파이널프로젝트 기획안

**과정명: 기업 요구사항 기반의 서비스 개발 프로젝트**

**기획안 작성일자 :**

|  |  |
| --- | --- |
| 팀 명 | 2조 : ordinary   * 일상적이고 평범한, 이용자와 밀접한 서비스를 만든다. |
| 프로젝트 조직  (구성원 및 역할) | 팀장 : 천승진  부팀장 : 황도연  팀원 : 권혁재  팀원 : 백혜수  팀원 : 주혜린 |
| 프로젝트 주제 및 개요  해결하고자 하는 문제  최종 산출물의 청사진 | |  |  | | --- | --- | | 문제점 인식 | 1. 기존 채점 서비스의 경우 키워드, 유사어, 띄어쓰기 등을 제대로 판별하지 못하여 서술형 답안의 채점은 자동화되지 못 함. 2. 자동화 채점을 시도하더라도 결국, 재채점 과정을 거쳐야 하기 때문에 출제자 및 조교의 불필요한 노동력이 소모됨. 3. 서술형 답안을 채점하기 위해 동원되는 인력 및 인프라에 소요되는 비용 발생 비중이 큼. 4. 서술형 채점에 소요되는 시간이 크므로, 국가 공인 시험의 경우 객관식 유형 위주로 출제가 되는데 이렇듯 현실적으로 문제 유형의 다양화를 가져오기엔 어려움이 있음. 5. 채점 자동화가 상용화된 타사의 프로그램이 있긴 하지만, 서술형 자동 채점에는 미흡함이 있고 여러 제약 조건이 따름(제공사의 방식을 따라 해야만 함). | | 주제 | * 서술형 답안을 자동으로 채점해 주는 프로그램을 만들어 업무의 효율성, 예산(비용)의 절감, 채점자의 과도한 업무량 개선, 선생님과 학생 간의 유대 강화, 시험 문제 유형의 다양화 등 보다 편리한 교육 환경의 인프라를 구축한다. | | 최종 산출물(예시 – 채점 프로그램(채끝)) | * 오프라인에서 이뤄진 서술형 답안의 텍스트를 인식시킨 후, 자동 채점 프로그램을 거쳐 결과를 확인하는 방식. | | 최종 산출물(예시 – 웹(서비스 개발 부분)) – 체험 버전  블루프린트 | * ‘Give me answer file’ 버튼을 누르면 답안지 파일을 제출할 수 있음. * 자동으로 채점된 결과에 따라 바로 밑에 숫자로, 문제 수, 맞은 수, 틀린 수가 표기됨. * 하단에 있는 4개의 카테고리는 난이도를 표현할 수 있는 옵션임(언어 및 학년 선택). * 사용자가 직관적으로 인지할 수 있도록 UX/UI를 표현함. | | 기대효과(확산) | * 기존의 채점 서비스와 달리 자동화 시스템을 구축하여 재채점 과정을 최소화할 수 있음. * 재채점 과정이 줄어들게 되면, 시험 채점 전반의 소요 시간 및 비용이 감소됨. * 효율성이 제고되면, 재채점 과정을 도맡아하는 인력들의 업무 환경이 개선됨. * 교육자의 업무 환경이 개선됨에 따라 교육자는 학생들과의 관계 개선 및 수업 퀄리티 상승 등을 기대할 수 있음. * 서술형 자동 채점 프로그램이 안정화되어, 국가 공인 시험에 서술형 문제가 도입된다면 자연스럽게 여러 교육기관에서도 이를 도입하여 활용할 것이라 예측함. | |
| 프로젝트 수행 방향  수행 방법/도구 | |  |  | | --- | --- | | Category | NAME | | Tool | * Jupyter Notebook * Google Colab * Naver cloud * Notion | | 언어 | * Python | | 라이브러리 | * Pandas * Numpy * Matplotlib.pyplot * Scikit-learn * seaborn * tensorflow * Pytouch |  * 데이터 수집 방향  1. 