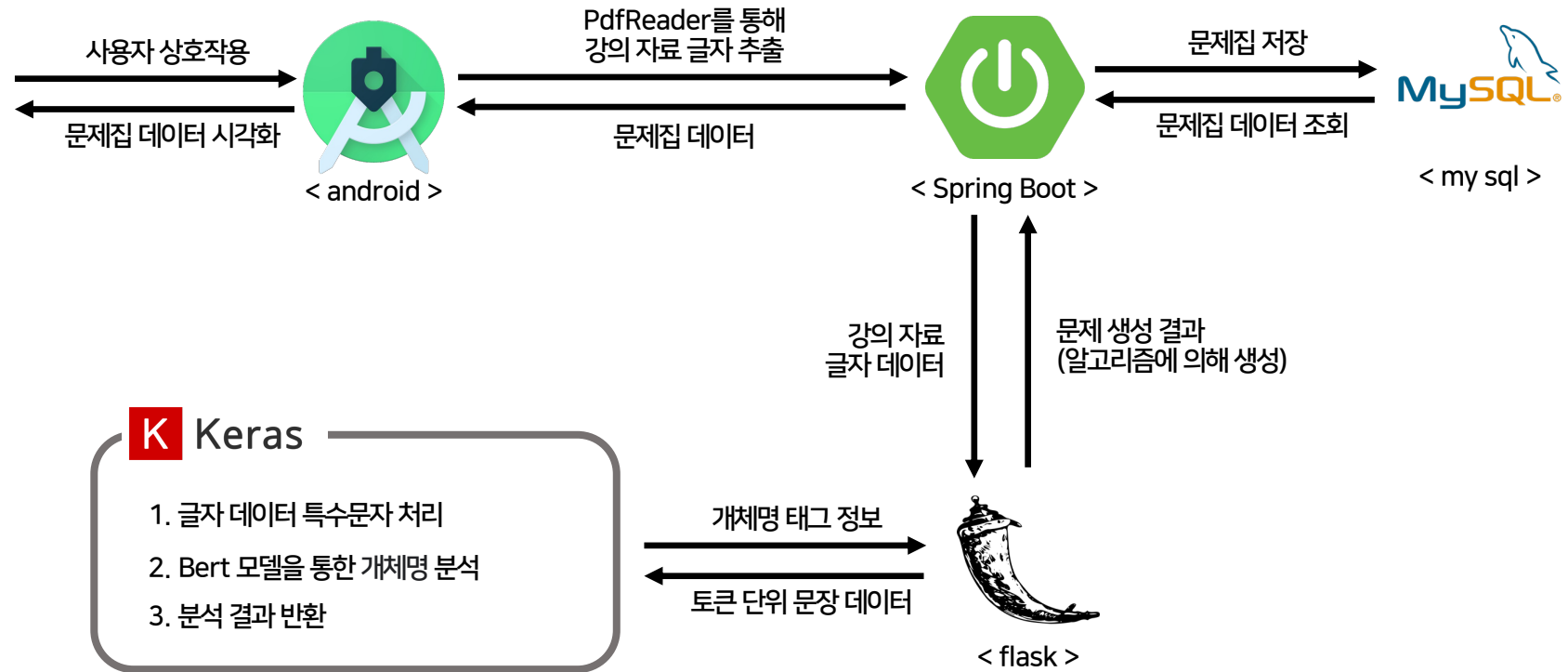
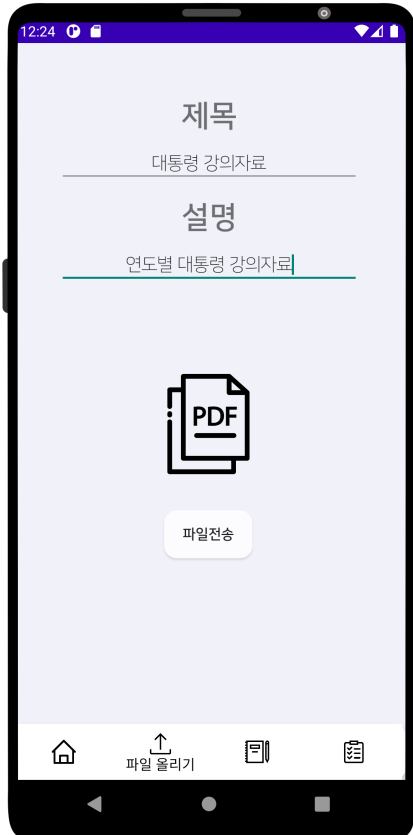
<기존의 학습 보조 어플리케이션 기능 표>

