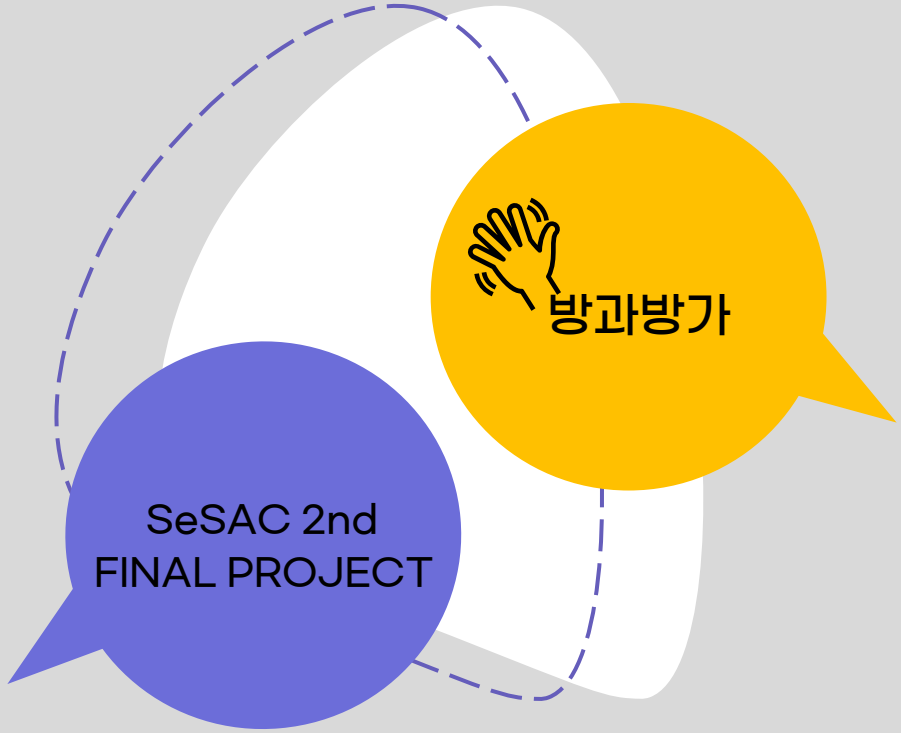
리뷰분석을 통한
개인 맞춤형
숙소 추천 서비스

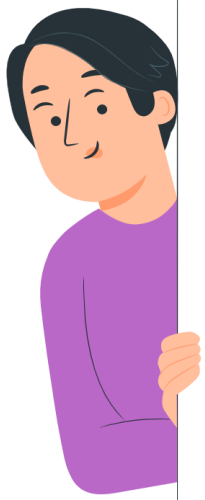
검색엔진과 추천시스템



목차

1. 팀소개
2. 프로젝트 소개 - 프로젝트 배경 및 목적 / 최종목표, 차별점 / 타임라인
3. 프로젝트 설계 - 서비스 플로우 / 아키텍처 / 엘라스틱서치
4. 프로젝트 진행 과정 - 데이터/ 추천모델 / 엘라스틱서치
5. 보완(문제 제기 & 문제 해결) - 엘라스틱 reranking / 추천고도화
6. 웹 시연 - 기능 / 웹구현 / AWS
7. 마무리 - 의의 / 한계점 / 향후계획

1. 팀 소개



방과방가

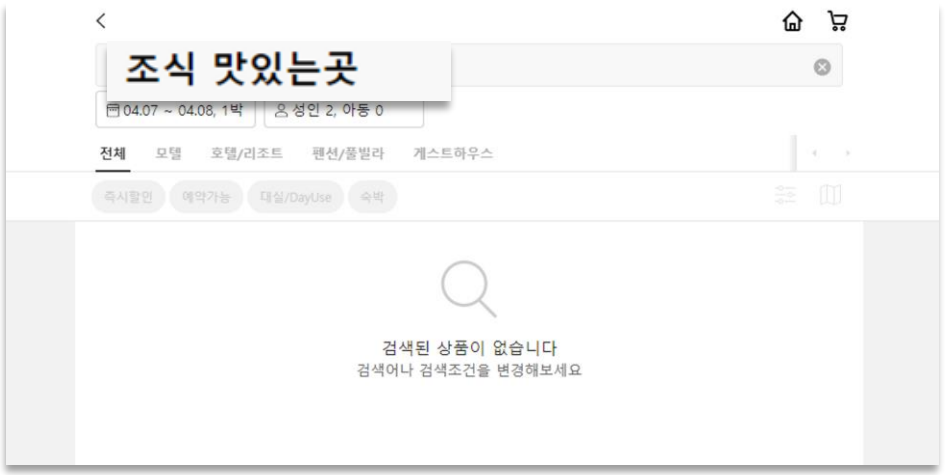
Room&Hi

: 내가 마음에 드는 방과 만날 수 있다.

