

사용자 친화적인 호텔 추천 챗봇 시스템

박정현, 이경민

숙박 서비스 시장의 성장

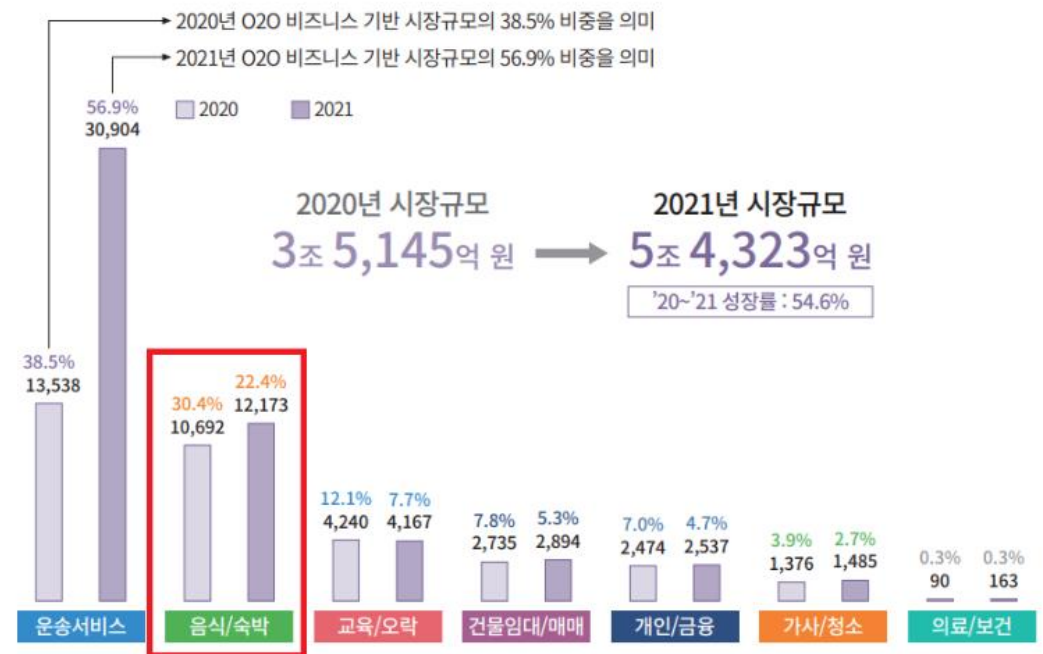
'20년~'21년 비즈니스 기반 시장규모

[단위:억 원]

