1. 프로젝트명 : 안숨은 글자찾기

2. 팀원(1조) : 김동주, 안성현, 이정빈, 이한빈, 허아현

3. 금주 진행 사항 요약

가) 금주 진행 기간 : 2022.10.04~2022.10.08

나) 금주 진행 진척률 (실제진행률 / 전체진행률) : 80% / 100%

다) 금주 진행 사항 핵심 요약

<2022.10.04(화)>

1. stt
2. 질의응답
3. 단어 빈도 추출 자격증별 어떤 단어가 많이 나오는지 추출

<2022.10.05(수)>

1. qna기능(stt-크롤링) 웹페이지에 올리기
2. ppt 작성
3. 타임스탬프 기능(유튜브 시간 이동)
4. 추천 동영상(업로더들이 영상에 달아놓은 키워드를 크롤링해서 뽑았고 키워드를 유튜브에 검색해서 추천동영상 검색)

<2022.10.06(목)>

1. stt
2. 질의응답
3. 단어 빈도 추출 자격증별 어떤 단어가 많이 나오는지 추출

-----------------------------------------------------------------------

<2022.09.23(금)>

1. EasyOCR 모델 추가 학습 진행상황(Anaconda prompt 환경)

가) 모델 학습

1) 빈 모델에 처음부터 다시 학습

- 커스텀 모델 생성 시 정확도가 잘 나오지 않음

2. 구현 진행상황

가) 데이터 수집, 정제

1) 텍스트 중복 제거 함수 수정 및 업데이트

나) 웹 구현 (플라스크)

1) 크롤링으로 유튜브 정보 가져오기

2) 전처리 모델(함수)과 웹페이지 연결

다) 버트(해시태그 유사 기능)

1) 빈도수별 키워드 추출 함수 생성

- 버트모델보다 단어 빈도로 키워드 추출하는 것이 낫다고 판단

3. 분석 진행상황

가) 개요 및 현황분석 (진행중)

나) easyocr과 tesseract 비교 (완료)

다) PPT 정리

★ 10/01(멘토링 3회차 질문사항)

1) PySceneDetect 이미지 저장하지 않고 넘기는 방법 문의

2) 서버 문의

24시간 돌려도 되는지

학습데이터 생성 갯수에 한계?

학습 어떻게 진행할지

빈 모델에 새로 학습?

원래 모델에 미세조정하기

미세조정한다면 어떤 학습 데이터를 사용할지 아이디어가 필요합니다…

모델 적용 후 오탈자 수정 가능 여부 확인

플라스크 구현 시간 단축을 위한 서버 사용법 문의

4. 차주 진행 사항 요약

가)특이사항 (팀원들 개인사정, 코로나19 등)

1) 9/24(토) 안성현 개인사정 불참

2) 9/28(수) 인공지능 AI 박람회 참석

3) 10/01(토) 안성현, 이정빈, 이한빈, 허아현 빅분기(필기)시험으로 오후 출석

나) 차주 진행 기간 : 2022.09.27~2022.10.01

다) 차주 진행 진척률 (예상진행률 / 전체진행률) : 65% / 100%

라) 차주 진행 사항 요약

1) 9/27 현황 및 환경분석 자료 확인 및 PPT

2) EasyOCR 학습모델 생성 및 정확도 향상 연구

3) Flask 웹구현 구체화