Univ.-Prof. Dr.-Ing. Matthias Boehm
Graz University of Technology
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMK endowed chair for Data Management

1 Data Management SS21: Exercise 01 – Data Modeling

Published: March 07, 2021 (last update: Mar 07)
Deadline: March 30, 2021, 11.59pm

This exercise on data modeling aims to provide practical experience in Entity-Relationship (ER) modeling, ER-relational mapping, and relational normalization. The expected result is a zip archive named `DBExercise01_<student_ID>.zip`, containing the partial results of the individual sub-tasks, submitted in TeachCenter.

1.1 ER Modeling (15/25 points)

Create an ER diagram in Modified Chen (MC) notation—including entity types, relationship types, attributes, cardinalities, and keys (create surrogate keys if natural unique identifiers are missing)—for managing the Summer Olympics (1896-2016) dataset[1]. It is up to you if you use existing tools for data modeling or draw this by hand. There are multiple correct ways of modeling this discourse, however the diagram should capture the following information:

- Olympic summer *games* (e.g., Summer 2016) are held—typically every four years—at one or multiple *host cities* (e.g., Rio, Brazil) in one or multiple countries, and are described by a year, a start date, and an end date.

- Participating *teams* in these games have a name and represent a specific *country*, identified by a three-letter NOC (National Olympic Committee) code. Special teams from multiple nations have dedicated codes. Countries are further described by their population, and GDP (gross domestic product) per capita.

- Each team is composed of multiple *athletes*, each described by an AID (globally unique across specific Olympic games), name, day of birth, height, weight, and gender. Athletes cannot be in multiple teams of a specific Olympic games.

- Athletes compete in potentially multiple *events* (e.g., Swimming Women's 4 x 100 metres Medley Relay)—which are classified into *sports* (e.g., Swimming)—of specific Olympic games, and might win Gold, Silver, or Bronze (1st, 2nd, 3rd rank) medals in these events.

Partial Result: `ERDiagram.pdf`

---

[1]Dataset link: `https://github.com/tugraz-isds/datasets/tree/master/summer_olympics`. A cleaned version of the dataset will be made available March 21, 2021 after additional data preparation and cleaning.

1.2 Mapping ER Diagrams into the Relational Model (10/25 points)

Create a relational schema for the ER diagram designed in Task 1.1 and bring it into third normal form, assuming the following additional functional dependencies:

$$\text{City} \rightarrow \text{Country}$$
$$\text{Games.BeginDate} \rightarrow \text{Games.Year}$$
$$\text{Games.EndDate} \rightarrow \text{Games.Year}$$
$$\text{Event.Name} \rightarrow \text{Event.SportsName}$$

This schema should include the relations and typed attributes, as well as all primary and foreign keys. Furthermore, the schema should also ensure that no athlete can win multiple medals in a single event of a specific Olympic games. It is up to you if you provide either a SQL script (`CreateSchema.sql`) with `CREATE TABLE` statements, or provide a text schema (`Schema.txt`) in the following text notation:

`<Table>(<Attribute 1>:<type>(PK), <Attribute 2>:<type>, ..., <Attribute n>(FK))`

Here, PK and FK indicate primary and foreign keys, where multiple attributes with (PK) represent a composite primary key. If an attribute is both foreign key and (part of) a primary key, use `<Attribute m>(PK,FK)`. Please, adhere to this notation with unchanged parentheses and delimiters, and limit yourself to common data types (i.e., `int`, `numeric(p,s)`, `char(n)`, `varchar(n)`, or `date`) because this sub-task is subject to automated grading.

Partial Result: `Schema.txt`, or `CreateSchema.sql`