kanonische Instanz

Sei Q eine konjunktive Anfrage der Form $ans(\vec{U}) \leftarrow R_1(\vec{U}_1), \ldots, R_n(\vec{U}_n)$ über einem Datenbank-Schema $\mathcal{R}$. Die *kanonische Instanz* $\mathcal{I}_Q$ zu Q wird wie folgt gebildet.

- $\mathcal{I}_Q$ ist eine Instanz zu $\mathcal{R} = \{R_1, \ldots, R_n\}$.
- Sei τ eine Substitution, die jeder Variablen X in Q eine Konstante a_X zuweist, die verschieden von allen anderen Konstanten a_Y zu den übrigen Variablen Y in Q ist.
- Sei $R \in \mathcal{R}$. $\mathcal{I}_Q(R)$ enthält gerade für jedes Literal im Rumpf der Form $R(t_1, \ldots, t_n)$ ein Tupel der Form $(\tau(t_1), \ldots, \tau(t_n))$; wir schreiben auch $\tau(R(t_1, \ldots, t_n)) \in \mathcal{I}_Q(R)$.

τ nennen wir *kanonische Substitution*.

Beispiel

$$
\begin{array}{rl}
Q: & ans(T) \leftarrow Sales(P,S,C), Part(P,T), Cust(C,A), Supp(S,A) \\
Q': & ans(T) \leftarrow Sales(P,S,C), Part(P,T), Cust(C,A), Supp(S,A), \\
 & \qquad Sales(P',S',C'), Part(P',T)
\end{array}
$$

$\mathcal{I}_Q$:

Sales

a_P	a_S	a_C

Part

a_P	a_T

Cust

a_C	a_A

Supp

a_S	a_A

$\mathcal{I}_{Q'}$:

Sales

a_P	a_S	a_C
$a_{P'}$	$a_{S'}$	$a_{C'}$

Part

a_P	a_T
$a_{P'}$	a_T

Cust

a_C	a_A

Supp

a_S	a_A

$$
\begin{array}{ll}
Q_1: & ans(\vec{U}) \leftarrow R_1(\vec{U}_1), \ldots, R_n(\vec{U}_n) \\
Q_2: & ans(\vec{V}) \leftarrow S_1(\vec{V}_1), \ldots, S_m(\vec{V}_m)
\end{array}
$$

Beweis "$\Rightarrow$":

$Q_1 \sqsubseteq Q_2$.
Betrachte die kanonische Instanz $\mathcal{I}_{Q_1}$ und sei τ die entsprechende kanonische Substitution.
Es folgt dann offensichtlich $\tau(ans(\vec{U})) \in Q_1(\mathcal{I}_{Q_1})$.
Wegen $Q_1 \sqsubseteq Q_2$ gilt auch $\tau(ans(\vec{U})) \in Q_2(\mathcal{I}_{Q_1})$.
Damit existiert eine Substitution ρ, so dass $\rho(S_i(\vec{V}_i)) = \tau(R_j(\vec{U}_j))$, $1 \leq i \leq m$, $j \in \{1, \ldots, n\}$ und $\rho(ans(\vec{V})) = \tau(ans(\vec{U}))$.
Die gesuchte Enthaltensein-Abbildung ergibt sich somit zu $\rho \circ \tau^{-1}$.

Korollar

Seien

$$
\begin{array}{ll}
Q_1: & ans(\vec{U}) \leftarrow R_1(\vec{U}_1), \ldots, R_n(\vec{U}_n) \\
Q_2: & ans(\vec{V}) \leftarrow S_1(\vec{V}_1), \ldots, S_m(\vec{V}_m)
\end{array}
$$

konjunktive Anfragen. Sei $\mathcal{I}_{Q_1}$ die kanonische Instanz zu Q_1 mit kanonischer Substitution τ.
Es gilt $Q_1 \sqsubseteq Q_2$ genau dann, wenn $\tau(ans(\vec{U})) \in Q_2(\mathcal{I}_{Q_1})$.

Beweis: Es bleibt zu zeigen, dass wenn $\tau(ans(\vec{U})) \in Q_2(\mathcal{I}_{Q_1})$, dann auch $Q_1 \sqsubseteq Q_2$.
$\mathcal{I}_{Q_1}$ enthält zu jedem Relationsbezeichner S_j in Q_2 eine nichtleere Relation. Das heißt, zu S_j existiert ein R_i, wobei $S_j = R_i$. Des Weiteren existiert eine Substitution ρ, so dass für $S_j(\vec{V}_j)$ gerade $\rho(V_j) \in \mathcal{I}_{Q_1}(R_i)$. $\rho \circ \tau^{-1}$ ist dann eine Enthaltensein-Abbildung von Q_2 nach Q_1.

Beispiel

Für die Anfragen aus dem Beispiel gilt:

$$ans(a_T) \in Q(\mathcal{I}_{Q'})$$

und

$$ans(a_T) \in Q'(\mathcal{I}_Q).$$

Satz

Seien Q_1, Q_2 konjunktive Anfragen. Es ist NP-vollständig zu entscheiden, ob $Q_1 \sqsubseteq Q_2$.

Minimalität

Konjunktive Anfragen können durch Streichen von Literalen im Rumpf minimiert werden.

Beispiel:

$$\begin{aligned} Q: \quad & ans(T) \leftarrow Sales(P, S, C), Part(P, T), Cust(C, A), Supp(S, A) \\ Q': \quad & ans(T) \leftarrow Sales(P, S, C), Part(P, T), Cust(C, A), Supp(S, A), \\ & \qquad Sales(P', S', C'), Part(P', T) \end{aligned}$$

Q ist eine zu Q' minimale äquivalente Anfrage.