학습, 보다 편리하고 효율적이게

온라인 교육의 확대로 인해 개인 간 학습 편차가 증가되면서¹⁾ 학습의 중요성이 대두되고 있다. 이에 다양한 교육관련 어플리케이션의 활용도가 증가하고있다.

기존의 어플리케이션들은 사용자에게 기출문제를 제공하거나 기존의 문제의 해답을 알려주는 형식이다. 이러한 방식은 다양한 유형의 문제를 제시하기에는 어려움이 있으며, 사전에 작성되어 있는 문제를 제공하기 때문에 새로운 유형의 문제를 접하기 힘들다는 문제점이 있다.

이러한 문제점들을 해결하고자 '혼공머신'을 통해 강의자료를 기반으로 하여 새로운 문제의 자동 생성 및 오답노트 관리 기능을 제공하여 **사용자의 교육 효과를 증가**시키는 것을 목표로 하였습니다.

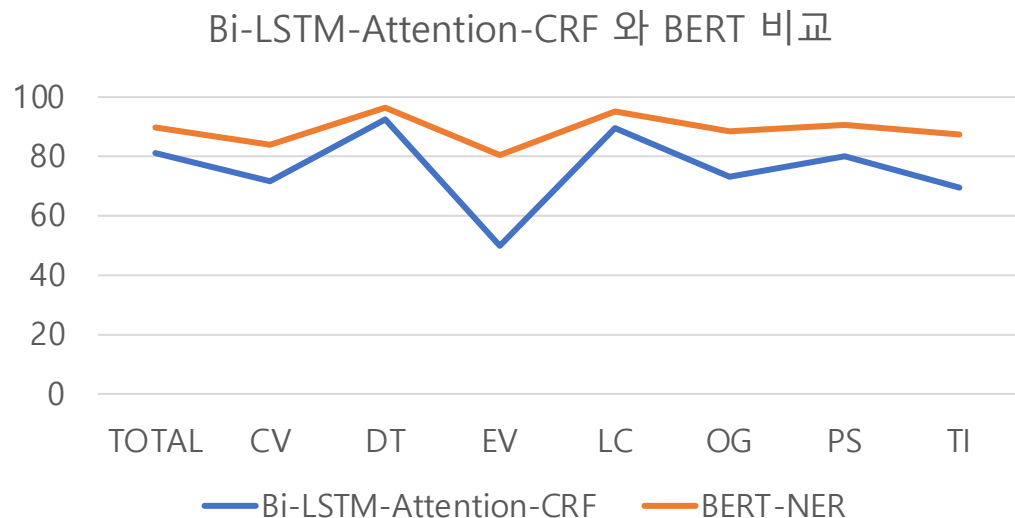


개체명 인식(NER)이란?

