

데이터 수집 및 저장 데이터 베이스 설계서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 데이터베이스 설계서
- 제출 일자 : 2025. 08. 22
- 깃허브 경로 : <https://github.com/skn-ai14-250409/SKN14-Final-2Team>
- 작성 팀원 :

개요	<ul style="list-style-type: none">• 소개• 시스템 개요
구조	<ul style="list-style-type: none">• 시스템 아키텍처• 요구사항 매트릭스
설계	<ul style="list-style-type: none">• 데이터 설계• 설계 근거• ERD

<p>소개</p>	<ul style="list-style-type: none"> - 목적 <ul style="list-style-type: none"> - 이 데이터베이스는 대화형 향수 추천 챗봇(ScentPick)이 수집한 향수/브랜드//메인 어코드/용량/가격 정보를 일관되게 저장하고, 사용자 대화 이력과 선호(즐거찾기·피드백)를 바탕으로 퍼스널라이즈된 추천을 제공하도록 설계되었습니다. - 범위 <ul style="list-style-type: none"> - 사용자 계정/세션/대화 이력 관리 <ul style="list-style-type: none"> ● 향수 메타데이터(브랜드, 라인, 부향률 등) ● 노트/메인 어코드 등 매핑(탑/미들/베이스) ● 외부 소스 수집(Fragrantica, 리테일러 가격/이미지 등) 및 출처 추적 ● 벡터 임베딩(향 설명, 메인 어코드 등)과 RAG 검색 연동 ● 사용자 즐거찾기, 리뷰
<p>시스템 개요</p>	<ul style="list-style-type: none"> - 시스템 역할 데이터 허브: 스크래핑으로 수집한 향수·노트·가격·리뷰·이미지 원천을 정규화해 저장 추천 근거 저장: 대화 중 사용된 후보/점수/근거 문장(소스 링크) 보존 → 재현 가능성 확보 개인화: 사용자 즐거찾기/리뷰/대화 로그 기반 개인화 - 주요 기능 <ul style="list-style-type: none"> ● 사용자 인증 및 로그인 시간 기록 ● 향수/노트/어코드/브랜드/가격/이미지 등 도메인 엔터티 관리 ● 세션별 대화 메시지 저장 ● RAG 검색을 위한 임베딩/소스 문서 관리

시스템 아키텍처	<div>- 데이터베이스 구조 관계형 DB (MySQL): 기준 마스터(향수/브랜드/어코드), 관계 테이블, 사용자/세션/대화/즐거찾기/가격</div> <div>벡터 저장소 (Pinecone : 향 설명·리뷰 요약 임베딩, 인덱스/버전 관리 RDB에는 벡터 ID/버전만 참조, 원문 텍스트는 규격화 테이블과 함께 관리</div> <div>- 적용한 데이터베이스</div> <div>- AWS의 RDS를 기반으로 MySQL DB를 적용함.</div> <div>- perfumes의 필요 칼럼들 / 키워드 사전을 PineconeDB에 벡터화 하여 저장함.</div> <div>- 전체 문서는 MySQL에 저장</div>																																							
요구사항 매트릭스	<div>요구사항 및 관련 테이블</div> <table><tr><td>요구사항</td><td>관련 테이블</td></tr><tr><td>사용자 로그인 정보 저장</td><td>users, oauth_accounts</td></tr><tr><td>사용자별 대화내역 기록</td><td>conversations, messages</td></tr><tr><td>향수/브랜드/노트/어코드 관리</td><td>perfumes, note_image</td></tr><tr><td>별점 / 즐거찾기 리스트 관리</td><td>ratings , favorites</td></tr></table>				요구사항	관련 테이블	사용자 로그인 정보 저장	users, oauth_accounts	사용자별 대화내역 기록	conversations, messages	향수/브랜드/노트/어코드 관리	perfumes, note_image	별점 / 즐거찾기 리스트 관리	ratings , favorites																										
요구사항	관련 테이블																																							
사용자 로그인 정보 저장	users, oauth_accounts																																							
사용자별 대화내역 기록	conversations, messages																																							
향수/브랜드/노트/어코드 관리	perfumes, note_image																																							
별점 / 즐거찾기 리스트 관리	ratings , favorites																																							
데이터 설계	<div>- 테이블 설명</div> <div>- users</div> <div>- 서비스 회원 기본 정보를 저장</div> <table><tr><th>컬럼명</th><th>데이터 타입</th><th>설명</th><th>NN (not null)</th></tr><tr><td>id</td><td>BIGINT</td><td>사용자 ID(PK)</td><td>True</td></tr><tr><td>email</td><td>VARCHAR(120)</td><td>사용자 이메일(Unique)</td><td>True</td></tr><tr><td>password_hash</td><td>VARCHAR(255)</td><td>비밀번호 해시값</td><td>True</td></tr><tr><td>name</td><td>VARCHAR(60)</td><td>이름</td><td>True</td></tr><tr><td>gender</td><td>VARCHAR(10)</td><td>성별</td><td>True</td></tr><tr><td>birth_year</td><td>INT</td><td>출생년도</td><td>True</td></tr><tr><td>created_at</td><td>DATETIME</td><td>생성 시각</td><td>True</td></tr><tr><td>updated_at</td><td>DATETIME</td><td>수정 시각</td><td>True</td></tr></table>				컬럼명	데이터 타입	설명	NN (not null)	id	BIGINT	사용자 ID(PK)	True	email	VARCHAR(120)	사용자 이메일(Unique)	True	password_hash	VARCHAR(255)	비밀번호 해시값	True	name	VARCHAR(60)	이름	True	gender	VARCHAR(10)	성별	True	birth_year	INT	출생년도	True	created_at	DATETIME	생성 시각	True	updated_at	DATETIME	수정 시각	True
컬럼명	데이터 타입	설명	NN (not null)																																					
id	BIGINT	사용자 ID(PK)	True																																					
email	VARCHAR(120)	사용자 이메일(Unique)	True																																					
password_hash	VARCHAR(255)	비밀번호 해시값	True																																					
name	VARCHAR(60)	이름	True																																					
gender	VARCHAR(10)	성별	True																																					
birth_year	INT	출생년도	True																																					
created_at	DATETIME	생성 시각	True																																					
updated_at	DATETIME	수정 시각	True																																					

