

Exercise sheet 1 (WS 2019)

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, typeset 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 achieved on the entire sheet.

Deadlines

until 30.10. 12:00pm Upload your solutions to TUWEL
13.11. 13:00pm Evaluation and feedback is provided in TUWEL

Consultation hours (optional)

In the week before the deadline there are consultation hours held by tutors. If you have problems understanding the topics of the exercise sheet or questions about the exercises, you are welcome to just drop in at these consultation hours. The tutors will gladly answer your questions and help you understand the subjects.

The goal of these consultation hours is to help you with **understanding the topics and specific tasks** of the exercise sheet. The tutors will not solve your exercises or check your answers before you hand them in.

Participation is completely voluntary — dates and locations of the consultation hours can be found in TUWEL.

Debriefing (optional)

A few days after you received your feedback and grading of this exercise sheet, there is the opportunity to go through the tasks in small groups (max. 25 persons). The (relative) small group size is intended to enable an active participation. Each of these groups will be held by an assistant. The specific agenda of these meetings will depend on the interests and question of the participants (i.e., you). The main objectives are answering your open questions and resolving your remaining issues regarding the exercises. Therefore, please have a look at your feedback and evaluation to identify such problems before you attend the class. When participating, dare to ask your questions – no question has a (negative) impact on your grade.

Participation is absolutely voluntary. Registration in TUWEL is required to keep the size of the groups small. Dates and locations can be found in TUWEL.

Further questions – TUWEL forum

If you have any further questions concerning the contents or organization, do not hesitate to ask them on TUWEL forum.

Exercise: EER-diagrams

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

NASA and ESA are planning a joint acquisition of their most important hardware components. In order to make the best possible decision in this sensitive matter, a database containing the most important information about the possible opportunities is needed. After an intensive worldwide selection process, you are hired to create the database and you receive the result of the requirements analysis as given on the next page.

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:

Both coffee machines and coffee capsules are products that have a price (PRICE) and a model number (MODELNR), where the model number is unique for each product. For coffee machines, the name (NAME), maximum water pressure (PRESSURE), and expected durability, measured in number of prepared portions (NUMBCAPS), are also collected for decision making. Additionally, it should also be noted if a coffee machine is an advancement of (at most one) older machine. For coffee capsules, the material (MATERIAL) of the capsule is stored, as well as the unique type of coffee capsule. Capsule types have a volume (VOLUME) and can be uniquely identified by the combination of size their (SIZE) and ID (CTID). A capsule type may be compatible with other capsule types. Each coffee machine supports at least one capsule type, but not more than ten.

Products are produced by manufacturers. For each manufacturer, the country in which he is taxable (TAXC) is noted, as well as the unique brand name (BRAND). Moreover, an arbitrary number of reviews are collected for each manufacturer. The reviews for each manufacturer are numbered consecutively (ID) and the rating is saved (TXT). Some coffee machines require licenses to support certain capsule types. In such cases, for each affected coffee machine it should be saved for which capsule types licenses of which manufacturers were purchased and the respective licence fee (FEE).

Products are sold by retailers. For each retailer, the name (NAME) and the website (WEBSITE) are known. Both the name and the website of all the retailers are different. Retailers often offer bulk discounts. The different bulk discounts of a retailer can be distinguished by the combination of label (LABEL) and allowed discount (DISCOUNT). There is, however, no coordination among the retailers in assigning labels and discounts. In order to benefit from a discount, a minimum quantity (QUANTITY) of the selected product has to be purchased. Discounts are only granted after buying at least four different products, and the minimum selling quantity of each product may vary. In addition, there may be coupon codes for a bulk discount. Such codes (CODE) are unique in combination with the discount.

For each coffee the aroma (AROMA), the acid (ACIDITY) as well as the body (BODY) is noted. The different coffees should be distinguished by their name (NAME). It is also stored if the coffee is a ristretto, lungo or espresso. Additionally, for each ristretto the roasting temperature (ROASTTEMP) and for each espresso the brew pressure (PRESSURE) is saved.

Finally, for each coffee capsule, the contained coffee is noted, as well as for each coffee machine which coffee they can prepare (each coffee machine can prepare at least one coffee).

