Exercise sheet 1 (WS 2020)

3.0 VU Data Modelling / 6.0 VU Database Systems

About the exercises

General information

In this part of the exercises you are asked to create a small database using EER-diagrams, transform EER-diagrams into a relational schema and make yourself familiar with relational algebra and relational calculus.

We recommend you to solve the problems **on your own** (it is a good preparation for the exam – and also for your possible future job – to carry out the tasks autonomously). Please note that if we detect duplicate submissions or any plagiarism, both the "original" and the "copy" will be awarded 0 points.

Your submission must consist of a single, typset PDF document (max. 5MB). **We do not accept PDF files with handwritten solutions.**

In total there are 8 tasks and at most 15 points that can be acchieved on the entire sheet.

Deadlines

until 05.11.	**12:00pm**	Upload your solutions to TUWEL
18.11.	13:00pm	Evaluation and feedback is provided in TUWEL

Further questions – TUWEL forum

If you have any further questions concerning the contents or organization, do not hestitate to ask them on TUWEL forum. **Under no circumstances should you post (partial) solutions on the forum!**

Changes due to COVID-19

Due to the ongoing situation with COVID-19 we will not offer in-person office hours for the exercise sheets. If you have technical issues, trouble understanding the tasks on this sheet, or other questions please use the TUWEL forum.

We also recommend that you get involved in the forum and actively discuss with your colleagues on the forum. From experience we believe that this helps all parties in the discussion greatly to improve their understanding of the material.

Exercise: EER-diagrams

Exercise 1 (Creating an EER-diagram) [3 points]

You are working in a startup that will revolutionize book marketing. After securing many millions in fuding, you decide to actually work on the project. You begin by designing the database on the basis of your current notes.

Draw an EER-diagram based on the available information (see next page). Use the notation presented in the lecture and the (min, max) notation. NULL values are not allowed and redundancies should be avoided. Sometimes it might be necessary to introduce additional keys.

A possible software for creating the EER-diagram is DIA (`http://wiki.gnome.org/Apps/Dia`, binaries at `http://dia-installer.de`; Attention: select ER in the diagram editor!). Of course you are also allowed to create the diagram with any other suitable software.

Description of the issues:

The crucial function of the database is to record books. For every book a title (TITLE) and a number of pages (PAGES) is stores, where the title identifies a book uniquely. Furthermore, for every book we save whether it specifically is a non-fiction, poetry, or a novel. Note that a book does not necessarily have to fall into one of these categories. For a novel we additionally store its genre (GENRE). Non-fiction books can be written about novels, but every non-fiction book can only be written about at most three different novels. It is also possible that a non-fiction book is not written about any novel. Books are written by authors. An author always has written at least one book. An author is uniquely identified by the combination of their name (NAME) and date of birth (DOB).

There can be multiple editions of each book. Different editions of a book can be distinguished by their edition number (EDNR). However, the edition number is not necessarily unique among different books. We also store a year (YEAR) value for each edition. An edition has at least one printing. The printings of an edition can be identified by their print number (PNR), which is again not necessarily unique over different editions. For every printing we record who does the printing (PRINTER). Additionally, some editions are special editions for which we also store a reason (REASON).

Editions are published by a publisher. A publisher has a name (NAME) and a budget (BUDGET) associated with them, where the name uniquely identifies a publisher. Every publisher has departments which are identified inside the publisher by a unique combination of area of responsibility (AREA) and their location (CITY). Note that the same area/location combination can occur in multiple publishers and that there are no departments without a publisher.

A publisher advertises, on a given date (DATE), to one or more target groups via up to two marketing channels. A marketing channels has a name (CHNAME) – which uniquely identifies the channel – and associated costs (COST). A marketing channel can also be a social network or a newspaper. For newspapers we additionally store the circulation (CIRC). For social networks the concrete platform (PLATFORM) is recorded. For target groups we record a description (DESC) as well as a age group (AGE), where the description is unique to the target group.

Target groups use social networks and are interested in at least one book. Every target group loves up to 5 authors. Regrettably, every target group also hates at least 1 (and arbitrarily many) authors. This motivates authors to use pseudonyms. Hence, we also record which author is an alias of which other author.

Exercise 2 (Semantics of EER diagrams) [1 point]

Consider the EER-diagram shown in Figure 1.