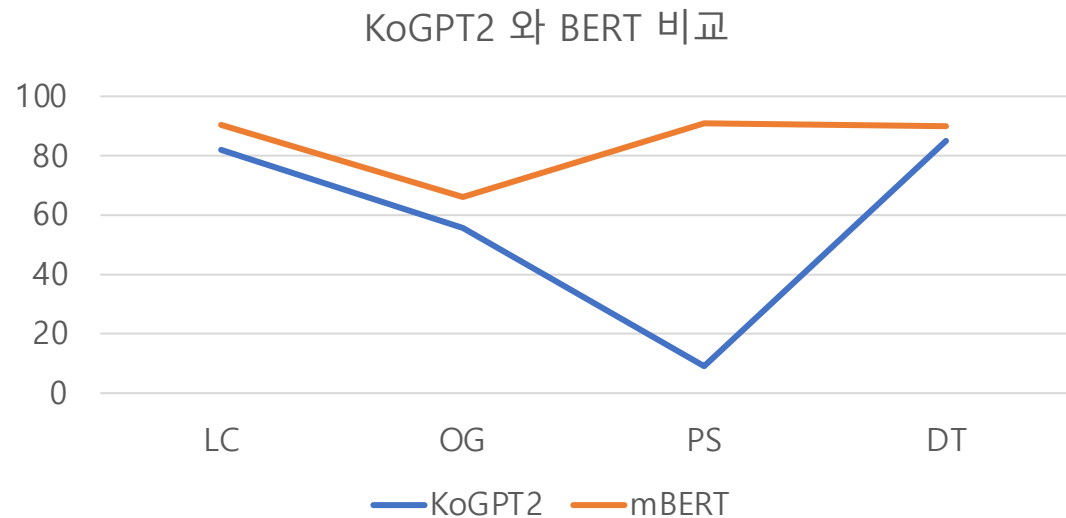
미리 정의해 둔 사람, 회사, 장소, 시간, 단위 등에 해당하는 단어(개체명)를 문서에서 인식하여 추출 분류하는 기법

BERT-multilingual 선정 이유:

- BERT 모델에 개체명 사전까지 활용한 모델의 F1-Score가 87.29점²⁾
- BERT 기반의 모델의 경우 기본적인 Multilingual 모델을 활용하여 3epochs 만 학습하였음에도 89.71의 F1-Score를 기록
- KoGPT2의 경우에도 BERT 보다 다소 낮은 레이블의 결과값을 보여줌³⁾



<Bi-LSTM-Attention-CRF와 BERT-multilingual의 성능비교>



<KoGPT2와 BERT-Multilingual의 성능비교>

예시 문장: 331년 고구려 고국원왕 즉위

1. 문장 데이터셋

- ('331년', 'DAT_B'), ('고구려', 'LOC_B'), ('고국원왕', 'PER_B'),
- ('즉위', 'EVT_B')

2. BERT 입력 생성 문장 최대 길이 8

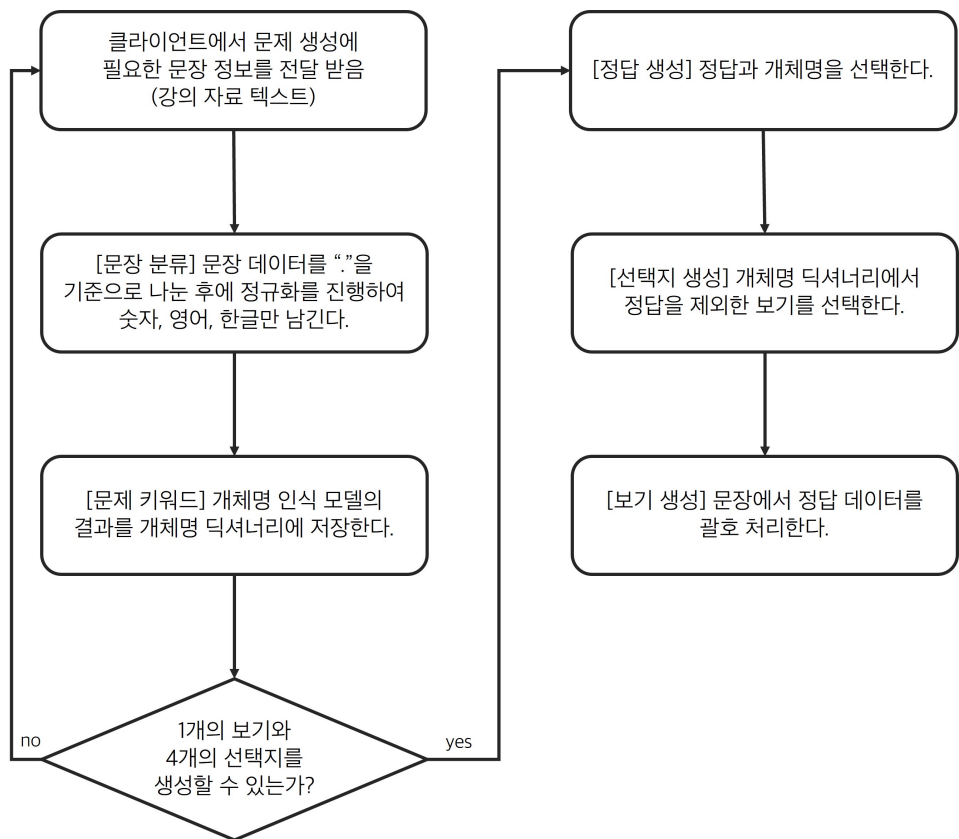
- 문장 인덱스: [1234, 234, 583, 32, 299, 299, 299, 299]
- 태그 인덱스: [2, 3, 4, 7, 29, 29, 29, 29]
- 마스크 데이터: [1, 1, 1, 1, 0, 0, 0, 0]

