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.

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.

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

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

a) Alle Landescodes von Ländern, in denen eine Sprache gesprochen wird, die auch in der Schweiz gesprochen wird.
b) Alle Landescodes von Ländern, in denen ausschliesslich Sprachen gesprochen werden, die in der Schweiz nicht gesprochen werden.
c) Alle Landescodes von Ländern, in denen nur Sprachen gesprochen werden, die auch in der Schweiz gesprochen werden.
d) 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.

Aufgabe 4 (Korrektheit der Join-Algorithmen) Die Definition des relationalen Join-Operators, $\bowtie$, auf den Folien ist nicht konstruktiv, sondern beschreibt "nur" *deklarativ*, aus welchen Tupeln das Ergebnis bestehen soll.

Beweisen Sie, dass die beiden folgenden Join-Algorithmen (a) und (b) korrekt sind, d.h. diese Menge von Tupeln liefern.

Gegeben seien Relationen $R(\bar{X})$ und $S(\bar{Y})$; berechnet werden soll $R \bowtie S$.

a) Nested-Loop-Join:

```
let Z̄ := X̄ ∩ Ȳ;
let T := ∅;
for r ∈ R do
    begin
      for s ∈ S do
        if r[Z̄] = s[Z̄] then
          begin
            μ := ein neues Tupel so dass μ(x) = r(x) für alle x ∈ X̄ und μ(y) = s(y) für alle y ∈ Ȳ;
            T := T ∪ {μ};
          end;
    end;
return T;
```

b) Loop-Index-Join: Hierbei ist ein Baumindex über den Join-Attributen $\bar{Z} := \bar{X} \cap \bar{Y}$ von S in der Datenbank vorhanden.
Nehmen Sie an, dass $\bar{Z}$ Schlüssel von S ist (dann erstellt ein Datenbanksystem diesen Index automatisch). Nehmen Sie weiter an, dass $\bar{Z}$ nur ein Attribut enthält, welches numerisch ist, und stellen Sie sich als Index einen Binärbaum wie (hoffentlich) in Info I besprochen vor. In den jeweiligen Baumknoten ist nicht nur der Wert, sondern auch eine Referenz auf das Tupel, das diesen Wert hat, enthalten.

```
let Z̄ := X̄ ∩ Ȳ;
let B := der Baumindex über S.Z̄;
let T := ∅;
for r ∈ R do
    begin
      if Suche nach r[Z̄] in B erfolgreich
        begin
          s = das vom Baumknoten referenzierte Tupel in S;
          μ := ein neues Tupel so dass μ(x) = r(x) für alle x ∈ X̄ und μ(y) = s(y) für alle y ∈ Ȳ;
          T := T ∪ {μ};
          end;
    end;
return T;
```

c) Wie ist die Komplexität der beiden Algorithmen?

Aufgabe 5 (Division mit Basisoperationen) Beweisen Sie, daß die in der Vorlesung angebene Darstellung der Division durch relationale Basisoperatoren als

$$r \div s = \pi[\bar{Z}](r) - \pi[\bar{Z}]((\pi[\bar{Z}](r) \bowtie s) - r)$$

mit $r \in \mathsf{Rel}(\bar{X})$, $s \in \mathsf{Rel}(\bar{Y})$ und $\bar{Z} = \bar{X} \setminus \bar{Y}$ äquivalent zu der gegebenen Definition

$$r \div s = \{\mu \in \mathsf{Tup}(\bar{Z}) \mid \mu \in \pi[\bar{Z}](R) \wedge \{\mu\} \bowtie s \subseteq r\}$$

ist.

Veranschaulichen Sie sich Ihre Überlegungen anhand des Beispiels "Geben Sie die Namen derjenigen Organisationen an, die auf jedem Kontinent mindestens ein Mitglied haben".

