

OJO
FINAL PROJECT

이민성, 최원희, 박성하, 김정명, 김태리

SNS기반
감성숙소 추천서비스

목차



프로젝트 개요

구성원 소개

주제 선정 배경

숙박서비스 현황

서비스 목표

서비스개발 과정

데이터 수집/처리/적재

데이터 자동화

데이터 분석

-EDA

-NLP

-군집분석

-유사도분석

데이터 분석 결과

서비스 소개

기능분해도

UI 설계도

서비스 소개

결론

기대 효과 및 시사점

개선점

구성원 역할

개발 후기



프로젝트 개요

구성원 소개

주제 선정 배경

감성숙소 정의

시장조사

문제 인식

프로젝트 목표



Data science

OJO

김정명

박성하

최원희



OJO

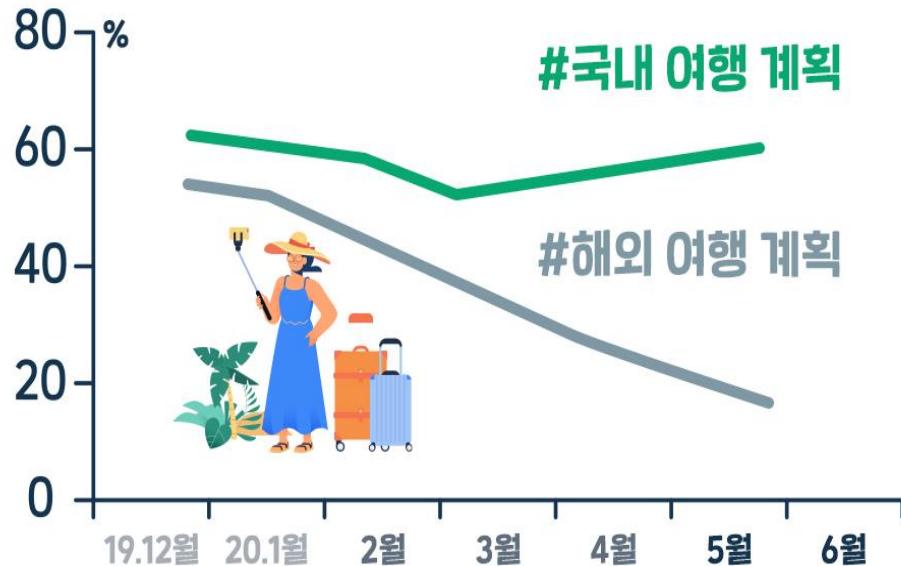
이민성

김태리

Data Engineering

주제 선정 배경

여행은 해외 대신 국내로

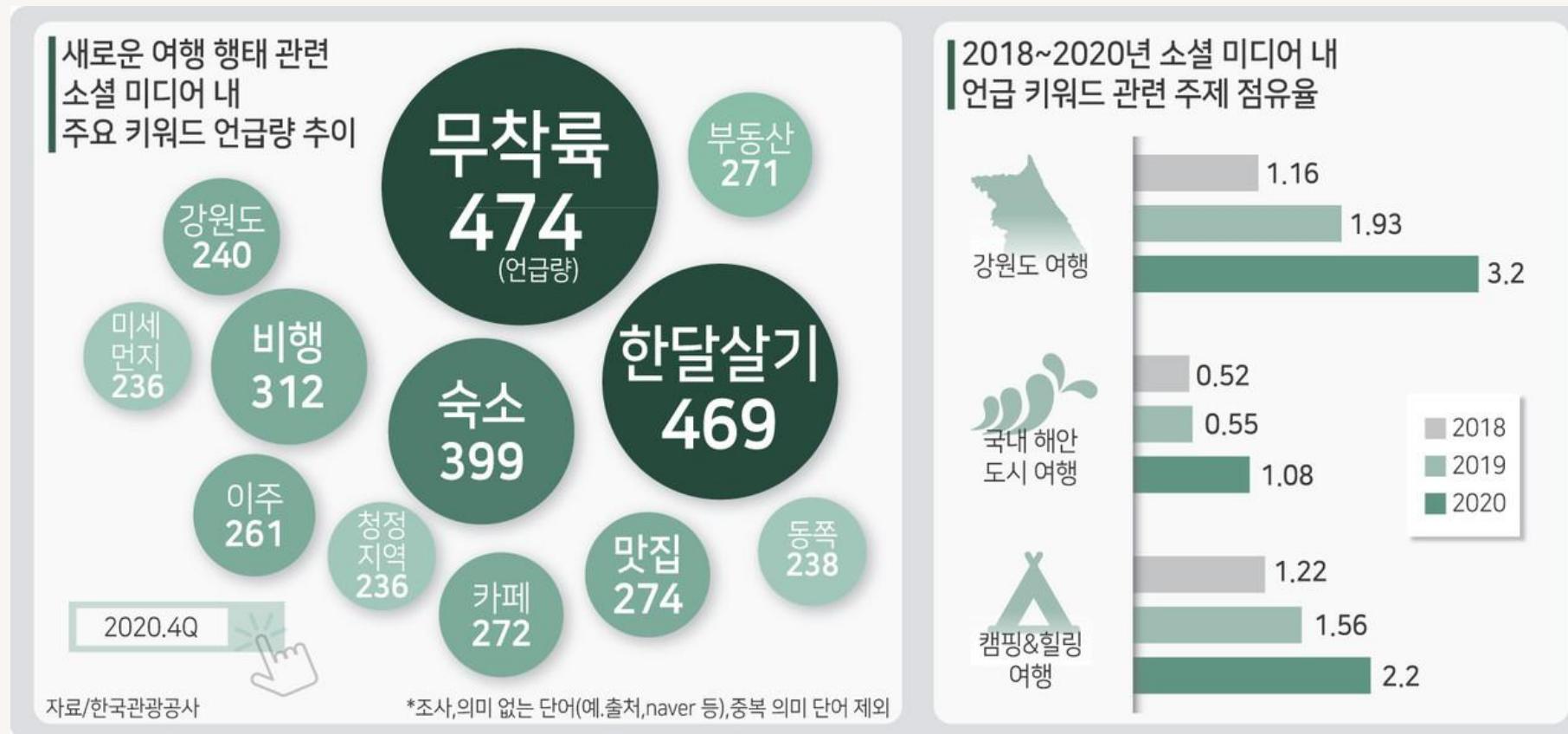


(출처 : Consumer Insight 주간 여행행태 및 계획 모니터링 조사 결과)

코로나 19로 발생한 여행 산업의 변동

- 여행트렌드분석 업체에 따르면
- 지난 상반기 국내여행 심리가 [사상최고치](#)를 기록(2015년 8월 이후)
- 국내 여행지 중 수도권을 제외한 모든 권역이 코로나 이전보다 [4~5%p](#) 상승
- 코로나19 속에서도 여전히 여행에 대한 관심도는 높음

주제 선정 배경



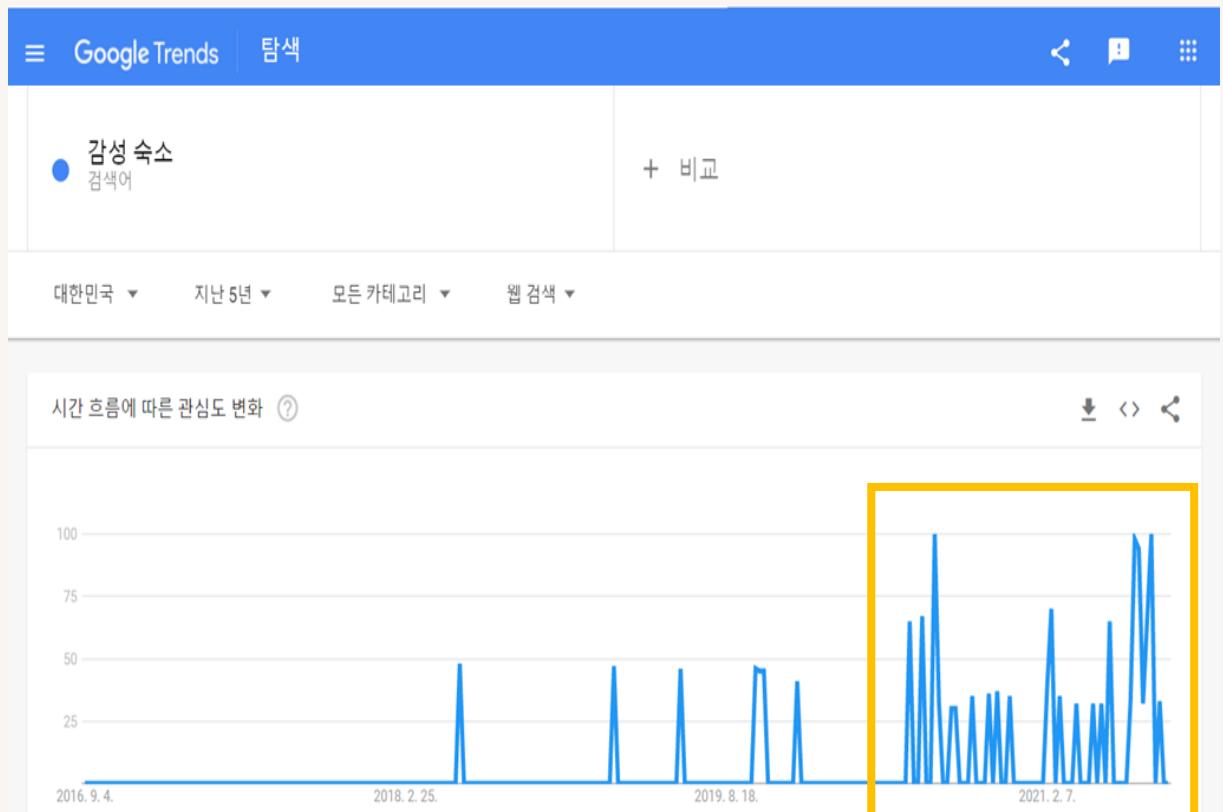
힐링여행에 대한 관심이 지속적으로 증가

주제 선정 배경



코로나로 인해 다양한 여행지에 대한 관광 수요 증가

주제 선정 배경



코로나로인해 국내 여행에 대한 관심도가 높아짐

“힐링”, “언택트”를 키워드로 한 여행형태가 출현

또한 소확행·워라밸 등 라이프 스타일의 변화에 따라

소규모 여행이 증가하고, 감성숙소에 대한 관심도가 높아짐

주제 선정 배경



출처: 호텔스닷컴



확실한 개성

개성있는 인테리어와 특색으로
이용자에게 독특한 경험과 느낌을 줄 수 있는 곳



목표 서비스층

20~30대 소규모 인원(2~4인)의 여행객으로
독립적인 여행을 희망하는 사람들



독특한 운영전략

SNS 마케팅을 활용하며, 인기관광지부터
비인기관광지까지 다양한 위치에 입지하는
숙박업체

숙박 서비스 현황

	예약	가격비교	분류기준	테마별 숙소 추천
호텔스컴바인	0	X	건물유형	시설
여기어때	0	X	건물유형	시설/지역
아고다	0	0	건물유형	시설
야놀자	0	X	건물유형	시설/지역
데일리호텔	0	X	건물유형	시설
꿀스테이	0	X	건물유형	시설
에어비앤비	0	X	건물유형	시설/지역/기능
스테이폴리오	0	X	건물유형	시설/디자인/기능



테마

적용하기

전체	<input checked="" type="checkbox"/>	갤러리	<input type="checkbox"/>
디자인투어	<input type="checkbox"/>	아웃도어	<input type="checkbox"/>
사색	<input type="checkbox"/>	BBQ	<input type="checkbox"/>
가족여행	<input type="checkbox"/>	영상	<input type="checkbox"/>
도심속휴식	<input type="checkbox"/>	로컬투어	<input type="checkbox"/>
정적인휴식	<input type="checkbox"/>	자연속의쉼	<input type="checkbox"/>
풀빌라	<input type="checkbox"/>	재생건축	<input type="checkbox"/>
수영장	<input type="checkbox"/>	파티하우스	<input type="checkbox"/>
커플여행	<input type="checkbox"/>	고액체험	<input type="checkbox"/>
나홀로여행	<input type="checkbox"/>	F&B	<input type="checkbox"/>
도보여행	<input type="checkbox"/>	여행자교류	<input type="checkbox"/>
워크숍	<input type="checkbox"/>	스파	<input type="checkbox"/>
문화예술	<input type="checkbox"/>	새로운영감	<input type="checkbox"/>
건축디자인	<input type="checkbox"/>	노천탕	<input type="checkbox"/>
오션뷰	<input type="checkbox"/>	다이닝룸	<input type="checkbox"/>
시티뷰	<input type="checkbox"/>	브라이덜샤워	<input type="checkbox"/>
액티비티	<input type="checkbox"/>	가구디자인	<input type="checkbox"/>
한옥	<input type="checkbox"/>	리스테이	<input type="checkbox"/>
요가	<input type="checkbox"/>	술로소피	<input type="checkbox"/>

예시 - 스테이 폴리오

다양한 카테고리로 감성숙소를 분류하였지만,
숙소의 종류는 200개가 안되는 것으로 확인됨.

