2022-2 데이터베이스(CSI3105) 프로젝트 최종 보고서

Members

임예원(2019145010) 김시열(2019145121)

김명준(2018126008) Kishen Ravi(2020127560)

이재욱(2019145122)

Divisions

Group A] Kishen Ravi 🡪 테이블 속성 도메인 스캔

이재욱 🡪 테이블 속성 편집

김명준 🡪 DB 연결 확인, 데이터 추가, 결과 조회

Group B] 임예원 🡪 결합 대상 테이블 검색, 다중결합, DB 연결 확인, 웹 클라이언트 디자인

김시열 🡪 단일결합

Differences from the first-report

1차 보고서의 경우 예시 데이터가 주어지지 않아 임의로 데이터와 레코드를 만들어 그에 대한 ERD와 Relational Schema, Sequence diagram 등을 작성하였고, 임의의 레코드들로 구성된 데이터의 CSV 파일을 스캔하는 기능을 구현하였다.

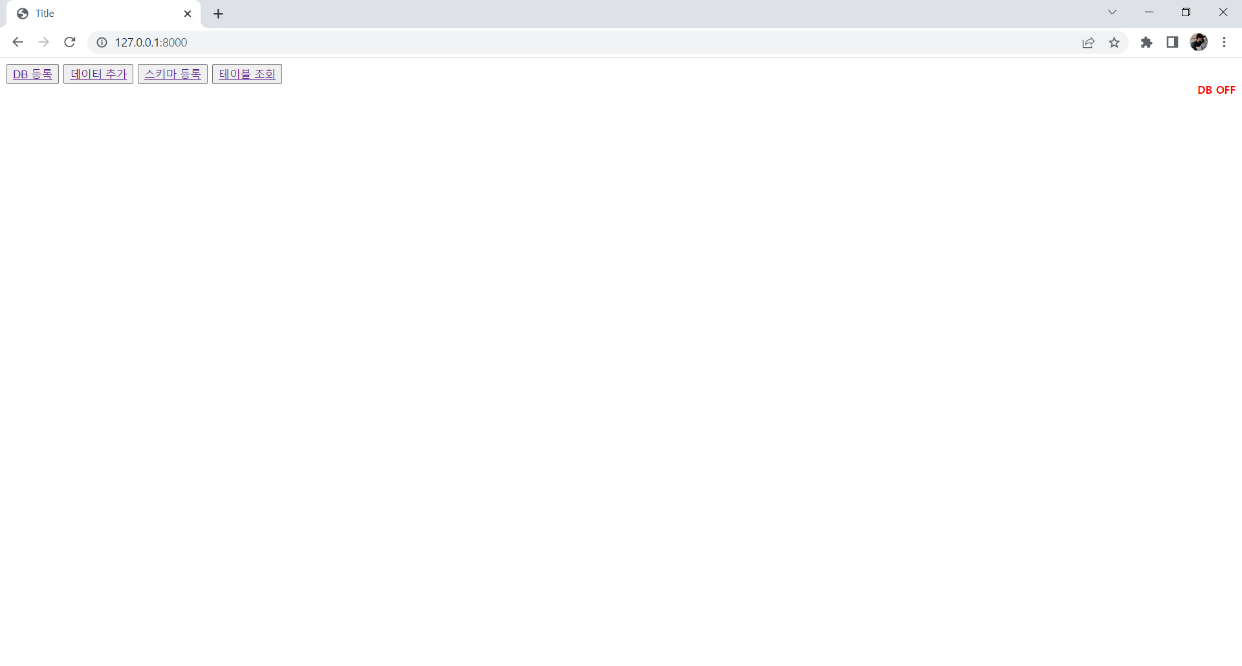
그러나, 최종 시스템의 경우에는 1차 보고서와 달리 예시 데이터 및 UI가 주어졌다. 주어진 예시 데이터를 바탕으로 CSV 파일 업로드 기능을 다시 설계하였다. 그리고, 도메인스캔, 속성편집, 결합, 결과 조회 등 프로젝트에 요구되는 기능을 바탕으로 최종 시스템을 구현하였다. 또한, 주어진 예시 UI를 변형하여 최종 시스템의 UI에 적용하였다.

1차 보고서와 동일하게 언어와 프레임워크는 python과 Django를 사용하였다.