Aufgabe 6 (Definition der Division) Betrachten Sie (i) die Definition der relationalen Division aus der Vorlesung:

$$r \div s = \{\mu \in \mathrm{Tup}(\bar{Z}) \mid \mu \in \pi[\bar{Z}](r) \wedge \{\mu\} \times s \subseteq r\}$$

sowie (ii) die kürzere Definition

$$r \cdot\!|\!\cdot s = \{\mu \in \mathrm{Tup}(\bar{Z}) \mid \{\mu\} \times s \subseteq r\}\ .$$

Sind (i) und (ii) äquivalent, bzw. warum nicht? Wo liegt das Problem?

Aufgabe 7 (Tupeloperatoren vs. Relationale Operatoren) In der Vorlesung wurden auf einzelnen Tupeln nur die Operatoren Projektion $\pi[\bar{X}](\mu)$, Selektion $\sigma[\alpha](\mu)$ und Renaming $\rho[A \to B](\mu)$ definiert. Die relationalen Operatoren wurden dann auf Basis dieser Operatoren definiert, wobei für das Join nur eine deklarative, auf tupelbasierter Projektion aufbauende Definition gegeben wurde.

- Geben Sie die Definition des relationalen Joins an: "Sei $r \in \mathrm{Rel}(\bar{X})$ and $s \in \mathrm{Rel}(\bar{Y})$. Dann ist $r \bowtie s = \{$*was gehört hier hin?*$\}$.
- Überlegen Sie, wie ein Join-Operator für Tupel $\mu \in \mathrm{Tup}(\bar{X})$, $\nu \in \mathrm{Tup}(\bar{Y})$, also $\mu \bowtie \nu$, definiert werden kann, und geben Sie darauf basierend eine Definition des relationalen Join-Operators an.
- Kann man eine entsprechende Definition auch für die Division angeben?

Aufgabe 8 (Äquivalenzen: Join, Division, Differenz) Seien $R(\bar{X}), S(\bar{Y})$ Relations-Schemata. Zeigen oder widerlegen Sie:

(a) Sei $\bar{X} \cap \bar{Y} = \emptyset$.

$$(R \bowtie S) \div S \equiv R.$$

(b) Sei $\bar{X} = \bar{Y}$ und $\bar{Z} \subseteq \bar{X}$.

$$\pi[\bar{Z}](R - S) \equiv \pi[\bar{Z}]R - \pi[\bar{Z}]S.$$

Aufgabe 9 (Algebra: Minimale- und Maximale Anzahl von Tupeln) Die Relationen $R(\bar{X})$ und $S(\bar{Y})$ enthalten n bzw. m Tupel. Wie groß ist die maximale und minimale Anzahl von Tupeln, die das Ergebnis folgender Operationen (bei geeigneten $\bar{X}, \bar{Y}$) enthalten kann?

a) $R \cup S$
b) $R \bowtie S$
c) $\sigma[C](R) \times S$, für eine Bedingung C
d) $\pi[\bar{Y}](R) - S$
e) $R \div S$

Aufgabe 10 (Transitive Hülle) Gegeben sei eine Relation $R(A, B)$. Skizzieren Sie einen Algorithmus, der, bestehend aus Operationen der relationalen Algebra und einer while-Schleife, die transitive Hülle der Relation R berechnet.

Hinweis: Die transitive Hülle einer Relation R, bezeichnet als R^*, ergibt sich wie folgt: betrachte z.B. eine Relation $R(von, nach)$ von Flugverbindungen. R^2 ist dann die Menge aller Verbindungen, die über eine Zwischenlandung zustandekommen, etc; R^n sind also diejenigen, Verbindungen, die sich aus n Teilverbindungen zusammensetzen. Die unendliche Vereinigung $R \cup R^2 \cup R^3 \cup ...$ für $R \to \infty$ wird dann als R^* bezeichnet. In einer endlichen Datenbasis benötigt man nur endlich viele Schritte um diese zu berechnen. Ein anderes beliebtes Beispiel ist die aus $Kind(x, y)$ berechnete Vorfahren-Relation.