Exercise 2 (Semantics of EER diagrams)

[1 point]

Consider the EER-diagram shown in Figure 1. It illustrates an arbitrarily chosen variant of the so called *property graph* model. (Background: property graphs are a further popular data model, especially suitable for storing and processing network data. However, to solve this problem, there is no knowledge about property graphs necessary. This task is only about the information presented in the EER diagram.)

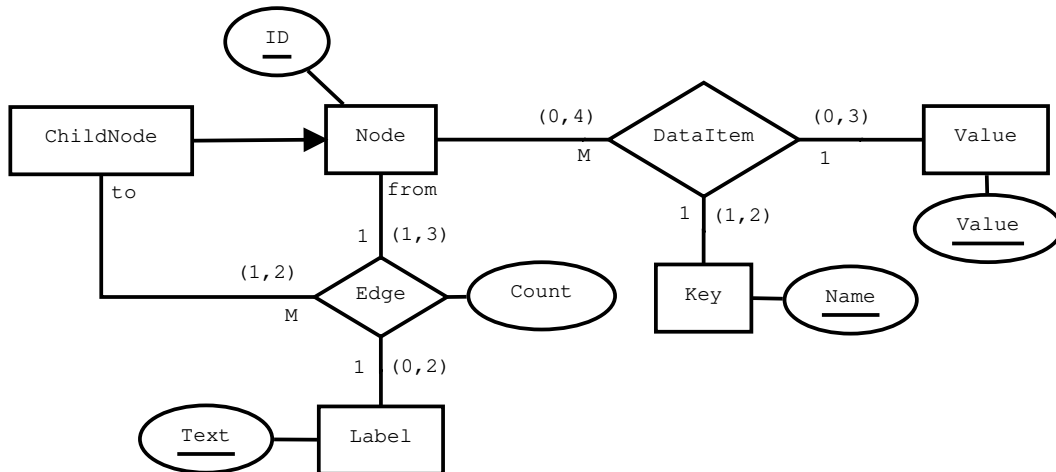


Figure 1: EER-diagram for exercise 2

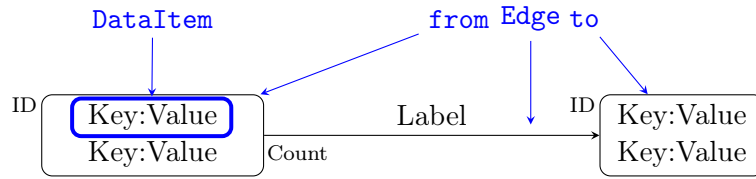
- In the ER diagram, both the notation by means of functionalities, as well as the (min, max) notation is used.
(note: this is not commonly done in practice.)

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, omitting which notation 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.
- Suppose you get the graph shown in Figure 2, and now you want/have to verify if it represents a valid instance of a property graph as described by the EER diagram.

Describe at least seven (7) violations of the rules presented in the EER diagram. (*Hint:* Specify only violations described by facts in the EER diagram. Ignore things that “make no sense” but not declared in the EER diagram.)

The “mapping” between the entity and relationship types described in the EER diagram and the specific instances in figure 2 is shown in the following graphic:



Note the following points:

- Nodes without incoming edges are always entities of type `Node`, but not of type `ChildNode`.
- Nodes with incoming edges are always of type `ChildNode`.
- Entities of type `Label`, `Key`, and `Value` are identified by the value of their key attributes `Text`, `Name`, and `Value`.
- Semicolons “;” separate different entities of the same type represented by their key.
- Elements shown in the above graphic blue are additional explanations and not part of the graph.