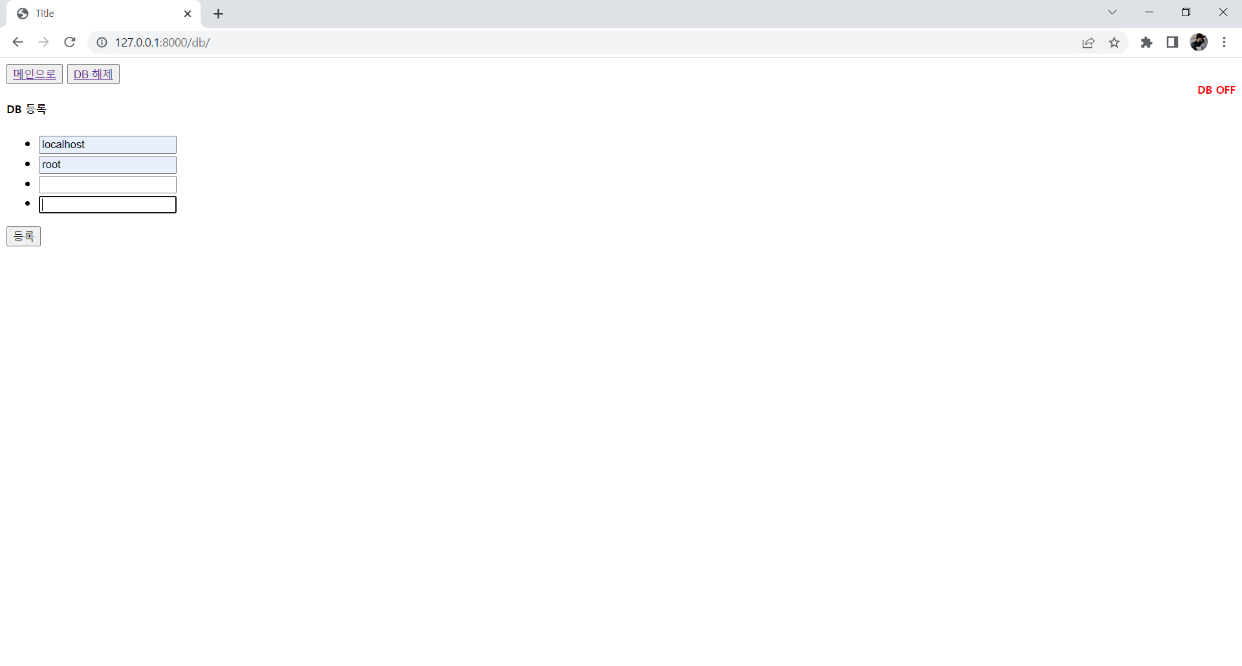
Process for the system

프로젝트 초기의 시스템은 다음과 같다.

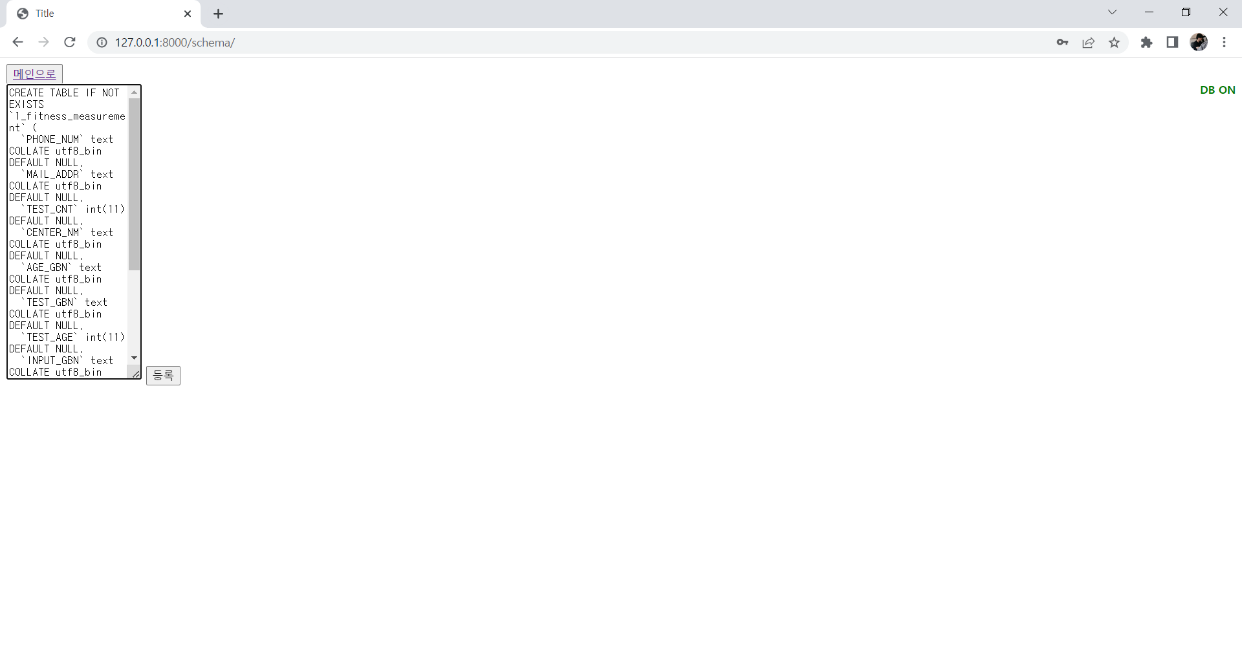
먼저, DB 연결 및 연결 여부 확인 기능을 로컬 서버 상에서 구현하였다.



