

Datenbanken
Wintersemester 2020/21
 Prof. Dr. W. May

2. Übungsblatt: Algebra

Besprechung voraussichtlich am 9./10./16./17.12.2020

Aufgabe 1 (Relationale Anfragen an Mondial: Bedingungen) Geben Sie Ausdrücke der relationalen Algebra für die folgenden Anfragen an die Mondial-Datenbank an:

- a) Die Namen aller Städte, die mehr als 1.000.000 Einwohner haben.
- b) Die Namen aller Städte, die mehr Einwohner als Neuseeland haben.
- c) Die Namen aller Städte, in denen mehr als 25% der Bevölkerung des jeweiligen Landes leben.

Für spätere Übungsblätter:

- Geben Sie dieselben Anfragen in SQL an.

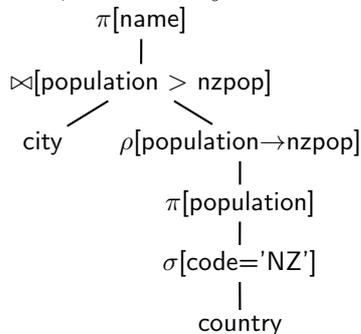
a) Einfache Selektion:

$$\pi[\text{name}](\sigma[\text{population} > 1000000](\text{city}))$$

b) Die Bevölkerungszahl von Neuseeland bekommt man als

$$\pi[\text{population}](\sigma[\text{code}='NZ'](\text{country})) .$$

In der Selektionsbedingung in Teil (1) ist aber nur eine Konstante erlaubt. Man benötigt ein Join, in dem man jedes Land mit dem Ergebnis der Subquery joint und den Vergleich durchführt:



Anstatt des “vollen” Join kann man auch ein Semijoin nehmen, da das Ergebnis des rechten Teilbaumes nur zum Vergleich verwendet wird, und keine Werte daraus weiter oben noch gebraucht werden.

Hinweis: in SQL kann man die Subquery in die WHERE-Klausel einbauen:

```

SELECT name
FROM city
WHERE population > (SELECT population FROM country WHERE code='NZ')
```

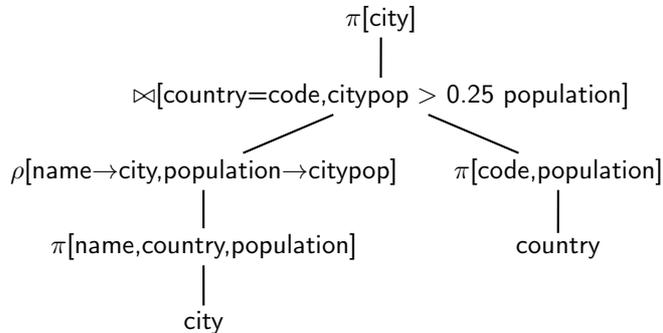
oder als Join formulieren:

```

SELECT name
FROM city, (SELECT population as nzpop FROM country WHERE code='NZ')
WHERE population > nzpop

```

c) Theta-Join (Bedingung: keine gleichnamigen Attribute!) mit Zusatzbedingung:



Aufgabe 2 (Äquivalenz von Ausdrücken) Gegeben seien folgende Relationen:

- $R(A,B,C)$
- $S(A,E,F)$
- $T(A,H)$

Die Wertebereiche aller nicht namensgleichen Attribute seien voneinander verschieden. Gegeben sei nun folgender relationaler Ausdruck:

$$\pi[E, H](\sigma[B = 10]((R \bowtie T) \bowtie S))$$

Sind die folgenden Ausdrücke äquivalent zu obigem Ausdruck? Begründen Sie Ihre Antwort.