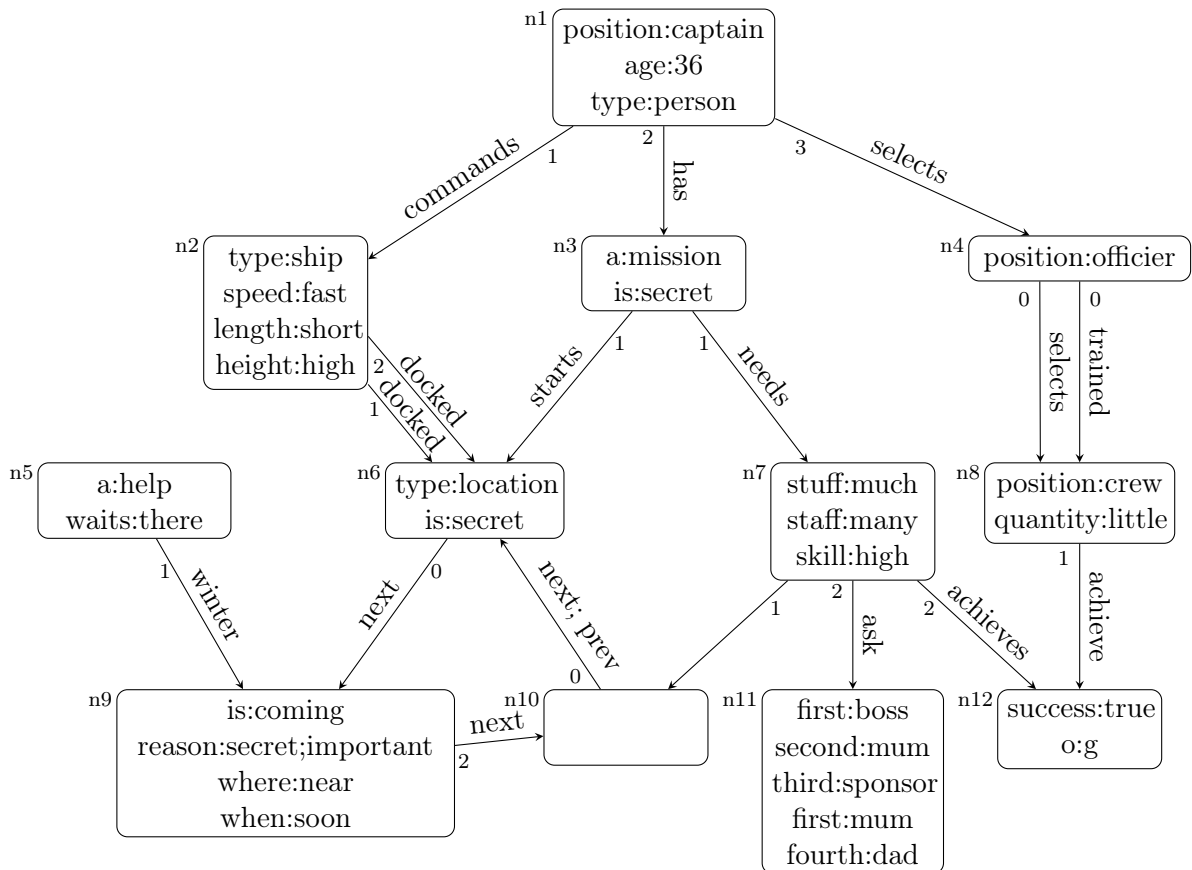


Figure 2: Task specification for Exercise 2: Example graph to be checked by the EER diagram in Figure 1

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

Construct a relational schema according to the EER-diagram given in Figure 3. 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.

