3. Übung zur Vorlesung "Datenbanken" im Sommersemester 2007

– mit Musterlösungen –

Prof. Dr. Gerd Stumme, Dipl.-Inform. Christoph Schmitz

14. Mai 2007

Hinweis: Wir schlagen vor, die Aufgaben in der Reihenfolge 1, 3, 4, 2 zu bearbeiten.

Aufgabe 1

Man betrachte die folgenden drei Relationen:

Klasse	
KID	LID
11a	WAG
10b	AMR

Lehrer	
LID	LName
WAG	Wagner
AMR	Meier A.

Schüler		
SID	SName	KID
8501	Gabi Müller	11a
8601	Rolf Peters	11a
8701	Rolf Peters	10b
8702	Inge Lang	10b

1. Geben Sie zu den folgenden Aufgaben Ausdrücke in relationaler Algebra für folgende Anfragen an und nennen Sie die Ergebnismengen:

   - Gesucht sind alle Schüler der Klasse 11a.

     $$\sigma_{KID=\text{‘11a’}}(\text{Schüler})$$

   - Gesucht sind alle Namen der Schüler, die die Klasse 11a besuchen.

     $$\pi_{SName}(\sigma_{KID=\text{‘11a’}}(\text{Schüler}))$$

   - Gesucht sind alle Schüler mit Geburtsjahr 1985. (Ihre SID habe die Form 85xx.)

     $$\sigma_{8500 \leq SID \wedge SID < 8600}(\text{Schüler}))$$

2. Bilden Sie das Kreuzprodukt (Lehrer $\times$ Klasse) $\times$ Schüler. Markieren Sie die vermutlich relevanten Tupel. Warum sind diese relevant?

	Lehrer. LID	Lehrer. LName	Klasse. KID	Klasse. LID	Schueler. SID	Schueler. SName	Schueler. KID
✓	WAG	Wagner	11a	WAG	8501	Gabi Müller	11a
	WAG	Wagner	10b	AMR	8501	Gabi Müller	11a
	AMR	Meier A.	11a	WAG	8501	Gabi Müller	11a
	AMR	Meier A.	10b	AMR	8501	Gabi Müller	11a
✓	WAG	Wagner	11a	WAG	8601	Rolf Peters	11a
	WAG	Wagner	10b	AMR	8601	Rolf Peters	11a
	AMR	Meier A.	11a	WAG	8601	Rolf Peters	11a
	AMR	Meier A.	10b	AMR	8601	Rolf Peters	11a
	WAG	Wagner	11a	WAG	8701	Rolf Peters	10b
	WAG	Wagner	10b	AMR	8701	Rolf Peters	10b
	AMR	Meier A.	11a	WAG	8701	Rolf Peters	10b
✓	AMR	Meier A.	10b	AMR	8701	Rolf Peters	10b
	WAG	Wagner	11a	WAG	8702	Inge Lang	10b
	WAG	Wagner	10b	AMR	8702	Inge Lang	10b
	AMR	Meier A.	11a	WAG	8702	Inge Lang	10b
✓	AMR	Meier A.	10b	AMR	8702	Inge Lang	10b

Interessant sind diejenigen Tupel, wo die Attribute gleichen Namens (Lehrer.LID und Klasse.LID usw.) die gleichen Werte haben → natürlicher Join.

3. Wie lauten die jeweiligen Ergebnismengen natürlicher Joins von Lehrer mit Klasse und Klasse mit Schüler?

LID	LName	KID
WAG	Wagner	11a
AMR	Meier A.	10b

KID	LID	SID	SName
11a	WAG	8501	Gabi Müller
11a	WAG	8601	Rolf Peters
10b	AMR	8701	Rolf Peters
10b	AMR	8702	Inge Lang

4. Geben Sie einen Ausdruck an, der alle Schüler ausgibt, deren Klassenlehrerin Frau Wagner ist, und berechnen Sie das Ergebnis.

$$\pi_{SID,SName,KID}(\sigma_{LName=\text{`Wagner'}}(\text{Lehrer} \bowtie \text{Klasse} \bowtie \text{Schüler}))$$

5. Ein Schüler, von dem nur der Name Rolf Peters bekannt ist, berichtet, sein Lehrer habe zuhause einen PC mit einer relationalen Datenbank. Gesucht sind die Namen der in Frage kommenden Lehrer. Mit welchem Ausdruck können diese Namen berechnet werden?

$$\pi_{LName}(\sigma_{SName=\text{`Rolf Peters'}}(\text{Lehrer} \bowtie \text{Klasse} \bowtie \text{Schüler}))$$

6. Wie lautet die Ausgabe eines Equi-Joins der Schüler-Relation mit sich selbst über das SName-Attribut? Welchen Sinn könnte eine solche Tabelle haben?