서비스 목표



서비스 타겟

코로나 시대 속

안전하면서도 자유로운 여행을 떠나고 싶은
20대 ~30대 여행객을 위한 숙소 정보 공유

목표

sns데이터를 기반으로 텍스트분석을 진행하여
감성숙소를 테마별로 군집화하고,
안전한 여행을 지향하며 숙소 추천서비스를 구현



서비스 개발 과정

데이터 수집/처리/적재

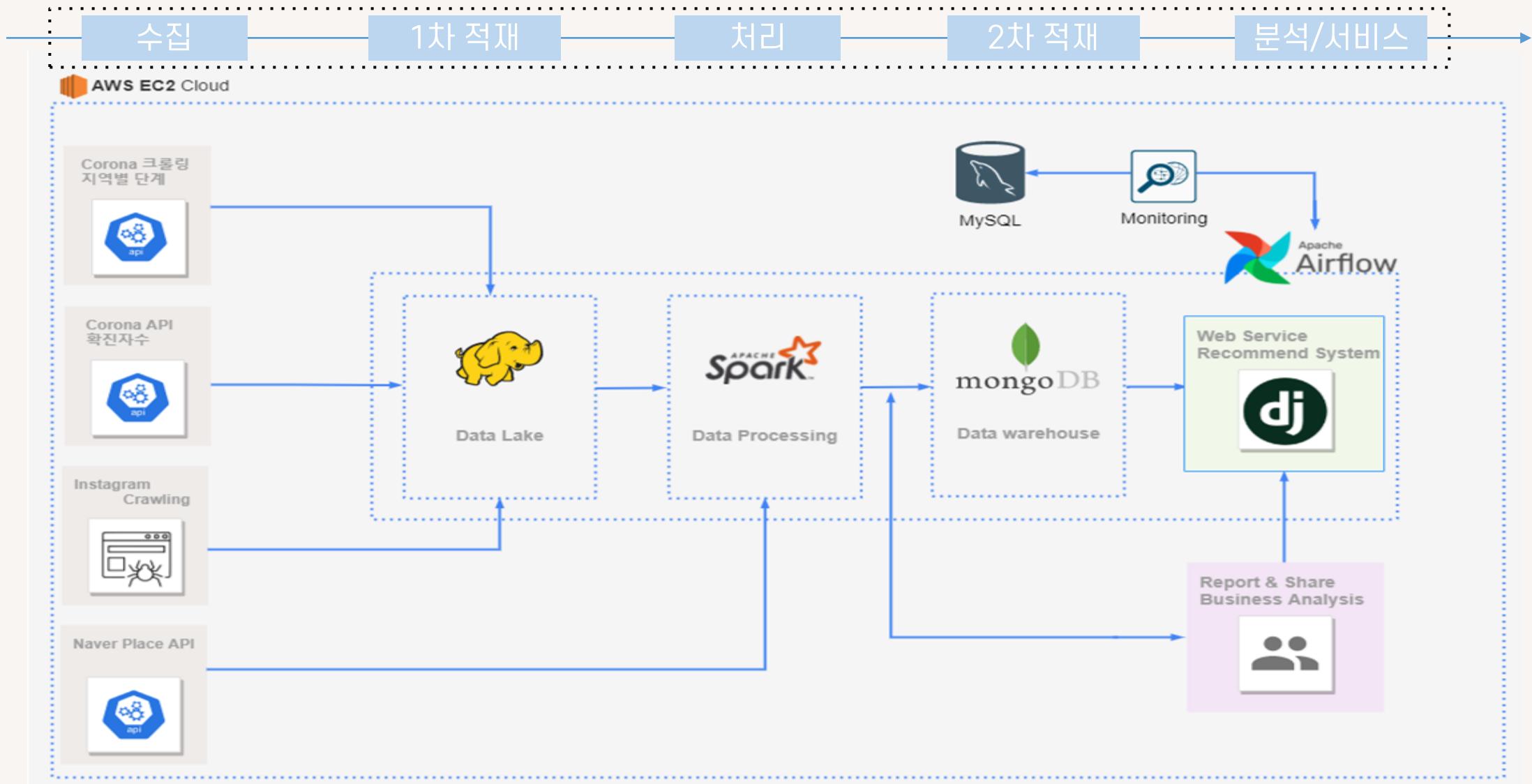
데이터 자동화

데이터 분석

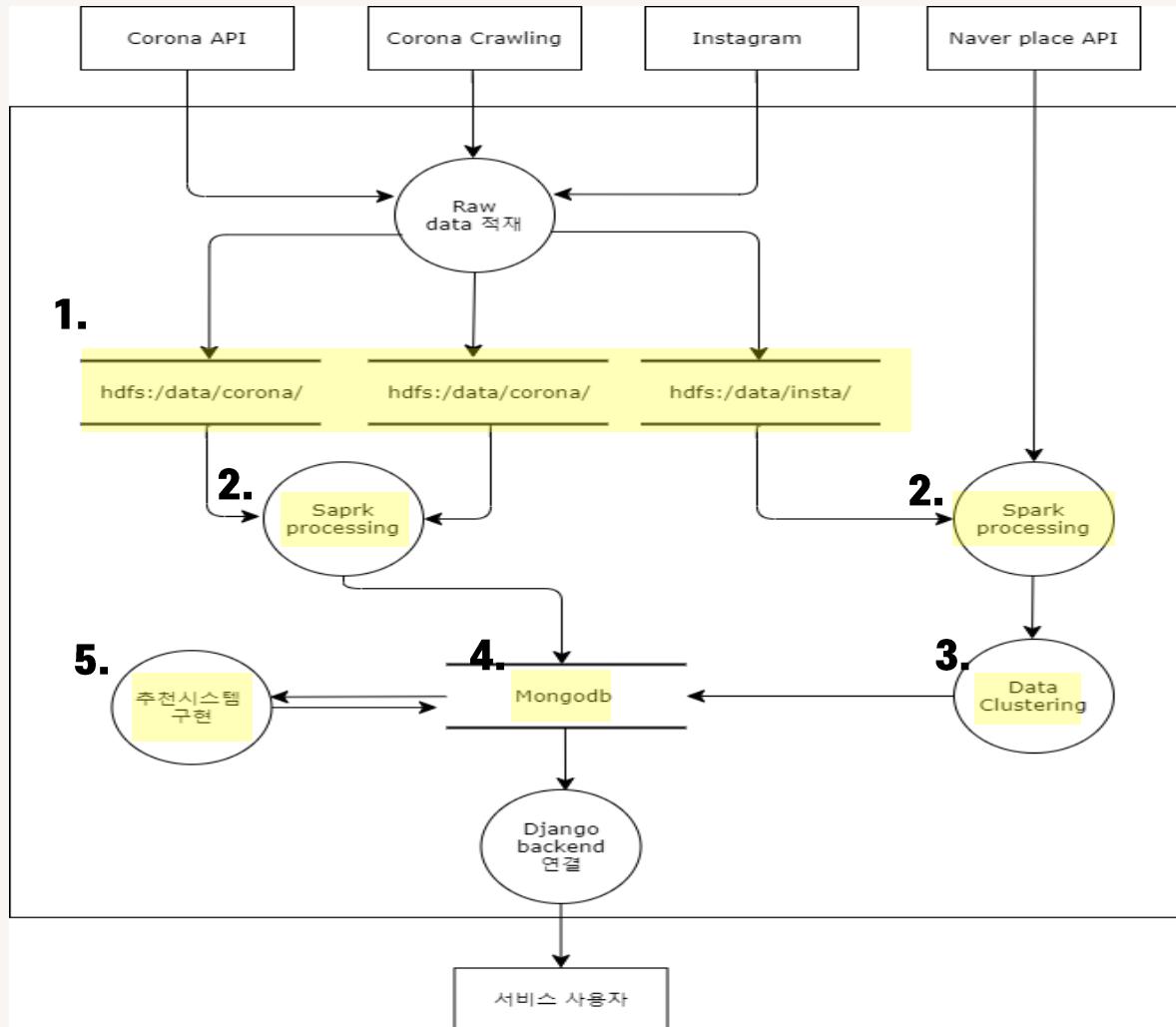
데이터 분석 결과

시스템 아키텍처

Work Flow



DFD(Data Flow Diagram)



1. 하둡 : 수집 데이터 저장
2. 스파크 : 수집 데이터 전처리
3. 클러스터링 모델 : 숙소 카테고리 분류
4. 몽고database : 분류된 숙소 정보 저장
5. 추천시스템 : Django 내에서 몽고database 연결

데이터 수집

인스타그램 게시물 크롤링

1. 기준

감성숙소 검색 시 [인기](#) 카테고리 상위 7개의 감성숙소 공유 계정(전국 단위)

2. 계정

@gamsung curation/ @daily_gamsung / @hi.stay.tour/

@sookso.diary/ @sookso.hada/ @gamsung.bnb / @rest_behappyhere

코로나 데이터

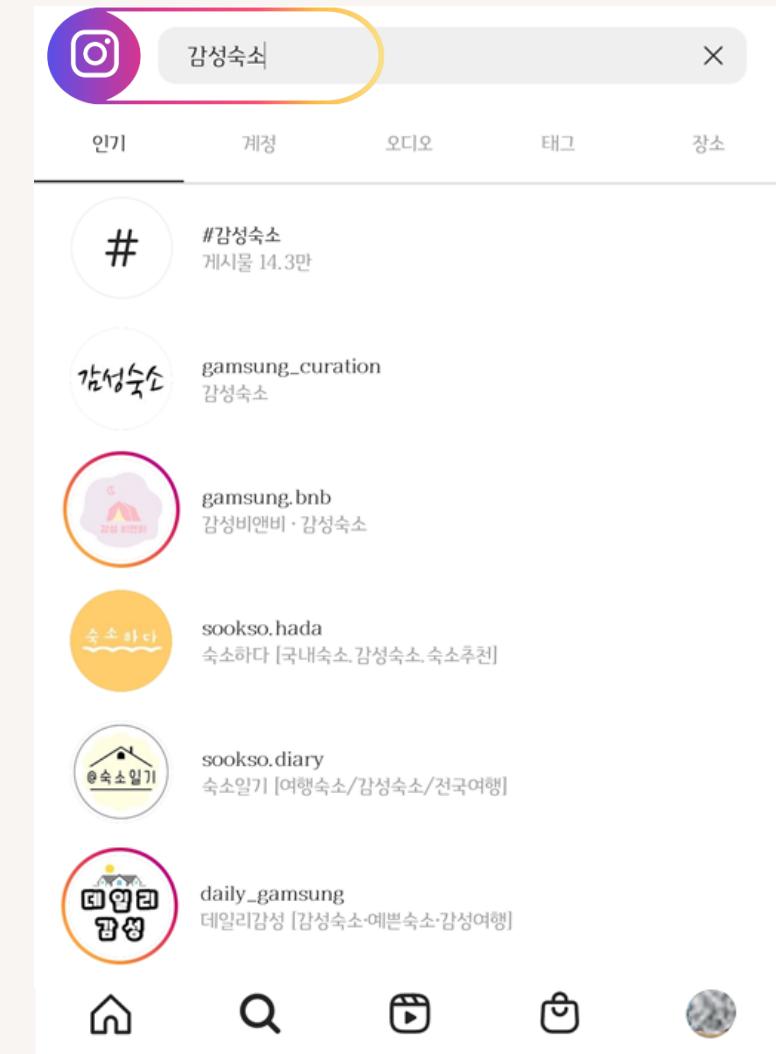
1. 보건복지부 코로나19 감염 현황 open api

2. 코로나 지역별 거리두기 단계 크롤링

네이버 데이터

1. 네이버 숙소정보 리스트 API

2. 네이버 블로그 '감성숙소' 키워드 게시글 크롤링 약 4000 건



데이터 수집



썸네일 이미지 크롤링

해당 이미지 숙소 대표 사진으로 서비스 활용 예정
숙소별 통합 content에 W2V 기법 적용하여,
숙소별 Feature Extraction 및 카테고리 도출
코사인 유사도 함수를 적용하여,
최적의 유사도 함수 선별 예정