영어 주관식, 서술형 문제(영어 기반 채점 서비스를 만들기 위해 수집). 2. 형태소, 단어 분석의 정확도를 높이기 위해, 품사별로 정리된 국어 사전 데이터 수집. 3. 프로젝트에 앞서 프로그램이 정상적으로 작동하는지 시연해 보기 위해 중학생용 사회, 과학 문제 데이터셋 수집. 4. 원활한 프로젝트 진행을 위해 선행된 비슷한 계열의 프로젝트 및 논문을 수집.  * ENL Study  1. 딥러닝을 이용한 자연어 처리 입문 2. 데이콘, 캐글 등에서 자연어 처리 관련 코드 리뷰 3. 컴퓨터 비전에 활용될 모델 – OCR, CRNN 등  * 전처리  1. 구조 탐색 2. EDA 3. 데이터 시각화  * 모델링  1. 컴퓨터 비전 : OCR, CRNN, EAST, EASYOCR 등 텍스트를 인식하는 다양한 모델들을 살펴보고, 최적화된 모델을 사용. 핵심은 오프라인에서 작성된 글씨를 컴퓨터가 인식하는 것. 2. 자연어 처리 : Word2vec, BERT, GPT-3 등의 모델을 모두 도전해 보고, 성능 비교. 가장 우수한 성능의 모델을 채택.  * 성능 개선  1. 하이퍼파라미터 수정 작업을 거치면 작업 수행.  * PPT 및 발표 준비 * 최종 발표 |
| 프로젝트 추진 일정 | **8월**    **9월**    \* 1주차(8/22~8/26) - 주제 확정, 관련 자료 탐색, 관련 코드 실행, 추가 과제 적용 여부 논의  2주차(8/29~9/2) - 기획안 작성 완료, 데이터 취득 및 전처리. 사용할 모델링 기법(BERT, GPT-3 등) 공부  3주차(9/5~9/9) - 전처리 완료, 사용할 모델링 기법 공부 완료  4주차(9/12~9/16) - 모델링 기법 코드 작성 및 실행, 웹 서비스 구현 작업  5주차(9/19~9/23) - 모델링 기법 실행 완료, 성능 평가지표 작성, 웹 서비스 구현 작업  6주차(9/26~9/30) - 성능 평가지표 완료, 서비스 구현 검증 작업, 발표 준비 및 마무리 |
| 참고 자료 | |  |  |  | | --- | --- | --- | | Name | Tags | 비고 | | 데이터마케팅코리아 | 참고 사이트 | - | | 영작시험지제작프로그램 | 참고 사이트 | - | | 제 7회 투빅스 데이터 분석 컨퍼런스 - 투빅스쿨 (딥러닝을 이용한 수능 영어영역 풀이) | 참고 사이트 | - | | 자연어 처리 주제(참고 용) | 관련 문헌 | Github | | 학생들의 서술형 답안을 자동으로 채점해 선생님들의 반복적 채점 작업을 효율적으로 줄여주는 프로젝트 | 참고 코드 | Github | | 한영 번역 알고리즘 | 참고 코드 | Github | | 네이버 사전 | 참고 사이트 | - | | 영단어 33000개 모음 제공 사이트 | 참고 사이트 | - | | Sklearn – 텍스트를 위한 머신러닝 | 관련 문헌 | 위키독스 | | WWIT - 윗 | 참고 사이트 | - | | Figma | 참고 사이트 | - | | 서비스 블루프린트 | 참고 사이트 | - | | 에듀허브 – 학원관리프로그램 출결관리 OMR 채점 | 참고 사이트 | - | | DACON – 자연어 기반 기후기술분류 AI 경진대회 | 참고 사이트 | - | | DACON – 월간 데이콘 한국어 문장 관계 분류 경진대회 | 참고 사이트 | - | | 영어기반 컴퓨터자동채점 모델과 기계번역을 활용한 서술형 한국어 응답 채점 | 관련 문헌 | - | | 딥러닝 OCR, CRNN | 관련 문헌 | - | |