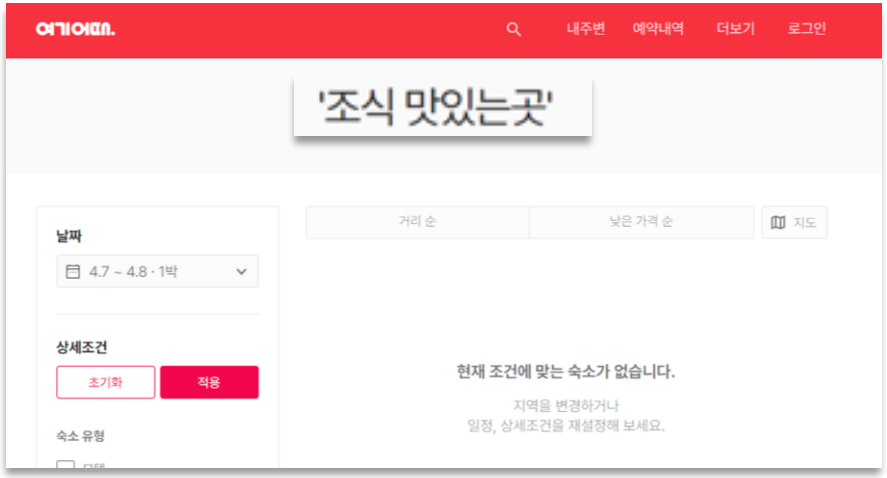
김다나 이지수 고준호 이선영 김경연

2. 프로젝트 소개

배경



국내 숙박 플랫폼 1위 야놀자



국내 숙박 플랫폼 2위 여기어때

검색피로도 관련
많은 양의 리뷰를 나타내는 이미지 -> 많은 리뷰들 속에서 내가 원하는 숙소를 하나하나 찾아야한다는 수고스러움이 나타나야함

최종목표

- 기존 숙박 플랫폼은 조식이 맛있는 호텔을 찾기 위해서는 조식에 대한 후기를 일일이 찾아봐야 했음

- 하지만 우리 서비스는 '조식 맛있는 호텔' 을 검색하면 됨 일일이 이 호텔 저 호텔을 찔러볼 필요가 없음.

소비자의 니즈를 정량화한 데이터와 리뷰로 제공함으로써
소비자의 검색피로도를 최소화하고 소비자에 맞게 정보를 전달하는 것

Project Timeline

Tasks	Feburary	March	April
1. 프로젝트 기획 및 사전 조사	2.15 - 2.21		
2. 프로젝트 설계	2.20 - 2.21		
3. 데이터 수집 및 전처리	2.21 - 3.7		
4. 추천 시스템 구현	2.28 - 3.8		
5. 웹 구현 및 AWS	3.6 - 3.16		
6. TEST 1차 - 보완		3.16 - 3. 22	
7. TEST 2차 - 보완			3.23 - 3. 31
8. 최종 TEST 및 배포			4.1 - 4.6

3. 프로젝트 설계

서비스플로우

애니메이션으로 단계별 강조 해주기 –
시스템 아키텍처랑 합치는 방법 으로 !!

Room&Hi

영등포에서 무료조식에다가 가성비 좋은 호텔 알려줘

검색엔진 (+STT)

사용자

ElasticSearch

"영등포" 에서 "무료조식에다가 가성비 좋은 호텔 알려줘"

지역 필터링 + 문장 벡터화 (SBERT)

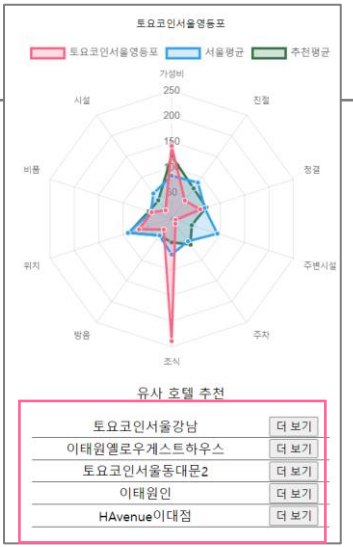
문장 유사도 계산

리뷰 문장 벡터화(Sbert) 저장

리뷰 데이터

호텔 등급, 지역, 날짜, 리뷰, 리뷰 수, 이미지, url

No.	호텔	등급	보러가기	더보기
1	토요코인서울영등포	3	B. × ×	더 보기
2	영등포데쟈르	모텔	× yo	더 보기
3	코트야드메리어트서울타임스퀘어	4	B. yo	더 보기
4	영등포호텔갤럭시	모텔	× yo ×	더 보기
5	영등포부띠크HotelSB	3	B. ×	더 보기



유사 호텔 추천

토요코인서울강남	더 보기
이태원엘로우게스트하우스	더 보기
토요코인서울동대문2	더 보기
이태원인	더 보기
HAvenue이대점	더 보기

카테고리별 긍부정 리뷰 제공

추천 시스템

지역	영등포구	선택안함				
등급	☐ 전체	☐ 5성	☐ 4성	☑ 3성	☐ 모텔	☐ 게스트하우스
가성비	상관없음	덜 중요	보통	중요	**매우중요**	
친절	상관없음	덜 중요	**보통**	중요	매우중요	
청결	상관없음	덜 중요	**보통**	중요	매우중요	
주변시설	**상관없음**	덜 중요	보통	중요	매우중요	
주차	**상관없음**	덜 중요	보통	중요	매우중요	
조식	상관없음	덜 중요	보통	중요	**매우중요**	
방음	상관없음	덜 중요	**보통**	중요	매우중요	
위치	상관없음	덜 중요	**보통**	중요	매우중요	
비품	상관없음	덜 중요	**보통**	중요	매우중요	
시설	상관없음	덜 중요	**보통**	중요	매우중요	

