

회 의 록

1. 회의 개요

프로젝트 명	융합프로젝트		
회의 주제	주제 정하기		
회의 일자	22/04/12		
참가 인	성낙원, 조성현, 조완제, 임지윤, 홍동수, 홍지수	담당자	5조

2. 회의 내용

Title	Contents
회의내용 및 결과	<p>9:00~11:00 : 새로운 주제 관련 회의 11:00~12:00 : 주제 관련 강사님께 브리핑 및 도서 주제로 90% 확정</p> <p>13:00~16:00 :</p> <ul style="list-style-type: none"> - 사이언스팀 : 책들의 유사도 분석 테스트 - 엔지니어링팀 : 최종 서비스에 포함될 기능 및 필요 데이터에 대한 회의 <p>16:00~17:30 : 사이언스팀과 엔지니어링팀 진행한 내용 공유 및 향후 방향에 관해 회의</p> <p>사이언스 세부내용 :</p> <p>1.Q&A :</p> <p>Q1. 고전 스테디셀러 sample1이랑 비슷한 카테고리 별로 이루어진 sample2로 돌려봤는데, 유사도가 매우 낮게 나와 어떤 방식으로 접근하면 좋을지 조언을 얻고 싶습니다.</p> <ul style="list-style-type: none"> - 책 소개 뿐만 아니라 다른 요소 같이 고려(장르, 작가)해야 할 것 같고, 추천으로 가는게 어떤가. - 장르 내에서의 규칙을 찾아라. <p>Q2. soynlp 라이브러리 중에 wordextractor 로 응집력을 계산해서 복합명사를 다루어보려고했는데, 단어가 응집되지 않습니다.</p> <ul style="list-style-type: none"> - soynlp보다 konlpy나 NLTK를 활용해 보아라. <p>Q3. 단어 빈도수 기반이 아니고, 문장의 흐름이나 분위기를 읽어내는 방법이 있나요?</p> <ul style="list-style-type: none"> - ‘워드임베딩’을 하면 캐치할 수 있다 - 문서 벡터를 이용한 추천 시스템(Recommendation System using Document Embedding) 참고 <p>2. 코사인 유사도 측정 결과 비교 결과 :</p> <p>-서양고전문학 30개, 히가시노게이고 소설 + 자기계발서 20개, 코사인 유사도 비교 ○ 가장 높은 유사도 : 고전문학 0.14, 소설+자기계발서 0.24</p> <p>3. 기타:</p> <p>-한국어 형태의 해결 방향 : konlpy 의 okt의 형태소 분석기 활용 방법</p> <p>엔지니어링 세부내용 :</p> <ol style="list-style-type: none"> 1. 추천서비스 <ul style="list-style-type: none"> ○ 키워드(관심사) 기반 추천 ○ 유저가 읽었던 책 기반 추천 ○ 시간대에 따른 추천(낮/밤) ○ sns언급량 기준 추천 ○ 짧은 책 추천 2. 검색기능

	<ul style="list-style-type: none"> ◦ elasticsearch <p>3. 커뮤니티</p> <ul style="list-style-type: none"> ◦ 내정보페이지(히스토리, 찜한 목록 등) ◦ 리뷰 등록 또는 별점 매기기 기능 <p>4. 추가적으로 공부할 기능</p> <ul style="list-style-type: none"> ◦ elasticsearch ◦ airflow ◦ spark streaming ◦ 분산저장
기타사항	<p>한글 텍스트분석 공부 + 도전 + 각종 코드 실험</p> <p>이동현 강사님: 엔지니어링이 쓸 수 있는 기술들을 추가적으로 사용하세요. 예를들어, Airflow, kafka, sparkstreaming, Elastic Serach와 같은 것들. 아니면 전처리를 많이 할 수 있는 데이터를 사용할 것</p>