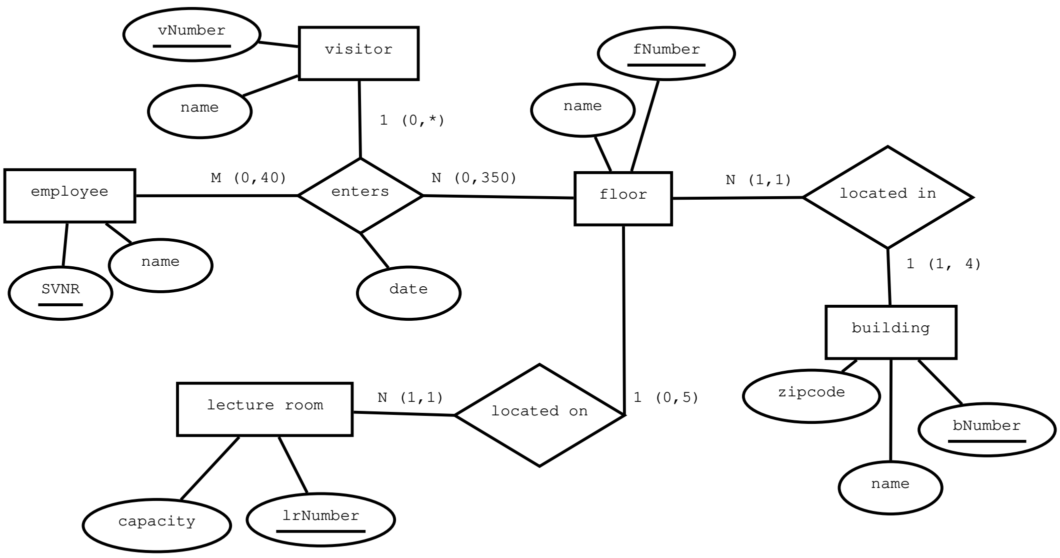

Figure 1: EER-diagram for exercise 2

In the ER diagram, both the notation by means of functionalities, as well as the (min, max) notation is used.
(*note:* this is not done in practice.)
Therefore, the diagram contains more information compared to the use of only one notation.

- Specify a specific relationship in the diagram where omitting one of the two notations causes a loss of information.
- For the chosen relationship type explain which notation, when omitted, leads to the loss of information.
- Explain briefly in your own words which information can no longer be represented.
- Provide a concrete example of the lost information, i.e. for the type of relationship you have chosen, specify an instance that violates (at least) one condition expressed by the omitted notation, but satisfies all requirements by the remaining notation.

Exercise 3 (Construct a relational schema) [2 points]

Construct a relation schema according to the EER-diagram given in Figure 2. `NULL` values are not allowed (you can assume that all attributes specified for an entity type exist for all entities of this type, i.e., the definedness of all attributes is 100%). Create as few relations as possible without introducing any redundancies. For each relation clearly mark the primary keys by underlining the corresponding attributes and display foreign keys in italics.

