캡스톤 디자인- 중간점검

‘채점해드림’(가제)#프로젝트 방향 변경으로 제목 변경 예정

팀 이름 : 문차장

팀원 : 20193494문서형, 20190101장세환, 20193038차선우

# 목차

* 현재 진행 상황
* 진행 방식
* 수정된 Proposal
* 기술 스택 및 프로젝트 스케줄

# 진행 상황

1. 전체 진행 상황

백엔드 설계 완료

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

메인화면 뷰, 회원가입 뷰, 로그인 뷰 UI/UX 디자인 완료

메인화면, 회원가입 화면 기능 구현 완료

로그인 뷰 기능 구현중

문자 인식 기반 코드 분석 및, 불필요한 부분 제거 및 최적화 작업 진행중

1. Front-end

프론트 엔드는 리액트 create-react-app 이용해서 프로젝트 생성 후 리액트 내장 서버와 스프링을 연결했습니다.

UI 디자인을 완료했습니다.

메인 페이지는 swiper.js 사용해서 카드 슬라이드 뷰로 만들었습니다.

회원 가입 페이지는 form tag 이용해 입력 칸을 만든 후 정보를 서버로 넘기도록 구현했습니다.

로그인 페이지 기능 구현 예정입니다.

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

1. Back-end

- 프로젝트 API 설계 문서를 작성하였습니다.

- 데이터베이스 설계 문서를 작성하였습니다.

- 회원가입, 로그인 기능 구현 :

Spring Security + JWT 방식의 로그인, 회원가입 구현 부분을 작성하였습니다. 회원가입이 정상적으로 진행되고, 로그인을 진행했을 때 Refresh token 이 DB 에 저장되고 Access토큰을 이용해 유저 정보를 받아오는 부분까지 API 테스트 완료하였습니다.

1. 문서 인식

영어 OCR 기술에 대해 분석하고 프로젝트에 적용할 수 있게 적용 중입니다. 기반 코드 (<https://github.com/JaidedAI/EasyOCR>)을 활용하여 코드에 대한 이해를 대부분 완료하였으며, 현재 불필요한 부분을 삭제하는 작업을 진행중입니다. 해당작업이 끝날 시, 저희의 작업은 항상 a4용지에 정해진 형식으로 들어오는 점을 활용하여 최적화 작업을 진행할 예정입니다.

1. CM-Mentoring

6주차 CM 멘토링은 예상보다 조금 늦어져 10/16(월) 오후 7시에 진행하기로 되어 있습니다. 해당 시간에는 다음과 같은 내용에 대하여 조언을 받을 예정입니다.

* Spring 서버와 Django 서버간의 연동에서 주의해야할 점
* 데이터베이스에 이미지를 저장하는 방법이 권장되지 않는다고 들었는데, 처리 방법에 대한 조언
* 특정 프로젝트를 사용할 때 라이선스 문제에 대한 조언
* 머신러닝을 이용하여 헷갈리는 한글(ex. 관 & 판)에 대해 처리를 한 번 더 진행할 예정인데, 이 방법이 유의미한 과정일지에 대한 조언
* 이미지를 주고 받는 방식에 대한 조언

# 진행 방식

1. Git hub 주소 : <https://github.com/CAU-CAPSTONE-2023-Team0204/VOCA>

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

1. 칸반 프로젝트 보드를 이용한 issue 관리

텍스트, 스크린샷, 소프트웨어, 번호이(가) 표시된 사진

자동 생성된 설명

# 프로젝트 소개

채점해드림은 영어단어 공부를 위한 웹 서비스입니다. 학생에 초점이 맞추어진 기존의 서비스들과 다르게, 학생뿐만 아닌 교사에게도 편리하고 유용한 서비스를 제공하자는 생각에서 출발하였습니다.

# 프로젝트 동기

저희 조는 학원에서 조교 아르바이트를 해보았다는 공통점이 있습니다. 이에 대해 불편함을 해소할 방법을 찾던 도중 공통적인 문제점으로 매우 단순한 작업인 시험지 채점에, 너무 많은 노동과 시간이 들고 있음을 꼽았습니다. 단순히 답지와 시험지를 비교하기만 하면 되는 업무임에도, 학생의 숫자가 많을수록 업무에 들어가는 시간이 많음으로 자동화를 할 수 있는 방법에 대해 고민해 보았습니다.

채점 서비스가 가장 필요한 타겟층은

1. 페이지 안에 문제의 집적도가 높을수록 수요가 높을 것이다.

2. 채점에 여러가지 해석이 필요하지 않을수록 수요가 높을 것이다.

3. 주기적으로 시험을 보는 과목에서 수요가 높을 것이다.

라는 점에 착안하여 채점해드림은 영어학원 교사들의 영어 단어 시험 업무를 도와주는 오프라인 영어 단어 시험 출제, 채점, 관리 서비스를 준비해 보았습니다.