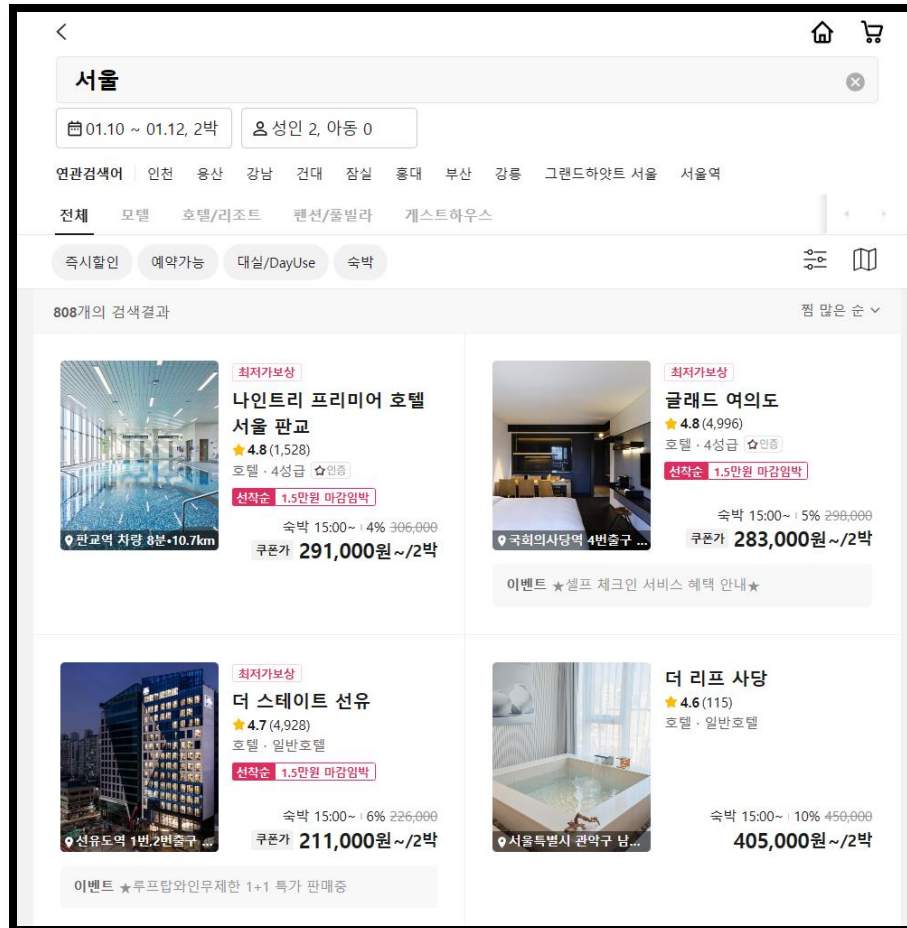
대분류		비즈니스 매출액		'20~'21년 증감률
		2020	2021	
1	운송 서비스	13,538	30,904	128.3%
2	음식점 및 숙박 서비스	10,692	12,173	13.9%
3	오락, 스포츠, 문화 및 교육 서비스	4,240	4,167	△1.7%
4	건물 임대, 중개(매매) 및 유지보수	2,735	2,894	5.8%
5	개인 미용, 금융 및 보험, 기타 서비스	2,474	2,537	2.5%
6	세탁, 청소 및 가사 서비스	1,376	1,485	7.9%
7	의료 및 보건 서비스	90	163	81.1%
합계		35,145	54,323	54.6%

'20년~'21년 비즈니스 기반 시장규모

[단위:억 원]



추천 시스템의 필요성

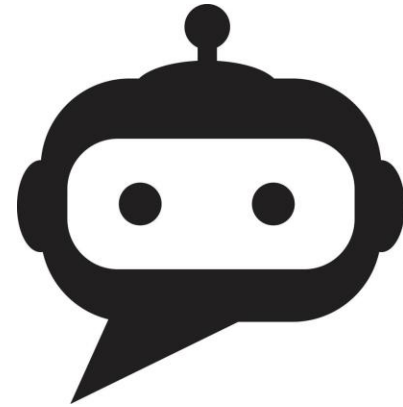


수많은 호텔 중,
나에게 딱 맞는 호텔은?

호텔 추천 챗봇 소개



"깨끗하고 친절한 호텔을 원해"



"신라스테이 구로"를 추천합니다. 이 호텔은
...

