



Übung zur Vorlesung *Einsatz und Realisierung von Datenbanksystemen* im SoSe18

Alexander van Renen, Maximilian E. Schüle (i3erdb@in.tum.de)
<http://db.in.tum.de/teaching/ss18/impldb/>

Blatt Nr. 04

Hinweise Dieses Übungsblatt ist für die nächsten zwei Wochen gültig.

Die Datalogaufgaben können auf <http://datalog.db.in.tum.de/> getestet werden. Auf der Seite können Sie dann unter *examples* einen entsprechenden Datensatz laden. Am Ende der EDB könnt ihr dann neue IDB Regeln definieren und diese dann in dem Query Eingabefeld abfragen. Zusätzlich zu dem in der Vorlesung vorgestellten Syntax hier noch eine kurz Übersicht der Vergleichsoperatoren: $X < Y$, $Y > X$ (kleiner, größer), $X = < Y$, $X > = Y$ (kleiner gleich, größer gleich), $X = Y$, $X \setminus = Y$ (gleich, ungleich), $not(pred(X, Y))$ (existiert nicht $pred(X, Y)$).

Hausaufgabe 1

Gegeben sei die nachfolgende *KindEltern*-Ausprägung für den Stammbaum-Ausschnitt der griechischen Götter und Helden:

KindEltern		
Vater	Mutter	Kind
Zeus	Leto	Apollon
Zeus	Leto	Artemis
Kronos	Rheia	Hades
Zeus	Maia	Hermes
Koios	Phoebe	Leto
Atlas	Pleione	Maia
Kronos	Rheia	Poseidon
Kronos	Rheia	Zeus

Formulieren Sie folgende Anfragen in Datalog und testen Sie unter (<http://datalog.db.in.tum.de/>):

- Bestimmen Sie alle Geschwisterpaare.
- Ermitteln Sie Paare von Cousins und Cousinen beliebigen Grades. Die Definition finden Sie auf Wikipedia.
- Geben Sie alle Verwandtschaftspaare an. Überlegen Sie sich eine geeignete Definition von Verwandtschaft und setzen Sie diese in Datalog um.
- Bestimmen Sie alle Nachfahren von Kronos. Formulieren Sie die Anfrage auch in SQL, so dass sie unter PostgreSQL ausführbar ist (online testen unter: <http://sqlfiddle.com> mit der Datenbank PostgreSQL statt MySQL, das Schema Textfeld können sie leer lassen, müssen aber trotzdem auf 'Build Schema' drücken). Sie können die Daten als Common Table Expression definieren und dann nutzen:

```

WITH RECURSIVE
kindEltern(vater,mutter,kind) as (
  VALUES
    ('Zeus', 'Leto', 'Apollon'),
    ('Zeus', 'Leto', 'Artemis'),
    ('Kronos', 'Rheia', 'Hades'),
    ('Zeus', 'Maia', 'Hermes'),
    ('Koios', 'Phoebe', 'Leto'),
    ('Atlas', 'Pleione', 'Maia'),
    ('Kronos', 'Rheia', 'Poseidon'),
    ('Kronos', 'Rheia', 'Zeus')
),
parent(eltern,kind) as (
  select vater, kind from kindEltern UNION
  select mutter, kind from kindEltern
)
select * from parent where eltern='Zeus'

```

Hausaufgabe 2

Bleiben wir bei dem bekannten Universitätsschema:

```

Assistenten(PersNr, Name, Fachgebiet, Boss)
hoeren(MatrnNr, VorlNr)
pruefen(MatrnNr, VorlNr, PersNr, Note)
Vorlesungen(VorlNr, Titel, SWS, gelesenVon)
Professoren(PersNr, Name, Rang, Raum)
voraussetzen(Vorg, Nachf)
Studenten(MatrnNr, Name, Semester)

```

Formulieren Sie folgende Anfragen in Datalog und testen Sie sie:

- Geben Sie alle *Professoren* an, die mindestens eine Prüfung abgehalten haben.
- Übersetzen Sie folgenden Ausdruck des Domänenkalküls in Datalog. Machen Sie sich der Bedeutung des Ausdrucks bewusst.

$$\{[t] \mid \exists v,s,g([v,t,s,g] \in \text{Vorlesungen} \wedge \exists v2([v,v2] \in \text{voraussetzen} \wedge \exists s2,g2([v2,\text{Wissenschaftstheorie},s2,g2] \in \text{Vorlesungen})))\}$$

- Joinen Sie nachfolgende Datalog-Anfrage so, dass Titel ausgegeben werden. Was bedeutet diese Anfrage?

```

geschwisterVL(N1,N2):-voraussetzen(V,N1),voraussetzen(V,N2), N1<N2.
nahverwandtVL(N1,N2):-geschwisterVL(N1,N2).
nahverwandtVL(N1,N2):-geschwisterVL(M1,M2),voraussetzen(M1,N1),
voraussetzen(M2,N2).