- oauth_accounts

- 사용자의 소셜로그인 정보 저장

컬럼명	데이터 타입	설명	NN (not null)
id	BIGINT	OAuth 계정 ID (PK)	True
user_id	BIGINT	FK → users.id	True
provider	VARCHAR (40)	공급자명 (google/naver/kakao)	True
provider_user_id	VARCHAR (120)	공급자 시스템 내 사용자 ID	True

- conversations

- 사용자의 대화목록 정보를 저장

컬럼명	데이터 타입	설명	NN(not null)
id	BIGINT	대화 ID (PK)	True
user_id	BIGINT	FK → users.id	True
title	VARCHAR(40)	사용자 대화 제목	True
started_at	DATETIME	대화 시작 시각	True
updated_at	DATETIME	마지막 업데이트 시각	True

- messages

- 사용자의 대화내역 정보를 저장

컬럼명	데이터 타입	설명	NN (not null)
id	BIGINT	메세지 ID (PK)	True
conversation_id	BIGINT	FK → conversations.id	True
role	VARCHAR(10)	화자 구분 (user/assistant/tool)	True
content	TEXT	메시지 본문	True
created_at	DATETIME	생성 시각	True

- favorites

- 사용자의 즐겨찾기 내역 저장

컬럼명	데이터 타입	설명	NN(not null)
user_id	BIGINT	FK → users.id	True
perfume_id	BIGINT	FK → perfumes.id	True
created_at	DATETIME	즐거찾기 등록 시각	True

- ratings

- 사용자의 별점 저장

컬럼명	데이터 타입	설명	NN(not null)
id	BIGINT	평점 ID (PK))	True
user_id	BIGINT	FK → users.id	True
perfume_id	BIGINT	FK → perfumes.id	True
score	INT	평점 값	True
created_at	DATETIME	생성 시각	True
updated_at	DATETIME	수정 시각	True

- note_images

- note별 이미지,

컬럼명	데이터 타입	설명	NN(not null)
id	BIGINT	노트 이미지 ID (PK)	True
category	VARCHAR(50)	노트가 속한 카테고리	True
note_name	VARCHAR(50)	노트 이름	True
image_url	VARCHAR(500)	노트의 이미지 url	True

- perfumes

- 향수와 관련된 정보들을 저장.

컬럼명	데이터 타입	설명	NN(not null)
id	BIGINT	향수 ID (PK)	True
brand	VARCHAR(50)	브랜드명	True
name	VARCHAR(50)	향수명	True
eng_name	VARCHAR(100)	향수명 영어	True
size_ml	DECIMAL(6,1)	용량 (ml)	True
price_krw	INT	가격(KRW)	True
detail_url	VARCHAR(500)	상세 페이지 URL	True
description	TEXT	향수 설명	True
concentration	VARCHAR(30)	부향률(EDP/EDT 등)	True
main_accords	VARCHAR(300)	메인 어코드(띄어쓰기 구분)	True
top_notes	VARCHAR(300)	탑 노트 (띄어쓰기 구분)	True
middle_notes	VARCHAR(300)	미들 노트 (띄어쓰기 구분)	True
base_notes	VARCHAR(300)	베이스 노트 (띄어쓰기 구분)	True
notes_score	JSON	향수에 해당되는 노트별 점수	True
season_score	JSON	향수의 계절별 연관도	False
day_night_score	JSON	향수의 주/야 연관도	False
created_at	DATETIME	생성 시각	True
updated_at	DATETIME	수정 시각	True

설계 근거	<ul style="list-style-type: none"> - 데이터 무결성 보장 - 외래 키(FK)를 활용하여 테이블 간 관계를 명확히 정의함. - 확장성 고려 <ul style="list-style-type: none"> - 대화·추천 로그 분리로 장기 보관/분석/ 확장 가능. - 가격 정보 api로 가져옴으로써 가격 변동 및 상품 변화에 유연하게 대처 - 보안 <ul style="list-style-type: none"> - 사용자 민감정보는 최소 수집·암호화, 소셜 로그인 지원.
ERD	<pre> erDiagram images --o{ users : "has" users --o{ oauth_accounts : "has" users --o{ conversations : "has" users --o{ favorites : "has" users --o{ ratings : "has" users --o{ perfumes : "has" conversations --o{ messages : "has" favorites --o{ ratings : "has" images { bigint id PK text note_name text image_url text note_imgs } users { bigint id PK text email text password_hash text name text gender int birth_year timestamp created_at timestamp updated_at } oauth_accounts { bigint id PK bigint user_id FK text provider text provider_user_id } conversations { bigint id PK bigint user_id FK text title timestamp started_at timestamp updated_at } favorites { bigint id PK bigint user_id FK bigint perfume_id FK timestamp created_at } ratings { bigint id PK bigint user_id FK bigint perfume_id FK int score timestamp created_at timestamp updated_at } perfumes { bigint id PK text brand text name text eng_name float size_ml int price_krw text detail_url text concentration text main_accords text top_notes text middle_notes text base_notes text description json notes_score json seasons_score json day_night_score timestamp created_at timestamp updated_at } messages { bigint id PK bigint conversation_id FK text role text content text model timestamp created_at } </pre>