고객이 원하는, 맞춤형 호텔 추천 서비스

호텔 추천 챗봇 소개

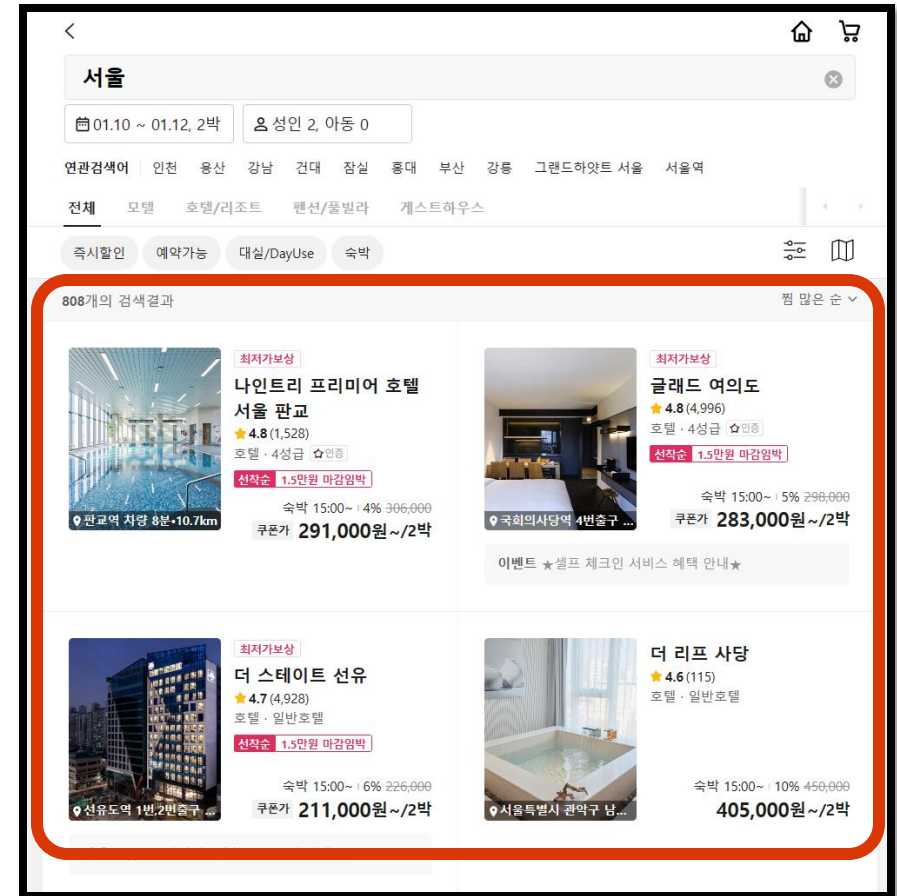


챗봇과의 대화를 통해 사용자의 취향을 파악하고,
리뷰 데이터에 근거해 가장 적절한 호텔을 추천

호텔 데이터 크롤링



유명 숙박 업체 '야놀자'



모든 호텔 정보 크롤링

리뷰 데이터 크롤링



유명 숙박 업체 '야놀자'

후기

후기 (166)

★ 4.7 / 5

최근 12개월 누적 평점

후기정책 ①

친절도 4.7
청결도 4.6
편의성 4.8
위치만족도 4.7

후기 수 및 평가 기준 안내 ②

객실 전체

최근 작성순 ▼

포토후기만 보기 ☒

숙소 선정 후기가 최근 작성순 후기 상단에 최대 3개까지 노출됩니다.

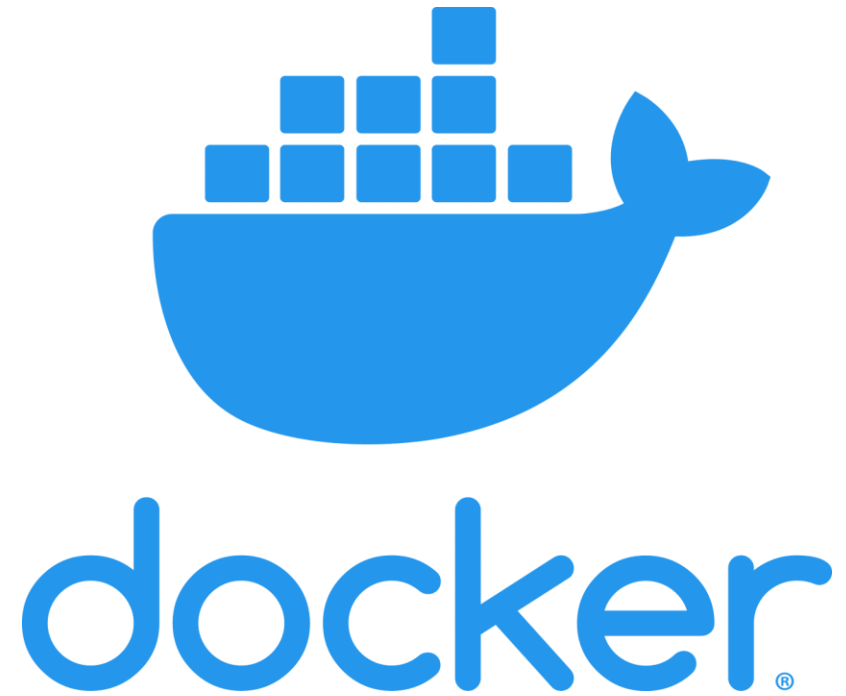
★★★★★
잔혹한인*
1A TYPE. FLAT SUITE / 주차불가
늘 믿고 오는곳입니다.
한국에서 정착하기전 꼭 오는 곳이에요
3일 전

