1. ppt 자료수집 + 배경

속성을 선택하게 -> 3,4 로 줄였지만

음식의 맛

음식의 질 -> 음식의 속성 2

접근성 -> 시설및 편의성 1

-=> 그 상위에 있는 거만 나온다라는 문제점

가시성있게

해시태그에서 선택하는걸 하면

홍콩반점 -> 중국집 위주로 해줬구나 ~~ (음식의 속성에 맞게 해줬구나..)

4개 음식의 속성부분(음)

구석구석의 추천

1. 관광, 음식
2. 해시태그를 보여줘야 하고
3. 리뷰 속성에 대한 것도 해시태그 형태로 보여줘야 겠죠?
4. 투어

구석구석 해시태그

마지막 우리가 관광지를 추천하는 것은 각 리뷰 속성(이런점이 좋았어요를 범주화)에 따라 추천

-> 하지만 구석구석 해시태그와 중분류(체험, 테마, 등) 관광지를 선택한 후 추천하면?

-> 우리는 리뷰 속성에 따라

1. 관광 대 음식

2.3 . 1) 각 대분류에 따른 리뷰 속성(해시태그로 -> 네이버 리뷰를 범주화한 것)

2) 중분류 선택 (테마, 체험, 역사 등등)

4. (잘모르겠음) 관광지 선택?

5. 그다음 뭐 리스트든 관광지 5개든 정보와 함께 보여주기

채그로

-> 만약 해시태그가 # 쌈밥 # ~~~~ 이렇다면

-> 사용자는 이것을 보고 아 쌈밥 약간 어 좋아 이런 쪽으로 해주겠지? 라고 생각할텐데

우리는 그 해시태그에 따라 쌈밥과 관련된 음식점을 추천해주는 것이 아니라 리뷰 속성, 음식의 속성, 편의성, 가격등을 기준으로 추천을 해줌

-> 우리가 하기로 한 과정으로 간다면 채그로는 한식, 그래 한식으로 필터링해서 추천

근데 결론은 음식의 맛이나 편의성 등이 유사한 걸로 추천

우리가 가지는 필터는

궁금한 점 ? 이런점이 좋아요 범주화를

태영님 : 피쳐 속성을 선택하게 해서 보여주는 게 좀 더 좋지 않을까라는 생각

순서를 리스트 형식으로 보여주고, 사용자가 직접 선택하게 하는 방향

수린님: 피쳐 속성으로 선택하는게 더 괜찮, 수치화된 데이터가 피쳐 속성이기 때문에

확정

1. 관광 음식으로 나눔
2. 중분류로 나누기(테마, 체험, )
3. 리뷰 속성을 해시태그화
4. 순위로 리스트 형식

* 속성 기준으로 ,

to do

* 스트리밋에 어떤 거 들어가야 할지
* 시설 및 분위기, 음식의 맛 -> 좀 그럴싸한 해시태그로

-> 앱구현 3명 - 가희님과 성원님

-> 발표준비 (틀을 만들어 놓자) - 발표 정리

- PPT(성원님 X), 일요일 OR 월요일

분류 정확도

- 텍스트 데이터 분석

- 사진과 정보

- 분류 확인

우선 순위. 리뷰 가지고 오는 크롤링.

API 사용해서 사진가져오기

PPT 틀 만들기

-> 태영님, 수린님, 지우

**# 프로젝트 소개**

* 프로젝트 주제
* 사용자 텍스트 분석을 통한 관광지 추천 시스템
* 주제 선정 배경
* 코로나 19 이후 관광 수요 증가
* 기존 여행지 추천 프로그램의 한계 > 대부분 숙박, 유료 관광지 예매 또는 여행 패키지 판매를 위한 추천 프로그램이다 > 따라서 자유 여행을 즐기는 관광객은 본인이 여행 계획을 짜기 위해 인터넷에서 다양한 매체를 뒤져야 하는 불편함이 존재 > 이러한 불편함 해소에 도움을 줄 수 있는 추천 프로그램을 개발해보자
* 그렇다면 개인 관광객은 여행 계획을 짤때 어떤 경로를 활용하는가?
* 1순위 - 블로그/커뮤니티(18%) / 2순위 - 가족/친구(16.6%) / 3순위 - 유튜브(15.9%)
* 해석 >> 주로 인터넷 환경에서 정보를 찾는데 블로그랑 커뮤니티가 1위
* 따라서 우리는 네이버 블로그 해시태그 및 네이버 지도의 사용자 리뷰를 가져와서 이를 토대로 분석을 진행해 보겠다.
* ppt 첨부할 내용

1. 배경

* 연도 별 관광 추이
* 타겟을 서울로 잡은 이유 (서울의 관광객 수가 가장 많음)
* STP
* 성원님 내용 참고
* 우리가 만들 프로그램이 무엇인지를 보고

(이런 프로세스로 가면 약간 가볍게 우리가 무엇을 할지)

개인으로 여행하는 방문객들의 정보를 얻는 방법 -> 인터넷 기반

인터넷 커뮤니티의 내용을 추출한 후 그 로 인해 분석 -> 프로젝트 시작

1. 구현 방식 및 프로세스 제시

* 목차 제시 WBS, 개개인 역할

1. 데이터 수집 방법

* 데이터 수집 방법 코드 + 데이터 형식(크롤링, API(위도경도))

1. 전처리

* 텍스트 분석(분류 방법) + 논문
* 네이버 서브클래스 기준 분류로 할때 일치율
* 단어 빈도 등등을 했는데 이 과정을
* 빈도 수에 대한 히스토, 바그래프, 워드클라우드
* EDA(박스 플랏, 왜도 첨도 등등)
* EDA에서 발생한 문제점을 해결하려고 노력함 로그변환 등등을 함

(MIN-MAX가 차이가 너무 크다는 문제 )

-> 로그변환 후 군집 시행 -> 공통적인 속성을 찾으려 했는데,

결론 ) 해시태그 분류하여 우리의 분류기준에 맞춰해줌

근데 이건 임의라 논문이나 근거를 토대로 분류

그거를 더 보완하고자 네이버 서브클래스 기준 분류로 할때 일치율

1. 모델링

* 도식화된 프로세스(우리가 고민한거)
* 프로세스 진행하며 주요깊게? 고민한 거 위주로

1. 앱 시현
2. 결론

* 진행시 셀프 디스(아쉬운 점을 우리는 다 알아 근데 우린 시간도 없고, 배운 것도 없다)
* 아쉬운점

1. 추후 과제

* 아쉬운점(성능평가를 위해서는 금전적인 이슈로 재정상태 악화)
* 추후 보완 과제를 좀 꼼꼼히
* 사용자 평가를 받는 게 추가 과제

1. 개발 후기 및 느낀점
2. 논문 참고

12일 날 제출 기획안 + PPT 체출

PPT