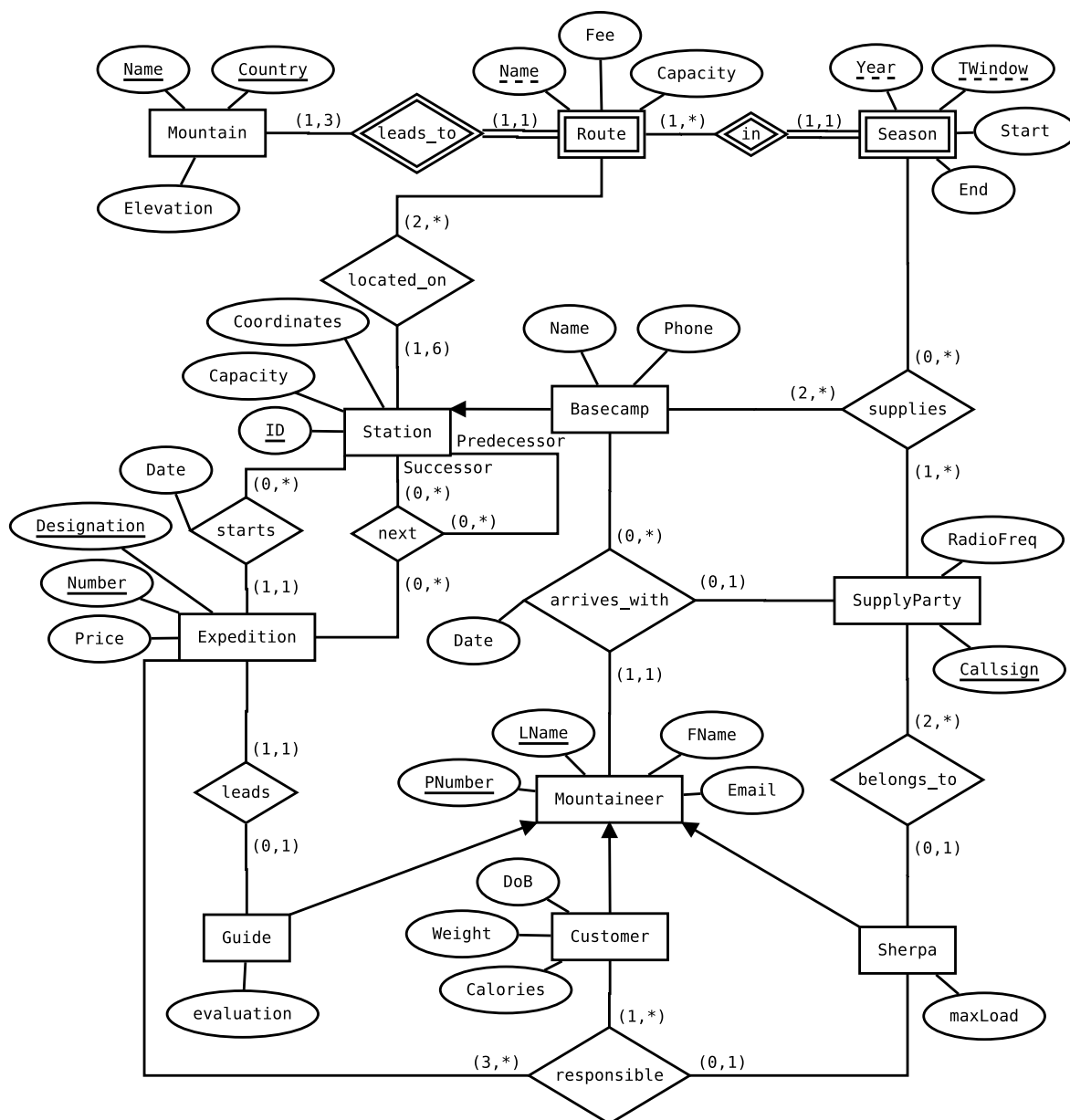


Figure 3: EER-diagram for exercise 3

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 L^AT_EX commands are listed as well.

Exercise 4 (Evaluation)

[0.4 points]

Consider the four relations below.

Drug		
name	manufacturer	substance
Hairloss	TU	Exam
Tiredness	Uni	Task
Success	TU	Knowledge

Box		
name	boxsize	price
Hairloss	2	19
Hairloss	5	20
Tiredness	1	5
Success	42	18,70

Branch-office	
address	bmanager
FH	H.A.S. Ended
GM	I.S. Done
Trei	R.E. Used
Fav	F. Inished
EI	C.O. Pied

instore			
address	name	boxsize	amount
Fav	Tiredness	1	5
GM	Hairloss	5	7
FH	Hairloss	2	13
Fav	Success	42	1
Trei	Tiredness	1	8

Provide the results of the following queries over these relations.

(a)

$$\pi_{\text{substance, bmanager}} \left(\pi_{\text{substance}}(\sigma_{\text{manufacturer}=\text{"Uni"}}(\text{Drug})) \times ((\sigma_{\text{boxsize}<2}(\text{Box}) \bowtie \sigma_{\text{amount}>6}(\text{instore})) \bowtie \text{branch-office}) \right)$$

(b)

$$\{f \mid f \in \text{branch-office} \wedge \forall l \in \text{instore} (l.\text{address} = f.\text{address} \rightarrow \neg(\exists m \in \text{Drug} (m.\text{name} = l.\text{name} \wedge m.\text{manufacturer} \neq \text{"TU"})))\}$$

Exercise 5 (Equivalences)

[2 points]

Consider the following pairs q_i, q_j of expressions in Relational Algebra over the relational schemas $R(\underline{A}BC)$, $S(B\underline{D}E)$ and $T(\underline{A}DF)$.