- $\pi[E, H](\sigma[B = 10](R)) \bowtie (\pi[A, E](S) \bowtie T)$
- $\pi[E, H](\sigma[B = 10](((\pi[B](R)) \bowtie (\pi[A, E](S))) \bowtie (\pi[A, H](T))))$
- $\pi[E, H](\pi[A, B](\sigma[B = 10](R)) \bowtie ((\pi[A](S)) \bowtie T))$

a) ist äquivalent:

Der ursprüngliche Ausdruck:

$$\pi[E, H](\sigma[B = 10]((R \bowtie T) \bowtie S))$$

Join ist assoziativ:

$$\pi[E, H](\sigma[B = 10](R \bowtie (T \bowtie S)))$$

Attribut B existiert nur in Relation R , daher kann Selektion vor Join ausgeführt werden:

$$\pi[E, H](\sigma[B = 10]R \bowtie (T \bowtie S))$$

Von S werden nur die Attribute A (im Join mit T) und E (in der abschließenden Projektion) benötigt, man kann die Projektion also auch gleich auf S ausführen:

$$\pi[E, H](\sigma[B = 10]R \bowtie (T \bowtie (\pi[A, E]S)))$$

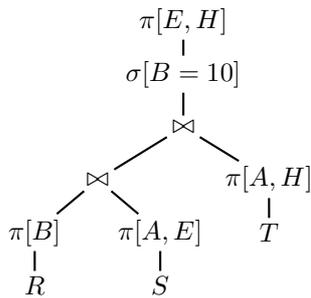
Join ist auch kommutativ:

$$\pi[E, H](\sigma[B = 10]R \bowtie ((\pi[A, E]S) \bowtie T))$$

Der letzte Ausdruck zeigt dass die Äquivalenz gilt.

Hinweis: in diesem Ausdruck werden alle π und σ so früh wie möglich ausgeführt.

b) nicht äquivalent:



$\pi[B](R)$ eliminiert das für den Join mit S und T wichtige Attribut A , hier wird stattdessen das Kreuzprodukt ausgeführt.

$\pi[A, E](S)$ ist wie in (1) gezeigt zulässig – F wird nicht gebraucht.

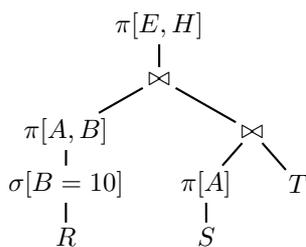
$\pi[A, H](T)$ ist nutzlos (Identität), stört aber keinen.

Der Ausdruck ist syntaktisch in Ordnung, hat aber ein anderes Ergebnis (eine Obermenge, da ein Join-Kriterium wegfällt).

Dazu gibt man dann ein (kleines) Beispiel an, z.B. einen Datenbankzustand \mathcal{S} mit $\mathcal{S}(R) = \{(a_0, b, c)\}$, $\mathcal{S}(S) = \{(a_1, e, f)\}$, $\mathcal{S}(T) = \{(a_1, h)\}$.

Die Auswertung des ersten Ausdrucks, $Q_0(\mathcal{S}) = \emptyset$, während die Auswertung des Ausdrucks in (a), $Q_a(\mathcal{S}) = \{(\perp, \langle \rangle)\}$ ergibt.

c) nicht äquivalent:



$\pi[A](S)$ eliminiert das Attribut E , welches in der Ausgabe enthalten sein soll. Der Ausdruck ist damit syntaktisch nicht zulässig.

Aufgabe 3 (Relationale Anfragen an Mondial: Schweizer Sprachen) Geben Sie Ausdrücke der relationalen Algebra für die folgenden Anfragen an die Mondial-Datenbank an:

- Alle Landescodes von Ländern, in denen eine Sprache gesprochen wird, die auch in der Schweiz gesprochen wird.
- Alle Landescodes von Ländern, in denen ausschliesslich Sprachen gesprochen werden, die in der Schweiz nicht gesprochen werden.
- Alle Landescodes von Ländern, in denen nur Sprachen gesprochen werden, die auch in der Schweiz gesprochen werden.
- Alle Landescodes von Ländern, in denen alle Sprachen gesprochen werden, die in der Schweiz gesprochen werden.

Für spätere Übungsblätter:

- Geben Sie dieselben Anfragen in SQL an.

- Join zur Formulierung einer "Auswahlbedingung". Verwende Tabelle `language(country,name,percent)`, z.B. (CH, "german", 65).