★★★★★
경주책방0*
1D TYPE. FLAT STUDIO / 주차불가
룸업그레이드에 따뜻하고 컨디션 좋아요 잘묵고갑니다
4일 전

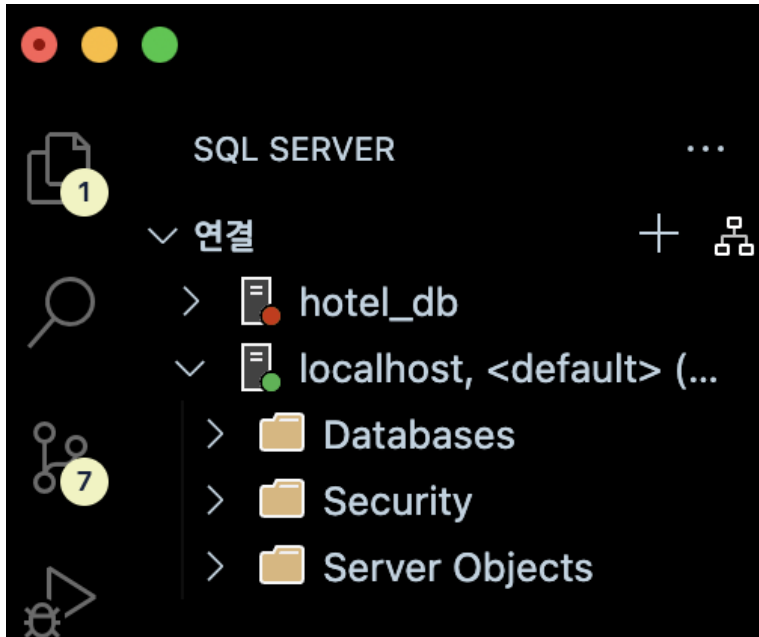
★★★★★
불꽃놀이그*****
1B TYPE. FLAT TRIPLE / 19시 체크인
항상 믿고 가는 곳
5일 전

모든 호텔 리뷰 크롤링





데이터베이스 구축



데이터베이스 구축



```
1 # 데이터베이스 연결 설정
2 server = "127.0.0.1"
3 username = "sa"
4 password = "Hotelchat44"
5
6 cnxn = pyodbc.connect(
7     "DRIVER={ODBC Driver 17 for SQL Server};SERVER="
8     + server
9     + ";UID="
10    + username
11    + ";PWD="
12    + password
13 )
14 cursor = cnxn.cursor()
```

<input checked="" type="checkbox"/>	Name	Image	Status	CPU (%)	Port(s)	Last started	Actions
<input checked="" type="checkbox"/>	sql 77d11bd675d	 mcr.microsoft.com	Running	2.7%	1433:1433	3 days ago	  

