Datenbanktheorie

1. Enthaltensein-Problem konjunktiver Anfragen
2. Minimierung und Chase
3. Auswertung von Verbundausdrücken

Literatur: Abiteboul, S., R. Hull und V. Vianu, *Foundations of Databases*, Addison-Wesley 1995.

Motivation: Anfrageoptimierung

Beispiel

Sei

Sales(PName, SName, CName),
Part(PName, Type),
Cust(CName, CAddr),
Supp(SName, SAddr).

Sind die folgenden zwei Anfragen Q, Q' äquivalent?

Q:

```
SELECT Type
FROM Sales NATURAL JOIN
     Part NATURAL JOIN
     Cust NATURAL JOIN Supp
WHERE CAddr = SAddr
```

Q':

```
SELECT Type
FROM Sales S1 NATURAL JOIN
     Part P1 NATURAL JOIN
     Cust NATURAL JOIN Supp,
     Sales S2, Part P2
WHERE CAddr = SAddr AND
      S2.PName = P2.PName AND
      P1.Type = P2.Type
```

Falls ja, dann ist Q der Version Q' offensichtlich vorzuziehen.
$\Longrightarrow$ Ziel ist eine *Minimierung* von Anfragen.

Enthaltensein-Problem

Formale Basis einer Anfragemininimierung.

Formalisierung: Konjunktive Anfragen

Eine *konjunktive Anfrage Q* zu einem Datenbank-Schema $\mathcal{R}$ hat die Form

$$ans(\vec{U}) \leftarrow R_1(\vec{U}_1), \ldots, R_n(\vec{U}_n),$$

wobei für $1 \leq i \leq n$

- die R_i Relationsbezeichner aus $\mathcal{R}$ und
- $\vec{U}$ und die $\vec{U}_i$ Vektoren von Variablen und Konstanten sind;
- die Variablen in $\vec{U}$ treten unter den Variablen in den $\vec{U}_i$ auf.
- Der Teil links vom $\leftarrow$ heißt *Kopf* der Anfrage, und der Teil rechts ihr *Rumpf*.

Beispiel

Sei

Sales(PName, SName, CName),
Part(PName, Type),
Cust(CName, CAddr),
Supp(SName, SAddr).

$$Q: \quad ans(T) \leftarrow Sales(P, S, C), Part(P, T), Cust(C, A), Supp(S, A)$$

Antwortmenge

- Die Antwortmenge *ans* zu Q ergibt sich durch folgende Kalkül-Anfrage:

$$\{\vec{U} \mid \exists X_1, \ldots, X_k (R_1(\vec{U}_1) \wedge \ldots \wedge R_n(\vec{U}_n))\},$$

  wobei die $X_1, \ldots, X_k$ alle Variablen im Rumpf der Anfrage Q sind, die nicht im Kopf auftreten.
- Die Menge der Antworten zu Q bzgl. einer Instanz $\mathcal{I}$ ist $Q(\mathcal{I})$.
- Für $\vec{U} \in Q(\mathcal{I})$ schreiben wir auch $ans(\vec{U}) \in Q(\mathcal{I})$.

Probleme

Seien Q, Q_1, Q_2 konjunktive Anfragen.

Enthaltensein: Gilt $Q_1 \sqsubseteq Q_2$, d.h., gilt $Q_1(\mathcal{I}) \subseteq Q_2(\mathcal{I})$ für jede Instanz $\mathcal{I}$?

Äquivalenz: Gilt $Q_1 \equiv Q_2$, d.h., gilt $Q_1 \sqsubseteq Q_2$ und $Q_2 \sqsubseteq Q_1$?

Minimierung: Finde zu Q_1 eine äquivalente konjunktive Anfrage Q_2, die höchstens so viele Literale im Rumpf hat wie Q_1 und die minimal ist in dem Sinn, dass jede zu Q_2 äquivalente konjunktive Anfrage Q_3 mindestens soviele Literale im Rumpf hat wie Q_2.

Q_2 heißt *minimal*.

Beispiel

Sei

Sales(PName, SName, CName),
Part(PName, Type),
Cust(CName, CAddr),
Supp(SName, SAddr).