좋아요, 텍스트, 태그 크롤링

텍스트에서 숙소명 추출 후 리스트화
숙소명 리스트 작성 후 네이버 장소 api 를 활용하여 실제 주소 가져오기
숙소명과 주소 매칭 후 content 텍스트 통합



데이터 1차 적재



Hadoop HDFS

목적:

DataLake의 목적으로

크롤링한 데이터를 원본상태로 적재

(이로 인해, 버전관리가 가능)

장점:

HDFS는 parquet파일의 형식으로

Hadoop 에코시스템과 원활한 데이터 흐름이

가능하기 때문에 쉽게 SPARK와 연동

HDFS 적재 데이터 목록

```
ubuntu@ip-172-31-0-68: ~/DE
(pytho... n3) ubuntu@ip-172-31-0-68:~/DE$ hdfs dfs -ls /data/insta
Found 14 items
-rw-r--r-- 1 ubuntu supergroup    79536 2021-09-30 13:41 /data/insta/daily_gamsung 2021-09-30.parquet
-rw-r--r-- 1 ubuntu supergroup    78484 2021-10-01 10:28 /data/insta/daily_gamsung 2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup   298535 2021-09-30 13:42 /data/insta/gamsung.bnb 2021-09-30.parquet
-rw-r--r-- 1 ubuntu supergroup   296515 2021-10-01 10:28 /data/insta/gamsung_bnb_2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup  336501 2021-10-01 10:28 /data/insta/gamsung_curation2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup  339058 2021-09-30 13:41 /data/insta/gamsung_curation_2021-09-30.parquet
-rw-r--r-- 1 ubuntu supergroup   81118 2021-10-01 10:28 /data/insta/hi_stay_tour2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup   82381 2021-09-30 13:42 /data/insta/hi_stay_tour_2021-09-30.parquet
-rw-r--r-- 1 ubuntu supergroup  109430 2021-10-01 10:28 /data/insta/rest_behappyhere2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup  110716 2021-09-30 13:42 /data/insta/rest_behappyhere_2021-09-30.parquet
-rw-r--r-- 1 ubuntu supergroup  327071 2021-10-01 10:28 /data/insta/sookso_diary2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup  331172 2021-09-30 13:42 /data/insta/sookso_diary_2021-09-30.parquet
-rw-r--r-- 1 ubuntu supergroup  125500 2021-10-01 10:28 /data/insta/sookso_hada2021-10-01.parquet
-rw-r--r-- 1 ubuntu supergroup  126786 2021-09-30 13:42 /data/insta/sookso_hada_2021-09-30.parquet
(de... python3) ubuntu@ip-172-31-0-68:~/DE$ hdfs dfs -ls /data/corona
Found 6 items
-rw-r--r-- 1 ubuntu supergroup    8538 2021-09-30 13:44 /data/corona/coronaAPI_2021-09-30
-rw-r--r-- 1 ubuntu supergroup    8549 2021-10-01 10:46 /data/corona/coronaAPI_2021-10-01
-rw-r--r-- 1 ubuntu supergroup    8559 2021-10-02 11:13 /data/corona/coronaAPI_2021-10-02
-rw-r--r-- 1 ubuntu supergroup    4460 2021-09-30 13:44 /data/corona/coronaStage_2021-09-30
-rw-r--r-- 1 ubuntu supergroup    4476 2021-10-01 11:00 /data/corona/coronaStage_2021-10-01
-rw-r--r-- 1 ubuntu supergroup    4476 2021-10-02 00:28 /data/corona/coronaStage_2021-10-02
(de... python3) ubuntu@ip-172-31-0-68:~/DE$
```

데이터 날짜별 버전관리

데이터 처리



SPARK 장점

RDD 형식의 빠른 데이터 처리

DataFrame API를 사용한 데이터 처리

다양한 확장자 (csv, parquet, json) 데이터 처리 가능

전처리 진행

1. 숙소 이름 추출 (name data)
2. Name data가 없는 index 제거
3. 계정 별 숙소 리스트 취합
4. Like int형으로 변환 후 취합
5. 통합 숙소 리스트를 기준으로 content 병합
6. Like 수 취합
7. Content 일괄 전처리 (토큰화, 불용어 제거)
8. 7개의 계정 DataFrame 취합
9. Place 컬럼을 Naver Place API 값으로 대체

Pandas dataframe 이용한 insta 데이터 처리

```
tot_dataset = daily_gamsung
for df in [gamsung_bnb, gamsung_curation, hi_stay_tour, rest_behappyhere, sookso_diary, sookso_hada]:
    tot_dataset = tot_dataset.append(df)

if tot_dataset.like.dtype == 'O':
    tot_dataset.like = tot_dataset.like.apply(lambda x : convert(x))

dg_cnt = Counter(tot_dataset.name)      # Counter 모듈로 name data count
li = []
for name, cnt in dg_cnt.items():        # 중복된 것만 뽑기 (cnt가 1이상이면 중복)
    if cnt > 1:
        li.append(name)

# 4. 행 가져와서 like집계, content, tag list append
for i in li:                          # 중복된 name 반복
    ex = tot_dataset.loc[tot_dataset.name == i] # 반복된 name의 행을 가져온다
    ex.reset_index(drop=True, inplace=True)     # index reset

    content = ex.iloc[0,:].content
    date = ex.iloc[0,:].date
    like = ex.iloc[0,:].like
    place = ex.iloc[0,:].place
    tags = ex.iloc[0,:].tags
    imgUrl = ex.iloc[0,:].imgUrl
    name = ex.name[0]
    overlap = ex.iloc[0,:].overlap

    for i in range(len(ex)-1):
        like += ex.iloc[i+1,:].like
        tags += ex.iloc[i+1,:].tags
        overlap += ex.iloc[i+1,:].overlap
```

데이터 처리

최종 데이터셋 : 720개의 숙소 정보를 취합

index	content	date	like	tags	imgUrl	name	overlap	place	wv	category
0	[석주재프 라이빗, 한 옥, 주재, 석주, 프라 이빗, 한옥, 하루, 텁, 조용, 힐 링,...]	2021-09-27	22	[#데일리감성밀양', '#데일리감성_밀양경 상남도로', '#데일리감성_경남 ...]	https://scontent-ssn1- 1.cdninstagram.com/v/t51...	석주 재	0	경상남 도 밀양 시 산외 면 금곡 2길 50	[0.052009687, -0.011980195, -0.0013845726, -0....]	0
1	[컴온인, 동 남아, 이국, 분위기, 컴 온, 컴, 동 남아, 이국, 분위기, 가 득, 담...]	2021-09-26	20	[#데일리감성여수', '#데일리감성_여수', '#데일리감성_전남 ...]	https://scontent-ssn1- 1.cdninstagram.com/v/t51...	컴온 인	0		[-0.11794634, -0.42266873, -0.1407647, -0.2249...]	0
2	[문, 스톤, 문, 스톤, 문, 스톤, 해수욕장, 근처, 깔끔, 신축, 카페, 바베...]	2021-09-26	13	[#데일리감성강화', '#데일리감성_강화경 기도로', '#데일리감성_경기 ...]	https://scontent-ssn1- 1.cdninstagram.com/v/t51...	문스 톤펜 션	0		[-0.15490063, -0.13003348, 0.15267421, -0.2060...]	0
3	[고등어, 민 박, 동쪽, 로맨틱, 고 등어, 민박, 고등어, 민 박, 푸르다, 빛, 바...]	2021-09-24	13	[#데일리감성제주', '#데일리감성_제주 ']	https://scontent-ssn1- 1.cdninstagram.com/v/t51...	고등 어민 박	0	제주특 별자치 도 제주 시 구좌 읍 해맞 이해안 로 1026	[0.04686849, -0.23572925, 0.0094889635, -0.419...]	9

데이터 2차 적재



NoSQL의 장점을 활용하여 정형화되지 않은 데이터를 저장
이미지, 리스트, 특수문자와 같은 데이터의 저장공간으로 활용

MongoDB – Django Backend 연결

Django는 MySQL과 PostgreSQL과 ORM방식으로 주로 사용되지만

Django (Django + Mongo) 를 이용하여

1. ORM 방식으로 Login 기능 구현

Pymongo를 이용하여

2. mongodbManager.py 생성 후 MongoDB를 이용

```
> use ojo_db
switched to db ojo_db
> show collections
__schema__
auth_group
auth_group_permissions
auth_permission
auth_user
auth_user_groups
auth_user_user_permissions
cluster
corona
django_admin_log
django_content_type
django_migrations
django_session
homepage_user_info
insta
insta_tot
matrix
>
```

corona
: 코로나 관련 데이터

homepage_user_info
: Django 웹페이지 user 정보

insta
: 수집 인스타그램 데이터

matrix
: tf-idf 유사도 행렬

MongoDB 적재

```
switched to db ojo_db
> db.cluster.find()
{ "_id" : ObjectId("615b0fc716e08c5b2061e978"), "content" : 결과, "createdAt" : "2021-08-10T10:45:28.000Z", "updatedAt" : "2021-08-10T10:45:28.000Z" }
```

Workflow 자동화



Airflow 사용 목적

SPARK 형식의 데이터 처리 (SparkSubmitOperator)

python 형식의 데이터 처리(pythonOperator)

Bash명령어의 스케줄링 기능(BashOperator)

따라서, Pipeline을 자동화하기 위해 사용

MySQL 사용 이유

기존의 sqliteDB(default)는 직렬 처리만 가능

병렬처리를 위해 MySQL과 airflow 연동

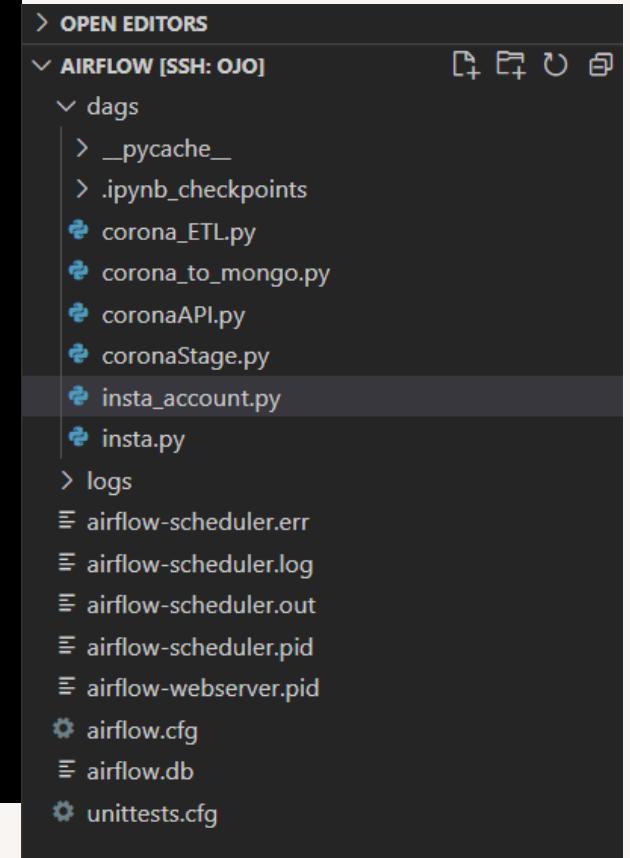
```
ubuntu@ip-172-31-0-68: ~
Reading table information for completion of table and column name
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+-----+
| Tables_in_airflow |
+-----+
| alembic_version   |
| chart              |
| connection         |
| dag                |
| dag_pickle          |
| dag_run             |
| import_error        |
| job                |
| known_event         |
| known_event_type    |
| kube_resource_version |
| kube_worker_uuid    |
| log                |
| sla_miss            |
| slot_pool           |
| task_fail            |
| task_instance        |
| task_reschedule      |
| users               |
| variable             |
| xcom                |
+-----+
21 rows in set (0.00 sec)

mysql>
```