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

위는 구인 사이트의 영어학원 채점 아르바이트 구인 글들입니다. 2023.09.22 기준, “알바몬” 어플에는 <채점 조교>를 구하는 모집글이 서울 기준 35건, 경기도까지 기준을 넓히면 102건이 올라와 있습니다. 기존 서비스들의 해결 시도에도 불구하고 채점에 대한 수요가 여전히 많습니다.

또한 프로젝트 진행을 위해 현재 학원에서 조교를 하고 있는 3명(앞으로 계속 더 많은 인원에게 실시할 예정입니다.)에게 인터뷰를 진행했습니다.

이XX (용산구 J영어학원 조교)

1. 우리 학원은 학생들을 관리하는 학원이다 보니, 학생들마다 스케줄이 다르고, 진도가 다르다. 맞춤형 시험지를 빠르게 만들어주는 기능은 유용하다. 또한 채점하는 시간을 줄이면 다른 학생들의 질문을 더 많이 받아줄 수 있고 효율적일 것 같다.

장XX(대치동 O영어학원 조교)

1. 학원 조교의 인력을 많이 줄일 수 있을 것 같다. 오-영어학원의 경우 한 수업 당 150~200명의 학생들이 있다. 수업 시작 전 단어시험을 보고, 다음 수업 쉬는 시간 전까지 조교 4명이서 한시간동안 시험지만 채점하고 있어야 한다. 이렇게 시간이 오래 걸리고 단순한 업무를 대체할 수 있으면 대체하고 싶다.

남XX(송파구 2영어학원 조교)

1. 다른 것보다도 재시험지를 자동으로 만들어 주는 기능은 정말 편리할 것 같다. 학생들에게 틀린 단어들에 대해 재시험을 시킬 때, 할 수 없이 같은 시험지로 계속 보게 되는데, 순서를 외워서 답을 적는 등 효율이 떨어진다. 맞춤형으로 자동으로 시험지를 만들어 주는 기능은 매우 도움이 될 것 같다.

저희는 영어학원 선생님들, 조교님들의 고충을 해결하는 서비스를 제공하고자 합니다.

# 프로젝트 세부사항

주요기능은 다음과 같습니다.

교사 서비스

1. 영어 단어 시험 출제 보조 서비스

* 출제자는 4가지 유형에 대해 별도의 문서 작업 없이 간단하게 문제를 출제할 수 있습니다.
* 출제자는 대상 학생들의 나이와 수준을 입력하여, 해당 학년 추천 영어 단어와 문제를 추천받을 수 있습니다.
* 출제자는 대상 학생들의 영어 교과서 종류를 입력하여, 해당 시험 범위의 단어와 문제를 추천받을 수 있습니다.
* 문제 선정이 끝나면, 자동으로 시험지를 pdf 파일로 생성하여 줍니다.

1. 영어 단어 시험 채점 서비스

* 생성된 시험지에 수기로 작성된 답안을 스캔 혹은 업로드하면 자동으로 채점합니다.
* 채점된 정보와, 해당 시험의 통계는 채점해드림을 통해 학생들에게 개별 통보됩니다.

1. 학생 관리 서비스

* 학생들을 그룹으로 묶어, 해당 시험 및 전달사항에 대해 관리를 용이하게 도와줍니다.
* 해당 시험의 데이터를 종합하여 출제자에게 학습에 도움이 되는 통계를 제공합니다.
* 각각의 학생들의 이전 시험 정보들을 저장하여 취약점, 성취도 등의 자료를 제공합니다.

학생 서비스

1. 영어 단어 추천 서비스

* 학생들은 채점해드림을 통해

1. 데일리 테스트 서비스

* 학생들은 학습 정도에 따른 데일리 테스트가 주어집니다. 해당 테스트를 진행함으로써 랭킹을 올릴 수 있습니다.

# 사용자 시나리오

1. 영어 학원 선생님은 매일 학생들의 영단어 암기를 확인하기 위해 영단어 시험을 출제해야 한다.

2. 선생님은 채점해드림에 로그인한다.

3. 선생님은 해당 분반의 정보와 학생 명단을 입력하여 새 클래스를 생성한다.

4. 선생님은 채점해드림에 그날에 테스트해야 할 영단어를 입력한다.

5. 채점해드림이 해당하는 영단어 시험지를 만들어주고 선생님은 출력하여 사용한다.

6. 학생들은 시험지에 답안을 쓴다.

7. 시험이 종료되면 선생님이 시험지를 걷는다.

8. 선생님은 학생들의 시험지를 스캔 혹은 촬영하여 채점해드림에 업로드한다.

9. 채점해드림이 시험지에 학생들의 답안을 추출하여 채점을 수행한다.

10. 채점 결과를 통해 학생이 틀린 문제를 따로 분류해 주고, 학생들의 성적 통계를 계산한다. (학생들의 평균 점수, 편차, 높은 오답률을 보인 단어 등 )

11. 이후 틀린 문제 데이터에 따라 채점해드림이 영단어 시험 맞춤 출제를 지원한다.