그 후, DB를 등록하는 기능을 구현하였다.



DB 등록이 완료되었을 경우, 데이터 추가, 스키마 등록, 테이블 조회 등의 기능을 추가하였다.



그 이후 프로젝트에서 요구하는 기능인 속성도메인 스캔 기능, 속성 편집 기능, 단일 결합, 다중 결합 등의 기능을 구현하였다.

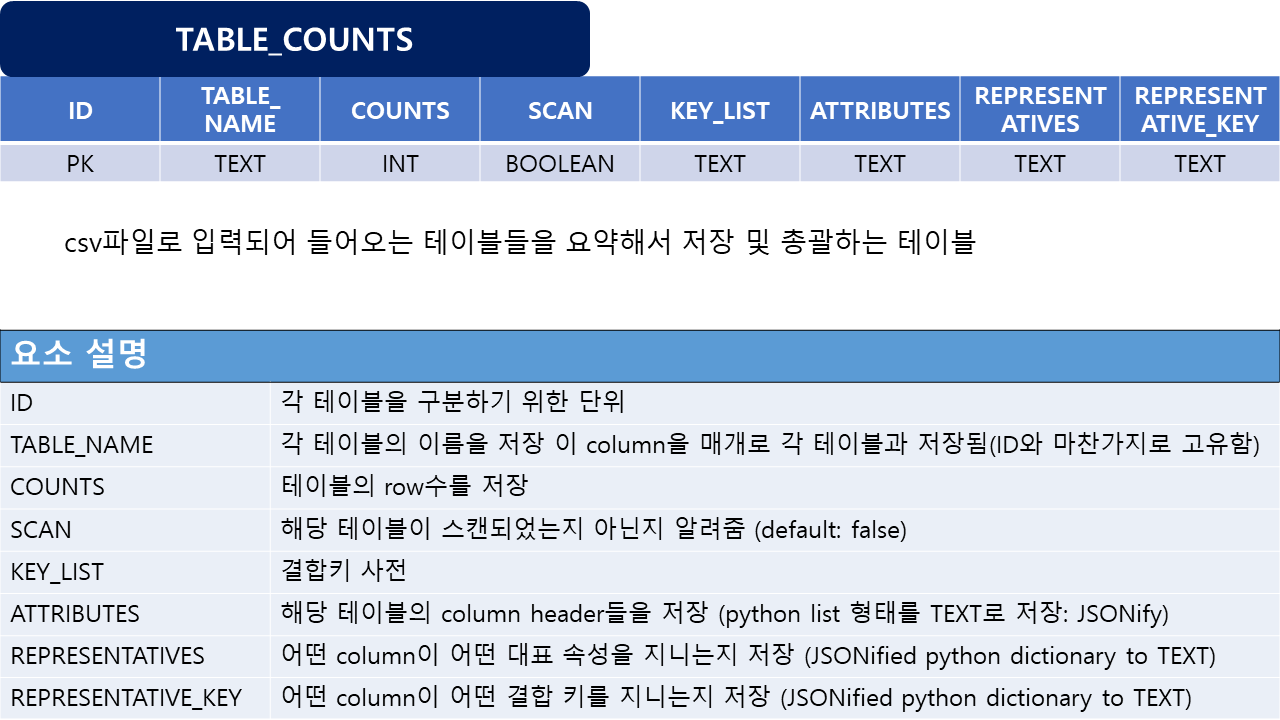
단일 결합

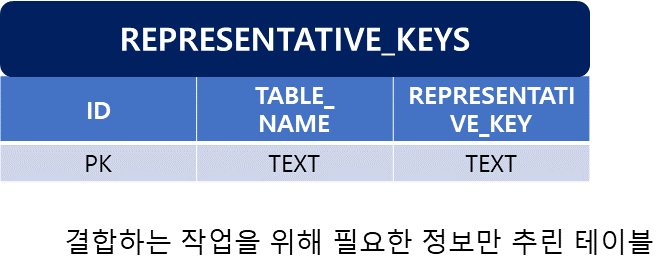
결합키 속성이나 대표속성, 그리고 이름으로 테이블을 검색하여 조건에 맞는 테이블들을 띄우고 그 테이블들 중에서 base table이 될 테이블과 join table을 선택하여 inner join을 실시하고 결과를 보여준다.