호텔 테이블

RESULTS					
	HOTEL_ID	NAME	LOCATION	RATING	REVIEW_COU...
1	1000095518	레스케이프 호텔	서울특별시 중구 ...	4.7	1081
2	1000099457	신촌 메이저 호텔	서울특별시 마포...	4.5	23
3	1000099624	노원 리츠호텔	서울특별시 노원...	4.1	158
4	1000099647	포포인츠 바이 쉐...	서울특별시 구로...	4.7	3324
5	1000100245	호텔 그레이스리 ...	서울특별시 중구 ...	4.8	4223
6	1000100307	트레블로지 명동 ...	서울시 중구 수표...	4.2	1183
7	1000100675	그랜드 인터컨티...	서울특별시 강남...	4.8	690
8	1000101062	신천(잠실새내) ...	서울특별시 송파...	4.4	303
9	1000101565	라마다 호텔앤스...	서울특별시 중구 ...	4.5	1665
10	1000101862	오라카이 대학로 ...	서울특별시 종로...	4.8	3044
11	1000102261	나인트리 프리미...	서울특별시 종로...	4.7	2714
12	1000102483	안다즈 서울 강남	서울시 강남구 논...	4.9	196
13	1000102490	반얀트리 클럽 앤 ...	서울특별시 중구 ...	5	18
14	1000102538	트레블로지 명동 ...	서울특별시 중구 ...	3.8	109