```

Hausaufgabe 3

Geben Sie Datalog Regeln an, die Studenten (Namen angeben) finden, die von einem Prüfer geprüft worden, der selbst nicht die geprüfte Vorlesung gehalten hat. Das korrekte Ergebnis für diese Anfrage ist *Russels Prüfling, Carnap*. Führen Sie die Anfrage im Datalog Tool aus!

Hausaufgabe 4

Definieren Sie das Prädikat $\text{sg}(X, Y)$ das für “same generation” steht. Zwei Personen gehören zur selben Generation, wenn Sie mindestens je ein Elternteil haben, das derselben Generation angehört.

Verwenden Sie beispielsweise die folgende Ausprägung einer ElternKind Relation. Das erste Element ist hier das Kind, das Zweite ein Elternteil.

```
parent(c, a).
parent(d, a).
parent(d, b).
parent(e, b).
parent(f, c).
parent(g, c).
parent(h, d).
parent(i, d).
parent(i, e).
parent(f, e).
parent(j, f).
parent(j, h).
parent(k, g).
parent(k, i).
```

- Definieren Sie das Prädikat in Datalog.
- Demonstrieren Sie die naive Ausführung des Prädikats.
- Erläutern Sie das Vorgehen bei der seminaiven Auswertung.

Gruppenaufgabe 5

Ist folgendes Datalog-Programm stratifiziert?

$$\begin{aligned} p(X, Y) & :- q_1(Y, Z), \neg q_2(Z, X), q_3(X, P). \\ q_2(Z, X) & :- q_4(Z, Y), q_3(Y, X). \\ q_4(Z, Y) & :- p(Z, X), q_3(X, Y). \end{aligned}$$

Ist das Programm sicher – unter der Annahme, dass p, q_1, q_2, q_3, q_4 IDB- oder EDB-Prädikate sind?

Hausaufgabe 6

Gehen Sie von folgender kombinierter Fragmentierung der in Abbildung 1 dargestellten Relation *Professoren* aus:

- Zuerst erfolgt eine vertikale Fragmentierung in

$$\begin{aligned} \text{ProfVerw} & := \Pi_{\text{PersNr, Name, Gehalt, Steuerklasse}}(\text{Professoren}) \\ \text{Profs} & := \Pi_{\text{PersNr, Name, Rang, Raum, Fakultät}}(\text{Professoren}) \end{aligned}$$

- Das Fragment Profs wird weiter horizontal fragmentiert in

$$\begin{aligned} \text{TheolProfs} & := \sigma_{\text{Fakultät} = \text{'Theologie'}}(\text{Profs}) \\ \text{PhysikProfs} & := \sigma_{\text{Fakultät} = \text{'Physik'}}(\text{Profs}) \\ \text{PhiloProfs} & := \sigma_{\text{Fakultät} = \text{'Philosophie'}}(\text{Profs}) \end{aligned}$$

Professoren						
PersNr	Name	Rang	Raum	Fakultät	Gehalt	Steuerklasse
2125	Sokrates	C4	226	Philosophie	85000	1
2126	Russel	C4	232	Philosophie	80000	3
2127	Kopernikus	C3	310	Physik	65000	5
2133	Popper	C3	52	Philosophie	68000	1
2134	Augustinus	C3	309	Theologie	55000	5
2136	Curie	C4	36	Physik	95000	3
2137	Kant	C4	7	Philosophie	98000	1

Abbildung 1: Beispielausprägung der um drei Attribute erweiterten Relation *Professoren*

Übersetzen Sie aufbauend auf dieser Fragmentierung die folgende SQL-Anfrage in die kanonische Form.

```
select Name, Gehalt, Rang
from Professoren
where Gehalt > 80000;
```

Optimieren Sie diesen kanonischen Auswertungsplan durch Anwendung algebraischer Transformationsregeln (Äquivalenzen).

Hausaufgabe 7

Für die Rekonstruierbarkeit der Originalrelation R aus vertikalen Fragmenten R_1, \dots, R_n reicht es eigentlich, wenn Fragmente paarweise einen Schlüsselkandidaten enthalten. Illustrieren Sie, warum es also nicht notwendig ist, dass der Durchschnitt aller Fragmentschemata einen Schlüsselkandidaten enthält. Es muss also nicht unbedingt gelten

$$R_1 \cap \dots \cap R_n \supseteq \kappa,$$

wobei κ ein Schlüsselkandidat aus R ist.

Geben Sie ein anschauliches Beispiel hierfür – am besten bezogen auf unsere Beispielrelation *Professoren*.

