**멘토링 회의록**

| **프로젝트명** | 안숨은그림찾기 | | |
| --- | --- | --- | --- |
| **일 시** | 2022년 09월 24일  09:00∼17:50 | **참 석 자** | 데이터에듀 :  정기영 연구원 |
| **장 소** | 한국품질재단  한가람IT 교육센터  강의실(202호) | 1조 : 김동주, 안성현,  이정빈, 이한빈,  허아현 |

| **멘토링 내용 (요약)** |
| --- |
| **【 멘토링 받은 내용 】**   * 아이스브레이킹 (근황토크) * 공부해두면 도움 되는 것   + 크롤링, API(네이버, 다음 등), JSON파일 정제   + <https://aiopen.etri.re.kr/>   + 아나콘다 환경 * 진행상황 공유 및 현황 브리핑   + <기영님질문>   + 중복제거가 무엇인가?   + 실시간인가? 구현 시간이 오래걸리니까 분석된건 저장해두고 불러오면 시간이 단축됨. 실시간이 아니라 스토리지에서 분석끝난 데이터를 가져오는 것(DB), JSON파일로 저장해두고 사용자는 파일만 불러와서 볼 수 있음   + 실시간일때 느리니까 유튜브조회 크롤링만 실시간하고 나머지는 파일링   + 유사도측정 함수 자체의 수치 조정이 가능한지 알아보기?   + 유튜브 크롤링할때 셀레리움을 썻는지, 유튜브크롤링 패키지를 썻는지(하루만번제한있음) : 정빈님이 유튜브 API를 녹여서 연습해보면 좋을 것 같음   + STT도 한번 고려하면 어떨까? 해봤다 정도라도(youtube-dl)   + 빈도로 했을때는 너무 불필요한 단어가 많이 나올 수 있으므로 차라리 너무 많이 나온 단어를 제외하거나 적게 나온 단어를 추출해보기, 불용어사전, 다항모델을 이용해보기(실무에 굉장히 도움이 됨) // 키버트가 좋다고 판단한 이유는 남겨두기(PPT)   + STT : 타임스탬프가 다 남고, 자막은 큰 타임스탬프만 남겨도 되는 방향도 추천함 (두가지를 잡으면 고민하는 점이 해결될 것)   + 기술별 마감 기한 정리해서 공유   + 영상추천을 추가하면 어떨까? / 카카오아레나 경진대회 1등 솔루션(서적)에 상품분류 추천 1등한 사람의 사례가 있음   + 책 검색은 기능은 없더라도 인터페이스에 추가하는 것을 추천   + 인터페이스도 사용자가 편리하게 의견을 내서 새롭게 만들어보는 것은 어떨까? 탭도 여러개 나누어서 인터페이스 설계도 고려해보기   + 두잇장고 : 퍼블리싱방법 있으니 참고하여 시도해보기   + 불필요한 텍스트를 제거하는 알고리즘이 있을까?   + <추천내용>   + 링크나 다운만 받아져있다면 다른 모든 영상이 가능한 장점이 있다   + 사용자가 사용할 것인가/필요한가? 고민이 필요함 PPT에 어떻게 녹일 수 있을지   + 동주님께 추천 : 위키독스에 키워드추출도 있지만 문장 요약이 있음 코드만 복붙해서 구현해봤으면 좋겠음(ex 세줄요약), 또는 댓글 긍부정(긍정%, 부정%) 댓글긍부정과 좋아요싫어요 비교   + 인식 텍스트 정리 어떻게 할 것인지 : 가져온 텍스트로 새로운 UI구성한다던지, 문단 정리 * <해야할 것> * 텍스트를 어떻게 배치할 것인지 * 사용자에게 어떻게 보여줄 것인가     **【 요청 사항 】**  ○ 질의사항  1) 서버 24시간 사용 가능한지   * 사용가능함   2) 결과가 제대로 나오지 않을 때 학습데이터 생성을 어떻게 하면 좋을지   * 이사님과 회의 후 말해주신다고 함 |
|