

회 의 록

1. 회의 개요

프로젝트 명	융합프로젝트		
회의 주제	주제 정하기		
회의 일자	22/04/13		
참가 인	성낙원, 조성현, 조완제, 임지윤, 홍동수, 홍지수	담당자	5조

2. 회의 내용

Title	Contents
회의내용 및 결과	<p>-09:00~14:00 : 강사님께 컨펌받기 위한 준비</p> <p>◦ 사이언스 : 참고 : https://wikidocs.net/102705</p> <ol style="list-style-type: none">1. Cleansing & Normalization -> (정규식, 특수문자제거, 불용어제거, 정규화(너무낮은빈도를 지우는것))2. ckonlpy : add_dictionary3. Stop words4. Okt morphs5. tokenizer6. counter vectorize -> Word2Vec<ul style="list-style-type: none">- LSA, HAL, Word2Vec, Glove, NNLM, RNNLM, FastText 이런 방법중에 일단 Word2Vec으로 선택함.- document term matrix -> DTM- 코사인 유사도를 측정해서 유사도가 높은순위의 책을 추천함. <p>◦ 엔지니어링 : - 이미지 :</p>

1. 책 추천 기능

- 책 추천은 우리가 직접 번역한 기준으로 필터링한 책들을 유사성 특성을 통해 추천한다.

✓ 기준: ① 리뷰수 (양라단), ② 개미상

- 2) 평점 4.0 이상 (모든 플랫폼의 평점 평균)
- 3) 스토리가 흥미하는 책?

} 최소 1000개 이상의 책을 가져올 수 있는 기준은
도야금지 활용하기

- 엔지니어링 데이터 수집 눈омер

(△ 현재로서는 알라딘이 가장 유망)

첫번째: ① 해당 책의 수 많은 플랫폼의 API를 활용해서 모든 책의 정보를 수집한다

· 수집하는 정보: 책제목, 저자이름, 책내용, ISBN, 책평가, 별점 등등..

② ISBN or (책제목, 저자이름)을 다른 API에 입력해서 추가정보를 가져온다.

· 수집하는 추가정보: 별점, 그 외에 플랫폼마다 다른 정보라든지 그 플랫폼에서만 구할 수 있는 정보

③ 표준으로 하나의 테이블로 만든다. <all_books>

예시)	ISBN	책제목	저자	책내용	별점 1	별점 2	별점 3	...
	1000	고고의 이노	최은영	~~~	4.5	3.8	4.2	
	1023	코라벌	도토령(?)	~~~	4.7	4.6	4.9	
	6666							

각각의 플랫폼에서
기여한 별점

④ 최종 테이블에서 추천 기준 (참고 ✓)을 기반으로 필터링한 추천 테이블 생성
<all_books>

예시) <rcm_books>

ID	ISBN	책제목	저자	책내용	평균별점	...
1	1000	고고의 이노	최은영	~~~	4.16	
2	1023	코라벌	도토령(?)	~~~	4.73	
3						

:

두번째:

① 각각의 API마다 수집할 수 있는 모든 책의 정보를 가져와서 별개의 테이블로 만든다.

예시) <aladin>

ISBN	책제목	저자	별점	리뷰	...
1000					
1023					

<elevenstreet>

ISBN	책제목	저자	별점	장르	...
1000					
1034					

<national library>

ISBN	책제목	저자	외국	보관수	...
1000					
1023					

JOIN을 통해 하나의 테이블로 생성
(ISBN 또는 (책제목, 저자)을 기준으로)

이후 단계는 1번 방법 반복

Outer/Inner
연결

- 14:00~15:00 : 강사님께 준비 내용 발표

○ 피드백 세부 내용 :

- 엔지니어링

- 테이블 최대한 정규화 해놓고 후에 필요에 따라 합칠 것

- 플랫폼마다 테이블 만들어서 join하면 속도가 느려질 수 있으니, 이런 부분 감안해서 ERD 및 DB 구성하기

- 사이언스

- 개인화 된 추천을 위해 출생년도, 연령대, 성별, 직업, 전공에 대한 기본적인 정보 필요

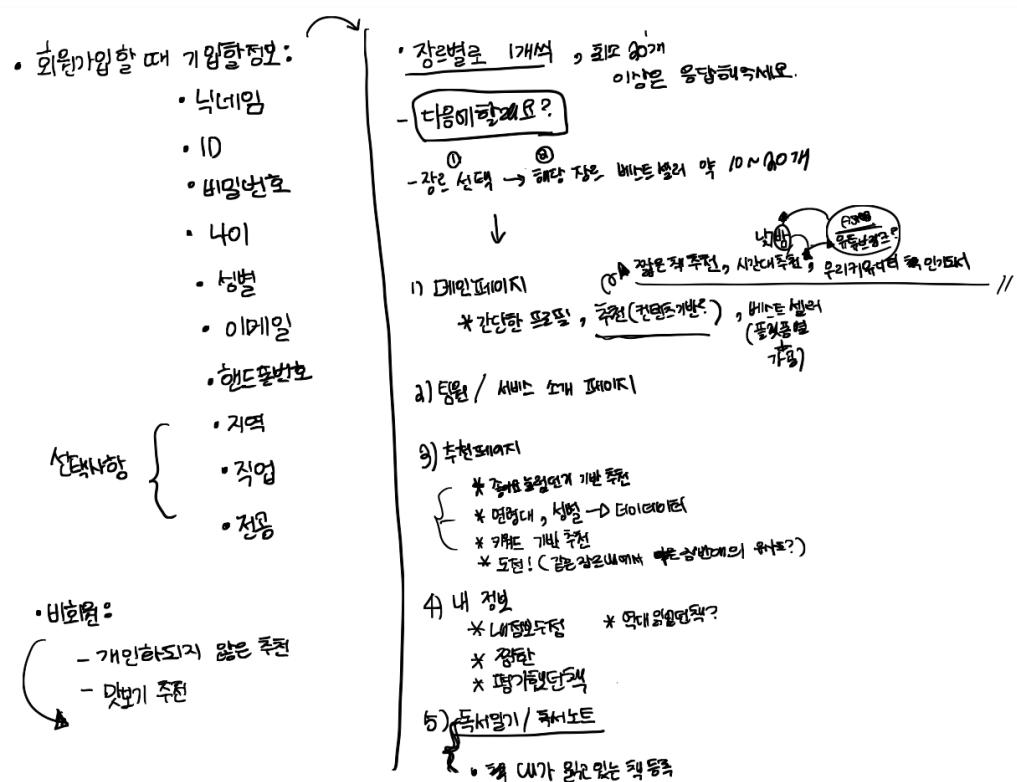
- 책을 카테고리 및 장르 별로 분류 (대분류_소분류_토픽으로 분류 > 토픽 모델링)

- 텍스트 기반의 군집분석 (클러스터링) 혹은 분류

- 벤치마킹 - 기존의 서비스와의 차별점 제시

- 처음에는 컨텐츠 기반의 서비스로 시작하지만, 나중에 유저기반 데이터로 서비스하기 위해 일반적으로 사용되는 유저 선호도 조사 설문을 도입하는 방안 고려

-15:00~17:00 : 화면 정의서 구상
이미지 :



17:30~17:50 어벤져스 역할 분배

성현-아이언맨,
낙원-캡틴,
동수-블랙위도우,
지윤-헐크,
완제-호크아이,
지수-토르

CONFIRMED (づ。o_~o。)づ. (o♥_♥o).
ADVICED (o_~o✿).

기타사항

내일 목표 :

- 화면구상서
- 기획서
- WBS