| 3팀 | | 5주차 진행 현황 (말하는 일기장) |
| --- | --- | --- |
| 2022.09.26 ~ 2022.09.30 |

o 팀원 : 강동현, 정영현, 방수진, 전아영, 성혜경, 지동주, 조혜리

**ㅇ금주 진행 사항**

| \* 금주 진행 기간 : 2022.09.26~2022.09.30 |
| --- |
| \* 금주 진행 진척률 (실제진행률 / 전체진행률) : 75% / 100% |
| \* 금주 진행 사항 요약  ㅇ 주제 선정 (말하는 일기장)   * 음성 데이터 -> 텍스트 데이터 -> 텍스트 요약 -> 감성 분석, 키워드 분석 * 어플 및 웹 서비스 구현   ㅇ 현재 날씨에 따라 배경화면 변경   * openweathermap api   ㅇ google speech api   * STT (speech to text) * TTS (text to speech)   ㅇ 감성 분석   * 감성 분석을 위한 단어의 감성 분류(현재 진행률 100%)   + AI hub 제공 : 학습데이터 단어 14000개, 문장 10000개 감정 분류 * KoBERT : 임시로 진행된 학습데이터 추가 시 정확도가 높아지는 것으로 확인 * 정확도 ( train : 약 88 / test : 약 61) * 데이터 증강 예정   ㅇ 키워드 분석   * KoNLP * WordCloud   ㅇ 어플 및 웹 서비스 구현   * django 웹 구현 (MVT모델 기반) * 사용자 DB 관리(oracle 기반)   ㅇ 최종 ppt 제작 |
| \* 특이사항  1. 모델을 위한 GPU 사용 및 코랩에서 지원하는 gluonnlp 사용을 위해 코랩에서 모델 구현   * 코랩에서의 모델과 로컬의 django 연동 필요   + 해결방안( 1번 진행 중단, 2번으로 진행 완료 )  1. 코랩에서 django 실행 후 모델 구현   from google.colab.output import eval\_js  print(eval\_js("google.colab.kernel.proxyPort(8000)"))   * 1. 마이크 문제   2. db 연동 문제(DB 연동 시 oracle은 지원하지 않는 듯함(?))  1. 로컬에서 django 실행 후 코랩에서 모델의 결과를 로컬로 다운로드 후 연동    1. google drive 로컬 pc로 연동    2. google API로 연동   2. 도메인 설정 및 https 로 권한이 증명된 웹사이트만 허용 가능(보안문제)   * SST 마이크 허용 문제 * TTS 스피커 허용 문제 * 현재 위치 기반 날씨 배경화면 - 위치 기반 허용 문제 * 웹 호스팅 필요   + 금액 발생 예상   + 호스팅 DB 연동시 oracle은 금액이 많이 발생하므로 다른 DB 사용 고려 필요   3. 감성 분석   * LSTM의 한계를 확인   - 초기 모델 : 감성 2개로 분류  - 개선 모델 : 감성을 20개로 분류  - 해당 감성분류를 세분화 하였을 때 정확도가 낮아지는 것으로 확인  --> 최종 모델 선정 : Kobert  - 초기 모델 : 감성 5개로 분류 (train: 약 60)  - 개선 모델 : 감성을 20개로 분류 (train : 86, test :72)  - 해당 감성분류를 세분화 하였을 때 정확도가 증가하는 것으로 확인   * LSTM 모델과 비교 시 Kobert가 더 좋은 모델임을 확인하여 모델 선정   4. 요약 : KoBART의 한계점 발생 (실패)   * KoBART 모델 사용 시 뉴스 및 사실 전달 기반 글의 요약 정확도는 높은 편이나 일상어로 이루어진 글은 정확도가 떨어짐을 확인   → 일상어 요약을 위해 파인튜닝을 시도했으나 인코더-디코더 모델로 양방향 학습이 필요하며, 학습시킬 사전의 양이 턱없이 부족 |
| \* 차주 진행 기간 : 2022.10.4~2022.10.7 |
| \* 차주 진행 진척률 (예상진행률 / 전체진행률) : 90% / 100% |
| \* 차주 진행 사항 요약 : 최종 ppt 제작, Kobert 모델 정확도 확인, 학습데이터 증강 |