텍스트 분석 기반 인천광역시 세대별 여행 추천 SERVICE

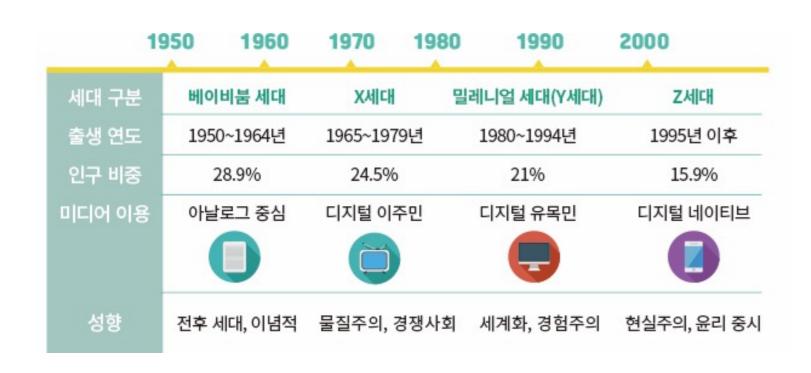






안동수 장하영

01 제안배경 및 인천e지 개선점



각 세대별 추구(선호)하는 성향이 다름



인천 e지 앱에는 ai 추천 서비스는 존재하지만, 세대별 추천 여행서비스가 없음.

=> 세대별로 선호하는 여행지를 텍스트 데이터로 분석하여 세대별 여행 추천서비스 아이디어 제안

02 데이터 추출 및 전처리 과정

1. 세대별 관련 키워드 선정(검색 필터링 시 이용)

19세미만 : 청소년, 방학

20-30대: 90년대생, 청년

40-50대 : 인천여행,중년

60대이상 : 인천여행, 어르신



2. 네이버 Api를 활용하여 세대별 블로그 데이터 추출 (키워드 포함된 데이터 추출)

[python_code]

query = urllib.parse.quote(input("검색 질의: 인천여행|청소년|방학 ")) sort = 'sim' #유사도순으로 정렬

관련도순 1000개 blog_data 추출

NAVER Developers

NAVER



<수집 및 활용데이터>

- 인천 여행 관련한 네이버 블로그 리뷰 데이터

3. 불용어 사전

-불용어 리스트를 만들어 관련없는 단어들 제거

의존명사: 것, 듯, 쪽 등

관련없는 단어: 이용, 아침, 추천, 일본, 서쪽 등



4. 형태소 분석을 통한 명사 추출

Noun

- Konlpy 패키지의 Mecab을 이용하여 명사추출

Ex)

세번째로 간 곳은 인천 송도센트럴파크공원... 여행은 정말 좋은 취미다.

→ 송도, 공원, 여행, 취미

Why?

세대별로 관심을 갖는 카테고리를 찾아보기 위해

02 데이터 추출 및 전처리 과정

- 파이썬 라이브러리를 활용한 인천 여행지 구글 맵리뷰 데이터 크롤링
 - 크롬 드라이버와 파이썬 selenium과 pandas를 활용한 인천 여행지 구글 맵리뷰 데이터 수집
 - (카테고리별 관광지) 관련성순 top 100 리뷰 데이터 추출





WEB Sé Selenium Pandas





- 2. 형태소 분석을 통한 명사 동사 형용사 추출
 - 관광지의 #해쉬태그 추출을 위해 동사 명사 형용사 추출
 - Okt 형태소 분석기 사용







리뷰데이터 리스트 형태로 저장

['우리나라에 우주박물관이 이렇게 잘 만들어졌었나 하고 놀랐습니다. 시설도 깔끔하고 내부에 아기자기하게 로켓모형등이 잘 전시되어 있습니다.'

'주차장도 잘되어 있고 입구부터 동선을 야외,실내 바닥에 선으로 유도해 두었습니다 실내에 어린이 체험용기구가 6개 정도 있는데 잘 만들었습니다']

<수집 및 활용데이터>

- 구글 관광명소 리뷰 Data





우리나라에 우주박물관이 이렇게 잘 만들어졌었나 하고 돌랐습니다. 시설도 깔끔하고 내부에 아기자기하게 로켓모형등이 잘 전시되어 있습니다

['우리나라', '우주', '박물관', '만들어지다', '하다', '놀라다', '시설', '깔끔하다', '내부', '아기자기하다', '로켓', '모형', '등', '자다', '전시', '되어다', '있다']

03 분석

💬 텍스트 데이터 분석

세대 별 선호하는 여행 키워드가 다름



세대 별로 카테고리(키워드) 선정

10대(19세미만)



20대-30대



40대-50대



60대 이상

Keyword

체험 프로그램 활동 교육 문화

활동적인 프로그램, 체험, 교육, 문화 활동과 관련된 여행 키워드 선호

10대 카테고리:

[농어촌체험, 액티비티, 교육, 자연, 문화활동]

Keyword

카페 추억 친구 경험 사진 맛집 문화

친구들과 추억을 쌓을 수 있고 맛집과 카페 그리고 사진 찍기 좋아하는 장소를 선호

20-30대 카테고리 : [맛집, 카페, 액티비티, 사진명소, 시장, 문화]

Keyword

카페 자유 친구 섬 사진 가족 공원 꽃

친구, 가족들과 함께 자연(꽃, 공원, 섬)을 자유롭게 투어하는 것을 선호

40-50대 카테고리 : [카페, 사진명소, 자연, 가족나들이]