Hausaufgabe (wird nicht in der Übung besprochen)

Beweisen Sie für die Beispielrelationen $R : \{[A, B, C]\}$ und $S : \{[C, D, E]\}$ folgende Eigenschaften der Join- / Semi-Join-Operatoren:

$$\begin{aligned} R \bowtie S &= R \bowtie (\Pi_C(R) \bowtie S) \\ R \bowtie S &= (\Pi_C(S) \bowtie R) \bowtie (\Pi_C(R) \bowtie S) \end{aligned}$$

Hausaufgabe (wird nicht in der Übung besprochen)

Gegeben das folgende Schema der EDB¹:

```
Product(maker, model, type).
PC(model, speed, ram, hd, price).
Laptop(model, speed, ram, hd, screen, price).
Printer(model, color, type, price).
```

¹Inspiziert von http://people.inf.elte.hu/sila/DB1English/exercise06_products.pdf.

Beantworten Sie in Datalog und testen Sie unter (<http://datalog.db.in.tum.de/>):

- a) What PC models have a speed of at least 3.00 GHz?
- b) Which manufacturers make laptops with a hard disk (hd) of at least 100 GB?
- c) Find the model number and price of products (of any type) made by manufacturer B.
- d) Find the model numbers of all color laser printers.
- e) Find those manufacturers that sell Laptops, but not PC's.
- f) Find those hard-disk sizes that occur in two or more PC's.
- g) Find those pairs of PC models that have both the same cpu speed and RAM. A pair should be listed only once, e.g., list (i,j) but not (j,i).

Hausaufgabe (wird nicht in der Übung besprochen)

Schreiben Sie zu dem Ubahn-Netz Beispiel auf der Datalog Seite (unter Examples) folgende Anfragen in Datalog:

1. Erstellen Sie den Stationsplan für den U-Bahnhof Fröttmanning, der alle Station, die ohne umsteigen erreichbar sind, auflistet.
2. Erstellen Sie für Garching-Forschungszentrum einen Plan, der alle erreichbaren Stationen, die minimale Anzahl an Umstiegen und Stops auflistet. Beschreiben Sie Ihren Ansatz ausführlich.

Hausaufgabe (wird nicht in der Übung besprochen)

Nun fügen wir der EDB folgende Einträge hinzu²:

```
ModelParts(model,partname)
Part(partname,marker)
ConsistsOf(partname,partname)
```

part ist hierbei ein Bauteil eines Geräts, **marker** ist der Hersteller des Bauteils. **ModelParts** verbindet ein Modell aus den ursprünglichen Daten mit seinem/seinen Bauteilen. **ConsistsOf** beschreibt die Hierarchische Beziehung zwischen Bauteilen.

'kompaktes' Beispiel:

```
ModelParts(workstation,mainboard-hl7).
ModelParts(workstation,hdd30g).
Part(mainboard-hl7,asuz).
Part(gpu7700,nvidio).
Part(hdd30g,sealgate).
Part(transistor,foxcom).
Part(motor,enginesUnited).
Part(wire,theWireCompany).
Part(magnet,theMagnetCompany).
ConsistsOf(hdd30g,transistor).
```

²Inspiziert von http://people.inf.elte.hu/sila/DB1English/exercise06_products.pdf.

```
ConsistsOf(hdd30g,motor).
ConsistsOf(motor,wire).
ConsistsOf(motor,magnet).
...
```

Beantworten Sie in Datalog:

- l) Find all models containing parts made by **sealgate**.
- m) Find all models which contain two different parts by the same maker (regardless of where in the hierarchy).

Hausaufgabe (wird nicht in der Übung besprochen)

Die Produktdaten einer Firma werden in einer deduktiven Datenbank mit folgenden Relationenschema gehalten:

- Bauteil(**Bauteiltyp**, Gewicht, KonstrukteurID)
- besteht_aus(**Bauteil**, **Komponente**, Menge)
- Konstrukteur(**KonstrukteurID**,Name,Geburtsdatum)

Die Relation *Bauteil* beschreibt das Gewicht eines Bauteiltyps und gibt den Konstrukteur an, der diesen Bauteiltyp entworfen hat. Die Relation *besteht* gibt an, aus welchen und jeweils wievielen Einzelkomponenten ein Bauteil besteht. In *Konstrukteur* sind die persönlichen Daten zu den Konstrukteuren gespeichert.

Formulieren Sie die folgenden Anfragen in Datalog:

- a) Geben Sie alle Bauteile an, aus denen ein Fahrgestell besteht.
- b) Geben Sie alle Bauteile an, an denen der Konstrukteur Schmidt direkt oder indirekt (er hat eine Komponente davon entworfen) beteiligt ist.