MySQL과 Airflow 연동 결과

Airflow DAGS .py 목록



Workflow 자동화

Airflow DAG 설계

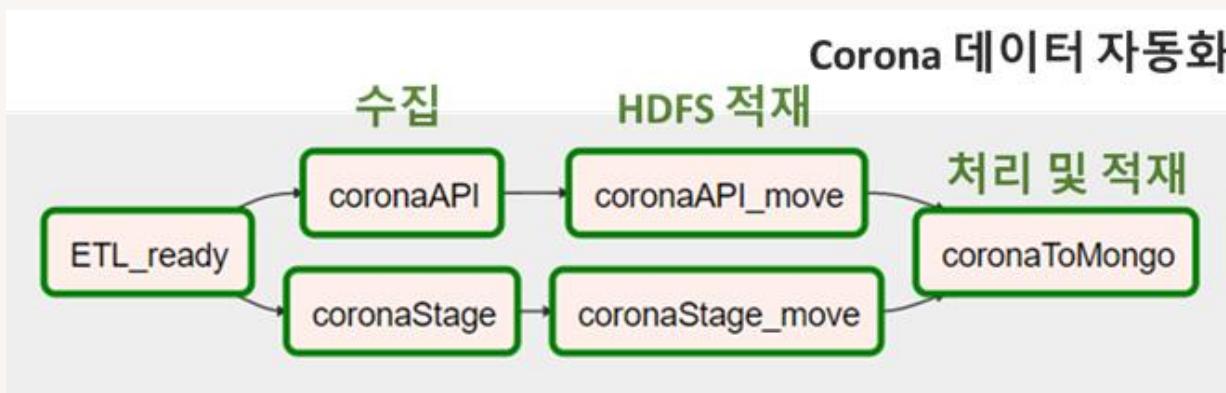
DAGs

Search:

	DAG	Schedule	Owner	Recent Tasks	Last Run	DAG Runs
	corona	0 10 * * *	airflow		2021-10-03 00:00	
	insta	@once	airflow		2021-10-05 00:00	



