

프로젝트 공지  
406.426B 데이터관리와 분석  
2019년 2학기

본 프로젝트의 목적은 주어진 요구사항들을 만족하는 데이터베이스를 설계 및 구현한 후, 해당 데이터베이스에 대한 질의를 수행하고, 이를 통하여 DB 마이닝을 실행하는 것이다.

프로젝트는 다음과 같이 3단계로 구성되어있다.

Project #1) Conceptual DB design

팀 구성일 : 9월 26일

발표일: 10월 15일

Project #2) DB implementation & query processing

팀 구성일 : 10월 22일

발표일: 11월 14일

Project #3) DB mining

팀 구성일 : 11월 19일

발표일: 12월 12일

본 프로젝트는 팀 별로 진행되며, 각 팀은 매 프로젝트마다 임의로 4명으로 구성된다. 단, 모든 조가 똑같이 4명씩 구성될 수 없는 경우, 몇 조는 예외적으로 3명 또는 5명으로 구성된다. 3명 또는 5명으로 구성되는 조는 최소한으로 한다.

각 프로젝트 별로 각 팀은 정해진 발표 일에 4분 이내로 결과 발표를 해야 하며, 발표일 0시까지 보고서 및 발표 자료를 ETL에 업로드하고(한 조당 한 명만), 업로드 한 자료로 해당 발표 일에 발표하여야 한다. 또한 발표 일에 보고서 및 발표 자료 하드카피를 담당 교수에게 제출해야 한다. (하드카피의 내용과 ETL에 제출된 내용은 정확히 같아야 한다.)

## 프로젝트 #1: Conceptual DB design

사이트 A는 가입자들에 대한 정보와 그들의 관심사, 그리고 특정 관심사를 공유하는 모임들에 대한 정보를 저장하고 이를 가입자들에게 제공한다. 본 프로젝트는 사이트 A가 사용하는 DB에 대한 ER diagram과 관계 스키마를 도식화하고, 그에 대한 설명을 보고서로 서술 및 발표하는 것을 목표로 한다. 사이트 A의 DB는 아래의 requirement들을 만족해야 한다.

- (R1) 사이트 A는 도시의 정보를 가지고 있다. 여기에는 사이트에서 부여한 고유의 id(정수), 도시 이름(최대 30자의 문자열), 주(state)(최대 30자의 문자열), 나라(최대 30자의 문자열), 우편번호(정수), 위도(실수), 경도(실수), 멤버 수(정수)가 저장되어야 한다. 우편번호, 위도, 경도는 빈 칸일 수 있다.
- (R2) 사이트 A는 주제에 대한 정보를 가지고 있다. 여기에는 사이트에서 부여한 고유의 id(정수), 주제 이름(최대 30자의 문자열), 주제에 대한 설명(문자열), 멤버 수(정수)가 저장되어 있다. 주제에 대한 설명은 빈 칸일 수 있다.
- (R3) 사이트 A는 사용자에게 그룹을 소개하며 정보를 제공해야 한다. 따라서 사이트 A는 그룹에 대한 정보를 가지고 있어야 한다. 여기에는 사이트에서 부여한 고유의 id(정수), 주제의 id(정수), 그룹 이름(최대 30자의 문자열), 그룹 생성일(yyyy-mm-dd hh:mm:ss), 그룹 소개글(문자열), 평균 평점(실수), 공개 여부(정수형, 공개라면 1, 비공개라면 0), 승인 필요 여부(정수형, 승인이 필요하다면 1, 필요없다면 0), 다른 가입자들에게서 받은 ‘좋아요’의 수(정수), 멤버 수(정수)가 저장되어 있다. 그룹은 하나의 주제만 가질 수 있으며, 한 그룹에 그룹장은 단 한 명 뿐이다. 하지만 한 사람은 여러 그룹의 그룹장을 맡을 수 있다. 그룹 소개글은 빈 칸일 수 있다.
- (R4) 사이트 A는 가입자들에 대한 정보를 가지고 있다. 여기에는 사이트에서 부여한 고유의 id(정수), 속해있는 모든 group들의 id, 가입자 이름(최대 30자의 문자열), 가입 날짜(yyyy-mm-dd hh:mm:ss), 최근 접속일(yyyy-mm-dd hh:mm:ss), 자기 소개글(문자열), 활동 여부(정수형, 활동 중이라면 1, 아니면 0), 가입자에게 ‘좋아요’를 남긴 가입자의 수(정수), 해당 가입자를 팔로우하는 가입자의 수(정수), 가입자의 친구의 수(정수)

수), 가입자가 속한 그룹의 수(정수)가 저장되어 있다. 한 명의 가입자는 활동 중인 도시를 하나만 설정할 수 있다. 자기 소개글은 빈 칸일 수 있다.