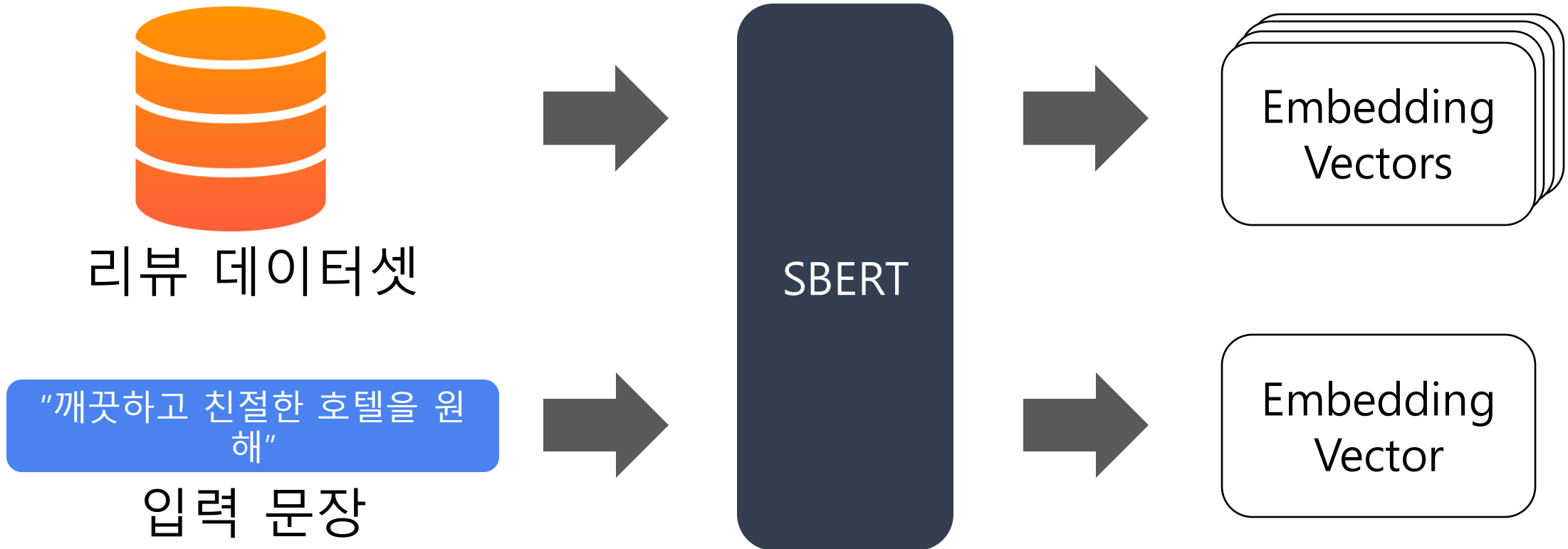
리뷰 테이블

RESULTS				
	HOTEL_ID	REVIEW_ID	RATING	COMMENT
1	1000099457	3506	5	교통이나 위치는 ...
2	1000099457	3507	5	잘쉬다갑니다
3	1000099457	3508	3	그냥그런 호텔입...
4	1000099457	3509	5	조앗다네요옹
5	1000099457	3510	5	저렴하고 좋아요
6	1000099457	3511	3	침구류 청결은 별...
7	1000099457	3512	5	서비스가 너무 좋...
8	1000099457	3513	5	깨끗하고 넓고 따...
9	1000099457	3514	5	특실 이용했는데 ...
10	1000099457	3515	5	깔끔하네요
11	1000099457	3516	5	좋습니다 !!
12	1000099457	3517	5	너무 편하고 좋았...
13	1000099457	3518	4	좋아용...
14	1000099457	3519	5	좋아요 맘에 들어요

KoBERT를 통한 감성 분석



SBERT를 통한 문장 임베딩



입력 문장과 리뷰 데이터 간의 유사도



$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

추천 스코어

$$\text{recommand hotel} = \arg \max \frac{1}{N} \times \sum_{i=1}^N \text{similarity}(\text{emb}(S_{input}), \text{emb}(S_i)) \times \text{sentiment}_i$$

추천 스코어

$$\text{recommand hotel} = \arg \max \frac{1}{N} \times \sum_{i=1}^N \text{similarity}(\text{emb}(S_{input}), \text{emb}(S_i)) \times \text{sentiment}_i$$

추천 스코어

$$\text{recommand hotel} = \arg \max \frac{1}{N} \times \sum_{i=1}^N \text{similarity}(\text{emb}(S_{input}), \text{emb}(S_i)) \times \text{sentiment}_i$$

추천 스코어

$$\text{recommand hotel} = \arg \max \frac{1}{N} \times \sum_{i=1}^N \text{similarity}(\text{emb}(S_{input}), \text{emb}(S_i)) \times \text{sentiment}_i$$

추천 스코어

$$\text{recommand hotel} = \arg \max \frac{1}{N} \times \sum_{i=1}^N \text{similarity}(\text{emb}(S_{input}), \text{emb}(S_i)) \times \text{sentiment}_i$$

추천 스코어

$$\text{recommand hotel} = \arg \max \frac{1}{N} \times \sum_{i=1}^N \text{similarity}(\text{emb}(S_{input}), \text{emb}(S_i)) \times \text{sentiment}_i$$

두 호텔 간 스코어 비교



3성급 ☆인증

호텔 그레이스리 서울

📍 시청역 8번출구 도보 5분 >

★4.8 (4,214) 숙소답변(3,717) >

★★★★★ 숙소 선정

풍덩캔*****

내 맘대로 9시간스테이 - 스탠다드 더블+주차불가

항상 깨끗하고 친절합니다.

깔끔하고 청결해 여자친구가 가장 좋아하는 곳 입니다. 자주 오는 곳이지만 여기는 관리자 분이 청결부분을 많이 신경 쓰신것 같아 항상 안심하고 편안하게 지내다 갑니다....

호텔 A



1성급 ☆인증

호텔 켈리년 건대 1호점

📍 어린이대공원역 5번출구 도보10분 >

★2.7 (11) 숙소답변(10) >

★☆☆☆☆

안압지오리***

디렉스 더블

벽지엔 온통 먼지가 몽글몽글

화장실은 곰팡이가 여기 저기

조명이 너무 어두워 안을 잘 살피지 못하고 아침에 화장실 창문열고 기겁하는줄

호텔 B

두 호텔 간 스코어 비교

	호텔 A	호텔 B
리뷰와 입력 문장 간의 유사도의 평균	0.3746	0.1020
리뷰 감성 분석 결과의 평균	0.9058	0.5189
추천 스코어	0.3394	0.0529