Keyword

맛집 힐링 공원 가족 산책 코스 역사

가족들과 공원을 산책하며 힐링을 느끼고 역사 코스 등을 선호



60대이상 카테고리 : [맛집, 역사, 시장, 자연, 힐링]

03 분석



여행지 카테고리 별 선호하는 해쉬태그 선정 (히트맵 사용)

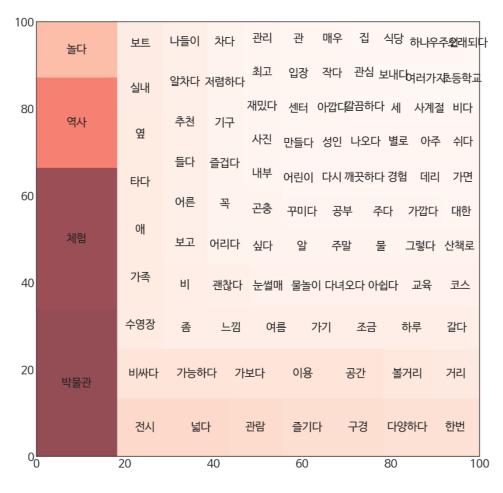
액티비티



액티비티 관련한 관광명소 크롤링 후 데이터 분석

탑승하고 즐기다 # 재미있는 놀이 # 시원하다

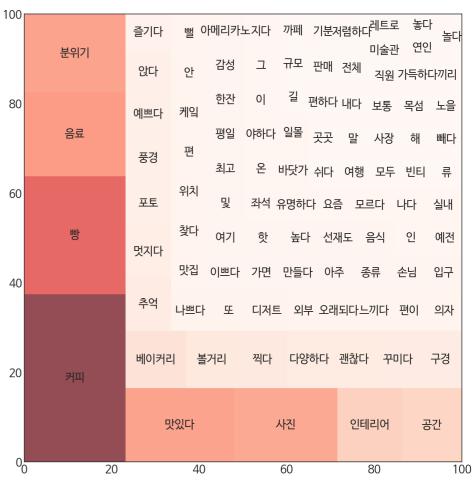
교육



교육 관련한 관광명소 크롤링 후 데이터 분석

체험하며 배우는 # 알찬 볼거리 # 다양한 박물관

카페



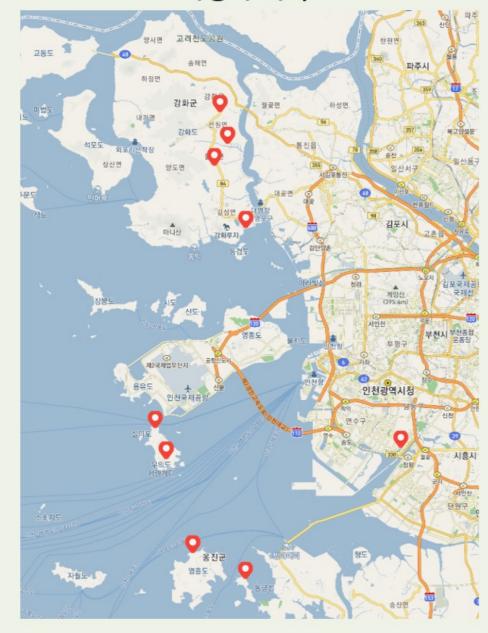
카페 관련한 관광명소 크롤링 후 데이터 분석

분위기 있는 # 사진찍기 좋은 # 인테리어가 예쁜

04 추천 서비스 대시보드

인천, 어디까지 가봤니?

여행지 위치



10 2030 4050 60

농어촌체험

액티비티

교육

자연

문화활동

- ✓ 늘솔길공원 양떼목장
- ✓ 선재도 어촌체험 휴양마을
- ✓ 강화도 자연체험농장
- ✓ 도래미마을
- ✓ 큰무리어촌체험마을
- ✓ 영암어촌계 체험장
- ✓ 박정희 천연염색 체험장
- ✓ 무의 까치놀섬마을
- √ 불은농촌문화센터

테마별 키워드

가족과 함께 즐기는 # 캐는 재미가 있는 # 놀이터

10대 워드클라우드



Best Review



늘솔길공원 양떼목장



선재도 어촌체험 휴양마을



강화도 자연체험농장

04 추천 서비스 대시보드 시현영상



시현영상 링크 -> https://drive.google.com/file/d/1rHwoPEA3ACDHVDBRXpxmUmOA-fjqcNos/view?usp=sharing

05 실현가능성 및 기대효과, 한계점

■ 실현가능성 :

텍스트 데이터를 기반으로 만들었기에 실용적이고 실현 가능성이 높다고 보고 있음.

■ 기대효과 :

세대별 선호하는 서비스를 알맞게 제공 -> 그로 인하여 인천e지 추천 서비스에 대한 만족도 향상 기대

■ 한계점 :

장소리뷰가 많은 지역이 있고, 적은 지역이 있어 크롤링에 한계가 있었음. 블로그 데이터만 사용하여 분석에 아쉬움이 조금 있음. 인스타그램 데이터와 같은 sns 데이터를 더 수집한다면 보다 정확도 높은 분석이 가능할 것 같다.

Thank you