- Verify, whether the two expressions are equivalent (i.e. whether they produce the same result on all possible instances of the relational schemas). You can assume that no NULL-values may occur in any instance of the schemas.
 - 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 relational schemas 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) $q_1: (\sigma_{A=D}(R \bowtie S)) \bowtie T$ and
 $q_2: (R \bowtie_{R.A=S.D \wedge R.B=S.B} S) \bowtie \sigma_{A=D}(T)$
- (b) $q_3: \pi_{AD}(((\pi_B(R) - \pi_B(S)) \bowtie R) \bowtie S) \cap \pi_{AD}(T)$ and
 $q_4: \pi_{AD}(T) - \rho_{D \leftarrow B}(\pi_{AB}(R \bowtie S) \cup \rho_{B \leftarrow D}(\pi_{AD}(T)))$
- (c) $q_5: R \bowtie (\pi_{AF}(T) \cap \rho_{A \leftarrow E, F \leftarrow D}(\pi_{DE}(S)))$ and
 $q_6: (R \bowtie \pi_{AF}(T)) \cap (R \bowtie \rho_{A \leftarrow D, F \leftarrow E}(\pi_{DE}(S)))$
- (d) $q_7: \rho_{A \leftarrow Q.A, B \leftarrow Q.B}(\pi_{Q.A, Q.B}(\sigma_{\theta}(\rho_Q(\pi_A(R) \times \pi_B(R)) \times (R \times T))))$
where $\theta = (Q.A = R.A \wedge Q.B = R.B) \vee (Q.A = T.A \wedge Q.B = T.F)$ and
 $q_8: (\pi_{AB}(R) \cup \rho_{B \leftarrow F}(\pi_{AF}(T)))$

Exercise 6 (Answer Sizes)

[2 points]

Consider the relational schemas $R(\underline{A}B)$, $S(\underline{A}BC\underline{D})$, and $T(\underline{A}C\underline{E})$ 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) $q_1: (\pi_A(R) \cup \pi_A(T)) \cup (\pi_A(R) \cap \pi_A(T))$ (where $|R| = 4$ and $|T| = 7$)
- (b) $q_2: ((\sigma_{C=4 \wedge D=2}(S) \bowtie S) \bowtie T) \bowtie T$ (where $|S| = 5$ and $|T| = 7$)
- (c) $q_3: \rho_{F \leftarrow B}(R) \bowtie \sigma_{A=4 \wedge B=2}(S)$ (where $|R| = 3$ and $|S| = 2$)
- (d) $q_4: \pi_{AF}(\rho_{F \leftarrow A}(T) \times R) - \pi_{AF}(\sigma_{F > A}(\rho_{F \leftarrow A}(S) \bowtie R))$ (where $|R| = 2$, $|S| = 3$ and $|T| = 3$)

Exercise 7 (Primitive Operations)

[1 point]

- (a) Express the operator \bowtie using only the primitive operators of Relational Algebra. To achieve that, use two relations, R and S . Schema R contains the attributes $(R_1, \dots, R_r, G_1, \dots, G_g)$ and schema S contains the attributes $(S_1, \dots, S_s, G_1, \dots, G_g)$. Express $R \bowtie S$ using primitive operators only.
- Hint:* Your expression may also contain constant relations. Keep in mind that r , s and g are not constants. A constant relation is for instance: $\{(1)\}$.
- (b) Translate the query $R \bowtie S$ for the relations R and S with the schemas from exercise (a) into both, tuple calculus and domain calculus (“Domänenkalkül”).
- (c) Translate the following query over the relations R , S and T first into a query in Relational Algebra which uses primitive operators only, and then into the tuple- and domain calculus. R and S have the same schema, as in the previous exercise (a) and T has the schema $(T_1, \dots, T_t, G_1, \dots, G_g)$.