터미널

python3 - chatbot + ▢

```
Analyzing selected hotels: 0%|          | 0/2 [00:00<?, ?it/s]
호텔 ID: 1000100245, 평균 유사도: 0.3746998906135559, 평균 감성 점수: 0.90580909
50229338, 추천 스코어: 0.3394065688218574
Analyzing selected hotels: 50%|          | 1/2 [01:32<01:32, 92.62s/it]
호텔 ID: 10048715, 평균 유사도: 0.10209819674491882, 평균 감성 점수: 0.518986830
6571787, 추천 스코어: 0.0529876195444585
Analyzing selected hotels: 100%|          | 2/2 [01:35<00:00, 47.76s/it]

추천 호텔 ID: 1000100245, 이름: 호텔 그레이스리 서울
127.0.0.1 - - [14/Dec/2023 17:01:30] "POST /chat HTTP/1.1" 200 -
```

추천 스코어 비교

깨끗하고 친절하다는 평이 많
던
호텔 A의 스코어

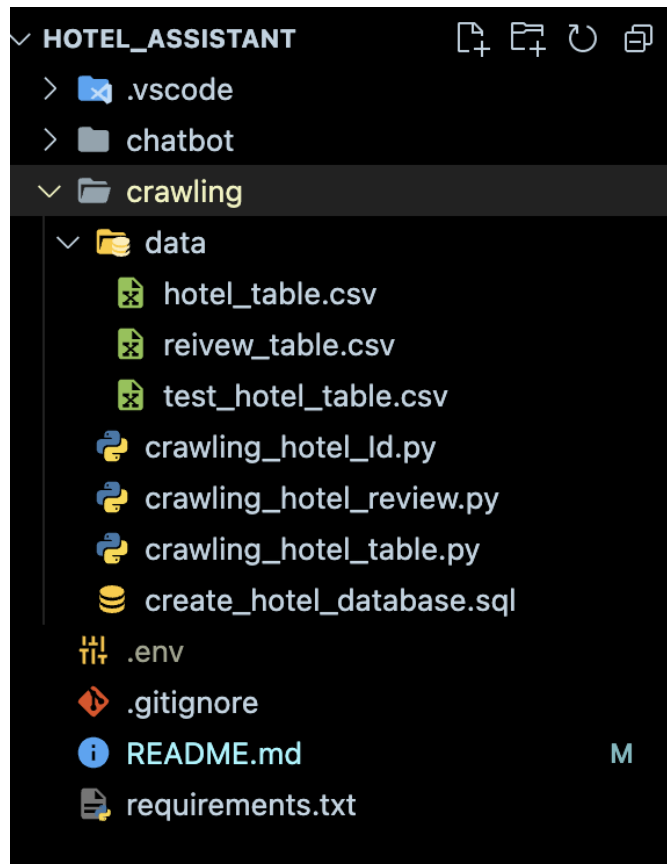


더럽고 냄새 난다는 평이 많던
호텔 B의 스코어

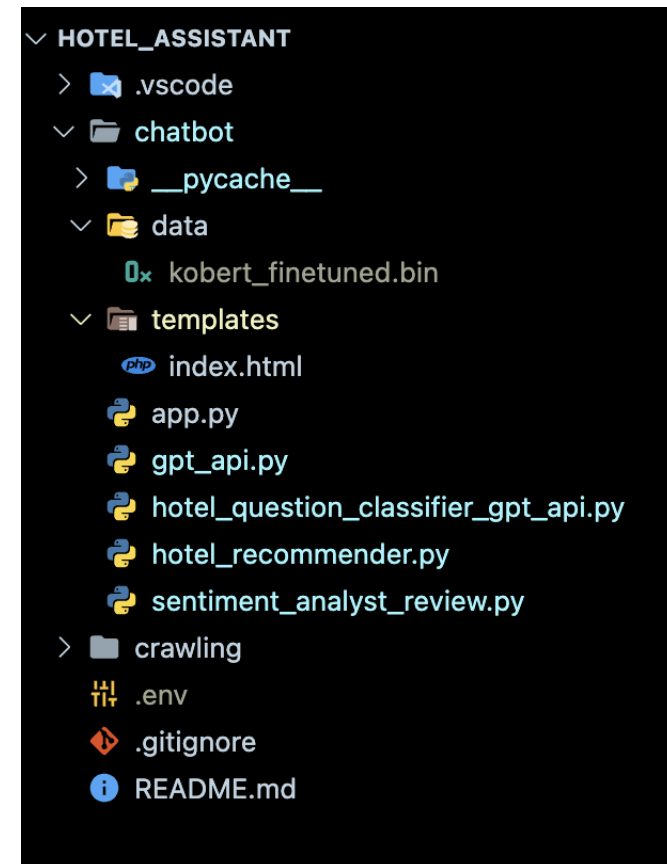
호텔 A가 B에 비해 호텔 스코어가 약 6.4배 높은 것을 확인

전체 디렉토리 구조

크롤링



챗봇



프로젝트 전체 구동 과정

- Crawling
 - crawling_hotel_id.py: 호텔 고유 ID 값 추출 후 HOTEL 테이블 생성
 - crawling_hotel_table.py: 호텔 정보 추출 후 HOTEL 테이블에 저장
 - crawling_hotel_review.py: 호텔의 리뷰 데이터 추출 후 REVIEW 테이블에 저장
- Chatbot
 - app.py + index.html: Flask 웹 앱 구현
 - Hotel_question_classifier_gpt_api.py: 사용자가 입력한 질문이 호텔에 관련된 질문 인지 GPT4 API로 구분하고, 결과값을 1,0으로 리턴