3. 개체명 인식 모델 만들기

- 독립 변수: 문장 인덱스, 마스크 데이터
- 종속 변수: 태그 인덱스

	개체명 범주	태그	정의
1	PERSON	PER	실존, 가상 등 인물명에 해당 하는 것
2	FIELD	FLD	학문 분야 및 이론, 법칙, 기술 등
3	ARTIFACTS_WORKS	AFW	인공물로 사람에 의해 창조된 대상물
4	ORGANIZATION	ORG	기관 및 단체와 회의/회담을 모두 포함
5	LOCATION	LOC	지역명칭과 행정구역 명칭 등
6	CIVILIZATION	CVL	문명 및 문화에 관련된 용어
7	DATE	DAT	날짜
8	TIME	TIM	시간
9	NUMBER	NUM	숫자
10	EVENT	EVT	특정 사건 및 사고 명칭과 행사 등
11	ANIMAL	ANM	동물
12	PLANT	PLT	식물
13	MATERIAL	MAT	금속, 암석, 화확물질 등
14	TERM	TRM	의학 용어, IT관련 용어 등 일반 용어를 총칭

< 개체명 태그 정보 >



< 객관식 생성 순서도 >

문제 생성 예시

- 1388년 이성계가 위화도 회군을 통해 정권을 장악
- { 'PER_B': ['이성계', '태종', '세종대왕', '연산군'], 'LOC_B': ['위화도'], 'DAT_B': ['1388년'] }
- 정답: 이성계, 개체명: 'PER_B'
- 선택지: 태종, 세종대왕, 연산군
- 1388년 {}가 위화도 회군을 통해 정권을 장악

[생성 문제]

문제: 다음 보기에 가장 알맞은 선택지는 무엇인가?

보기: 1388년 {}가 위화도 회군을 통해 정권을 장악

- 태종
- 이성계
- 세종대왕
- 연산군

문제 생성 예시

- 1979년 10월 26일 박정희 대통령 피살 사건. 1987년 노태우 6 29 선언. 2017년 3월 10일 박근혜 대통령 탄핵 사태. 1973년 8월 8일 김대중 납치 사건.
- { '-': ['피살', '사건', '탄핵', '사태', '1973년', '납치', '사건'],
'AFW_B': ['선언'], 'CVL_B': ['대통령', '대통령'],
'DAT_B': ['1979년', '2017년', '3월'],
'DAT_I': ['10월', '26일', '10일', '8월', '8일'],
'NUM_B': ['1987년', '6', '29'],
'PER_B': ['박정희', '노태우', '박근혜', '김대중']}
- 정답: [1,0,2,3]
- 선택지: [[2, 1, 0, 3], [3, 0, 2, 1], [3, 1, 2, 0]]

