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

1 Data Management SS 2020: Exercise 01 – Data Modeling

Published: March 14, 2020 (last update: March 14, 2020)
Deadline: March 31, 2020, 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 PDF-file named **DBExercise01_<studentID>.pdf** submitted in TeachCenter.

1.1 ER Modeling (10/25 points)

Create an ER diagram in Modified Chen (MC) notation—including entity types, relationship types, attribute types, cardinalities, and keys[1]—for managing bibliography data. It's 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, but the diagram should capture the following information:

- A *person* has a unique ID AKey, a unique name, may have multiple aliases, may have a website, and is affiliated with at most one *institution* (e.g., a university, research lab, or company). A *person* can act as both, an author of publications, or as an editor of *conferences* (e.g., a PC chair). Editors can organize as many *conferences* as they like (meaning also none). The number of publications a *person* might publish is also not limited.
- An *institution* has a—not necessarily unique—name, and is located in a specific country. Many *persons* can be affiliated with the same *institution*. Furthermore, a subset of *institutions*—so-called degree-granting institutions such as universities—can award master and PhD degrees. Accordingly, an *institution* can be linked to zero or many *theses*.
- We consider two types of publications: *papers* and *theses*. A *paper* has a unique ID PKey, a title, an ordered list of up to 64 authors (i.e., *persons*), a page range, and is either published in a *journal* (i.e., as a journal article) or published in the proceedings of—and presented at—a *conference*. Due to double submission policies, a single paper cannot appear in multiple *journals* or *conferences*. A *thesis* has a unique ID TKey, a title, exactly one author, a publication year, a type (master or PhD), a number of pages, and an ISBN number. Additionally, each *thesis* is associated with exactly one *institution*.
- A published *journal* issue has a journal short name, a journal title, a volume number, an issue number, and a year of publication. Volumes typically cover all issues of a year but might not be aligned with calendar years. A journal issue contains at most 256 *papers*.
- A *conference* is an event of paper presentations, described by a short name, a title, a year, and takes places at a certain location (city, country). The conference proceedings—which refers to a printed or electronic book—receives an ISBN number. Such a conference can be organized by zero to 32 editors (i.e., *persons*) and covers between one and 512 *papers*.

[1] Use surrogate keys in case of missing unique identifiers or strings longer than 4 letters.

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

Create a relational schema for the ER diagram designed in Task 1.1. This schema should include the relations and typed attributes, as well as all primary and foreign keys. It's up to you if you provide a SQL script, or use the following text notation:

<Table>(<Primary key>:<type>, <Attribute>:<type>, ..., <Foreign key>:<type>)

1.3 Relational Normalization (5/25 points)

Bring the relational schema from Task 1.2 into third normal form by listing any necessary schema changes. Furthermore, please explain in detail—with specifics of your particular schema—why your modified schema is now in 3rd normal form.

1.4 Extra Credit (+5 points)

(a) Provide an additional list of all relationship types from Task 1.1 in (min,max)-notation using the following notation:

```
<entity1> (min,max) - <relationship> - (min,max) <entity2>
```

(b) Have a look at the given ER diagram in Figure 1. Identify and explain poor design choices made in this diagram. Additionally, suggest how this diagram could be improved, either by explaining necessary modifications or by drawing the improved ER diagram.

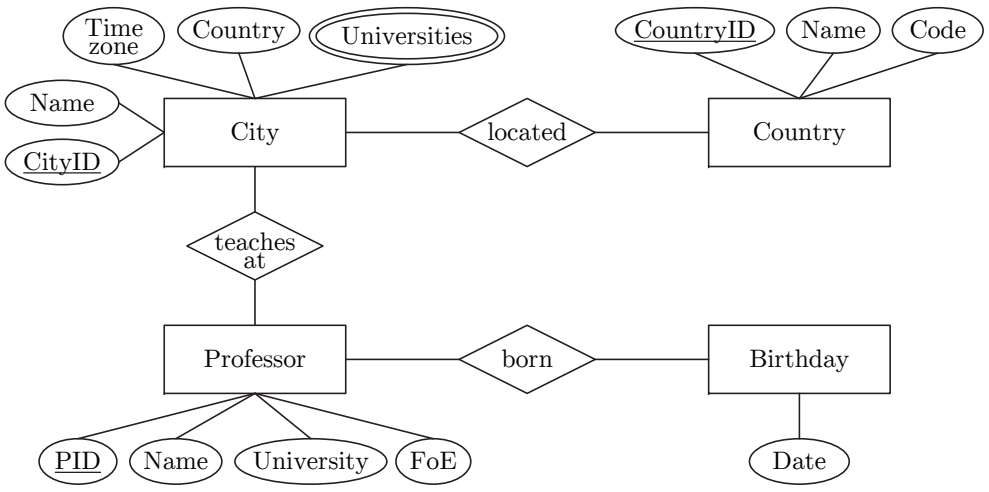

Figure 1: ER Diagram with Poor Design Choices.