Python

카테고리별 리뷰 긍부정 **라벨링**

카테고리별 숙소 **점수화**

	가성비	친절	청결	주차	조식	...	시설
호텔 1	23	57	79	5	88	...	65
호텔2	11	90	99	45	22	...	87
호텔3	99	36	43	55	54	...	32
호텔4	100	90	98	88	89	...	100
...	...	...	...	...	...	...	...
호텔 386	70	87	88	12	22	...	90

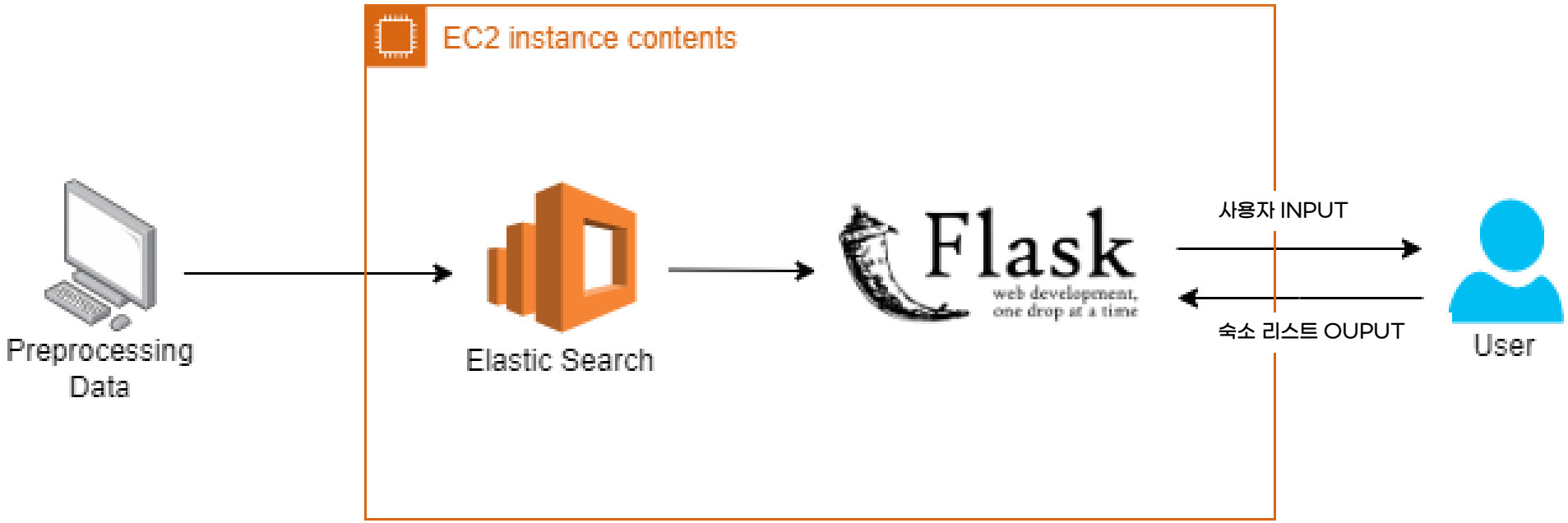
"영등포", "3성" 1차 필터링 + 카테고리별 가중치 반영하여 점수 계산

피어슨 상관계수
유사 호텔 추천

No.	호텔	등급	보러가기	더보기
1	토요코인서울영등포	3	B. × ×	더 보기
2	영등포GMS호텔	3	× yo	더 보기
3	더스테이트선유	3	B. yo	더 보기
4	호텔부티크9	3	× yo ×	더 보기
5	영등포부띠크HotelSB	3	B. ×	더 보기

시스템 아키텍처

이거 검색엔진이랑 추천시스템 나눠서 표기 하기



기술개요

엘라스틱서치란?

Apache Lucene기반의 Java 오픈소스 분산 검색 엔진으로 방대한 양의 데이터를 신속, 실시간으로 저장, 검색, 분석할 수 있는 검색엔진