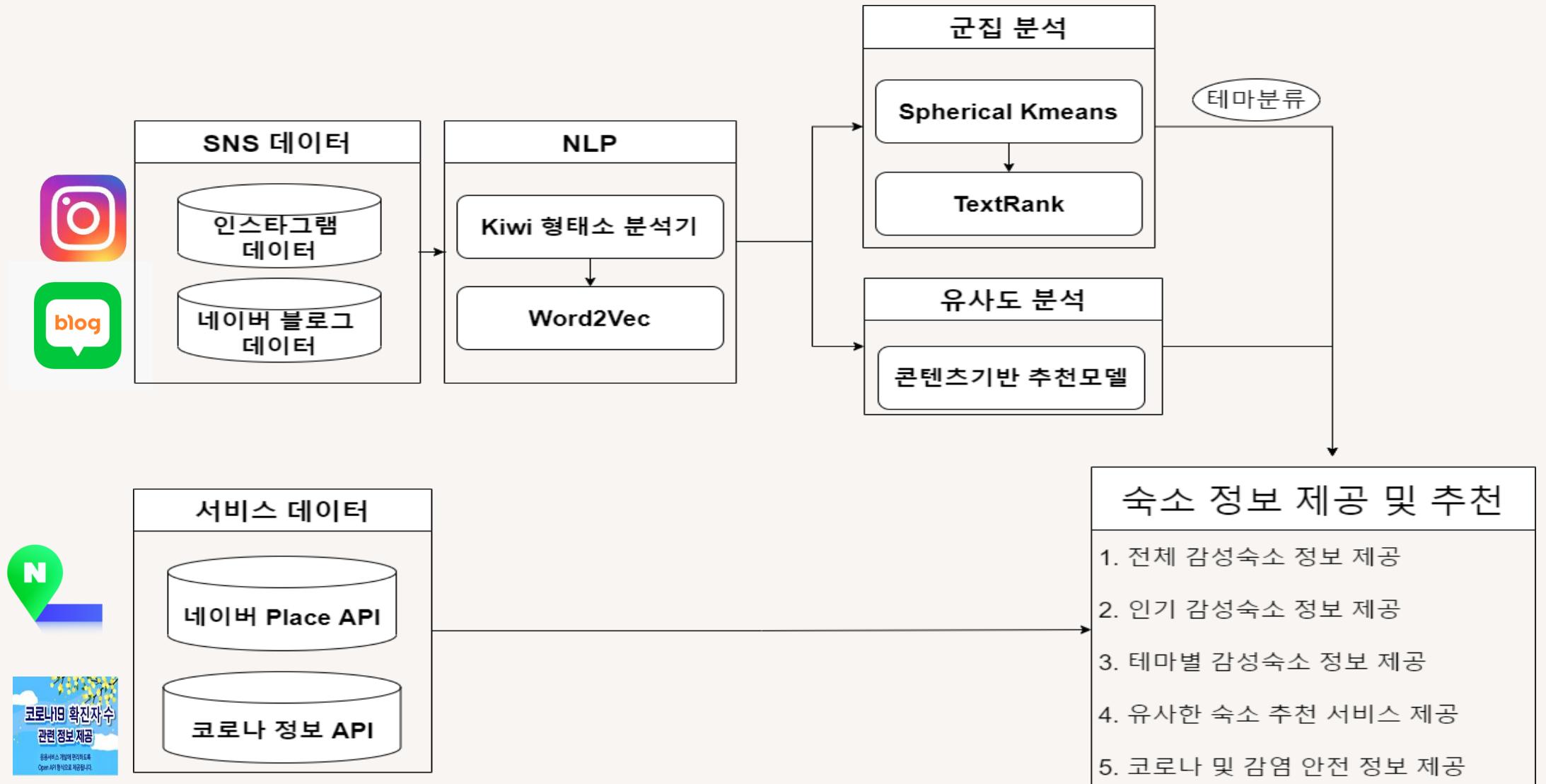
DAG :
corona
Graph view



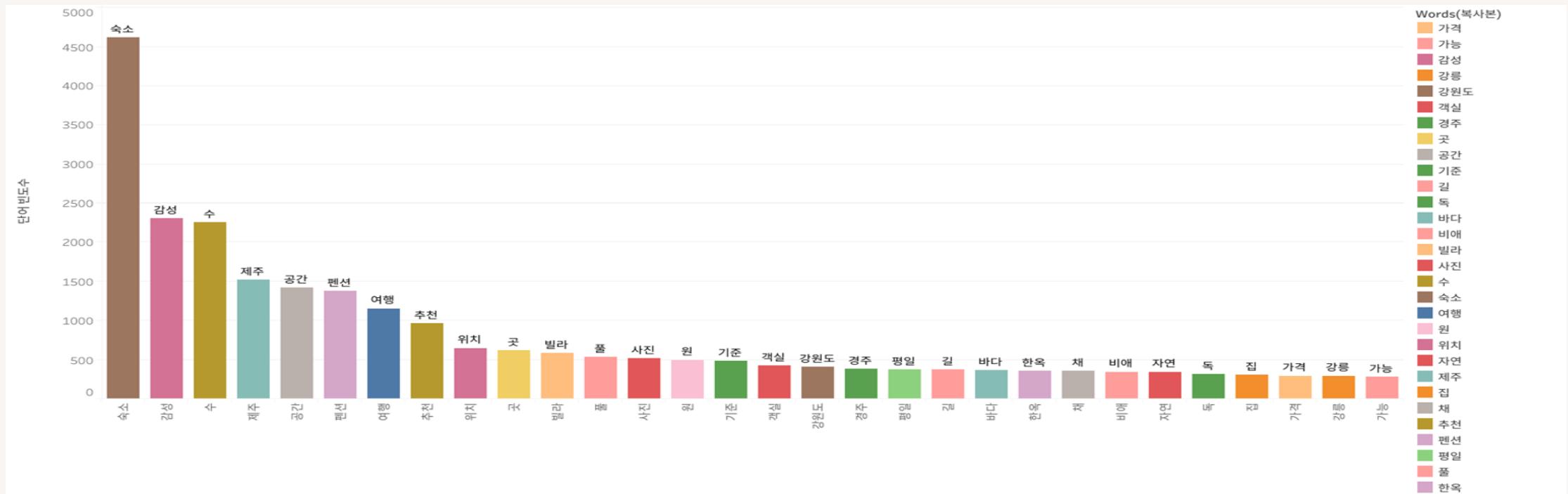
DAG : insta
Graph view



분석 프로세스



단어 빈도수 분포



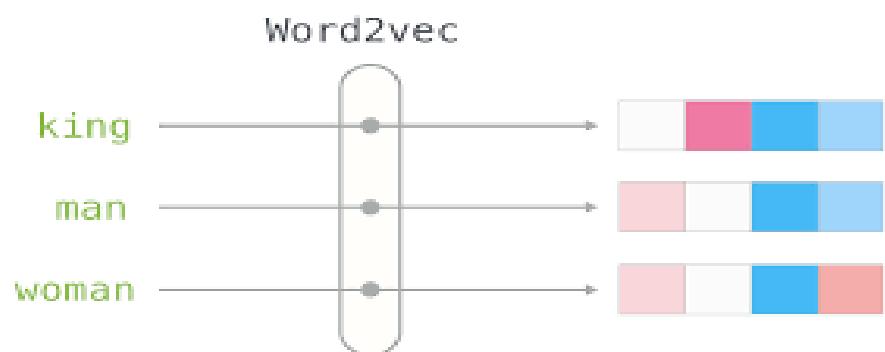
토큰화된 텍스트 데이터 상위 20개 단어 : 감성과 관련된 단어는 부족



텍스트 분석을 위한 임베딩기법

Word2Vec을 이용한 웹 문서 클러스터링 시스템 구현*

이현석, 안성훈, 이용환, 천명재, 박혁주, 박미화, 이용규
동국대학교 컴퓨터공학과-서울
e-mail : lazuki@dongguk.edu

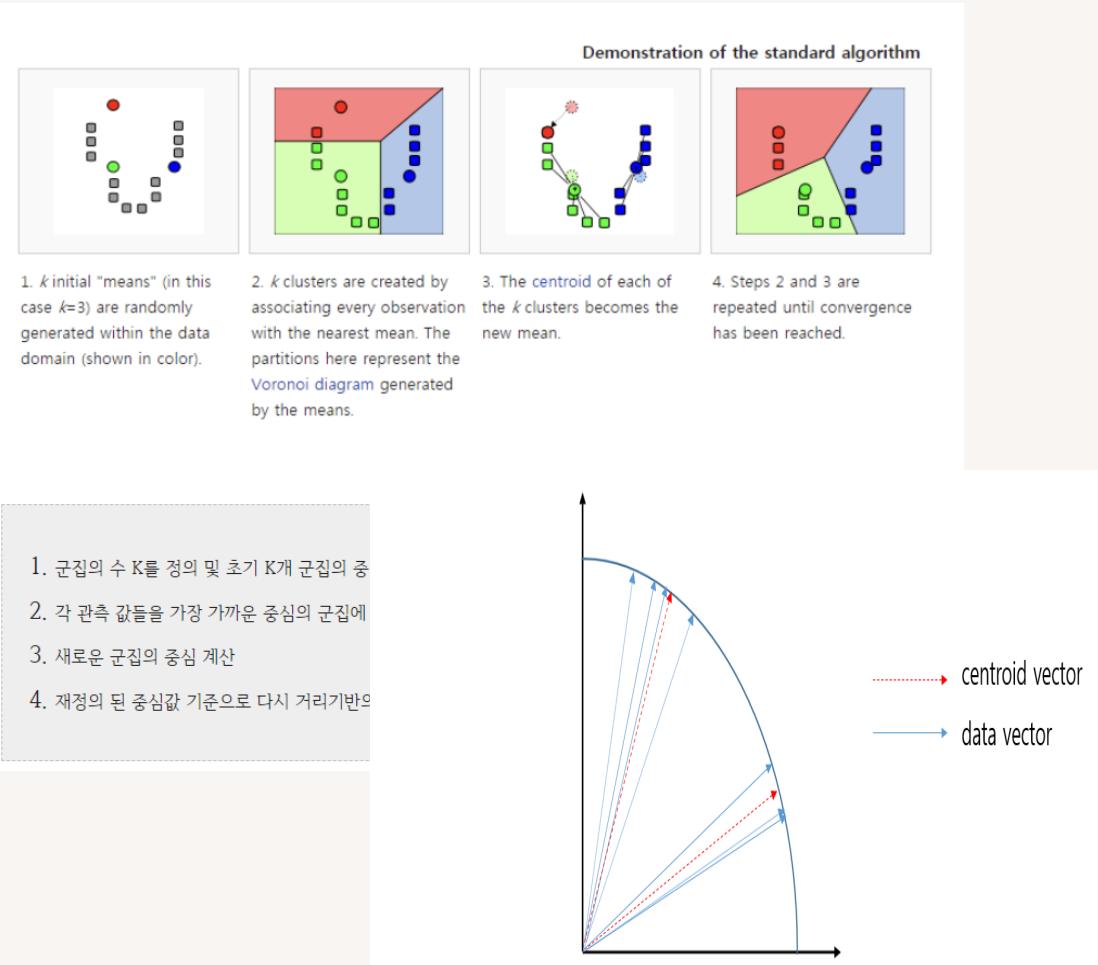


W2V의 강점 : 데이터 희소성, 유의어 해석

이쁘다, 예쁘다 등

유의어를 고려할 수 있는 W2V기법이 더욱 적절

Spherical K-means



가중치와 거리 정의

- 텍스트 분류에서 높은 성과를 보이는 코사인 거리법을 적용한 K-means 군집 분석 수행 > Spherical K-means

초기 군집수의 결정

- Elbow와 실루엣 기법을 활용하여 최적의 군집수 결정

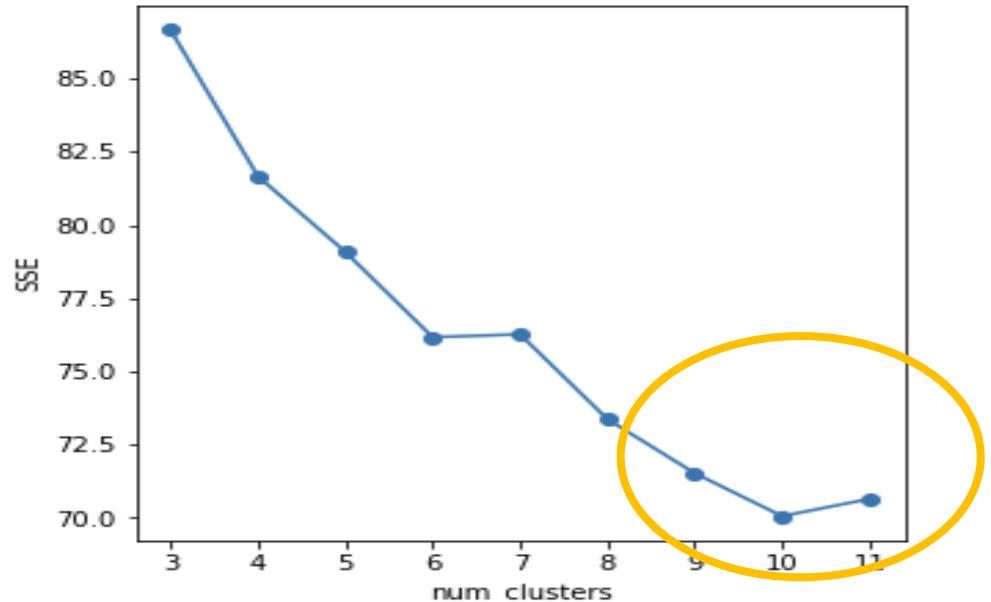
군집별 해석

- TextRank 기법을 적용하여, 각 군집별 키워드를 추출 후 군집 해석

$$TR(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} TR(V_j)$$

군집분석

Elbow를 통한 최적 군집수



- 6과 10 지점에서 두번의 경사 변화가 나타남
- 10 이후로는 SSE의 변동성이 높음.
- **최적 군집수를 10으로 설정**

TextRank기법을 활용한 결과해석

keywords
(::기록', 33.5657007490261),
(::국내', 32.57158090901199),
(::여행지', 16.609191914951744),
(::예쁘다', 12.539879124397292),
(::즐기다', 10.903731012174422),
(::국내외', 9.782780058418812),
(::만원', 7.925628220078781),
(::핫', 7.352853884107331),
(::수영장', 7.280217514661686),
(::그램', 7.2160691805341015),
(::가족', 7.081380631091912),
(::힐링', 7.0347946270245965),
(::슬', 6.882867149232523),
(::한국', 6.337599205678244),
(::인테리어', 5.438794658073363),
(::커플', 5.324179603940402),
(::느끼다', 5.260345657614331),
(::분위기', 4.654988711388162),
(::가족', 4.600031490255179),
(::프다', 4.486491807012084)]

- 문서 요약에 대표적인 기법인 TextRank를 활용하여 군집별 키워드 비율을 분석하고, 이를 토대로 군집별 카테고리 해석을 진행함.
- 예 : 커플, 가족, 즐기다, 수영장 = '다같이 놀기 좋은 숙소'

군집분석

분석 결과 - 테마 선정

- 1인 여행객을 위한 숙소**
- 커플 혹은 친구 여행을 위한 숙소**
- 독특한 경험을 할 수 있는 숙소**
- 한옥과 같은 분위기를 가진 숙소**
- 가족, 단체 여행객을 위한 숙소**
- 자연과 힐링을 경험할 수 있는 숙소**
- 자연과 체험을 할 수 있는 숙소**
- 편안하고 안락한 여행을 위한 숙소**
- 시원한 바다를 경험할 수 있는 숙소**

1	2	3	4
('데일리', 24.935294024485227)	('코', 1.20964809699917)	('국내', 37.78431356833906)	('한옥', 56.51945545236883)
('비수기', 12.057179996312124)	('취소', 1.20964809699917)	('기록', 34.84267868956896)	('마당', 7.403442834557912)
('민박', 0.752228407502203)	('터지다', 1.20964809699917)	('여행지', 24.570492920098797)	('매력', 6.447264433961667)
('카페', 8.499622505331669)	('시국', 1.116861413069044)	('예쁘다', 13.139589865372605)	('비수기', 6.117988092520302)
('예쁘다', 8.288733443602965)	('코로나', 1.116861413069044)	('在国内', 10.73574351127415)	('현대', 5.781463953153268)
('힐링', 8.074683432603463)	('빨리', 1.0259194138607315)	('수영장', 10.33780626499373)	('넓다', 5.509172554759226)
('사용', 7.9480404069551245)	('작년', 1.0259194138607315)	('국내외', 9.878034530146346)	('경상북도', 5.181394375970992)
('조용', 7.828400563750255)	('끌나다', 0.9362713155467332)	('힐링', 6.218013349765479)	('힐링', 5.074150221241952)
('바다', 7.72324593676768)	('해수', 0.9362713155467332)	('요기', 6.188651808399677)	('그램', 8.364006624503883)
('한옥', 7.684489809837887)	('이쁘다', 0.8474520132109842)	('핫', 5.936877701000646)	('가득', 8.330169465811244)
('숲', 6.479345568892888)	('바다', 0.8474520132109842)	('바다', 5.843712734285374)	('마을', 4.991818179341114)
('동남구', 0.8474520132109842)	('바베큐', 5.819150795495323)	('바베큐', 5.819150795495323)	('편안', 4.908045786964804)
('인테리어', 5.919118274260104)	('카페', 0.7590236988137503)	('아외', 5.418723052301292)	('바다', 6.949418827427136)
('분위기', 5.716973242765413)	('가든', 0.7590236988137503)	('인생', 5.2604355568061594)	('따뜻', 4.89127514523596)
('꾸미다', 5.341107391671329)	NaN	('삿', 4.939218175230081)	('실', 4.783000887587435)
('가득', 5.254572758219836)	NaN	('프다', 4.618005965277799)	('쉬다', 4.776649731628528)
('한적', 5.245813784827844)	NaN	('풀장', 4.5202381507176765)	('거실', 4.713629584570271)
('보내다', 4.921579403715745)	NaN	('스파', 4.134073711830698)	('문', 4.683058951345555)
('마음', 4.779946460355078)	NaN	('짬뽕', 3.930393526779876)	('여유', 4.60434552204709)
('운영', 4.776060267809384)	NaN	('인피니티', 3.684579053704483)	('조식', 5.68534682172055)
('책', 4.759083916506332)	NaN	('데이트', 3.6697512483511585)	('삿', 5.609764172302821)
	: object	: object	: object
		: object	: object

5	6	7	8	9
('가득', 13.066737423740788)	('바라보다', 22.47114557286466)	('수영장', 7.224090075581868)	('호캉', 19.96676123328397)	('바다', 30.1050026124984)
('인테리어', 12.166679787734356)	('바다', 20.28177819965627)	('야외', 5.6053399526928285)	('계획', 19.374903059527135)	('온선류', 25.960386458742136)
('따뜻', 10.017353819322413)	('소리', 16.106911305620482)	('숲', 5.484033633155842)	('가득', 16.872587398583467)	('수영장', 18.61295628282745)
('예쁘다', 9.951273554079085)	('비수기', 16.016684161179395)	('카페', 5.1528469215314505)	('수영장', 14.71594330231118)	('바라보다', 12.182121454063763)
('소품', 9.415921687682243)	('갖추다', 14.716089484896782)	('바다', 5.010606708292838)	('여유', 14.513251862279748)	('바베큐', 11.840513982884154)
('침실', 8.53309771471552)	('힐링', 14.560528065585256)	('깔끔', 5.0077109221345255)	('따뜻', 14.180644392509265)	('힐링', 9.417858798694747)
('거실', 7.224999156410808)	('꾸미다', 13.378995547344031)	('사용', 4.987651259339811)	('침실', 13.879458810533445)	('멋지다', 9.119443299527658)
('카페', 6.948643646604961)	('누리다', 12.807805865203774)	('해변', 4.827628378634301)	('분위기', 13.775211702252836)	('계획', 8.55128852730715)
('아기자기', 6.727351206010846)	('마을', 12.567199356042469)	('이쁘다', 4.794644331654773)	('넓다', 13.33351366863463)	('카페', 8.305664958802879)
('화이트', 6.187223848910862)	('조용', 12.118174137104905)	('글램핑', 4.744108251101518)	('거실', 13.264322034687913)	('힐링', 8.033348626646285)
('아늑', 6.008473076252454)	('주인장', 12.10758851040689)	('체험', 4.650313759080591)	('바베큐', 12.800774396230112)	('일출', 8.00701456521179)
('식물', 5.994942732950067)	('풍경', 12.002051765182847)	('리', 4.611774347578682)	('가족', 12.38320417691535)	('야외', 8.0052033021571)
('매력', 5.953408583841335)	('오롯', 10.947128322715054)	('바베큐', 4.565915745881867)	('마음', 12.344001666551884)	('노을', 7.95240707435879)
('주인장', 5.68094588343934)	('경남', 10.749917465110995)	('가족', 4.426454524852189)	('조식', 12.31185690685536)	('가족', 7.8797026477092365)
('편안', 5.624637172834014)	('경상북도', 10.618139493158058)	('파티', 4.348580012794538)	('마당', 12.20436466208069)	('프라이빗', 7.846586882313262)
('분위기', 5.57941277267268)	('바람', 10.607316768899269)	('친구', 4.082253681080293)	('정원', 11.261415698167063)	('아름답다', 7.844886555789189)
('감각', 5.44771388165695)	('가족', 10.190615152407672)	('걸다', 3.933700766837997)	('휴식', 11.203913842460208)	('개별', 7.000903615537254)
('햇살', 5.302178605704328)	('수영장', 10.11041979511465)	('예쁘다', 3.77668910487075)	('문', 10.839235655530707)	('인피니티', 8.857005568034965)
('힐링', 5.054289170051441)	('주변', 9.6732227120898)	('소품', 3.699449543263684)	('마을', 10.71412437536388)	('넓다', 6.775226734372409)
('계획', 5.050717335281296)	('친구', 9.465239798868707)	('주차장', 3.6914490118122996)	('편안', 10.682987777396256)	('하늘', 6.632056306853087)
object	object	object	object	object