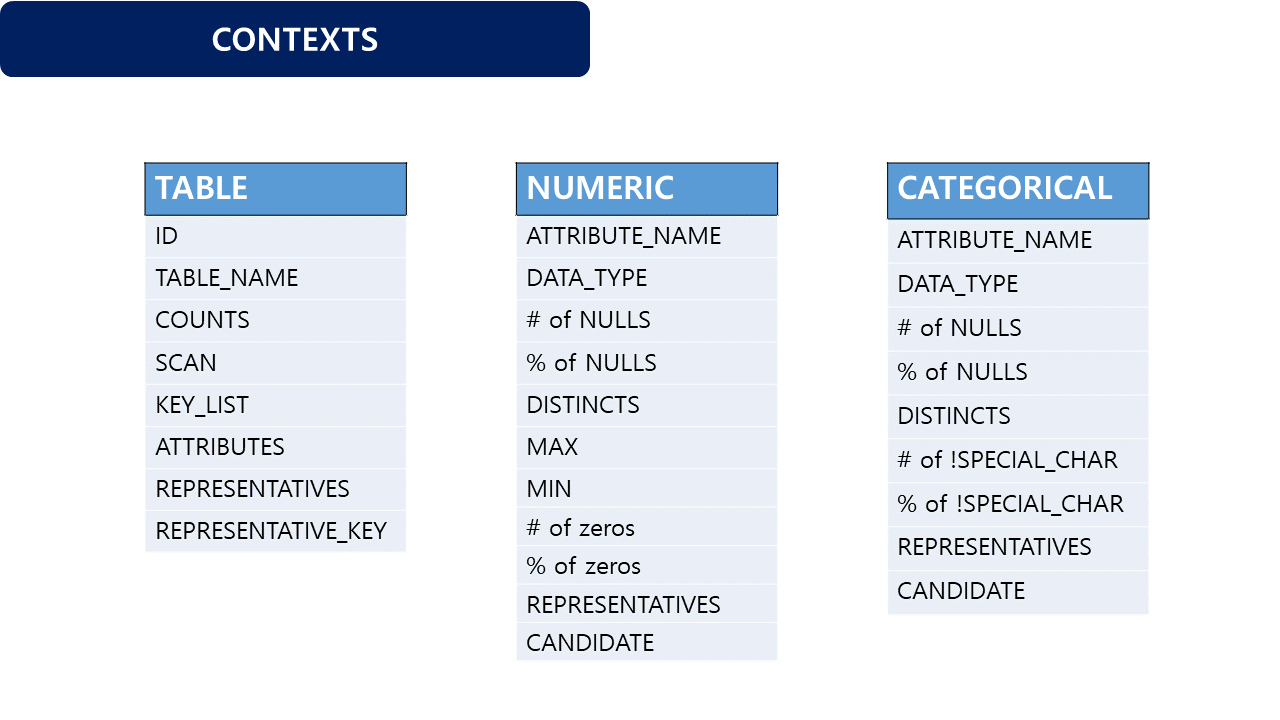
다중 결합

단일 결합 기능과 비슷한 방식으로 base table과 join table을 선택한다. 다중 결합 탭에서는 join table을 여러 개 선택할 수 있으며 결과로는 각 join table 에 대한 base table과의 inner join( base & join1, base & join2)의 결과를 보여준다.