S1.SID	S1.KID	SName	S2.SID	S2.KID
8501	11a	Gabi Müller	8501	11a
8601	11a	Rolf Peters	8601	11a
8601	11a	Rolf Peters	8701	10b
8701	10b	Rolf Peters	8601	11a
8701	10b	Rolf Peters	8701	10b
8702	10b	Inge Lang	8702	10b

Sinn: Erkennen von möglichen Namensverwechslungen. So gibt es z. B. einen Rolf Peters in der 11a, sowie einen weiteren Rolf Peters in der 10b.

Praktischer wäre, noch so zu selektieren, daß nur die Tupel mit S1.SID < S2.SID gewählt werden:

$$\sigma_{S1.SID<S2.SID}(\rho_{S1}(\text{Schüler}) \bowtie_{S1.SName=S2.SName} \rho_{S2}(\text{Schüler}))$$

S1.SID	S2.KID	SName	S2.SID	S2.KID
8601	11a	Rolf Peters	8701	10b

7. Der neu eingestellte Referendar ('MUE', 'Müller') wird in die Lehrer-Relation eingefügt, hat aber noch keine Klasse. Wie sieht das Ergebnis eines Outer-Joins von Lehrer mit Klasse aus? Welchen Vorteil hat hier der Outer-Join?

LID	LName	KID
WAG	Wagner	11a
AMR	Meier A.	10b
MUE	Müller	-

Vorteil: Abfragen aller verfügbaren Attribute der Relationen, die durch den Join kombiniert werden, ohne dabei Tupel auszublenden, für die nicht alle Attribute definiert sind.

Aufgabe 2

Gegeben seien folgende Relationen:

- Student(Matr#, Name) – Studenten und ihre Namen
- Professor(Prof#, Name, Rang) – Professoren mit Name und Besoldungsgruppe (z. B. 'C4')
- Vorlesung(Vorl#, Titel, Prof#) – Vorlesungen und ihre Titel sowie Nummer des Dozenten
- Hören(Matr#, Vorl#) – Welcher Student hört welche Vorlesungen (n:m)
- Voraussetzung(Vorl#, VoraussetzungVorl#) – Welche anderen Vorlesungen setzt eine Vorlesung voraus

1. Geben Sie einen Ausdruck in relationaler Algebra an, der die Studenten (MatrNr) ausgibt zusammen mit den Nummern der Voraussetzungen der Vorlesungen, die der jeweilige Student gehört hat.

   $\pi_{Matr\#,VoraussetzungVorl\#}(\text{Hören} \bowtie Voraussetzung)$

2. Geben Sie einen Ausdruck an, der die Studenten ausgibt, die eine Vorlesung hören, deren (direkte) Voraussetzung sie nicht gehört haben.

$$Student \bowtie \pi_{Matr\#}(\rho_{Vorl\#\leftarrow VoraussetzungVorl\#}$$
$$(\pi_{Matr\#,VoraussetzungVorl\#}(\text{Hören} \bowtie Voraussetzung)) - \text{Hören})$$

   Erläuterung:

   a) $\text{Hören} \bowtie Voraussetzung$: Verknüpfe Hören mit Voraussetzungen; das Ergebnis enthält u.a. die Matr# und die Vorl# der vorausgesetzten Vorlesung
   b) $\pi_{Matr\#,VoraussetzungVorl\#}(\ldots)$: Projiziere das Ergebnis auf Matr# und Vorl# der vorausgesetzten Vorlesungen
   c) $\rho_{Vorl\#\leftarrow VoraussetzungVorl\#}(\ldots)$: Benenne die Nummer der vorausgesetzten Vorlesungen um in Vorl#
   d) $(\ldots - \text{Hören})$: subtrahiere von den zu hörenden Vorlesungen die bereits gehörten
   e) $Student \bowtie \pi_{Matr\#}(\ldots)$: Projektion auf Matr# und Join mit Student, um auch den Namen zu erhalten

Aufgabe 3

1. Zeigen Sie, daß die folgenden Operationen die Mächtigkeit der relationalen Algebra nicht erhöhen, d. h. daß man sie mit den jeweils anderen Operationen der Algebra nachbauen kann!

   a) Semijoin: $R \ltimes S$ Der Semijoin ist die Projektion des Joins auf eine der beiden Relationen: $R \ltimes S = \pi_R(R \bowtie S)$ und $R \rtimes S = \pi_S(R \bowtie S)$.
   b) Schnitt: $R \cap S$; stellen Sie den Schnitt ohne Benutzung der Mengendifferenz $R - S$ (wie in der Vorlesung gezeigt) dar!