장점 1

inverted index(역색인)을 사용하는 방식으로 데이터 검색에서 관계형 데이터 베이스보다 빠른 성능

장점 2

전문검색을 통해 TF-IDF방식으로 유사도가 높은 데이터를 출력 가능

사용 이유

사용하는 데이터가 많고 검색해야할 쿼리가 많았기 때문에 RDBM 대신 엘라스틱서치를 선택

4. 프로젝트 진행

데이터

데이터 수집

서울 386개 숙박 시설

yanolja 310개

여기어때. 257개

Booking.com 157개

- 수집 데이터 -
날짜 . 리뷰 . 별점
총 397,152개

데이터 전처리

1. 특수문자, 개행문자 필터링
2. 날짜 형식 맞추기
   : '16분전', '4일 전' , '3개월 전' → yyyy-mm-dd
3. 불용어 제거
   :한국어 불용어 사전 + 서울,호텔,리뷰 …
4. 문장 분리 (KSS)
5. 맞춤법 검사 (Hanspell)

ht_id	date	star	review
1	2021.08.01	8	위치며 청결함이며 친절함 무엇 하나 부족함 없음
1	2021.08.01	8	다만 창문 없는 객실 이용하게 되어 조금 답답했음
1	2021.05.02	10	재미있게 잘 보내다 갑니다
1	2021.05.02	10	사실 창문 내다봤을 때 전망 괜찮은 데를 원했는데 딱 그곳으로 배정해 주셔
1	2021.05.02	10	그리고 와인이랑 샐러드도 넘넘 맛있게 잘 먹었어요
1	2023.02.14	10	서울 갈 때마다 이용합니다
1	2023.02.14	10	위치적으로도 좋고 친절합니다
1	2023.02.12	10	명동역과 가까운 위치에 있고 깔끔하고 구성비 있는 곳이에요

추천 모델 - 카테고리사전 수정전

* MeCab 명사 빈도수 추출 → 총 36개의 카테고리

카테고리	키워드(하위단어)
가격	구성비, 가성비,가격
직원	친절, 직원, 응대, 대응, 사장
프론트	프런트, 체크인, 체크아웃
통신	인터넷, 와이파이
분위기	분위기, 조명
시설/인테리어	인테리어, 디자인, 외관, 엘리베이터, 에레, 신추, 건물 , 시설
로비	로비, 입구
라운지	클럽라운지, 옥상, 루프탑
주차장	주차장, 주차, 발레, 차량
산책	산책, 공원
부대시설	편의시설, 부대시설
수영장/사우나/스파	수영장, 스파, 사우나, 수영

카테고리	키워드(하위단어)
휘트니스	피트니스
비즈니스	비즈니스, 출장
위치	위치, 접근성, 거리, 이동, 도보, 도심, 식당
교통	지하철, 자히철역, 공항, 버스, 대중교통, 교통
관광	쇼핑, 백화점, 관광, 관광지, 볼거리
가구	소파, 책상, 가구, 옷장, 탁자,전등
침대/가구	침대, 침구류, 베드, 트윈, 매트리스, 푹신, 시트, 침구, 침실, 커버
커튼/카페트	카펫, 블라인트, 커튼 , 카페트
가전/전자제품	에어컨, 티브이, 히터, 컴퓨터, 가습기, 청정기, 정수기, 전자레인지, 세탁기, 스피커, 가전, 전자제품
화장실	욕조, 욕실, 비데, 배수구, 목욕, 반신욕, 샤워, 화장질
비품	어메니티, 수건, 칫솔, 용품, 타올, 일회용품, 물품, 비누,린스, 휴지, 면도기, 제품, 세면도구, 비품
물	수압, 온도, 온수, 물

카테고리	키워드(하위단어)
객실	객실, 공간, 천장, 테라스, 조리, 취사
방음	소리, 방음, 옆방
온도	냉방, 중앙난방, 보일러, 외풍
바닥/벽	바닥, 벽지, 벽
창문	창문
룸컨디션	컨디션, 노후, 연식, 낙후, 모기, 룸컨디션
냄새	냄새, 담배, 하수구
전망	경치, 풍경, 야경, 뷰, 전망
편의용품	충전기, 콘센트, 슬리퍼, 생수, 옷걸이 ,편의용품, 드라이기
청결	청결, 먼지, 머리카락, 얼룩, 곰팡이, 정리, 자국, 물때, 쓰레기, 벌레,청소 , 깔끔, 깨끗
룸서비스	룸서비스
조식	조식, 아침식사, 뷔페

추천모델 - 감정분류 모델

모델 명	Accuracy
KoBERT	0.9122
KLUE-BERT-base	0.8912
KLUE-RoBERTa-large	0.9344
kykim/funnel-kor-base	0.9404
KoELECTRA-Base-v3	0.9443

	precision	recall	f1-score	support
0	0.95	0.96	0.96	1043
1	0.96	0.96	0.96	1062
2	0.90	0.88	0.89	624
accuracy			0.94	2729
macro avg	0.94	0.94	0.94	2729
weighted avg	0.94	0.94	0.94	2729

Classification report

라벨링 결과

ht_id	date	review	label
188	2020.02.27	미니바는 술 제외한 음료는 다 무료였고 어메니티 보디로션 향이 너무 좋네요	1
353	2021.11.21	위치 좋고 직원의 친절함은 정말 환상이었습니다	1
154	2020.02.25	침대가 4개 있었는데 위치 선정이 너무 좋았고 저렴한 가격 대비 너무 좋은 공간이었어요	1
302	2023.02.26	접근성이 좋고 주변 편의시설이 많아서 좋아요	1
225	2021.05.27	직원들이 상냥합니다	1