 - sentiment_analyst_review.py: 호텔 리뷰 데이터 셋을 통해 감성 분석 학습하여 모델을 저장한다.
 - hotel_recommender.py: 사용자가 입력한 질문을 임베딩하여 호텔별로 유사도와 감성점수 결합하여 추천점수를 도출하고, 가장 점수가 높은 호텔을 리턴
- Chat UI Flow
 - Flask run -> index.html -> 127.0.0.1:5000 -> 사용자 질문 입력 -> hotel_question_classifier_gpt_api.py(def is_hotel_related) -> 0 : GPT4 API 로 답변, 1: hotel_recommender.py (def recommend_hotel -> def get_sentiment_score) 추천점수 계산 후 호텔 이름을 리턴

챗봇 예시

호텔 추천 챗봇

나: 안녕

챗봇: 안녕하세요, 저는 호텔 추천 챗봇입니다. 원하시는 호텔을 말씀해 주세요.

나: 나는 깨끗하고 친절한 호텔을 원해.

챗봇: 네, 호텔 그레이스리 서울 호텔을 추천드립니다. 해당 호텔은 별점4.8점으로, 깨끗한 객실과 친절한 직원으로 유명합니다.

질문을 입력하세요...

보내기

업데이트 필요 사항

- 현재 사용자가 챗봇에게 호텔을 추천을 요청하면 hotel_recommender.py에서 추천 시스템이 구동되는데, 완전한 답변을 받기까지 시간이 많이 소요된다.
 - 임베딩을 비롯해서 점수 계산 로직을 개선해서 속도 향상이 필요하다.
- 챗봇 UI를 개선하여 사용자에게 좀 더 편리한 환경을 구성할 수 있다.
- 호텔 리뷰페이지에서 업데이트 되는 값들을 추가적으로 크롤링하여 데이터 테이블 신규화가 필요하다.
- 호텔 리뷰데이터를 감성 평가 훈련데이터로 사용할 때 리뷰 점수를 가지고 긍정과 부정 레이블로 나누었는데, 해당 부분에 로직을 개선해서 정확도를 높일 수 있다.
- 위치, 가격 등 시멘틱한 정보를 룰베이스 기반으로 처리하는 로직을 추가하여, 전반적인 모델의 성능을 높일 수 있다.
- 사용자에게 위치 데이터를 사용하여 호텔 위치 이동시간을 사용하여, 위치 기반 추천 시스템을 구현할 수 있다.