$$\pi_{T_1, \dots, T_t, G_1, \dots, G_g}(T \bowtie (\pi_{G_1, \dots, G_g}(R) - \pi_{G_1, \dots, G_g}(S)))$$

Exercise 8 (Formalizing Queries)

[3.6 points]

A cafe manages its sales data with the following schema (Primary keys are underlined, foreign keys are written in *italics*).

```

Barista      (FName, SName, BirthDat, Salary)
learned_from(TeacherFName: Barista.FName, TeacherSName: Barista.SName,
             StudentFName: Barista.FName, StudentSName: Barista.SName)
Customer     (CNr, Email)
Coffee       (Designation, Description, Price, Difficulty)
Purchase     (Date, CNr: Customer.CNr, Coffee: Coffee.Designation,
             FName: Barista.FName, SName: Barista.SName, Amount, Rating)
Beans        (Name, Country-of-origin, Aroma, Quality)
contains     (Coffee: Coffee.Designation,
             BeansN: Beans.Name, BeansC: Beans.Country-of-origin, Amount)

```

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

Express the queries described below in **Relational Algebra**, the **tuple calculus** (Tupelkalkül) and the **domain calculus** (Domänenkalkül).

Exercises marked with **, i.e. **Exercises (a), (b), (c) and (h)**, are *not mandatory*. You will not get any points for solving them. Exercises (a), (b) and (c) are simple exercises, in case you do not have any experience with the formulation of queries in Relational Algebra (or database queries in general) yet. Exercise (h) is more complex, in case you want another example to practice. You find the solutions for all non-mandatory exercises at the end of this document.

- (a) ** Find all Barista (FName, SName, BirthDat) who earn more than 2500.
- (b) ** Find all Barista (FName, SName, BirthDat) that have never received a 5-star rating (or better). Assume that the rating is stored as natural number.
- (c) ** For the customer with CNr 74656, find all countries from where the beans of her purchases between 01.10.2371 and 20.12.2371 originate. Assume that the relation **contains** only contains entries with values greater than 0 for the **Amount**-attribute, i.e. only entries for beans which are used in coffee are stored.
- (d) Find all Barista (list the attribute SName only) with a salary less than 1800, who have already received a 3-star rating (or better) for a coffee with the difficulty 5. Assume that both, the attribute **Rating** and **Difficulty** are stored as natural numbers.
- (e) Find the coffees (attribute **Designation** only) which contain beans from all countries stored in the database.
- (f) For all Barista find the countries (you can use countries that are stored in the database only) from which beans originate, that have never been used in a coffee sold by the Barista before. Provide a table with the columns (SName, Land), where “SName“ is the Surname of the Barista and “Land“ the name of the origin country.
- (g) Find the names of all beans that have been used in any coffee with difficulty 3 or higher, sold on 1.10.2019.
- (h) ** First, consider all teachers who have at least one student that earns more than 5000. Then for the solution find the list of ratings for all students of these teachers.

Sample Solutions for the additional exercises

- (a) ** Find all Barista (FName, SName, BirthDat) who earn more than 2500.

Relational Algebra:

$$\pi_{\text{FName, SName, BirthDat}}(\sigma_{\text{Salary} > 2500}(\text{Barista}))$$

Tuple Calculus:

$$\{[b.\text{FName}, b.\text{SName}, b.\text{BirthDat}] \mid b \in \text{Barista} \wedge b.\text{Salary} > 2500\}$$

Domain Calculus:

$$\{[fn, sn, bdat] \mid \exists s([fn, sn, bdat, s] \in \text{Barista} \wedge s > 2500)\}$$

- (b) ** Find all Barista (FName, SName, BirthDat) that have never received a 5-star rating (or better). Assume that the rating is stored as natural number.