숙소 추천을 위한 유사도 분석

TF-IDF 임베딩 - 코사인 유사도

[27] insta_REC('스튜디오노이')

item_idx	name	content	token_nolist	score
51	51 두모가	서쪽 바다 마을 개별 수영장 프라이빗 독채하우스 거리 밖 거리 나누다 거리 둘담 느...		0.212191
20	20 제주 청수리오즈네	청수리오즈 서쪽 청수리오즈 오즈 서쪽 감귤 과수원 반딧불 유명 조용 시골 주택 채우...		0.183733
42	42 코끼리잠	코끼리잠재 해수욕장 근처 조용 마을 조식 코끼리 잠 콘크리트 외관 끝다 코끼리 마음...		0.173669
407	408 클랭블루스테이	클랭블루스테이 서쪽 조용 마을 노을 클랭블루스테 총 이루어지다 전용 푸릇푸릇 잔디밭...		0.170440
74	74 세렝게티민박	세렝게티 민박 포구 조용 마을 팀 세렝게티 민박 서쪽 전통 특징 간직 현대 감각 탄...		0.167447
395	396 물리적공간	물리 조용 한적 마을 팀 아늑 물리 푸르다 들판 녹밭 배경 물리 화이트 톤 모던 미...		0.161436
371	372 스테이아하	스테이아하프라이빗 수영장 스테이아 스테이아 협 해수욕장 가깝다 주변 유명 카페 여행...		0.156405
417	418 월령지현	월령지현 서쪽 골목길 안쪽 자리 잡다 흙 나무 질감 살리다 독채스테 월령지 현 지난...		0.152602
36	36 시온스테이	시온스테이 애 조용 마을 추억 선물 독채 시온스테 가족 머무르다 따뜻 바라다 마음...		0.150229
454	455 모노토너스	모노토너스 서쪽 시골 마을 바다 모노토너 평형 또는 전용 바다 거실 테라스 샤워실...		0.144795

TF-IDF * 코사인 유사도의 최고 점수 : 0.212191

Word2Vec * 코사인 유사도의 최고 점수 : 0.925350

Word2Vec 임베딩 - 코사인 유사도

insta_REC('스튜디오노이')

item_idx	content	name	token_nolist	score
66	[에디토리얼제주, 제주다움, 오롯, 느끼다, 독채스테, 에디토리얼제, 에디토리얼제...]	에디토리얼제주	에디토리얼제주 제주다움 오롯 느끼다 독채스테 에디토리얼제 독채 거실 키...	0.925350
38	[워킹, 써클, 수영장, 탕, 깔끔, 인테리어, 독채, 워킹, 써클, 워킹, 써클,...]	워킹인써클	워킹 써클 수영장 탕 깔끔 인테리어 독채 워킹 써클 워킹 써클 하루 팀 숙소 안...	0.912723
70	[영스테이, 한적, 시골, 영스테, 마음, 편안, 머무르다, 뜻, 목조, 짓다, 언...]	영스테이	영스테이 한적 시골 영스테 마음 편안 머무르다 뜻 목조 짓다 언제나 온기 가득...	0.911793
82	[산내원림, 한적, 산, 독채풀, 산, 넓다, 정원, 이루다, 가까이, 느끼다, 한...]	산내원림	산내원림 한적 산 독채풀 산 넓다 정원 이루다 가까이 느끼다 한적 독채풀 빛나 컨셉...	0.910412
398	[투스프링베어제, 프라이, 빗, 투, 스프링, 베다, 오롯, 팀, 시원, 수영장, ...]	투스프링베어	투스프링베어제 프라이 빗 투 스프링 베다 오롯 팀 시원 수영장 나무 해외여행 느다...	0.909712
74	[세렝게티, 민박, 포구, 조용, 마을, 팀, 세렝게티, 민박, 서쪽, 전통, 특징, ...]	세렝게티민박	세렝게티 민박 포구 조용 마을 팀 세렝게티 민박 서쪽 전통 특징 간직 현대 감각 탄...	0.909282
395	[물리, 조용, 한적, 마을, 팀, 아늑, 물리, 푸르다, 들판, 녹밭, 배경, 물리, ...]	물리적공간	물리 조용 한적 마을 팀 아늑 물리 푸르다 들판 녹밭 배경 물리 화이트 톤 모던 미...	0.908653
443	[형제, 서쪽, 형제, 잔디밭, 수국, 굴나무, 어우러지다, 멋, 스럽, 외관, 눈...]	오형제	형제 서쪽 형제 잔디밭 수국 굴나무 어우러지다 멋 스럽 외관 눈 띠다 둘담 두르다...	0.907259
55	[동경, 둘담, 두르다, 쌓이다, 동경, 동경, 바닷가, 소리, 품다, 팀, 독채숙소, ...]	동경신촌	동경 둘담 두르다 쌓이다 동경 동경 바닷가 소리 품다 팀 독채숙소 화이트 우드 테...	0.905691
56	[나름, 따뜻, 독채, 나름, 나름, 고유, 방식, 모습, 마음, 조용, 살피다, ...]	나름의고요	나름 따뜻 독채 나름 나름 고유 방식 모습 마음 조용 살피다 뜻하다 우드 화이트 독...	0.905447

유사도 점수면에서 W2V기법이 TF-IDF보다 4배 이상 ↑

숙소 추천을 위한 유사도 분석

- 추천 결과 비교

숙소 ‘스튜디오노이’의 숙소 이미지



숙소 ‘스튜디오노이’의 특징 : 제주도 한경면, 독채, 정원, 돌담, 화이트 앤 우드톤, 넓은 공간 등

숙소 추천을 위한 유사도 분석

- 추천 결과 비교



-W2V 추천 결과 : 에디토리얼 제주 숙소



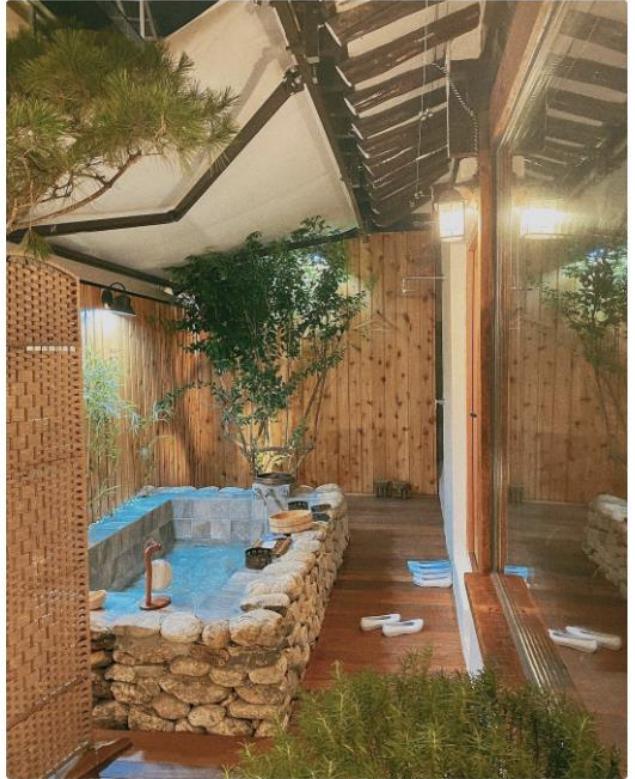
-TF-IDF 추천 결과 : 두모가 숙소



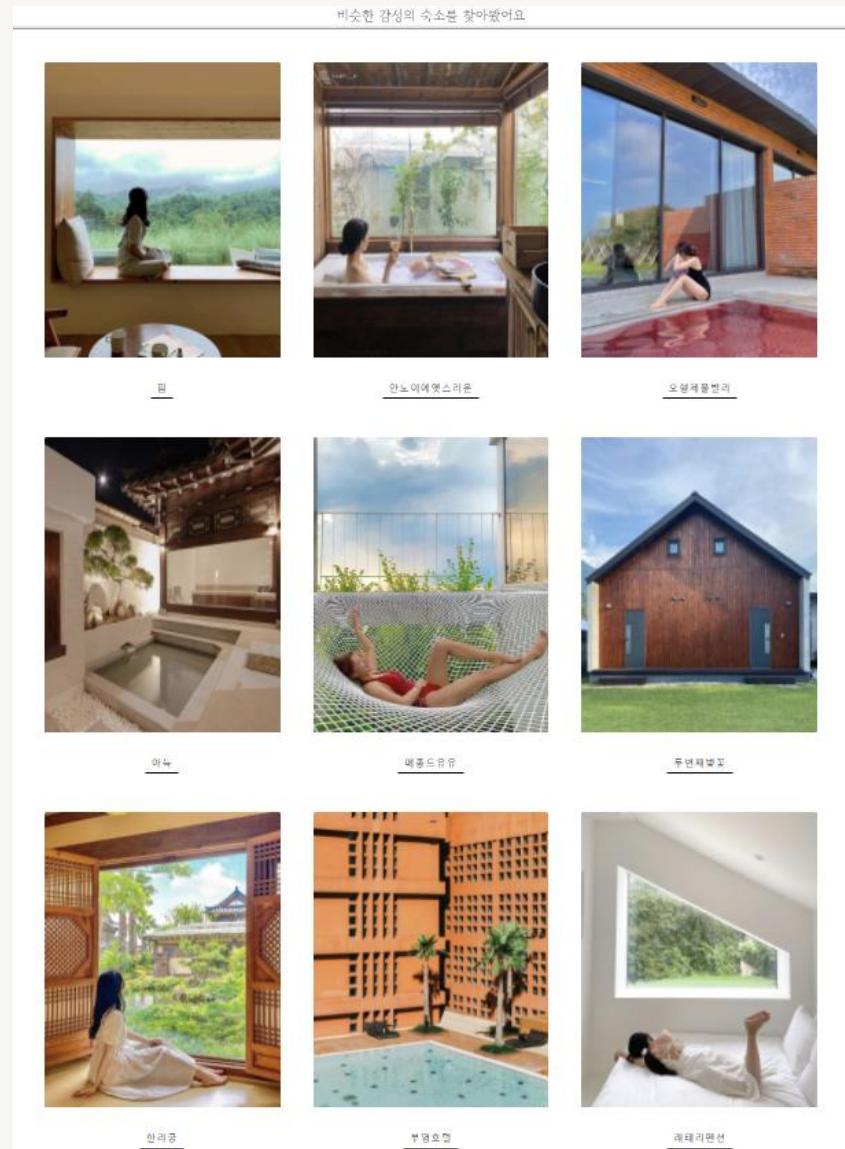
지역적 유사성은 TF-IDF가 높으나,

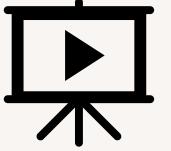
내외부적 공간디자인의 유사성은 W2V 추천 결과가 더 높을 것으로 판단하였다.

숙소 추천을 위한 유사도 분석



고즈넉한 느낌의 감성 한옥스테이 숙소 #전주 #다가도원 🌻 하루에 딱 한 텁만 받는 온천히 나만을 위한 특별함 ✨프라이빗한 독채공간의 한옥스테이 등 유리창 옆의 실내욕조와 실외에는 선녀탕이라고 불리는 노천탕으로 힐링 제대로 느낄 수 있어요 ♥ 밖에는 캠핑세트가 있어서 놀기 있는 저녁까지 즐길 수 있는 숙소 편안한 무릉도원 같은 한옥스테이 전주 다가도원에서 힐링 느껴보세요 🌟💡 주소전부 전주시 완산구 전주객사1길 95-18 📞 Tip◆ 입실 15:00~◆ 퇴실 11:00 🌿 숙소기록위치 : 전주 객리단길 근처에 있어서 접근성 좋음(지도에서는 확인되지 않음) 가격 : 평일 28만원 / 주말 35만원특징 : 성인 최대 2인만 가능하고 인원추가 ✗ , 노키즈존 , 노천탕에 실내탕까지 힐링제대로 즐길 수 있음 - 국내외 예쁜 숙소 정보는 #숙소기록 #숙소 제보는 DM 또는 이메일로 ❤️ #숙소기록 _국내#숙소기록_전주 #숙소 #숙소추천 #국내숙소 #여행 #여행지추천 #국내여행지추천 #국내여행 #여행스타그램 #커플여행 #가족여행 #감성숙소 #전주숙소 #전주숙소추천 #전주가볼만한곳 #전주여행 #전주감성숙소 #전주핫플 #전주핫플레이스 #전주풀빌라 #전주에어비엔비 #전주독채맨션 #전주펜션





서비스 소개

웹사이트 소개

UI 설계도

기능 분해도

기능분해도

기능 / 구성	URL(localhost:8000)	VIEW	TEMPLATES	REDIRECTION
팀소개/팀원소개	/	home	index.html	
회원가입	/register	register	register.html	/login
로그인	/login	login	login.html	/register
DB login 확인	/logcon	logcon		
DB 등록 확인	/regcon	regcon		
랜딩페이지	/choice	choice	choice.html	
인기숙소 둘러보기	/popular	showPop	popular.html	/choice
테마 둘러보기	/theme	showTheme	theme.html	/choice
전체숙소 둘러보기	/totalstay	totalstay	totalstay.html	/choice
테마별 분류 목록	<str:theme>/stay/	stayFilter	stayfilter.html	/choice
상세페이지 내용	<str:name>/detail/	stayDetail	staydetail.html	/choice

templates / html

```

> static
└ templates\homepage
  <> choice.html
  <> index.html
  <> login.html
  <> popular.html
  <> register.html
  <> relogin.html
  <> staydetail.html
  <> stayfilter.html
  <> theme.html
  <> totalstay.html
  ...