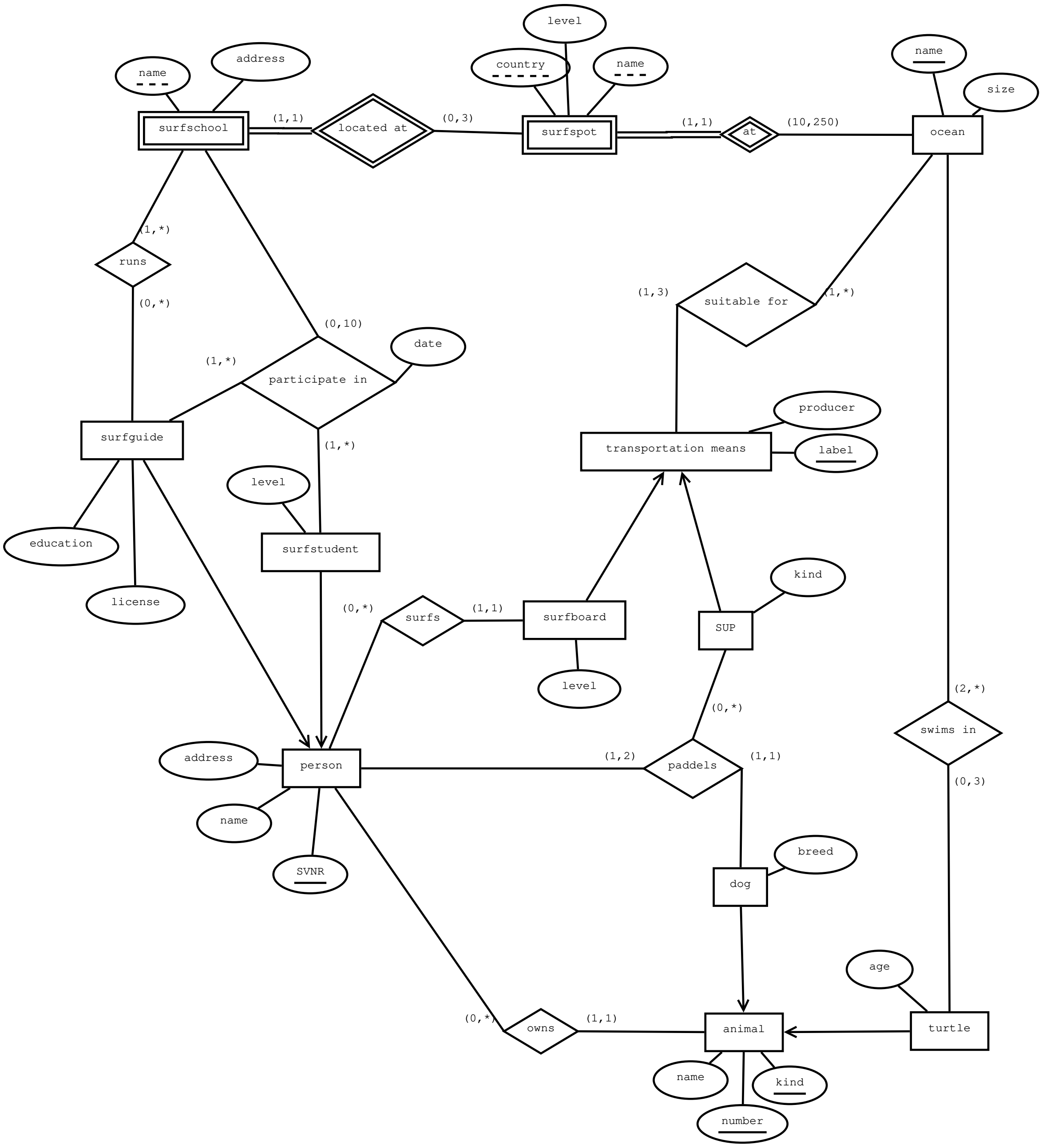

Figure 2: EER-diagram for exercise 3

Exercises: Relational Algebra - Relational Calculus

To help with typesetting the solutions to the following exercises, we compiled a list of the most important symbols for the Relational Algebra at `http://dbai.tuwien.ac.at/education/dm/resources/symbols.html`. You can copy and paste them into your Word/ LibreOffice/ OpenOffice/... document. In addition, the corresponding LaTeX commands are listed as well.

Exercise 4 (Evaluation) [0.5 points]

Consider the four relations below.

Band		
name	**genre**	**founded**
Cid Rim	TBD	2010
JSBL	Funk	2008
CSH	Indie	2010
Dorian Concept	Electronic	2006

Musician		
name	**instrument**	**studio**
Will	Guitar	False
Peter	Bass	True
Clemens	Drums	True
Dorian	Keyboard	True

partOf	
mname	**bname**
Clemens	JSBL
Clemens	Cid Rim
Will	CSH
Dorian	JSBL
Dorian	Dorian Concept

Song			
title	**by**	**length**	**genre**
Killer Whales	CSH	6:14	Indie
My Boy	CSH	2:52	Indie
Swnerve	Cid Rim	4:26	Electronic
Failures III	Dorian Concept	3:40	Jazz

Provide the results of the following queries over these relations.

(a)

$$\pi_{\text{title,genre}}\Big(\pi_{\text{title,length}}(\texttt{Song}) \times \big((\rho_{\text{mname}\leftarrow\text{name}}(\sigma_{\text{studio}=\textbf{True}}(\texttt{Musician})) \bowtie \texttt{partOf}) \rtimes \rho_{\text{bname}\leftarrow\text{name}}(\sigma_{founded\geq 2010}(\texttt{Band}))\big) \Big)$$

(b)

$$\begin{aligned}\{m.name, b.name \mid\ & m \in \texttt{Musician} \land b \in \texttt{Band} \land m.\text{instrument} \neq \text{'Bass'} \land \\ & \exists p \in \texttt{partOf}(\\ & b.\text{name} = p.\text{bname} \land m.\text{name} = p.\text{mname} \land \\ & (\forall s \in \texttt{Songs}(b.\text{name} = s.\text{by} \rightarrow s.\text{genre} = b.\text{genre})))\}\end{aligned}$$

Exercise 5 (Equivalences) [2 points]

Consider the following relation schemata $R(\underline{A}BC)$, $S(B\underline{D}E)$, $T(C\underline{D}F)$ and the pairs q_i, q_j of expressions in relational algebra below. For every pair of expressions:

- Verify, whether the two expressions are equivalent (i.e. whether they produce the same result on all possible instances of the relation schemata). You may assume that no `NULL`-values occur in any instance of the schemata.
- Justify your answer with a brief **explanation**.
- In case the two expressions are *not* equivalent, additionally provide a **counterexample**. (A counterexample consists of the concrete instances of the affected relation schemata and the results of both expressions over these instances.)
  In case one of the expressions is not a valid expression in relational algebra you do not have to provide a counterexample, hence in this case an explanation suffices.