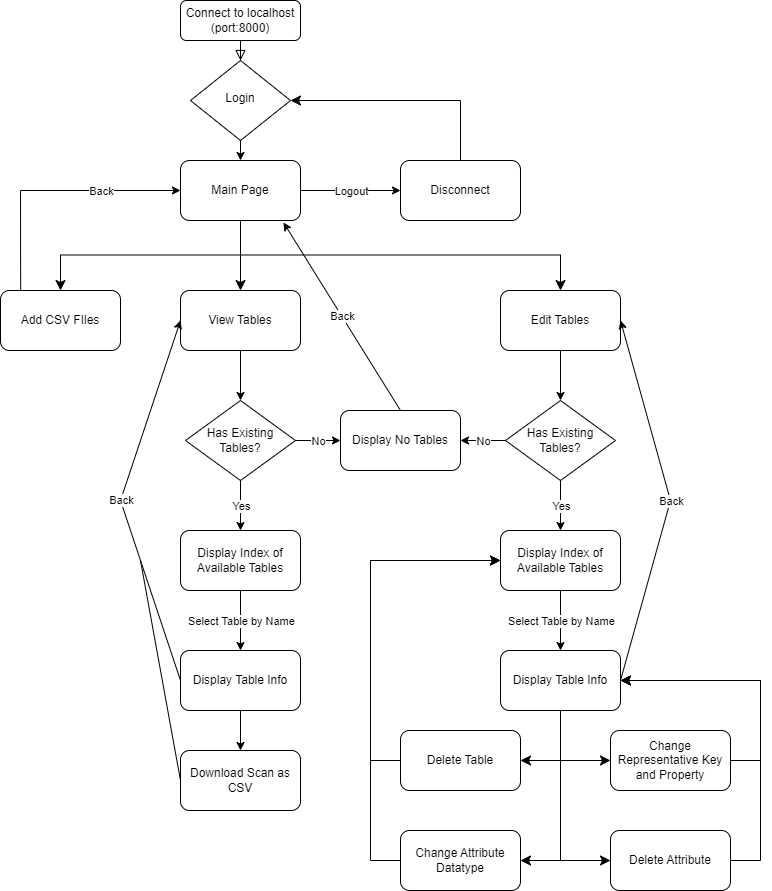
Prepared Table & Contexts for HTML Tags







The Flow Chart of Part A



Part A 설명

해당 파트에서 주된 기능은 테이블 접근, 열람, 스캔, 수정이다. Maria DB와 연계하기 위해 IP 주소, 호스트 등등의 정보를 입력하고 Django의 세션을 활용하여 다른 함수마다 해당 연결을 유지할 수 있다.

테이블 등록 시에 CSV파일을 업로드하는데 그 때 TABLE\_COUNTS에 그에 대한 정보를 추가하고 TABLE\_NAME을 매개로 새로 생성된 테이블을 TABLE\_COUNTS 쪽에서 참조할 수 있도록 한다. (Foreign Key 연결은 아님).

Table list는 TABLE\_COUNTS를 SELECT \* 해서 조회한다.

수치 해석을 할 때는 (DESC TABLE\_NAME)등의 쿼리를 통해 테이블의 여러 정보를 받아오고 DISTINCT, MAX, MIN의 방법을 활용했다.

범주 해석을 할 때는 정규표현식([^a-zA-Z0-9\u3130-\u318F\uAC00-\uD7AF\s])을 사용하여 특수문자를 구분했다. 또한 DISTINCT를 사용해서 NULL의 비율이나 상이 범주값을 어렵지 않게 구할 수 있었다.

테이블 내용 수정은 각각의 속성 삭제, (데이터 타입, 대표 속성, 대표결합키)의 변경 사항 저장, 테이블 삭제 기능을 지원한다. 속성을 삭제할 때는 연관된 테이블 모두를 UPDATE하는 것에 주의를 기울였다. 데이터 타입을 변경할 때는 STRING에서 INTEGER로 돌아올 때 정규표현식([^0-9])을 사용하여 숫자형식만 들어간 column을 걸러냈다. 대표 속성은 유저가 직접 입력하는 방식, 대표 결합키는 결합키의 후보에 적절한 요소만이 설정되도록 했다. 테이블의 삭제는 각 테이블을 지워버리는 것은 물론, TABLE\_COUNTS, REPRESENTATIVE\_KEYS 테이블 모두에서 그 데이터를 삭제한다.

DB 연결 해제는 세션을 지움으로써 구현했다.

Part B 설명

Part B의 단일/다중 결합 구현에는 도메인 스캔과 대표속성, 대표 결합키 설정이 완료된 테이블들의 정보가 JOINABLE\_TABLE에 삽입된다. 결합을 시도하면 JOINABLE\_TABLE에서 결합 가능한 테이블들을 사용자가 입력한 조건에 맞게 선택해서 표시하고 단일/다중 결합 여부에 따라 최대 2개/2개 이상의 테이블을 INNER JOIN 하여 그 결과를 표시한다. 결과물 TABLE은 두 TABLE의 결합을 의미하는 이름으로 DB에 저장되고, 결합 결과 요약 내용은 SINGLE/MIJLTI\_JOIN\_RESULT 라는 테이블에 저장된다. 결과 조회 페이지에서 이 RESULT 테이블에서 결합 결과 이력을 뽑아서 SINGLE/MULTI JOIN 을 나누어서 그 결과를 표시한다. 결합 결과물을 CSV파일로 다운로드 받는 기능은 결합 결과물과 결과 요약 테이블에서 SQL QUERY 를 통해 내용을 가져와 Pandas.DataFrame으로 만든 다음 to.csv 메소드를 이용해 다운로드 받을 수 있도록 하였다.