Satz

Sei $Q_1: \; ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})$ eine konjunktive Anfrage. Dann existiert eine minimale (Anzahl Teilziele) zu Q_1 äquivalente konjunktive Anfrage $Q_2: \; ans(\vec{V}) \leftarrow S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})$, so dass $\{S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})\} \subseteq \{R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})\}$.

Beweis

Sei Q_3 eine minimale äquivalente konjunktive Anfrage zu Q_1.
Es existieren dann Enthaltensein-Abbildungen θ von Q_1 nach Q_3, bzw. λ von Q_3 nach Q_1.
Seien o.B.d.A. $\{S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})\}$ gerade diejenigen Teilziele von Q_1, die Bild bzgl. λ sind und sei Q_2 die so gebildete konjunktive Anfrage.
(i) Es gilt dann offensichtlich $Q_1 \sqsubseteq Q_2$.
(ii) $Q_2 \sqsubseteq Q_1$ folgt mittels der Enthaltensein-Abbildung $\theta \circ \lambda$.
(iii) Die Minimalität ergibt sich aus der Tatsache, dass wegen der Verwendung von λ Q_2 nicht mehr Teilziele haben kann, als Q_3.

Algorithmus zur Minimierung einer konjunktiven Anfrage

Betrachte alle Enthaltensein-Abbildungen der Anfrage auf sich selbst und wähle diejenige aus, deren Bild eine Teilmenge mit minimaler Anzahl Literale ergibt.

Mengen von konjunktiven Anfragen

- Seien $Q_1, Q_2, \ldots, Q_n$ konjunktive Anfragen zu einem Datenbank-Schema $\mathcal{R}$ der Form

$$ans(\vec{U}) \leftarrow R_{i1}(\vec{U}_{i1}), \ldots, R_{in_i}(\vec{U}_{in_i}),$$

  wobei $n_i \geq 1$ für $1 \leq i \leq n$.
- Seien $P_1, P_2, \ldots, P_m$ konjunktive Anfragen zu demselben Datenbank-Schema $\mathcal{R}$ der Form

$$ans(\vec{V}) \leftarrow S_{j1}(\vec{V}_{j1}), \ldots, S_{jm_j}(\vec{V}_{jm_j}),$$

  wobei $m_j \geq 1$ für $1 \leq j \leq m$.
- Die einzelnen Q's und P's haben Antwort-Literale gleicher Stelligkeit im Kopf.

- Sei $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_n\}$ und $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ Mengen konjunktiver Anfragen.
- Es gilt $\mathcal{Q} \sqsubseteq \mathcal{P}$, $\mathcal{Q}$ ist in $\mathcal{P}$ *enthalten*, genau dann, wenn für jede Instanz $\mathcal{I}$ von $\mathcal{R}$ gilt:

$$\bigcup_{i=1,\ldots,n} Q_i(\mathcal{I}) \subseteq \bigcup_{j=1,\ldots,m} P_j(\mathcal{I}).$$

- Gilt $\mathcal{Q} \sqsubseteq \mathcal{P}$ und $\mathcal{P} \sqsubseteq \mathcal{Q}$, dann sind $\mathcal{P}$ und $\mathcal{Q}$ *äquivalent*, $\mathcal{P} \equiv \mathcal{Q}$.
- Beispiel:

$$\begin{array}{ll} Q_1: & ans(X,Y) :- E(X,X), E(X,Y) \\ Q_2: & ans(X,Y) :- E(X,W), E(W,Y) \\ Q_3: & ans(X,Y) :- E(X,Y), E(X,U), E(U,Y) \end{array}$$

  Es gilt $Q_1 \sqsubseteq Q_3 \sqsubseteq Q_2$.
  Darüberhinaus auch $\{Q_1, Q_2, Q_3\} \equiv \{Q_2, Q_3\}$.

Enthaltensein-Beziehung

Eine Menge $\mathcal{Q}$ ist in einer Menge $\mathcal{P}$ von Anfragen enthalten, wenn jede Anfrage in $\mathcal{Q}$ in mindestens einer Anfrage von $\mathcal{P}$ enthalten ist.

Ist dies auch notwendig?

Eine *Enthaltensein-Beziehung* ist eine Abbildung Ω von $\mathcal{Q}$ nach $\mathcal{P}$ mit der Eigenschaft:

wenn $\Omega(Q_i) = P_j$, dann auch $Q_i \sqsubseteq P_j$, wobei $1 \leq i \leq n$, $1 \leq j \leq m$.

Satz

Seien $\mathcal{Q}$, $\mathcal{P}$ Mengen konjunktiver Anfragen. $\mathcal{Q} \sqsubseteq \mathcal{P}$ genau dann, wenn eine Enthaltensein-Beziehung Ω von $\mathcal{Q}$ nach $\mathcal{P}$ existiert.

Beweis

(1) Es existiert eine Enthaltensein-Beziehung Ω. klar.
(2) $\mathcal{Q} \sqsubseteq \mathcal{P}$.
Angenommen es existiert ein Q_i so dass für alle P_j gerade $Q_i \not\sqsubseteq P_j$. Bilde die kanonische Instanz $\mathcal{I}_{Q_i}$ und sei τ die entsprechende kanonische Substitution.
Da $\mathcal{Q} \sqsubseteq \mathcal{P}$, folgt

$$\tau(ans(\vec{U})) \in \bigcup_{j=1,\ldots,m} P_j(\mathcal{I}_{Q_i}).$$

Damit existiert insbesondere $j', 1 \leq j' \leq m$, so dass $\tau(ans(\vec{U})) \in P_{j'}(\mathcal{I}_{Q_i})$.
Damit aber gerade $Q_i \sqsubseteq P_{j'}$, ein Widerspruch.