(a) $q_1\colon (\pi_C(T) - \sigma_{C>5}(R)) \cap \rho_{C\leftarrow B}(\pi_B(S))$ and
$q_2\colon \big(\sigma_{C>5}(\pi_C(T) - \pi_C(R))\big) \cap \rho_{C\leftarrow B}(\pi_B(S))$

(b) $q_3\colon S \bowtie (\pi_{CD}(T) \cap \rho_{D\leftarrow A}(\pi_{CA}(R)))$ and
$q_4\colon (S \bowtie \pi_{CD}(T)) \cap (S \bowtie \rho_{D\leftarrow A}(\pi_{CA}(R)))$

(c) $q_5\colon \pi_{AB}(\sigma_{A>B \vee A<B}(R)) \times \pi_{DF}(\sigma_{D>F \vee D<F}(T))$ and
$q_6\colon \pi_{ABDF}(\sigma_{(A>B \wedge D>F)\vee(A<B \wedge D<F)}(R \times T))$

Exercise 6 (Answer Sizes) [1.5 points]

Consider the relational schemas $R(\underline{A}B)$, $S(AB\underline{CD})$, and $T(AC\underline{E})$ and an instance of every schema, where there are $|R|$ tuples in the instance of R, $|S|$ tuples in the instance of S, and $|T|$ tuples in the instance of T.

- Provide the minimal and maximal size (= number of tuples) of the following expressions in Relational Algebra for the given values of $|R|$, $|S|$, $|T|$.
- Justify your answer.
- For both, the smallest and biggest possible answer size, provide concrete instances of the schemas (with R, S, and T having $|R|$, $|S|$, and $|T|$ tuples, respectively) over which the query returns an answer with the minimal/maximal number of tuples.

(a) $q_1\colon\ \rho_{B\leftarrow C}(T) ⟕ \sigma_{A=4\wedge B=2}(R)$ (with $|R| = 5$ and $|T| = 4$)

(b) $q_2\colon\ \pi_{EX}(\rho_{X\leftarrow A, Y\leftarrow C}(S) \times T) - \pi_{EX}(\rho_{X\leftarrow A}(T) \ltimes S)$ (with $|S| = 7$ and $|T| = 4$)

(c) $q_3\colon\ (\pi_A(\sigma_{A\neq 1}(R)) \cap \pi_A(\sigma_{A>3}(T))) \cup \rho_{A\leftarrow C}(\pi_C(T) \bowtie \pi_C(S))$
(with $|R| = 5$, $|S| = 3$ and $|T| = 3$)

Exercise 7 (Query Languages) [1 point]

Consider the relational schema $R(\underline{A}B)$, $S(AB\underline{CD})$, and $T(AC\underline{E})$.

In the following exercises you are given a query in one of the query languages from the lecture. Your task is to translate the query into the two other query languages that were discussed in the lecture.

(a) Translate the query

$$\pi_{A,C}(S \rtimes T)$$

into tuple relational and domain relational calculus.

(b) Translate the query

$$\{x \mid \exists y, z([x, y] \in R \land [y, x, z, 3] \in S)\}$$

into tuple relational calculus and relational algebra.

(c) Translate the query

$$\{u, v \mid u \in R \land v \in T \land \exists w \in S(w.C > v.C)\}$$

into domain relational calculus and relational algebra.

Exercise 8 (Formalizing Queries) [4 points]

A company manages information about the entries of its employees in a data base with the following schema (primary keys are underlined, foreign keys are written in *italics*).

```
employee    (persNr, SVNR, eName)
floor       (fName, name)
enters      (employeeNr: employee.persNr, floorName: floor.fName, date)
leaves      (employeeNr: employee.persNr, floorName: floor.fName, date)
building    (bName, zipcode)
located_in  (fName: floor.name, bName: building.bName)
lecture_room(lrNumber, name)
located_on  (lrNumber: lecture_room.lrNumber, floorName: floor.fName)
```

(In the following you may use suitable (unique) abbreviations for relations and table names.)

Express the queries described below in **relational algebra**, the **tuple relational calculus** and the **domain relational calculus**.

(a) Find the names, the personal number and the social security number (SVNR) of all employees, that are currently on the floor with name 'H1'.

(b) Currently there are students in the lecture room with name 'Gödel'. Find all employees that are currently located on a floor, on which also the students are located.

(c) Find all employees that were located in different buildings on the same day.