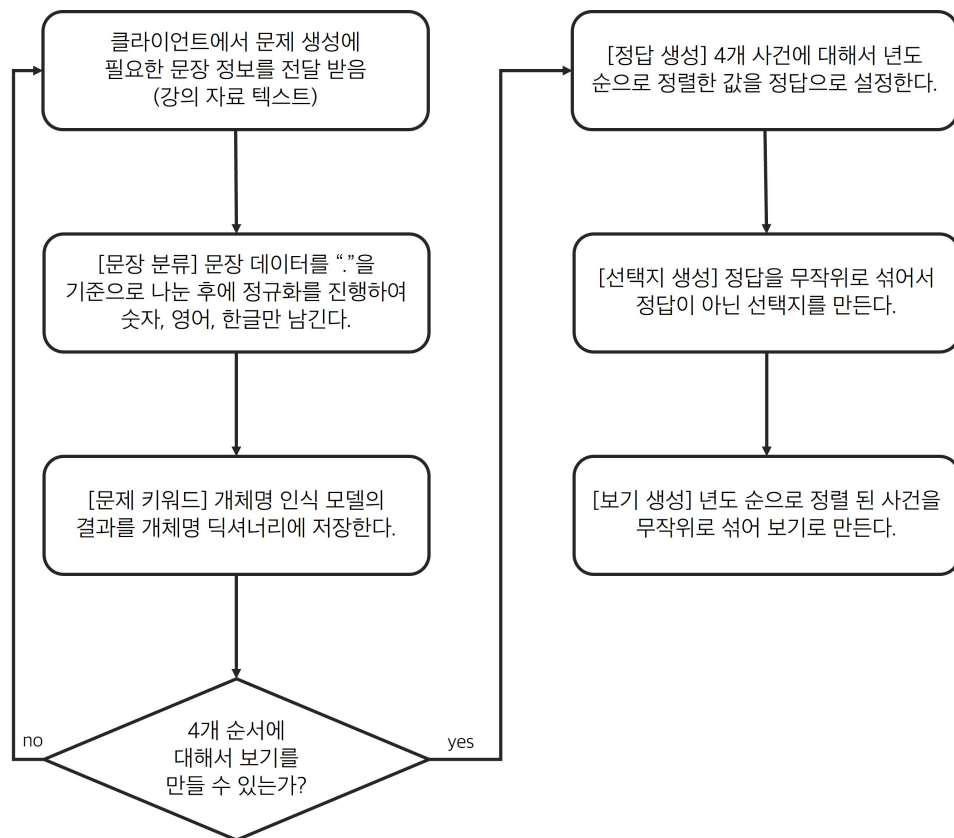
[생성 문제]

문제: 다음 사건의 순서로 알맞은 보기는 무엇인가?

보기:

0. 박정희 대통령 피살 사건
1. 노태우 6 29 선언
2. 박근혜 대통령 탄핵 사태
3. 김대중 납치 사건

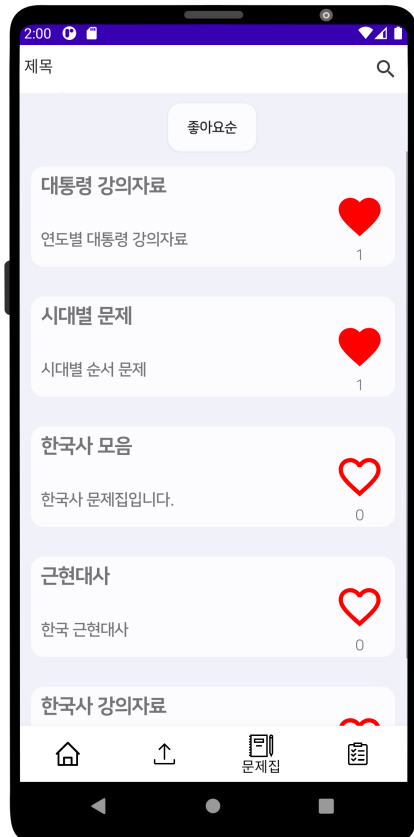
1. 2->1->0->3
2. 1->0->2->3
3. 3->0->2->1
4. 3->1->2->0