(R5) 사이트 A는 그룹장에 대한 정보를 가지고 있다. 여기에는 사이트에서 부여한 고유의 id(정수), 그룹장의 가입자로서의 id, 그룹장으로 활동 중인 모든 그룹의 id가 저장되어 있다.

(R6) 사이트 A는 가입자들이 그룹에 대해 남긴 ‘좋아요’를 저장하고 있다. 한 명의 가입자는 하나의 그룹에 대해 단 한 번만 ‘좋아요’를 누를 수 있다. 이 때, ‘좋아요’를 누른 가입자의 id, 그룹의 id가 저장되어야 한다.

(R7) 사이트 A는 만남 장소에 대한 정보를 가지고 있다. 여기에는 사이트에서 부여한 고유의 id(정수), 도시의 id, 만남 장소 이름(최대 30자의 문자열), 주소(문자열), 우편번호(정수), 위도(실수), 경도(실수), 평균 평점(실수), 받은 평가의 수(정수)가 저장되어 있다. 주소, 우편번호, 위도, 경도는 빈 칸일 수 있다.

(R8) 사이트 A는 그룹들이 개최하는 오프라인 모임에 대한 정보를 가지고 있다. 여기에는 사이트에서 부여한 고유의 id(정수), 해당 모임을 가지는 그룹의 id, 해당 모임의 만남 장소의 id, 모임 이름(최대 30자의 문자열), 첫 모임 생성 날짜(yyyy-mm-dd hh:mm:ss), 최근 업데이트 날짜(yyyy-mm-dd hh:mm:ss), 실제 모임 날짜(yyyy-mm-dd hh:mm:ss), 예상 모임 시간(정수), 비용(문자열), 모임 소개글(문자열)이 저장되어 있다. 모임 소개글은 빈 칸일 수 있다.

(R9) 사이트 A는 모임을 찾아오는 법(문자열)을 저장하고 있으며 이는 특징적인 요소와 이에 대한 세부 설명으로 나뉜다. 예를 들면 ‘산공 동문회를 오기 위해서는 지하철을 타고 서울대입구역 2번출구로 오면 됩니다’는 정보는 ‘산공 동문회’라는 모임의 id, ‘지하철’이라는 특징적인 요소(문자열)와 ‘서울대입구역 2번출구로 오면 됨’이라는 상세 정보(문자열)로 저장된다. 하나의 모임은 하나의 특징적인 요소에 대해 최대 하나의 기록을 가지고 있다. 하나의 모임은 다수의 찾아오는 법을 가질 수 있으며, 정보가 부족해 찾아오는 법이 저장되어 있지 않을 수 있다.

- (R10) 사이트 A는 모임에 초대된 사람의 수에 대한 정보를 저장하고 있다. 초대된 사람의 수에 대한 기록은 모임의 id, 사람 수의 제한(정수, 제한이 없다면 -1), 초대에 응한 사람의 수(정수), 응답 대기 수(정수)를 포함한다.
- (R11) 사이트 A는 가입자들 사이의 교류를 기록한다. 가입자들 사이의 교류는 ‘좋아요’, ‘팔로우’, ‘친구’의 세 가지로 나뉘어진다. ‘좋아요’는 해당 가입자의 취향, 리뷰 등이 마음에 들 때 그에 공감하는 의미로써 사용되곤 하며, ‘팔로우’는 해당 가입자가 리뷰, 팁을 남길 때에 알림을 오도록 하는 기능을 의미한다. ‘친구’는 일방적인 ‘좋아요’와 ‘팔로우’와는 달리 두 가입자의 합의 하에 생성되는 관계이다. 사이트 A는 ‘좋아요’, ‘팔로우’, ‘친구’의 기록을 따로 저장하며, 이들의 기록은 모두 2개의 가입자 id로 구성되어 있다. 그룹에 대한 ‘좋아요’와 사용자에게 대한 ‘좋아요’는 각각 별개의 기능이다.
- (R12) 길이가 따로 명시되지 않은 모든 문자열의 데이터는 최대 255자의 가변 길이를 가지고 있다고 가정한다. 또한 빈칸이 가능하다고 명시되지 않음은 빈칸이 허용되지 않음을 의미한다. 자료형이 명시되지 않은 데이터는 데이터에 적합한 자료형을 선정하고 그에 대한 근거가 보고서에 제시되어야 한다.

데이터베이스는 **redundancy**를 최소화하는 방향으로 설계되어야 한다. 또한, 제출할 보고서에는 문제 정의, ER 다이어그램, 관계 스키마, **constraint** 등이 포함되어야 한다. 관계 스키마의 경우 **primary key**와 **foreign key**, **domain** 및 **integrity constraint**가 명시되어야 한다.

#### 채점 기준 (절대 평가)

- 설계된 DB의 requirements 만족 여부: 90%  
ER 다이어그램(40%), 관계 스키마(40%), Constraints(10%)
- 보고서 품질: 5%
- 발표: 5%