      $R - S = \pi_R(R \bowtie S)$

      Es reicht aus, einen Self-Join der beiden Relationen zu machen. Da sie für die Differenz das gleiche Schema haben müssen, bedeutet das, daß nur diejenigen Tupel im Ergebnis sind, die in beiden Relationen R und S vorkommen. Dies projiziert auf R liefert das Gewünschte.

2. In der Vorlesung wurde folgender Ausdruck für die relationale Division angegeben, wobei das Schema von S in dem von R enthalten ist:

$$R \div S = \pi_{(R-S)}(R) - \pi_{(R-S)}((\pi_{(R-S)}(R) \times S) - R)$$

Erklären Sie, was die einzelnen Terme dieses Ausdrucks bedeuten und wieso er die relationale Division darstellt!

Wir nennen im Folgenden die Attribute in $R - S$ den R-Teil und den Rest den S-Teil in den Tupeln von R.

Der Ausdruck $\Pi_{(R-S)}(R) \times S$ entfernt den S-Teil der Tupel aus R und bildet das Kreuzprodukt davon mit allen Tupeln aus S. Es wird also eine Relation mit dem Schema von R gebildet, die aber zu jedem Tupel in R Tupel mit allen S-Teilen aus S enthält.

Davon wird R subtrahiert. Es bleiben also genau diejenigen Tupel übrig, deren S-Attribute nicht mit diesem R-Teil in R vorkommen.

Deren R-Teile von den R-Teilen des ursprünglichen R subtrahiert liefern das Gewünschte, nämlich die Tupel des R-Teils von R, die mit allen Tupeln aus S im S-Teil vorkommen.

Aufgabe 4

1. Formulieren Sie die Anfragen der Teile 1 und 4 von Aufgabe 1 im relationalen Tupelkalkül.

   Schüler der Klasse 11a:

   $$\{s \mid s \in \text{Schüler} \land s.KID = \text{'11a'}\}$$

   Namen der Schüler der Klasse 11a:

   Mit den auf den Folien zur Vorlesung vorgestellten Sprachmitteln ist diese Anfrage nicht anders als die vorige zu formulieren.
   Es gibt allerdings die Möglichkeit, mittels $[s.A_1, \ldots, t.A_k]$ neue Tupel aufzubauen, die Attribute (hier $A_1, \ldots, A_k$) von Tupelvariablen (hier: s, t) kombinieren:

   $$\{[s.\text{Name}] \mid s \in \text{Schüler} \land s.KID = \text{“11a”}\}$$

   Schüler des Geburtsjahres 1985:

   $$\{s \mid s \in \text{Schüler} \land s.SID >= 8500 \land s.SID < 8600\}$$

   Schüler von Frau Wagner:

   $$\{s \mid s \in \text{Schüler} \land \exists k \in Klasse(s.KID = k.KID \land \exists l \in Lehrer(l.KID = k.KID \land l.LName = \text{“Wagner”}))$$

2. Welcher der folgenden Ausdrücke ist sicher, welcher nicht? Was berechnen diese Ausdrücke?

a)

$$\{s \mid s \in Studenten \wedge \neg(\exists h \in hören(h.MatrNr = s.MatrNr))\}$$

Berechnet die Studenten, die Vorlesungen hören.

Sicher, weil das Ergebnis durch $s \in Studenten$ beschränkt wird. Berechnet diejenigen Studenten, die keine Vorlesungen hören.

b)

$$\{s \mid \neg(s \in Studenten \wedge \neg(\exists h \in hören(h.MatrNr = s.MatrNr))\}$$

Berechnet alle Tupel, die nicht ein Tupel der Relation *Student* sind, so daß dieser Student keine Vorlesung hört; gemeint ist also: alle Studenten, die eine Vorlesung hören.

Nicht sicher, weil das Ergebnis nicht auf die Domäne der Anfrage beschränkt wird.

Korrektur: der ursprüngliche Ausdruck

$$\{s \mid \neg(\exists h \in hören(h.MatrNr = s.MatrNr))\}$$

ist syntaktisch nicht korrekt, da s nicht an eine Relation gebunden wird.