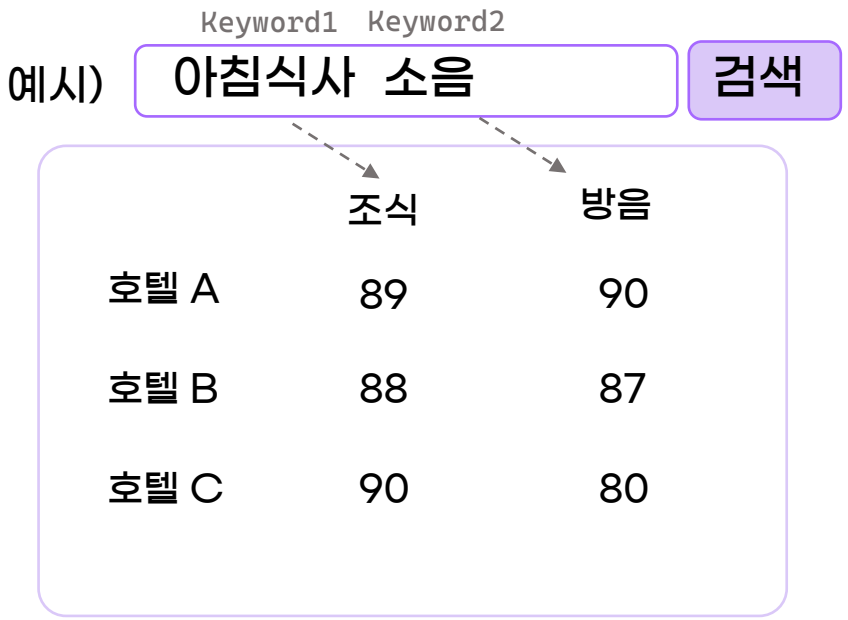
ht_id	date	review	label
190	2022.02.23	청소 및 물품 비치가 살짝 아쉬웠다	0
111	2020.06.04	실내 수영장까지의 이동경로가 불편	0
188	2022.01.15	샤워실 면 한쪽이 유리 문이라 좀 부담스러웠음	0
80	2022.12.27	천장 히터만 틀어지고 창문 쪽에선 찬바람이 너무 많이 들어와서 놀러 와서 감기 걸렸어요	0
246	2022.09.27	다만 저한테는 샤워기 키가 크신 분들한테는 불편하실 수 있습니다	0

ht_id	date	review	label
363	2022.02.27	다만 강남이라 주차는 이해해야 하죠	2
180	2022.08.27	중문 없는 게 이렇게 큰가 싶을 정도로 복도 소음도 크지만 거 빼면 다 좋아요	2
26	2023.01.03	역에서도 가깝고 조명이 조금 어둡긴 하지만 이용하는데 불편함 없어요	2
126	2022.03.01	시설이 오래되었지만 직원 서비스로 다 커버가 되네요	2
237	2022.07.04	위치 접근성 최곤데 침대 뜯어진 자국 있고 그런 것만 빼면 좋았어요	2

추천모델- 만족지수

$$\text{긍정비율} = \frac{\text{해당 카테고리 긍정 리뷰 수}}{\text{해당 카테고리 언급 리뷰 수}} \times 100$$

$$\text{만족지수} = \frac{\text{긍정비율}}{\text{긍정비율} + \text{부정비율}} \times 100$$



두 개의 카테고리 점수의 합이 가장 큰 숙박시설 10개를 추천

엘라스틱서치

리뷰데이터가 들어갈 필드와 타입 지정

```
es.indices.create(
    index='re_list',
    body={
        "mappings": {

            "properties": {
                "ht_id": {
                    "type": "keyword"
                },
                "date": {
                    "type": "date",
                    "format": "yyyy.mm.dd"
                },
                "review": {
                    "type": "text",
                },
                "label": {
                    "type": "integer",
                }
            }
        }
    }
)
```

고유값으로 검색 가능

날짜 범위 지정하여 검색 가능

일부단어, 유사도 검색 가능

숫자값으로 검색 가능

“아기랑 놀기 좋은 곳 "

```
12.538096 / 혼자 쓰기 딱 좋은 곳이에요
12.470367 / 위치도 좋고 무엇보다 아기와 함께하기 좋았어요
12.202542 / 위치도 매우 찾기 좋은 곳입니다
12.128616 / 새로 오픈한 곳이라 깔끔 간단한 취사가 가능 히노끼탕에서 놀기 좋음
12.0970955 / 광장시장 최고 낮에는 아기랑 수영장에서 즐거운 시간 보냈네요
11.885953 / 호캉스하기 딱 좋은 곳입니다
11.885953 / 깔끔하게  하루 머물기 좋은 곳입니다
11.8196335 / 가까운 곳에 카페랑 서브웨이 있어서 너무 좋았어요
11.814242 / 5명 호텔 숙박 되는 곳 찾기 쉽지 않은데 잘 놀고 갑니다
11.596999 / 대학로 근처라 놀기도 좋고요
```