위 과정을 수행하기 위해 필요한 기능을 정리해 보았습니다.

* 단어장 입력기능
* 단어장 불러오기 기능
* 시험 문제 형식 제공 기능
* 시험지 생성 및 내보내기 기능
* 단어장에 기반한 일반 출제
* 학생 통계에 기반한 맞춤 출제
* 답안지 이미지 인식 기능
* 채점 기능
  + - (단어일 경우) 유의어 판별 기능
    - (문장일 경우) 의미 유사도 판별 기능
* 시험 결과 통계 제공 기능
* 데일리 테스트 제공 및 랭킹 기능

# 구현 계획 및 고려사항

이전 기획과 마찬가지로 사진 업로드가 필요하므로 모바일, 컴퓨터 모두 사용가능한 웹으로 개발 예정입니다.

영어 단어 시험은 크게 두가지로 진행됩니다.

1. 한국어 뜻이 제공되고, 영어 철자를 맞추어야 하는 문제

영어 OCR은 현재 많은 연구가 이루어졌으며, 현재 99.5%이상의 높은 정확도를 보이고 있습니다.

1. 영어 단어가 제공되고, 한국어 뜻을 맞추어야 하는 문제

한국어 뜻을 맞추는 문제는 다음과 같은 순서로 해결할 예정입니다.

1. Wordnet을 이용하여 정답의 유의어를 찾습니다.
2. OCR을 이용하여 한국어 단어를 읽습니다.
3. 해당 단어를 영어로 번역한 후, 정답 단어와 유의어와 비교합니다.
4. 일치하는 단어가 있는 경우 정답을, 아닐 경우 오답을 반환합니다.

한글 OCR의 경우 다른 언어들에 비해 정확도가 떨어지는 편입니다. 이는 한 음운이 한 글자로 되어있어 구조가 단순하고 알파벳 26자 만을 인식하면 되는 영어와 비교하면, 초성-중성-종성으로 이루어진 한글은 조합의 가짓수가 11,172나 되기 때문입니다. 이중 자주 사용되는 글자만 추려도 2350개(EUC-KR)로 매우 많은 양입니다. 현재 상용화 된 프로그램인 ABBYY사의 OCR의 한글 인식률은 98%로 소개되고 있으며, 2018년 논문에 따르면 98.56%까지 증가하였다고 발표되었습니다. Github의 다른 프로젝트들은 95% 이상의 정확도를 가지고 있습니다.

한국어에는 각 글자마다 비슷하게 생긴 글자들이 있습니다.(ex. 관, 판) 이러한 글자들로 인해 정확도가 떨어지게 됩니다. 한국어 뜻을 맞추는 문제의 핵심은 한국어의 맞춤법 또는 한국어의 철자가 아닙니다. 이 점을 이용하여 채점하는 것에 있어서는 위의 연구 이상의 정확도를

방법은 다음과 같습니다.

1. 한국어 글자들을 모양을 기준으로 벡터 공간에 나열합니다.
2. 위의 4단계를 거쳤을 때 오답이 나왔다면, 벡터 좌표 상 가까운 글자로 변환 후 위의 3,4번 단계를 다시 진행합니다.
3. 벡터 좌표상 거리가 일정 거리 이하인 문자들에 대해서 반복합니다.

영단어의 뜻을 학생이 수기로 작성하는 경우 정답을 여러가지 방식으로 쓸 수 있다는 문제점이 있었습니다. 예를 들어 take an exam은 시험을 치다, 시험을 치르다, 시험을 보다 등 여러가지 방식으로 쓸 수 있습니다. 이를 해결하는 방법을 생각해보았습니다.

1. 한글 뜻을 번역기능을 사용하여 영단어로 바꿉니다.

2. 피시험자의 답변과 출제자의 답변을 영어로 비교합니다. 이 때 각 단어별로 동의어들도 확인하여 정확도를 높입니다. Python 모듈인 wordnet을 이용하여 영단어끼리의 동의어를 비교할 예정입니다.

3. 답이 단어가 아닌 구일 경우 유사도 비교 방법(ex. 맨해튼 유사도, 코사인 유사도) 등을 이용하여 최종적인 답안을 비교하여 일정 수치 이상일 경우 정답으로 처리합니다.

# 기술 스택 및 프로젝트 스케줄

**텍스트, 폰트, 스크린샷, 디자인이(가) 표시된 사진

자동 생성된 설명**

**팀 스케줄**

**텍스트, 도표, 라인, 번호이(가) 표시된 사진

자동 생성된 설명**

**프론트엔드(문서형)**

**텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명**

**백엔드(차선우)**

**텍스트, 번호, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명**

**텍스트 인식 및 채점(장세환)**

**텍스트, 스크린샷, 라인, 번호이(가) 표시된 사진

자동 생성된 설명**