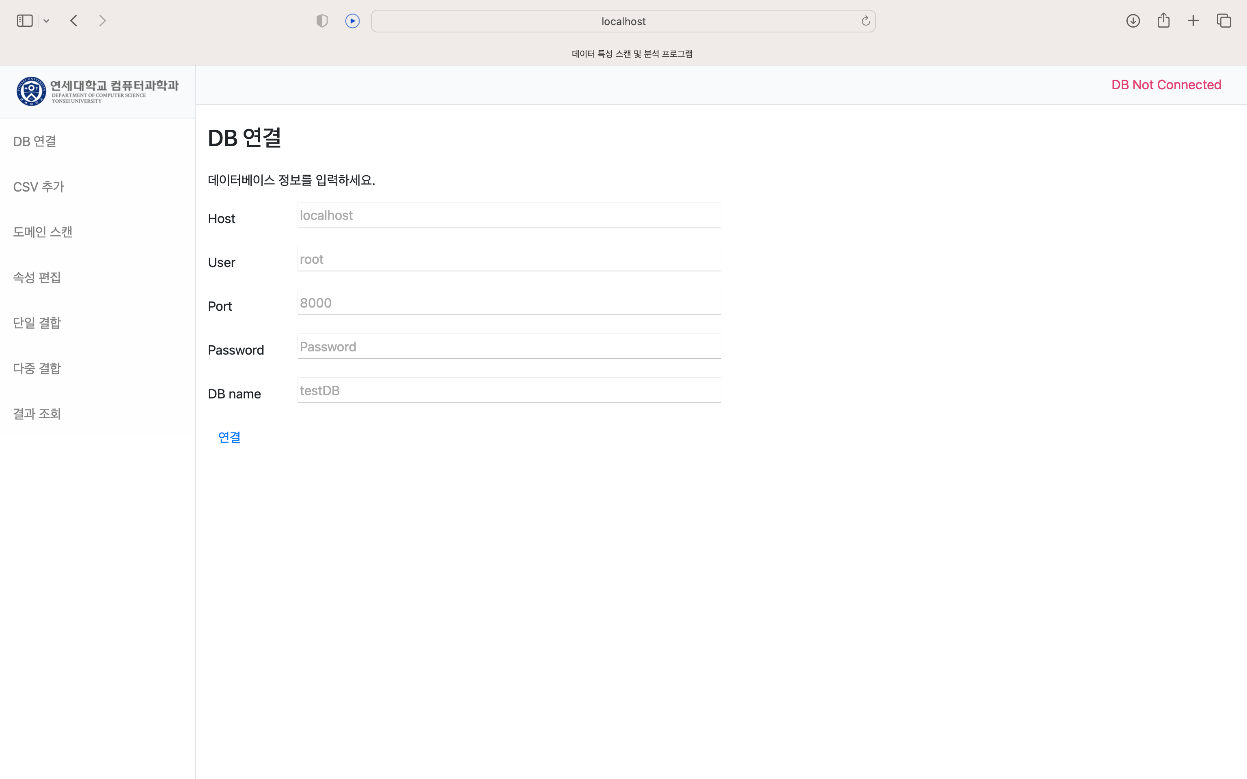
Result of the system(시스템의 결과)

메인 화면(DB연결 여부는 우측 상단의 DB ON/OFF로 확인 가능)