```

url.py

```

homepage > urls.py > ...
1  from django.conf.urls import url
2  from django.urls import path
3  from . import views
4
5  app_name = 'homepage'
6
7  urlpatterns = [
8      path('', views.home, name='home'),
9      path('register/', views.register, name='register'),
10
11     path('regcon/', views.regcon, name='regcon'),
12     path('logcon/', views.logcon, name='logcon'),
13     path('choice/', views.choice, name='choice'),
14     path('totalstay/', views.totalstay, name='totalstay'),
15     path('popular/', views.popular, name='popular'),
16
17     path('login/', views.login, name='login'),
18     path('theme/', views.showTheme, name='theme'),
19     path('<str:theme>/stay/', views.stayFilter, name='stay'),
20     path('<str:name>/detail/', views.stayDetail, name='staydetail'),
21 ]

```

UI 설계

숙소 추천 서비스 – 테마 선택 페이지

The UI design mockup illustrates a travel service theme selection page. The main content area includes:

- ① 도구상자 (Toolbox)
- ② background Image (Background Image)
- ③ Text (Text): sample('숙소 테마 선택하기')
- ④ (A dotted line indicating a scrollable area)
- ⑤ image keyword (Image Keyword): A grid of 6 items.
- ⑥ (A dotted line indicating a scrollable area)
- ⑦ 하단배너 (Bottom Banner)
- ⑧ top (Top button icon)

The sidebar contains a table titled "DEVELOP Description" with 8 rows, each corresponding to a numbered callout in the main content:

DEVELOP Description	
1	도구상자 구성요소 홈버튼 인기숙소 둘러보기 코로나 종합안전정보 좌측 상단 고정 형태로 처리
2	백그라운드 이미지 고정노출
3	상단 테마 이미지 & 텍스트 바
4	구성요소 키워드 속성에 맞는 이미지 노출 여행 속성에 맞는 키워드 제시
5	추천 기준이 될 수 있는 이미지 및 텍스트 선택 기능
6	여행 키워드 9개 노출 후 스크롤 처리
7	전 페이지 공통배너 처리
8	top버튼 기능 제공

웹 페이지 구현

모든 순간이 여행이 되도록



테마별 숙소
보러가기



인기숙소
보러가기



전체숙소
보러가기



결론

기대 효과 및 시사점

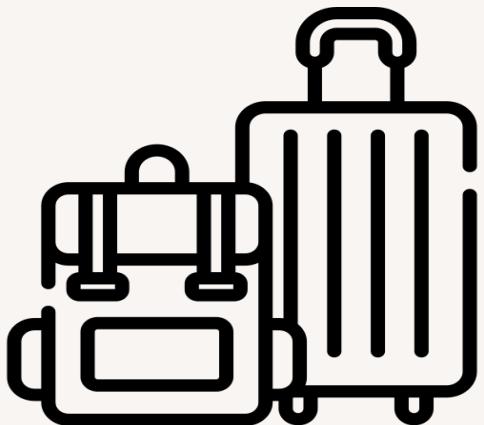
개선점

구성원 역할

개발 후기

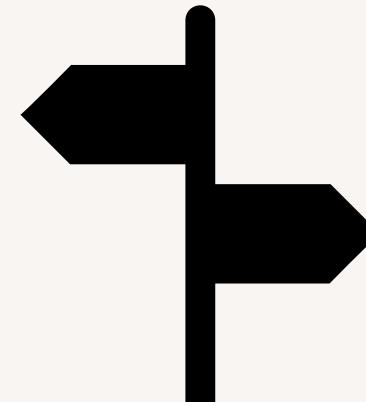
기대 효과 및 시사점

서비스 측면



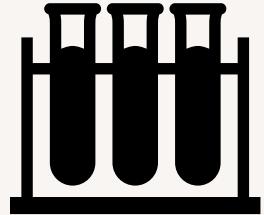
새로운 여행트렌드에 맞춘
여행산업 활성화 지원

연구 측면



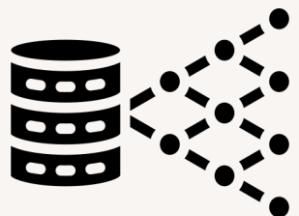
감성숙소에 대한 연구의
전초적인 역할 수행

개선점



선행연구의 미비

선행연구의 미비로, 보편화된 기준이 없어, 분석 결과의 신뢰도나 객관성 ↓



이미지 분석

이미지 분석을 활용하면 실제 공간에 대한 분석 정확도 ↑



개인화

개인의 취향과 특성을 반영할 수 있는 맞춤형 추천 시스템을 통해 서비스 만족도 ↑

개발 후기

역할: 팀장, 발표, PPT제작, 문서관리, 작업 스케줄링, 모델링, 분석모델 아키텍쳐 구성, 데이터 분석, 전처리

이민성

후기 : 새로운 분들과 새로운 프로젝트를 진행하면서, 많은 것을 배울 수 있었던 좋은 시간이었습니다. 쉽지 않은 경험이었지만 이 속에서도 여러 문제를 해결해가며 새로운 방법들을 다양하게 시도해볼 수 있었던 것은 좋았습니다. 프로젝트를 끝까지 함께 해준 팀원들과 옆에서 지속적인 관심과 도움을 주신 멘토님들, 김성환, 이경미 강사님께 감사인사 드리고 싶습니다.

역할:PPT/회의록 작업, 파이프라인 구축, 데이터수집 및 전처리, 웹사이트 제작, 서비스 구현, 아이디어 제시

최원희

후기 : 저에게 유익하고 만족스러운 경험이었던 본 교육 목표의 최종 프로젝트가 무사히 끝나서 정말 기쁩니다. 데이터 엔지니어링 전 과정을 통해 배운 개념과 프로그램을 직접 사용했다는 부분이 프로젝트 내에서 가장 뜻깊은 부분이였습니다. 크롤링과 open api부터 spark, hadoop, mongo, airflow를 이용한 백엔드엔드와 Django를 이용한 프론트엔드까지 함께하며 도움주신 ojo조원분들에게 감사하다는 말 드리고 싶습니다. 프로젝트 기간 동안 아낌없이 조언주신 이경미강사님과 김성환 강사님께도 감사드립니다.

김정명

역할: PPT/DFD 작업, 아키텍처 설계, 데이터 전처리, 웹사이트 제작, 서비스구현

후기 : 팀원들과 함께 마지막 프로젝트를 진행하면서 많은 것들을 배우고 재밌는 시간이였습니다. 프로젝트 진행 과정 중 많은 어려움이 있었지만 서로 분업을 통해 각자가 해결할 수 있는 부분에 대해서 열심히 한 결과 좋은 결과물이 만들어졌다고 생각합니다. 또한, 이번 교육과정으로 데이터 엔지니어에 대한 인프라 구축, Spark, Nosql, Airflow 등 많은 것들을 사용한 덕분에 앞으로 많은 도움이 될 것이라고 생각합니다. 마지막으로 고생하신 팀원 분들과 강사님들에게 감사하다는 말씀을 드리고 싶습니다.

박성하

역할: PPT 작업, 데이터 전처리, 데이터수집 및 웹사이트 제작, 서비스 구현

후기 : 조원들의 도움으로 다양한 툴을 사용하고 공부할 수 있는 기회가 되었습니다. 최종적으로는 실제 서비스 구현까지 연결되어 교육과정 유종의 미를 거두게 된 좋은 경험이 되었습니다.

김태리

역할: 문서 작업, 선행연구 조사, 데이터 전처리, 데이터 분석, 모델링

Reference

저자	참고자료
Anna Huang.(2008)	Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research 20 student conference (NZCSRSC2008), Christchurch, New Zealand, pages 49–56
Inderjit S Dhillon and Dharmendra S Modha. (2001)	Concept decompositions for large sparse text data 120 using clustering. Machine learning, 42(1):143–175,
전병국,안현청(2015)	사용자 리뷰 마이닝을 결합한 협업 필터링 시스템 *스마트폰 앱 추천에의 응용*. 한국지능정보시스템학회
임명진, 김판구,신주현(2020)	리뷰의 의미적 토픽 분류를 적용한 감성 분석 모델. Smart Media Journal / Vol.9, No.2 / ISSN:2287-1322
현지연, 유상이, 이상용(2019)	평점과 리뷰 텍스트 감성분석을 결합한 추천시스템 향상 방안 연구. 한국지능정보시스템학회, 지능정보연구 제25권 제1호
박호연, 김경재(2021)	BERT 기반 감성분석을 이용한 추천시스템, 한국지능정보시스템학회, 지능정보연구 제27권 제 1호 , 2021년, pp.1 – 15
소셜미디어 빅데이터 활용 국내관광 트렌드 분석 및 2019 트렌드 전망	한국관광공사, 한국빅데이터센터, 2028년
이현석, 안성훈 외 5명(2016)	Word2Vec을 이용한 웹 문서 클러스터링 시스템 구현, 한국정보처리학회, 2016년도 추계학술발표대회, pp.26 - 29
임근영, 조영복(2018)	딥러닝과 Char2Vec을 이용한 문장 유사도 판별, 한국정보통신학회, 한국정보통신학회논문지 Vol. 22, No. 10: 1300~1306, Oct. 2018
배영준, 장호택 외 2명(2018)	향상된 TextRank 알고리즘을 이용한 자동 회의록 생성 시스템, 한국정보전자통신기술학회논문지 v.11 no.5 , 2018년, pp.467 - 474
서봉원(2016)	콘텐츠 추천 알고리즘의 진화, 방송 트렌드&인사이트 , 2016. 04+05 VOL. 05
오재택, 이상용(2020)	Fuzzy-AHP와 Word2Vec 학습 기법을 이용한 영화 추천시스템, 한국디지털정책학회, 2020, vol.18, no.1, pp. 301-307 (7 pages)

Web UI Template

The image displays four screenshots of Bootstrap templates, each enclosed in a blue rounded rectangle and numbered 1 through 4.

- Screenshot 1:** A grayscale one-page Bootstrap theme named "GRayscale". It features a large header with the title "GRayscale" and a "GET STARTED" button. The URL is "Start Bootstrap".
- Screenshot 2:** A shop homepage template titled "Shop in style". It shows a grid of eight product cards with placeholder images and titles like "Fancy Product", "Special Item", "Sale Item", and "Popular Item". The URL is "Start Bootstrap".
- Screenshot 3:** A blog post template titled "Welcome to Blog Post!". It includes a sidebar with categories like "Web Design", "HTML", "CSS", and "Tutorials". The URL is "Start Bootstrap".
- Screenshot 4:** A modern business template titled "A Bootstrap 5 template for modern businesses". It features a large central image placeholder labeled "600 x 400". The URL is "Start Bootstrap".

