

데이터 수집 및 저장 데이터베이스 설계 문서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 데이터베이스 설계 문서
- 제출 일자 : 2025.02.13
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN06-FINAL-2Team.git>
- 작성 팀원 : 정유진

□ 사용 데이터베이스

- 로컬 데이터베이스: MySQL
- Vector Store: ChromaDB
- 원격 데이터베이스: AWS RDS(배포시)

□ 목차

- Vector Store
- 로컬 데이터베이스

1. Vector Store

1) 역할

- 웹툰/웹소설 텍스트 데이터를 임베딩 벡터로 변환하여 빠른 유사도 검색을 수행.

2) 데이터 저장 방식

- **Raw 데이터 파일(total.json)**을 읽어 각 콘텐츠를 Document 객체로 변환.
- **page_content**: 주요 메타데이터를 문자열로 결합.
- **metadata**: type, platform, title, genre, author, age_rating, price_type.

2. My SQL

1) USERS

```
CREATE TABLE USERS (  
  USER_ID INT AUTO_INCREMENT NOT NULL PRIMARY KEY,  
  EMAIL VARCHAR(255) NOT NULL UNIQUE,  
  PASSWORD VARCHAR(128) NOT NULL,  
  NAME VARCHAR(20) NOT NULL,  
  USERNAME VARCHAR(20),  
  BIRTHDAY DATE NOT NULL,  
  GENDER ENUM("Male", "Female") NOT NULL,  
  JOINED_DATE DATETIME DEFAULT CURRENT_TIMESTAMP NOT NULL  
);
```

속성	데이터 타입	설명
USER_ID	INT	사용자 고유 ID, 기본 키, 자동 증가
EMAIL	VARCHAR(255)	사용자 이메일 (유니크), 로그인 시 ID
PASSWORD	VARCHAR(128)	사용자 비밀번호
NAME	VARCHAR(20)	사용자 이름
USERNAME	VARCHAR(20)	사용자 닉네임
BIRTHDAY	DATE	사용자 생년월일
GENDER	ENUM("Male", "Female")	성별
JOINED_DATE	DATETIME	가입 날짜 (기본값: 현재 시각)

2) CONTENTS

```
CREATE TABLE CONTENTS(  
  CONTENT_ID BIGINT NOT NULL PRIMARY KEY,  
  CONTENT_TYPE VARCHAR(20) NOT NULL,  
  PLATFORM VARCHAR(50) NOT NULL,  
  TITLE VARCHAR(255) NOT NULL,  
  STATUS VARCHAR(50) NOT NULL,  
  GENRE VARCHAR(50) NOT NULL,  
  VIEWS INT,  
  RATING FLOAT,  
  LIKES INT,  
  KEYWORDS TEXT NOT NULL,  
  AUTHOR VARCHAR(100) NOT NULL,  
  ILLUSTRATOR VARCHAR(100),  
  ORIGINAL VARCHAR(100),  
  AGE_RATING VARCHAR(20) NOT NULL,  
  PRICE_TYPE VARCHAR(50) NOT NULL,  
  PRICE VARCHAR(20) NOT NULL,  
  EPISODE INT NOT NULL,  
  THUMBNAIL TEXT NOT NULL,  
  URL TEXT NOT NULL  
);
```

속성	데이터 타입	설명
CONTENT_ID	BIGINT	작품 고유 ID, 기본 키
TYPE	VARCHAR(20)	작품 종류 (예: 웹소설, 웹툰)
PLATFORM	VARCHAR(50)	작품 제공 플랫폼 (예: 카카오페이지, 네이버 웹툰 등)
TITLE	VARCHAR(255)	작품 제목
GENRE	VARCHAR(50)	작품 장르 (예: 판타지, 로맨스 등)
THUMBNAIL	TEXT	작품 썸네일 이미지 URL
URL	TEXT	작품의 상세 페이지 URL
AGE_RATING	VARCHAR(20)	작품의 연령 제한 (예: 전체이용가, 19금 등)

3) BASIC_TASTE

```
CREATE TABLE BASIC_TASTES (
  BASIC_ID INT NOT NULL PRIMARY KEY,
  BASIC_TYPE VARCHAR(20) NOT NULL,
  PLATFORM VARCHAR(50) NOT NULL,
  TITLE VARCHAR(255) NOT NULL,
  GENRE VARCHAR(50) NOT NULL,
  KEYWORDS TEXT NOT NULL,
  THUMBNAIL TEXT NOT NULL
);
```

속성	데이터 타입	설명
BASIC_ID	INT	기본 취향 고유 ID, 기본 키
BASIC_TYPE	VARCHAR(20)	기본 취향 종류 (예: 장르, 플랫폼 등)
PLATFORM	VARCHAR(50)	플랫폼 정보
TITLE	VARCHAR(255)	작품 제목
GENRE	VARCHAR(50)	작품 장르
KEYWORDS	TEXT	작품 키워드
THUMBNAIL	TEXT	작품 썸네일 URL

4) CONTENTS_USERS_MAPPING

```
CREATE TABLE CONTENTS_USERS_MAPPING (
  USER_ID INT NOT NULL,
  CONTENT_ID INT NOT NULL,
  FEEDBACK TINYINT,
  DELETED_DATE DATETIME
);
```

속성	데이터 타입	설명
USER_ID	INT	사용자 고유 ID (외래 키)
CONTENT_ID	INT	작품 고유 ID (외래 키)
FEEDBACK	TINYINT	사용자 피드백 (1: 좋아요, 0: 싫어요)
DELETED_DATE	DATETIME	피드백 삭제 날짜

5) BASIC_TASTE_USERS_MAPPING

```
CREATE TABLE BASIC_TASTES_USERS_MAPPING (
  USER_ID INT NOT NULL,
  BASIC_ID INT NOT NULL,
  FOREIGN KEY (USER_ID) REFERENCES USERS(USER_ID),
  FOREIGN KEY (BASIC_ID) REFERENCES BASIC_TASTES(BASIC_ID)
);
```

속성	데이터 타입	설명
USER_ID	INT	사용자 고유 ID (외래 키)
BASIC_ID	INT	기본 취향 고유 ID (외래 키)

6) ERD