5. 보완

문제제기랑 문제해결 페이지 구분하기

피드백 + 자체 테스트 했었다는 얘기로 풀고 시작하기

엘라스틱서치 - Reranking

문제 제기

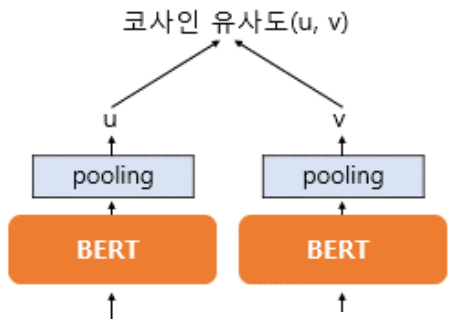
엘라스틱서치 기본 성능의 한계 : TF-IDF 검색에서는 중요 단어, 키워드가 정확하게 추출 되지 않음

문제 해결

SBERT(Sentence BERT) 모델을 통해 문맥에 따라 유사도를 검색하는 문장임베딩 수행

-> Fast Elasticsearch Vector Scoring plugin 사용

SBERT(Sentence BERT)



"깨끗하고 청결한 호텔 " "청결하면서 깨끗하기도 한 호텔 "

두 문장에서 문장벡터의 평균을 통해 유사한 문장임을 확인

"아기랑 놀기 좋은 곳 "

1.767511 / 저녁에 산책도 좋았고 아기랑 놀기 괜찮았어요

1.7231867 / 아기들 놀 수 있는 수영장 이 하나 있습니다

1.7157646 / 아기랑 같이 산책했는데 좋아하더라고요

1.7097986 / 위치도 좋고 무엇보다 아기와 함께하기 좋았어요

1.7081176 / 위치도 찾기 좋고 거리도 적당하고 애들 놀기에도 좋고 괜찮았어요

1.7040801 / 위치가 일단 아주 좋고 근처 놀 수 있는 곳들이 많다

1.6928089 / 아기 반려견과 같이 갈 수 있는 점도 좋았고 식사 수영장까지 포함돼 있어 좋았습니다

1.6909626 / 놀이공원이 같이 있어 좋아요

1.6898826 / 욕조가 크고 넓어서 아기가 놀기 좋았어요

엘라스틱서치 - Reranking

문제 제기

엘라스틱서치 기본 성능의 한계 : TF-IDF 검색에서는 중요 단어, 키워드가 정확하게 추출 되지 않음

문제 해결

지역, 지하철역, 호텔 성급 등과 같은 숙소의 특성을 1차적으로 필터링하여 검색 정확도 향상

-> 정규표현식을 통해 해당 단어가 있을 경우 필터링하여 검색 조건에 가중치 부여

최종 Mapping 형태

문장벡터화

필터링 데이터 추가

```
{"mappings": {
        "properties": {
            "ht_id": { "type": "keyword"},
            "
            "
            "cat_name": { "type": "keyword" },
            "date": { "type": "date", "format": "yyyy.mm.dd"},
            "review": { "type": "text"},
            "rev_vec": { "type": "dense_vector", "dims": 768},
            "label": {"type": "integer" },
             "grade": { "type": "keyword" },
            "gu": { "type": "keyword"},
            "station": {"type": "keyword"},
           "station2": {"type": "keyword"},
            "word1": {"type": "keyword"},
            "word2": {"type": "keyword"},
            "word3": { "type": "keyword"}
        }
    }
}
```

“ 영등포에서 친구와 함께 놀기 좋은 호텔”

ht_id	score	cat_id	date	review
213	440.61902	10	2023.01.27	여의도 한강공원 놀러 갔다가 친구들이랑 가기 딱 좋아요
288	430.20334	3	2022.03.27	여자친구 생일이라 겸사겸사 여자 친구랑 놀러 갔는데 방도 크고 깔끔하고 너무 좋았네요
216	415.21790	8	2023.02.24	근처에 먹거리 타운도 가까워서 놀다가 술 마시고 자고 가기 좋아요
307	414.53528	10	2023.02.25	역에서 가깝고 주변에 편의시설 등 많아서 이동하기 놀기 다 좋았어요
294	400.75708	1	2020.11.26	접근성이 좋고 구성비가 좋은 호텔이다
215	399.58618	3	2022.12.27	여자친구랑 갔는데 시설 깔끔하고 좋았습니다

추천모델 개선 - 카테고리 사전 수정후

문제 제기

개인의 선호를 구체적으로 반영하기 위해 카테고리 수를 확장하였지만, 그로 인해 카테고리에 포함되는 리뷰 수 부족 문제 발생

문제 해결

초기 36개에서 10개로 카테고리 수를 축소함으로써 카테고리별 리뷰 수를 확보

-> 개인의 세부적인 선호도는 엘라스틱 검색 엔진으로 대체 가능

카테고리	키워드
가성비	구성비,가성비,가격,저렴,비쌈,비용 등
친절	친절,직원,응대,대응,사장
청결	배수구,화장실,,하수구, 청결,먼지,머리카락,얼룩,곰팡이,,물때,쓰레기,벌레,청소,깔끔,깨끗 등
주변시설	쇼핑,백화점,관광,관광지,볼거리 등
주차	주차,발레,차량
조식	조식,아침 식사,뷔페
방음	소리,방음,옆방,소음,시끄,조용
위치	위치,접근성,거리,이동,도보,도심, 지하철,지하철역,공항,버스,대중교통,교통
비품	어메니티,수건,칫솔,용품,타월,일회용품,세면도구,비품, 생수, 충전기,드라이기 등
시설	분위기,인테리어, 엘리베이터, 피트니스,책상,가구,옷장,, 침대,침구류,침실,, 에어컨,티브이,벽지, 룸컨디션 등

