

팀 회의록

팀명	데미키들	참석자	강영민, 김성윤, 이성일, 이호건
일시	2023. 05. 30 (화) 22:00 ~ 23:30	장소	비대면 회의 (Google Meet)
회의 내용	<p>안건</p> <p>서점 데이터와 도서관 데이터를 비교한 결과 상관관계가 두드러지지 않고 우리 팀이 기대하는 양상이 나타나지 않는다.</p> <p>회의 과정</p> <p>원하는 기대효과: 이슈에 의해 도서 트렌드가 변경되었다.</p> <p>해결방안:</p> <p>여러 주제 분류가 실제로는 이슈와 연관될 가능성이 낮을 것 같다. 따라서 이슈와 반응이 직결될 것 같은 사회과학 주제를 탐구하자. 또한 기사데이터를 추출하는 과정에서도 사회과학안에서 경제라는 영역이 두드러진 데이터를 뽑을 수 있다 생각되어 '경제'를 세부주제로 정한다.</p> <p>회의</p> <p>문제: 이슈와 도서 데이터 간의 상관관계를 어떻게 측정할 것인가?</p> <p>해결방안:</p> <p>원하는 기간의 기사 데이터들을 불러와 키워드를 추출하고 가장 빈도가 높은 키워드를 해당 기간의 키워드로 설정한 후 해당기간 도서데이터와 비교한다. 이 때 도서 데이터 역시 원하는 기간의 책들의 줄거리에서 키워드를 추출한다.</p> <p>문제: 줄거리데이터는 어떻게 가져올것인가?</p> <p>해결방안:</p> <p>yes24라는 플랫폼에서 줄거리 데이터를 제공한다. 그 데이터를 활용하고 축소하여 진행한다. 줄거리 데이터 분류는 군집화를 통해 가장 빈도가 높은 키워드를 추출한다.</p> <p>문제: 키워드가 여전히 일치하지 않는다면 어떡할 것인가?</p> <p>해결방안:</p> <p>그 자체로도 유의미한 결과로 해석될 수 있다.</p>		

회의 결과 -

데이터 수집(네이버 기사 크롤링, yes24 데이터와 매칭)
모델 고도화 (선택사항)
도서관, 서점 데이터 처리 => 주제만 뽑아내자

증빙
사진