텍스트이(가) 표시된 사진

자동 생성된 설명

DB 등록

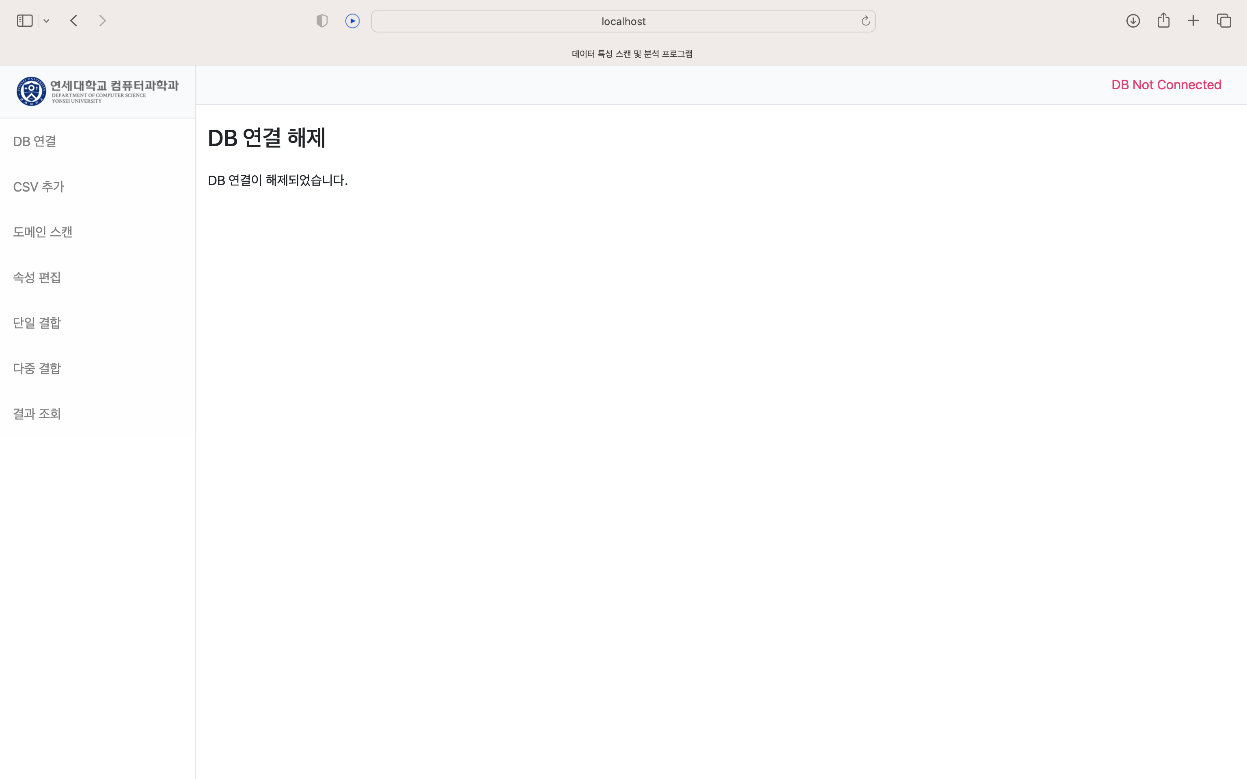


DB 등록 완료

텍스트이(가) 표시된 사진

자동 생성된 설명

DB 연결해제



CSV 추가

텍스트이(가) 표시된 사진

자동 생성된 설명

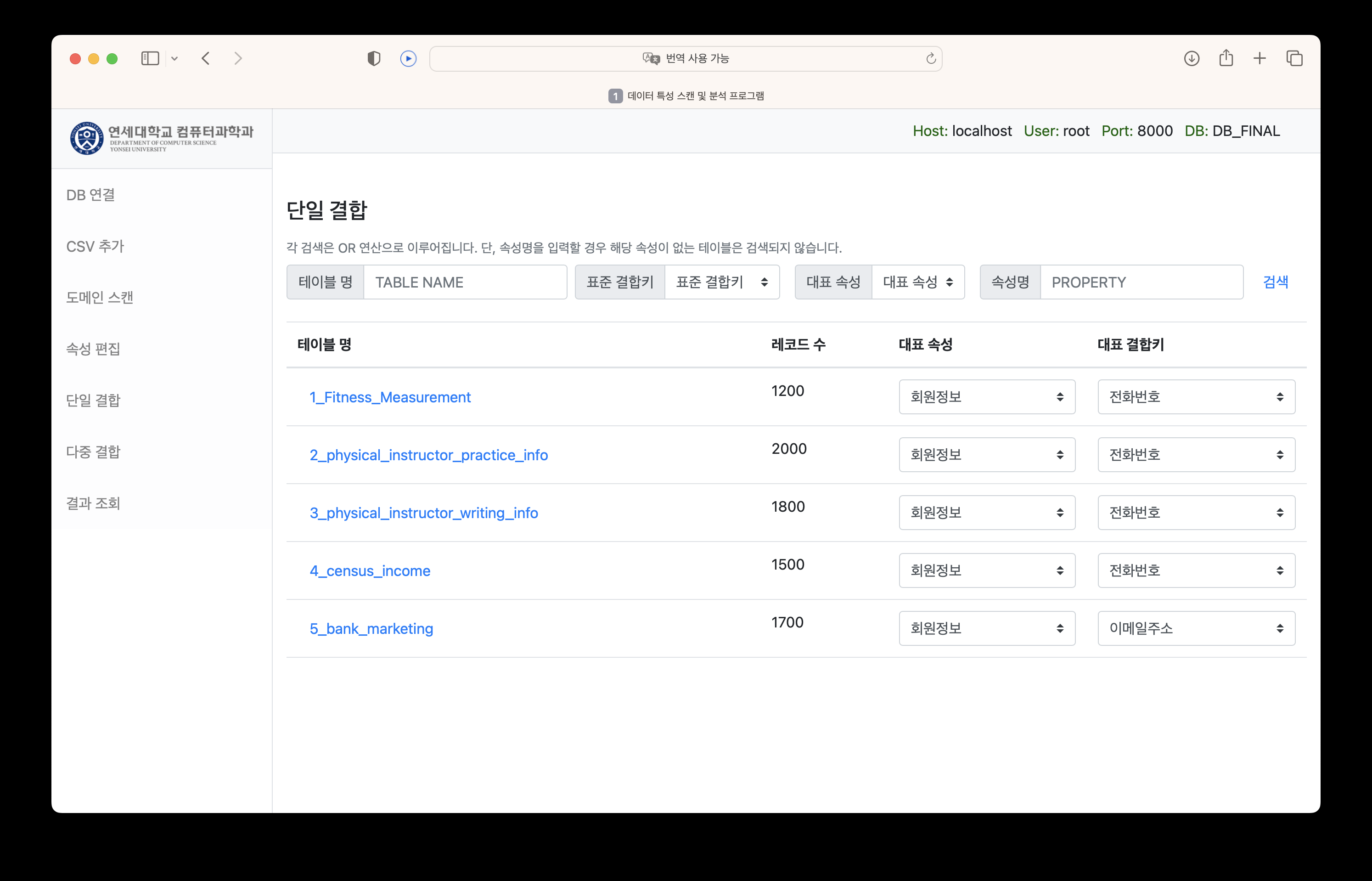
도메인스캔



속성편집



단일결합



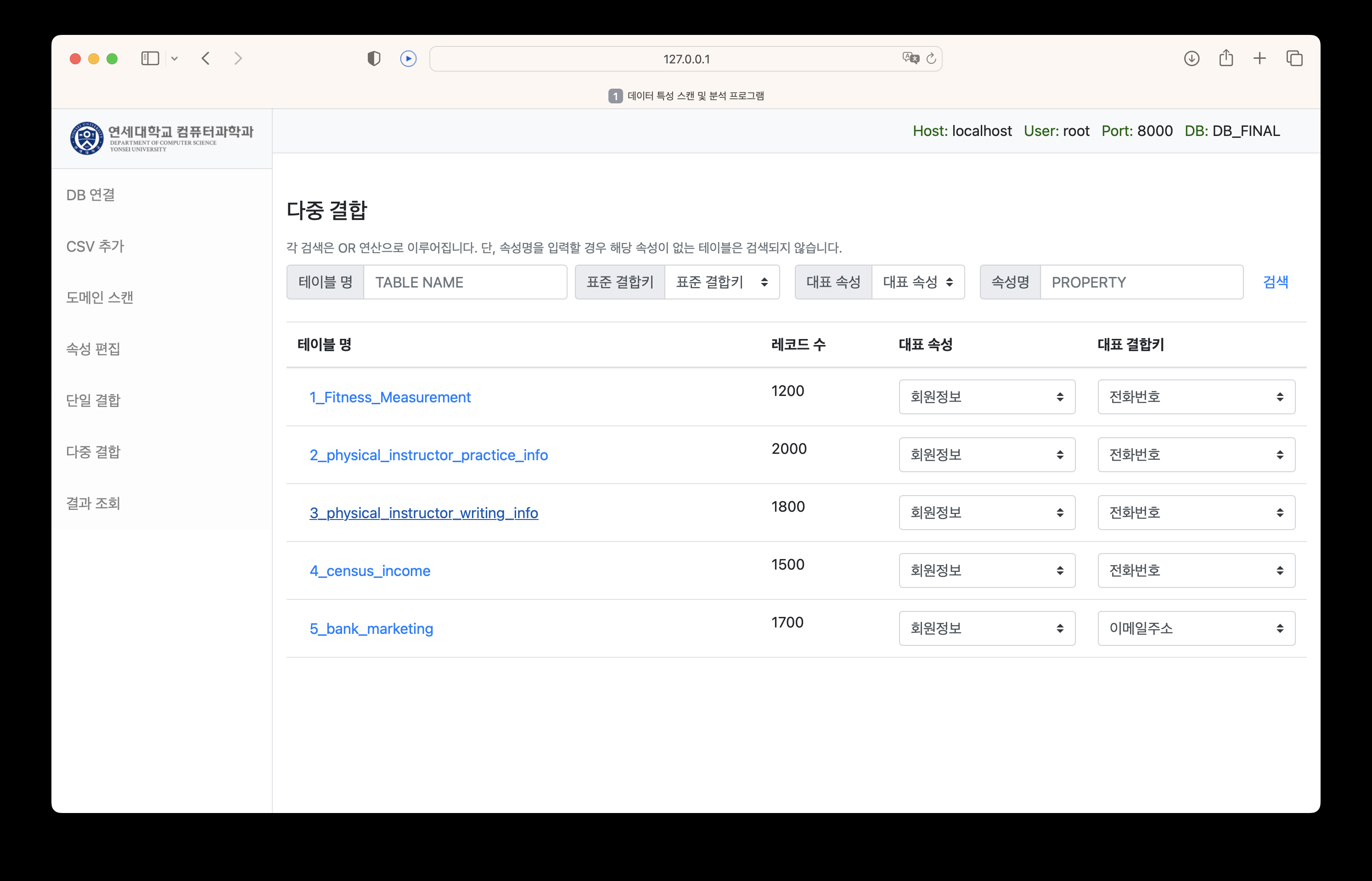
테이블이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

다중결합



테이블이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

결과조회

텍스트이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

Example Records used in Database(예시 데이터)

Ex) 1\_fitness\_measurement.csv



총 1200개의 레코드로 이루어졌다.

Ex) 2\_physical\_instructor\_practice\_info.csv



총 2000개의 레코드로 이루어졌다.

Ex) 3\_physical\_instructor\_writing\_info.csv



총 1800개의 레코드로 이루어졌다.

Ex) 4\_census\_income.csv



총 1500개의 레코드로 이루어졌다.

Ex) 5\_bank\_marketing.csv



총 1700개의 레코드로 이루어졌다.

총 5개의 csv파일과 sql파일을 예시 데이터로 받았다.

위 예시 데이터들을 이용한 DB를 dump하여 ‘data.sql’로 별도 첨부한다.