Die folgenden zwei Anfragen Q, Q' sind äquivalent:

$$Q: \quad ans(T) \leftarrow Sales(P, S, C), Part(P, T), Cust(C, A), Supp(S, A)$$

$$Q': \quad ans(T) \leftarrow Sales(P, S, C), Part(P, T), Cust(C, A), Supp(S, A), Sales(P', S', C'), Part(P', T)$$

Lemma

Seien

$$Q_1: \quad ans(\vec{U}) \leftarrow R_1(\vec{U}_1), \ldots, R_n(\vec{U}_n)$$
$$Q_2: \quad ans(\vec{U}) \leftarrow S_1(\vec{V}_1), \ldots, S_m(\vec{V}_m)$$

konjunktive Anfragen, wobei

$$\{R_1(\vec{U}_1), \ldots, R_n(\vec{U}_n)\} \supseteq \{S_1(\vec{V}_1), \ldots, S_m(\vec{V}_m)\}$$

Dann $Q_1 \sqsubseteq Q_2$.

Substitution

- Eine *Substitution* θ über einer Variablenmenge $\mathcal{V}$ ist eine Abbildung von $\mathcal{V}$ nach $\mathcal{V} \cup dom$, wobei *dom* ein relevanter Wertebereich.
- Der Definitionsbereich einer Substitution θ ist erweitert auf die Konstanten $a \in dom$ und Relationsbezeichner $R \in \mathcal{R}$, wobei $\theta(a) = a$, bzw. $\theta(R) = R$.
- Substitutionen verallgemeinern Variablenbelegungen.

Beispiel

Betrachte

$$Q: \quad ans(T) \leftarrow Sales(P,S,C), Part(P,T), Cust(C,A), Supp(S,A)$$

$$Q': \quad ans(T) \leftarrow Sales(P,S,C), Part(P,T), Cust(C,A), Supp(S,A), Sales(P',S',C'), Part(P',T)$$

und θ wie folgt:

X	P	P'	S	S'	C	C'	T	A
$\theta(X)$	P	P	S	S	C	C	T	A

Enthaltensein-Abbildung

Seien

$$Q_1: \quad ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})$$
$$Q_2: \quad ans(\vec{V}) \leftarrow S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})$$

konjunktive Anfragen.
Eine Substitution θ ist eine *Enthaltensein-Abbildung* (engl. *containment mapping*) von Q_2 in Q_1 genau dann, wenn Q_2 mittels θ in eine Teilmenge von Q_1 umgewandelt wird:

- $\theta(ans(\vec{V})) = ans(\vec{U})$,
- für jedes $i = 1, \ldots, m$ existiert ein $j \in \{1, \ldots, n\}$, so dass $\theta(S_i(\vec{V_i})) = R_j(\vec{U_j})$.

Beachte dass somit $S_i = R_j$! Beachte weiter, dass Relationsbezeichner in Q_1 und Q_2 jeweils mehrfach auftreten können.

Beispiel

Betrachte noch einmal

$$Q: \quad ans(T) \leftarrow Sales(P,S,C), Part(P,T), Cust(C,A), Supp(S,A)$$

$$Q': \quad ans(T) \leftarrow Sales(P,S,C), Part(P,T), Cust(C,A), Supp(S,A), Sales(P',S',C'), Part(P',T)$$

und θ wie folgt:

X	P	P'	S	S'	C	C'	T	A
$\theta(X)$	P	P	S	S	C	C	T	A

θ ist eine Enthaltensein-Abbildung.

Satz

Seien

$$
\begin{aligned}
Q_1 : \quad & ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n}) \\
Q_2 : \quad & ans(\vec{V}) \leftarrow S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})
\end{aligned}
$$

konjunktive Anfragen.
$Q_1 \sqsubseteq Q_2$ genau dann, wenn eine Enthaltensein-Abbildung θ von Q_2 nach Q_1 existiert.

$$
\begin{aligned}
Q_1 : \quad & ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n}) \\
Q_2 : \quad & ans(\vec{V}) \leftarrow S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})
\end{aligned}
$$

Beweis "$\Leftarrow$":

Es existiert eine Enthaltensein-Abbildung θ von Q_2 nach Q_1.
Sei $\mathcal{I}$ eine Instanz zu Q_1 und sei $\mu \in Q_1(\mathcal{I})$.
Es existiert dann eine Substitution τ, so dass $\tau(\vec{U_j}) \in \mathcal{I}(R_j)$, $j \in \{1, \ldots, n\}$ und $\mu = \tau(\vec{U})$.
Betrachte dann die Substitution $\tau' = \theta \circ \tau$ und weiter $\tau'(S_i(\vec{V_i}))$.
Es gilt dann $\tau'(\vec{V_i}) \in \mathcal{I}(S_i)$, $i \in \{1, \ldots, m\}$ und damit auch $\mu = \tau'(\vec{V})$.
D.h., $\mu \in Q_2(\mathcal{I})$.