수정 전 카테고리별 평균 리뷰 수 : 약 14,428 개

리뷰 수 3배 증가

수정 후 카테고리별 평균 리뷰 수 : 약 34,767 개

추천모델 개선 - 만족지수 산정 방법 수정

문제 제기

만족지수 계산 시 분모에 해당하는 호텔의 해당 카테고리 언급 리뷰 수에 차이가 발생
-> 만족지수의 객관성 부족 문제 발생

문제 해결

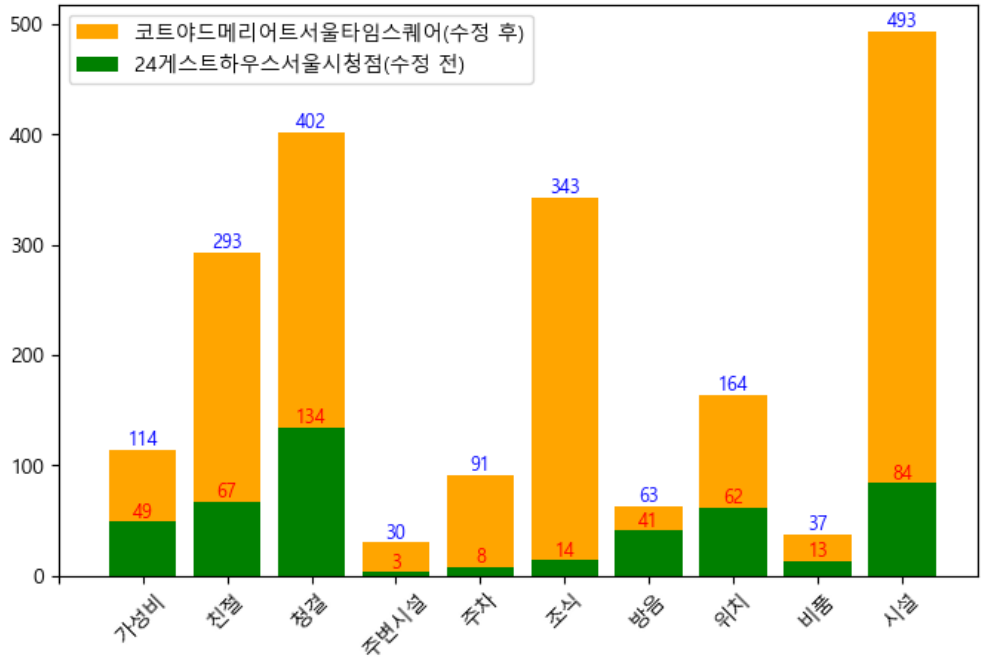
만족지수에 전체 호텔의 평균 카테고리 언급 리뷰 수 로 가중치를 줌으로써 만족지수의 객관성 확보

$$\frac{\text{카테고리 언급 리뷰 수}}{\frac{\text{카테고리별 총 언급 수}}{\text{카테고리 언급 호텔 수}}} \times \text{만족지수}$$

$$\text{긍정비율} = \frac{\text{해당 카테고리 긍정 리뷰 수}}{\text{해당 카테고리 언급 리뷰 수}} \times 100$$