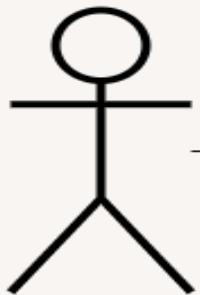
Bootstrap 템플릿 활용

1. 메인 화면
2. 감성숙소 테마 선택
3. 세부 숙소 선택
4. 해당 숙소 설명

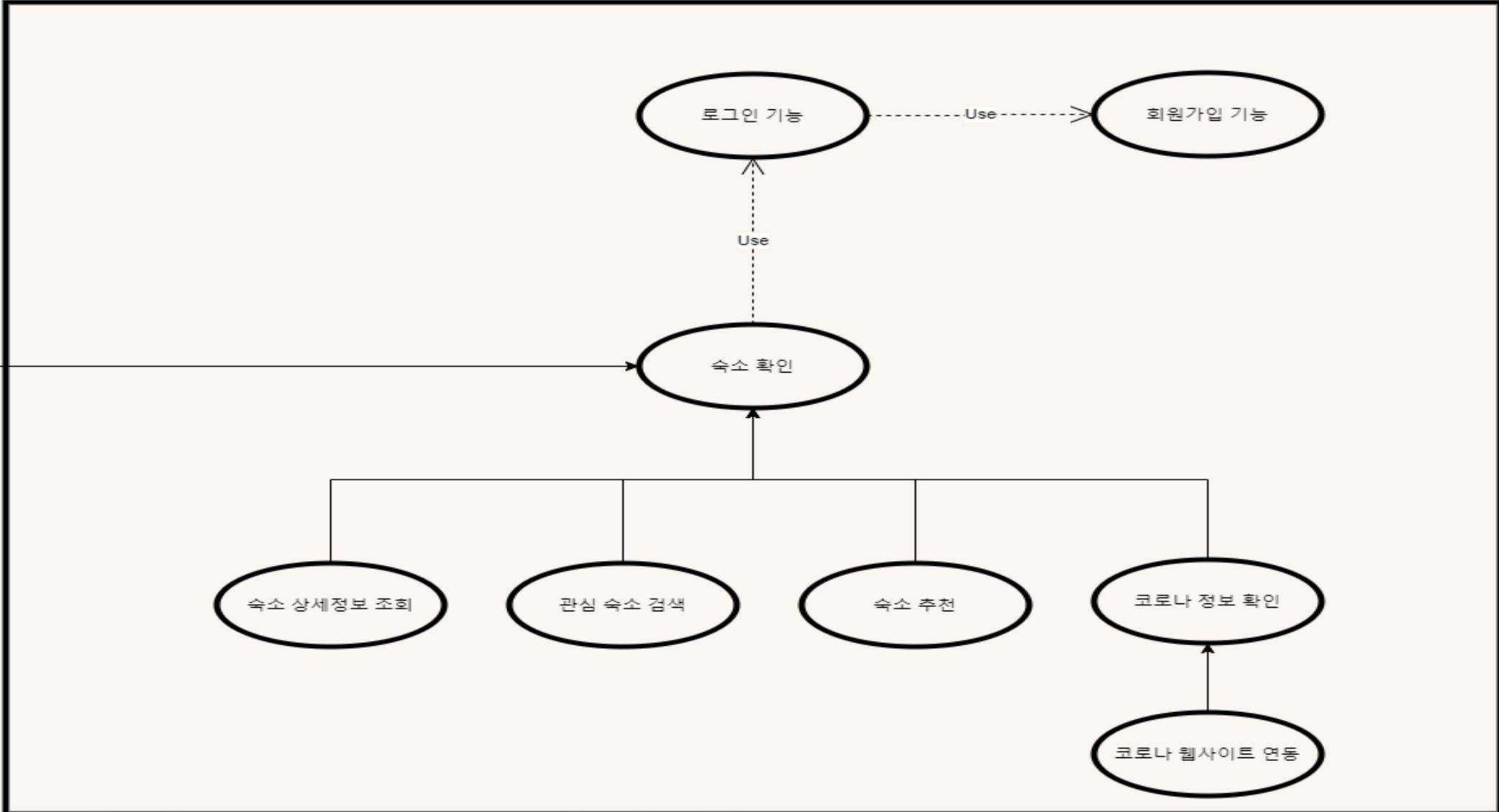
THANK YOU

ojo

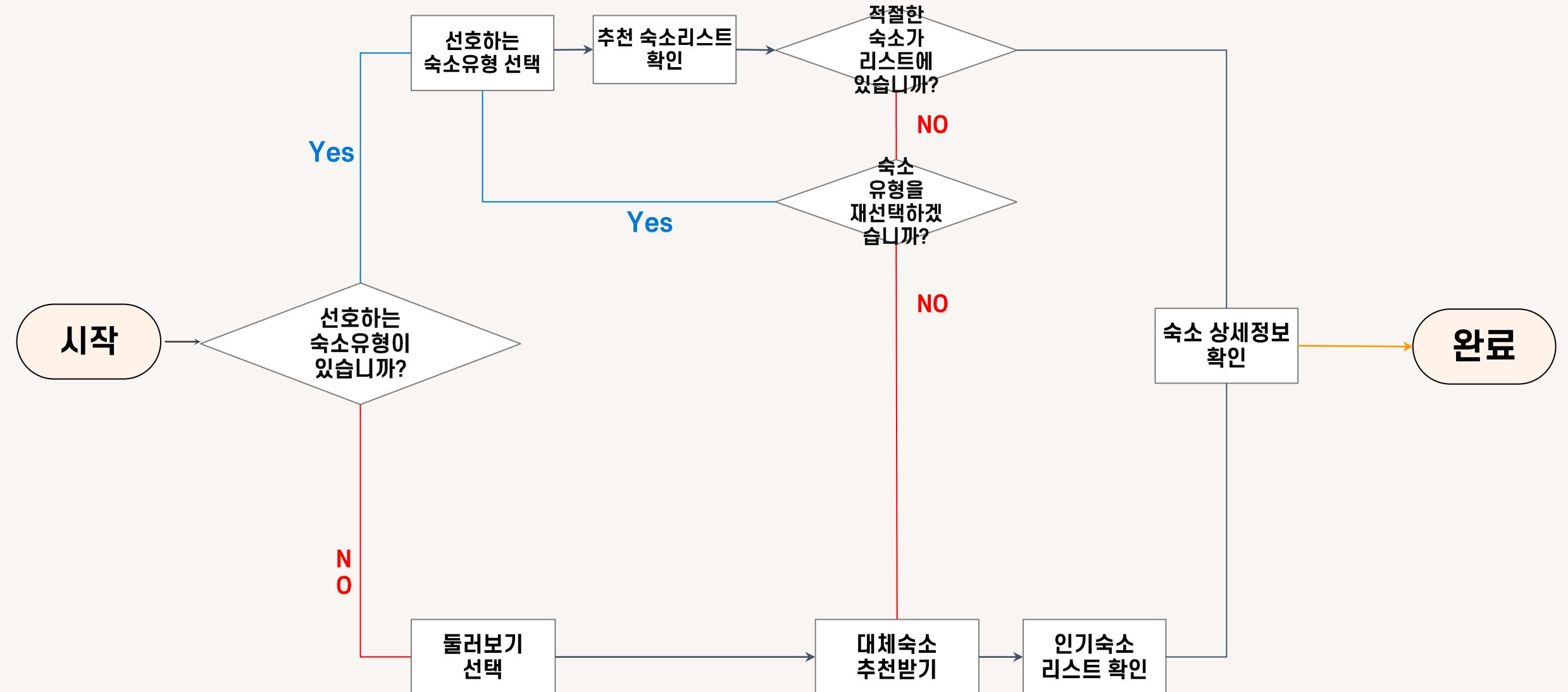
Usecase 다이어그램



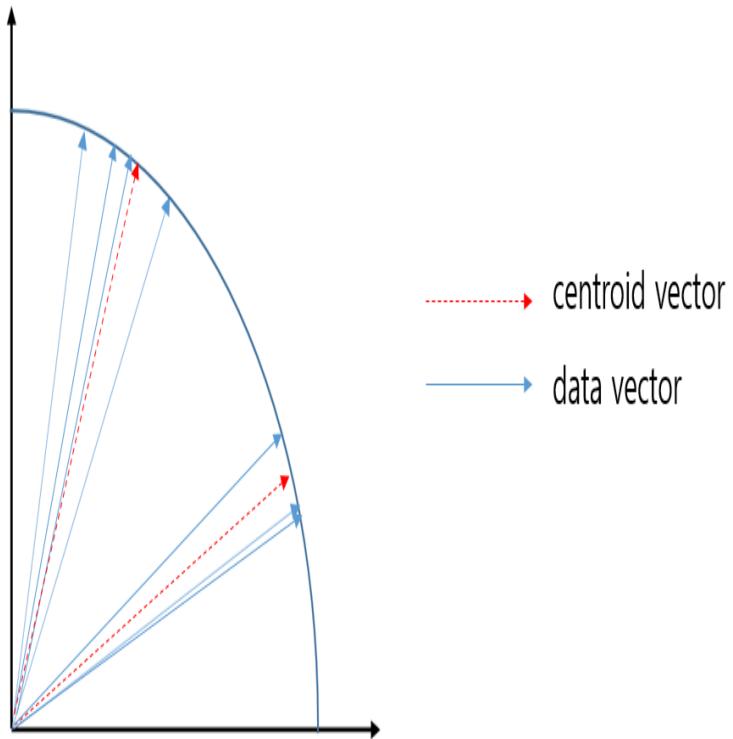
사용자



서비스 플로우

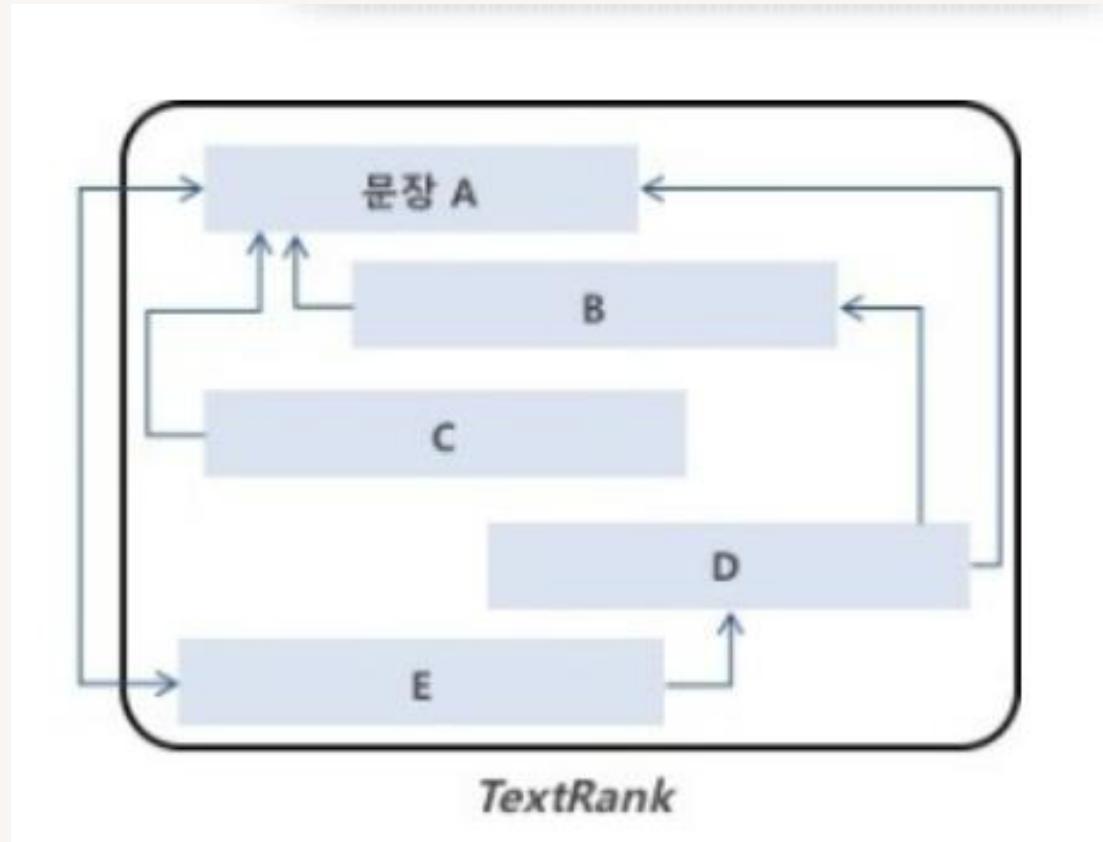


Spherical K-means



- 고차원의 sparse vectors 간의 거리 척도는 두 벡터에 포함된 공통 성분의 무엇인지를 측정하는 것이 중요하기 때문에 Jaccard distance, Pearson correlation, Cosine distance 와 같은 척도를 쓰면 거리가 잘 정의되나, Euclidean distance를 이용하면 안됨 (Anna Huang, 2008).
- Cosine distance 를 이용하면 모든 벡터가 unit vector 화 되기 때문에 k-means 군집화는 아래 그림처럼 각도가 비슷한 (단어 분포가 비슷한) 문서를 하나의 군집으로 묶는 의미로 해석할 수 있습니다. 이처럼 sparse vector 로 표현되는 문서 집합에 대한 k-means 는 Cosine distance 를 이용하는 것이 좋고, Cosine distance 를 이용하는 k-means 를 Spherical k-means라 함 (Inderjit et al., 2001).

군집분석



출처:<https://excelsior-cjh.tistory.com/93>

TextRank의 특징

- Co-occurrence: 단어의 동시 발생 빈도
- 그래프 : 텍스트의 어휘나 의미로 만든 그래프를 기반으로 함
- 텍스트 간 관계 중심 : 텍스트 간 관계를 중심으로하여 분석
- 대표적 방법론 : 문서요약, 키워드 추출의 대표적인 방법론임