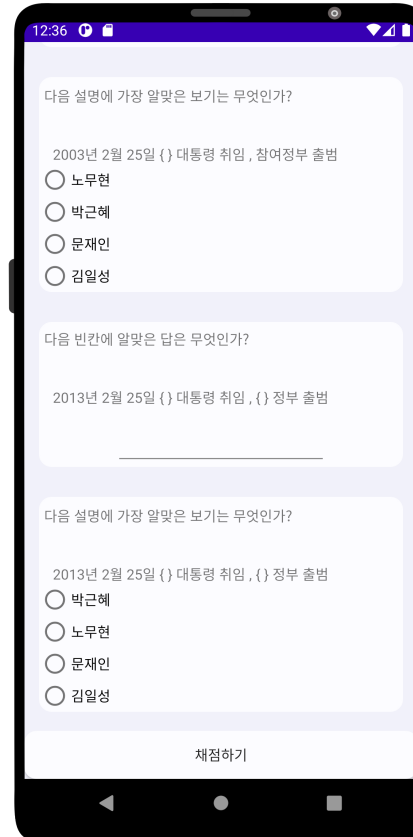
< 순서 생성 순서도 >



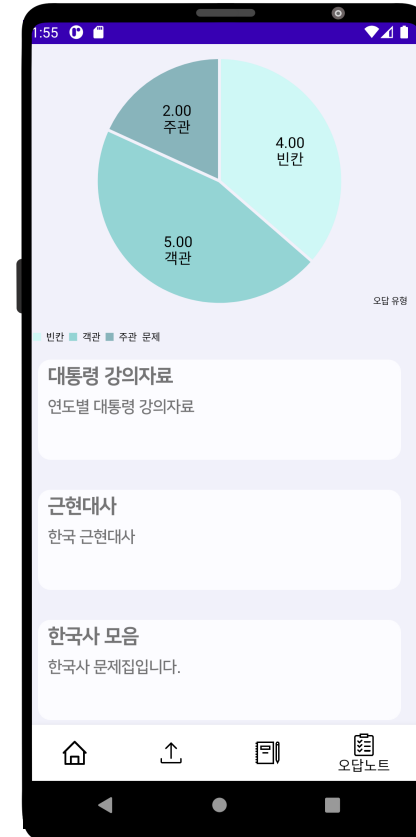
<업로드 화면>



<문제 집 화면>



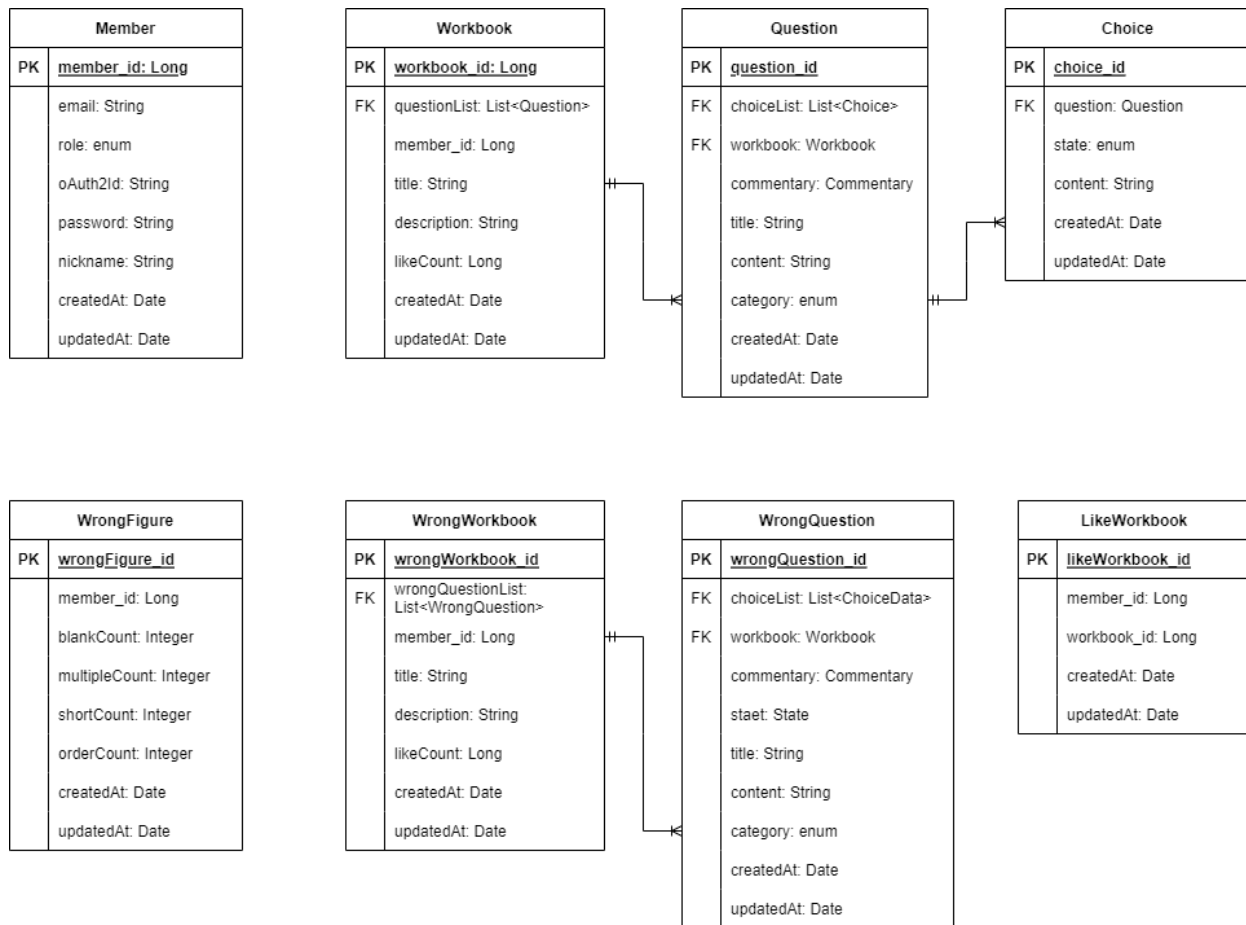
<문제 화면>



<오답노트 화면>



<오답문제 화면>



< 시스템 데이터베이스 >

서비스 요구사항

1. 문제집은 여러 문제를 가질 수 있다.
2. 문제는 여러 선택지를 가질 수 있다.
3. 오답 문제집은 여러 문제를 가질 수 있다.
4. 오답 문제는 여러 보기는 가질 수 있다.
5. 문제집은 회원만 생성할 수 있다.
6. 회원은 문제집 좋아요 기능을 사용할 수 있다.
7. 회원은 유형별 틀린 문제의 개수를 저장한다.

- [1] 김지연, "원격수업으로 학습격차 심화...중하위권 학습성취도 떨어져", 연합뉴스, 2021년 12월 26일자.
- [2] 황석현, 신석환, 최동근, 김성현, 김재은. (2019). BERT를 활용한 한국어 개체명 인식기. 2019년 추계학술발표대회 논문집 제26권 제2호, 820-822.
- [3] 김성현, 소영숙, 송치성, 한지윤. (2021). 한국어 개체명 인식 과제에서의 의미 모호성 연구. 제33회 한글 및 한국어 정보처리 학술대회 논문집, 203-208.