$$\text{만족지수} = \frac{\text{긍정비율}}{\text{긍정비율} + \text{부정비율}} \times 100$$

동일한 조건에서 1순위 숙소의 리뷰 수 차이

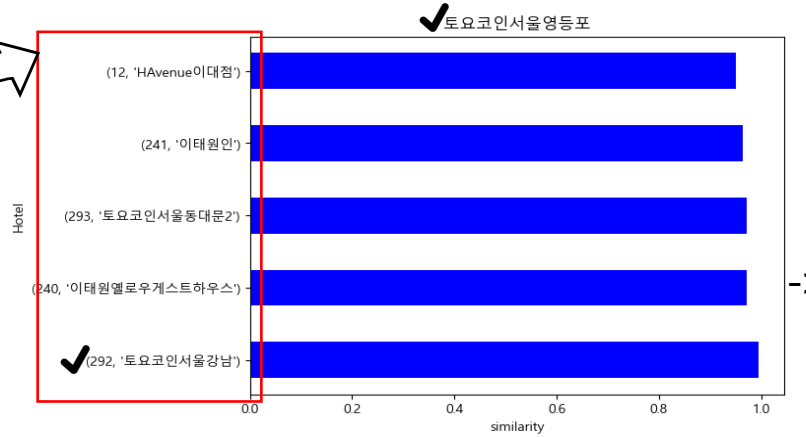


추천모델 개선-유사호텔 추천과 카테고리 가중치 적용

호텔별 카테고리 점수 테이블

ht_id	가성비	친절	청결	주변시설	주차	조식	방음	위치	비품	시설
1	66.8978	53.3862	38.9684	24.1121	3.7218	19.6568	14.2828	53.0949	9.5212	13.2588
2	79.7346	56.5584	27.5881	48.2243	0.0000	217.1516	12.1874	80.5917	7.6536	15.3840
3	1.8041	14.7865	8.4127	NaN	2.7914	16.3807	4.9268	6.7765	7.2508	8.0356
4	97.2546	85.0711	44.4523	16.0748	14.3557	2.8081	80.0375	56.5606	33.1656	32.0250
5	40.4117	61.9010	43.1160	16.0748	11.1655	12.6365	12.9954	20.2833	14.2818	16.5115
...	...	...	...	...	...	...	...	...	...	...
381	52.0734	78.9748	111.6245	8.0374	226.4048	NaN	43.6844	43.1482	67.7014	101.6982
382	15.3949	6.8906	5.9979	NaN	7.4437	NaN	3.2845	0.0000	0.0000	3.1308
384	8.4191	18.9430	14.1770	NaN	7.4437	0.0000	1.6423	1.4521	0.0000	6.1280
385	8.4191	1.6538	5.4007	NaN	NaN	NaN	3.2845	0.9681	11.3293	4.3048
386	18.9000	49.6697	39.7307	NaN	23.4476	42.7535	5.8226	55.0762	7.3850	46.5109

상관분석(Correlation Analysis)



피어슨 상관계수를 사용하여
해당 호텔과 유사한 호텔을 추천

-> 동일한 체인의 호텔끼리 상관계수 높게 나옴

가중치

상관없음	0
덜 중요	0.5
보통(기본값)	1
중요	1.5
매우중요	2

가성비	친절	청결	주변시설	주차	조식	방음	위치	비품	시설
x 2.	x 1.	x 1.	x 0.5	x 0.	x 2.	x 1.	x 1.	x 1.	x 1.5

카테고리 가중치의 기본값을 1(보통)으로 설정함으로써 사용자가 더 우수한 숙소를 추천받을 수 있도록 함

수정 내용 비교

	수정 전	수정 후
엘라스틱서치	리뷰데이터의 TF-IDF 기반 검색	- SBERT를 통한 문장벡터 유사도 기반 검색 - 필터링 조건 추가로 검색 정확도 향상
추천 알고리즘	36개의 많은 카테고리로 인해 카테고리별 충분한 리뷰 수 확보 어려움	카테고리 수를 10개로 축소함으로써 카테고리별 충분한 리뷰 수 확보 가능
추천 알고리즘	만족지수의 객관성 부족	전체 호텔의 평균 리뷰 수를 가중치로 반영하여 만족지수 객관성 향상

6. 웹 시연

웹구현 / 기능

7. 마무리

의의

- 소비자의 만족도 향상

- 평점과 리뷰가 불일치 하는 문장의 경우, 문장을 나눠 긍부정 라벨링을 진행함으로써 보다 정확한 감정모델링 결과를 얻을 수 있었다.

- 긍부정 지표들의 카테고리 로직을 통해 보여줌으로써 선택한 호텔이 업계의 평균인지, 상위인지에 대한 판단이 가능하다.

- 소비자의 효율적인 소비 촉진 및 시간 단축

- 기존보다 직관적으로 요약 정보를 나타내며 필요한 정보를 찾을 수 있게 하여 선별 용이성 제공

- 유지해야할 서비스와 개선해야할 서비스 피드백 제공

- 숙박리뷰 외에 다른 리뷰에도 적용 가능한 서비스

한계점

- 개인화추천에서 user데이터가 부족했기 때문에 2차원에서 군집화를 이룰 수 있었음(맞아?)

- 리뷰 데이터자체가 정성적 평가이기 때문에 완벽한 정보가 아니다

- 지표의 한계, 가격의 경우 날짜별 행사별 정확한 가격을 알 수 없는 상황에서 남겨진 리뷰는 정확한 지표가 아니다.

- 불균형데이터

  중립데이터인 라벨2에서의 결측치?로스값이 높음

  ko-electra Classification Report / 라벨 2에 대한 정확도가 가장 낮음

	precision	recall	f1-score	support
0	0.95	0.96	0.96	1043
1	0.96	0.96	0.96	1062
2	0.90	0.88	0.89	624
accuracy			0.94	2729
macro avg	0.94	0.94	0.94	2729
weighted avg	0.94	0.94	0.94	2729

향후계획

- 라벨2 데이터로 정보성 리뷰 추출 및 사용

- 주기적인 업데이트를 위한 스크래핑 및 전처리 자동화 시스템 구축

- 로그인 기능 추가로 개인검색기록 누적을 통해 추천 알고리즘 고도화

- 국내 및 해외 주요 지역 서비스 확대

References

감사합니다 !

우리 조 최고야