Relational Algebra:

$$\pi_{\text{FName, SName, BirthDat}}(\text{Barista} \bowtie (\pi_{\text{FName, SName}}(\text{Barista}) - \pi_{\text{FName, SName}}(\sigma_{\text{Rating} \geq 5}(\text{Purchase}))))$$

Tuple Calculus:

$$\{[b.\text{FName}, b.\text{SName}, b.\text{BirthDat}] \mid b \in \text{Barista} \wedge \neg(\exists k \in \text{Purchase}(k.\text{SName} = b.\text{FName} \wedge k.\text{SName} = b.\text{SName} \wedge k.\text{Rating} \geq 5))\}$$

Domain Calculus:

$$\{[fn, sn, bdat] \mid \exists s([fn, sn, bdat, s] \in \text{Barista} \wedge \neg(\exists d, cnr, cof, a, r([d, cnr, cof, fn, sn, bdat, a, r] \in \text{Purchase} \wedge r \geq 5)))\}$$

- (c) ** For the customer with CNr 74656, find all countries from where the beans of her purchases between 01.10.2371 and 20.12.2371 originate. Assume that the relation
- contains**
- only contains entries with values greater than 0 for the
- Amount**
- attribute, i.e. only entries for beans which are used in coffee are stored.

Relational Algebra:

$$\pi_{\text{BeansC}}(\sigma_{\text{Date} \geq 01.10.2371 \wedge \text{Date} \leq 20.12.2371 \wedge \text{CNr} = 74656}(\rho_{\text{Numbers} \leftarrow \text{Amount}}(\text{Purchase})) \bowtie \text{contains})$$

Tuple Calculus:

$$\{[c.\text{BeansC}] \mid c \in \text{contains} \wedge \exists p \in \text{Purchase}(p.\text{Coffee} = c.\text{Coffee} \wedge \\ p.\text{CNr} = 74656 \wedge \\ p.\text{date} \geq 01.10.2371 \wedge \\ p.\text{date} \leq 20.12.2371)\}$$

Domain Calculus:

$$\{[o] \mid \exists c, b, a, d, fn, sn, a2, r([c, b, o, a] \in \text{contains} \wedge \\ [d, 74656, c, fn, sn, a2, r] \in \text{Purchase} \wedge \\ d \geq 01.10.2371 \wedge \\ d \leq 20.12.2371)\}$$

- (d) ** First, consider all teachers who have at least one student that earns more than 5000. Then for the solution find the list of ratings for all students of these teachers.

Relational Algebra:

$$\pi_{\text{Purchase.FName, Rating}}(\sigma_{\phi}(\sigma_{\text{Salary} \geq 5000}(\text{Barista}) \times \rho_{s1}(\text{learned_from}) \times \rho_{s2}(\text{learned_from}) \times \text{Purchase}))$$

where ϕ is the following condition:

$$\phi = \text{s1.StudentFName} = \text{FName} \wedge \text{s1.StudentSName} = \text{SName} \wedge \\ \text{s1.TeacherFName} = \text{s2.TeacherFName} \wedge \text{s1.TeacherSName} = \text{s2.TeacherSName} \wedge \\ \text{s2.StudentFName} = \text{Purchase.FName} \wedge \text{s2.StudentSName} = \text{Purchase.SName}$$

Tuple Calculus:

$$\{[s.\text{FName}, k.\text{Rating}] \mid s \in \text{learned_from} \wedge p \in \text{Purchase} \wedge \\ s.\text{StudentFN} = p.\text{FName} \wedge s.\text{StudentSName} = p.\text{SName} \wedge \\ \exists l \in \text{learned_from}(\\ l.\text{TeacherFName} = s.\text{TeacherFN} \wedge l.\text{TeacherSN} = s.\text{TeacherSN} \wedge \\ \exists b \in \text{Barista}(\\ b.\text{FN} = l.\text{StudentFN} \wedge \\ b.\text{SN} = l.\text{StudentSN} \wedge \\ b.\text{Salary} \geq 5000))\}$$

Domain Calculus:

$$\{[svn, rat] \mid \exists bbn, bsn, bdat, sal([bfn, bsn, bdat, sal] \in \text{Barista} \wedge sal \geq 5000 \wedge \\ \exists lfn, lsn([lfn, lsn, bfn, bsn] \in \text{learned_from} \wedge \\ \exists ssn([lfn, lsn, sfn, ssn] \in \text{learned_from} \wedge \\ \exists dat, cnr, cof, amo(\\ [dat, cnr, cof, sfn, ssn, amo, rat] \in \text{Purchase}))))\}$$