Conjunctive Queries

Definition

A *conjunctive Query* Q over a database schema $\mathcal{R}$ is given as

$$ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n}),$$

such that for $1 \leq i \leq n$

- R_i a relation name in $\mathcal{R}$ and
- $\vec{U}$ and $\vec{U_i}$ vectors of variables and constants;
- any variable appearing in $\vec{U}$ appears also in some $\vec{U_i}$.
- Left to $\leftarrow$ is the *head* of the query, and to the right there is the *body*. The atoms in the body are also called *subgoals*.

Example

$$\begin{gathered}\textit{Sales(Part, Supplier, Customer)},\\ \textit{Part(PName, Type)},\\ \textit{Cust(CName, CAddr)},\\ \textit{Supp(SName, SAddr)}.\end{gathered}$$

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

Answer

- The set of answers *ans* to Q can be defined by a relational calculus query:

$$\{\vec{U} \mid \exists X_1, \ldots, X_k(R_1(\vec{U_1}) \wedge \ldots \wedge R_n(\vec{U_n}))\},$$

  where the $X_1, \ldots, X_k$ are exactly the variables in the body of Q, which do not appear in the head.
- The set of answers Q w.r.t. an instance $\mathcal{I}$ is denoted $Q(\mathcal{I})$.
- For $\vec{U} \in Q(\mathcal{I})$ we shall also write $ans(\vec{U}) \in Q(\mathcal{I})$.

Problemes

Let Q, Q_1, Q_2 be conjunctive queries.

Containment: $Q_1 \sqsubseteq Q_2$, i.e., $Q_1(\mathcal{I}) \subseteq Q_2(\mathcal{I})$ for any instance $\mathcal{I}$?

Equivalence: $Q_1 \equiv Q_2$, i.e., $Q_1 \sqsubseteq Q_2$ and $Q_2 \sqsubseteq Q_1$?

Minimization: Given Q_1, construct an equivalent query Q_2, which has as most as much subgoals in its body as Q_1 and is minimal in the sense, that any query Q_3 being equivalent to Q_2 has at least as much subgoals in the body as Q_2.

Q_2 is called *minimal*.

Example

$$
\begin{gathered}
\mathit{Sales(Part,\ Supplier,\ Customer)},\\
\mathit{Part(PName,\ Type)},\\
\mathit{Cust(CName,\ CAddr)},\\
\mathit{Supp(SName,\ SAddr)}.
\end{gathered}
$$

Equivalent queries:

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

Lemma

Let

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

be conjunctive queries, where

$$\{R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})\} \supseteq \{S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})\}$$

Then $Q_1 \sqsubseteq Q_2$.

Substitution

- A *substitution* θ over a set of variables $\mathcal{V}$ is a mapping from $\mathcal{V}$ to $\mathcal{V} \cup dom$, where *dom* a corresponding domain.
- We extend θ to constants $a \in dom$ and relation names $R \in \mathcal{R}$, where $\theta(a) = a$, resp. $\theta(R) = R$.

Example

Consider

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

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

and θ:

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

Containment Mapping

Given conjunctive queries

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

Substitution θ is called *containment mapping* from Q_2 to Q_1, if Q_2 can be transformed by means of θ to become Q_1:

- $\theta(ans(\vec{V})) = ans(\vec{U})$,
- for $i = 1, \ldots, m$ there exists a $j \in \{1, \ldots, n\}$, such that $\theta(S_i(\vec{V_i})) = R_j(\vec{U_j})$.

Example

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

θ:

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

θ is a containment mapping.

Theorem

Let

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

be conjunctive queries.
$Q_1 \sqsubseteq Q_2$ iff there exists a containment mapping θ from Q_2 to Q_1.

Proof "$\Leftarrow$":

There exists containment mapping θ.
Let $\mathcal{I}$ be an instance of Q_1 and let $\mu \in Q_1(\mathcal{I})$.
There exists a substitution τ, such that $\tau(\vec{U_j}) \in \mathcal{I}(R_j)$, $j \in \{1, \ldots, n\}$ and $\mu = \tau(\vec{U})$.
Consider a substitution $\tau' = \tau \circ \theta$ and further $\tau'(S_i(\vec{V_i}))$.
There holds $\tau'(\vec{V_i}) \in \mathcal{I}(S_i)$, $i \in \{1, \ldots, m\}$ and therefore also $\mu = \tau'(\vec{V})$.
D.h., $\mu \in Q_2(\mathcal{I})$.

Canonical Instance

Let Q be a conjunctive $ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})$ over a database schema $\mathcal{R}$. The *canonical instance* $\mathcal{I}_Q$ to Q is constructed as follows.

- $\mathcal{I}_Q$ is an instance of $\mathcal{R} = \{R_1, \ldots, R_n\}$.
- Let τ be a substitution, which assigns to any X in Q an unique constant a_X.
- For any literal $R(t_1, \ldots, t_n)$ in the body, insert a tupel of the form $(\tau(t_1), \ldots, \tau(t_n))$ into $\mathcal{I}_Q(R)$; we also write $\tau(R(t_1, \ldots, t_n)) \in \mathcal{I}_Q(R)$.
- No other tuples are inserted into $\mathcal{I}_Q(R)$.

τ is called *canonical substitution*.

Example

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

$\mathcal{I}_Q$:

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

Proof "$\Rightarrow$":

$Q_1 \sqsubseteq Q_2$.
Consider $\mathcal{I}_{Q_1}$ and the corresponding canonical substitution τ.
Then $\tau(ans(\vec{U})) \in Q_1(\mathcal{I}_{Q_1})$.
Because of $Q_1 \sqsubseteq Q_2$ further $\tau(ans(\vec{U})) \in Q_2(\mathcal{I}_{Q_1})$.
Thus, there exists a substitution ρ, such that $\rho(S_i(\vec{V}_i)) = \tau(R_j(\vec{U}_j))$, $1 \leq i \leq m$, $j \in \{1, \dots, n\}$ und $\rho(ans(\vec{V})) = \tau(ans(\vec{U}))$.
$\tau^{-1} \circ \rho$ is a containment mapping.

Corollary

Let

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

be conjunctive queries, $\mathcal{I}_{Q_1}$ the canonical instance to Q_1 with canonical substitution τ.
$Q_1 \sqsubseteq Q_2$, iff $\tau(ans(\vec{U})) \in Q_2(\mathcal{I}_{Q_1})$.

Proof: We show, whenever $\tau(ans(\vec{U})) \in Q_2(\mathcal{I}_{Q_1})$, then $Q_1 \sqsubseteq Q_2$.
For any S_j in Q_2, $\mathcal{I}_{Q_1}$ is not empty. Therefore, for S_j there exists a R_i, such that $S_j = R_i$. Further, there exists a substitution ρ, such that for $S_j(\vec{V_j})$ we have $\rho(V_j) \in \mathcal{I}_{Q_1}(R_i)$. $\rho \circ \tau^{-1}$ is a containment mapping from Q_2 to Q_1.

Example

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

and

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

Minimality

To minimize a conjunctive query we can try to delete subgoals in its body.

Example:

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

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

Q is a minimal query equivalent to Q'.

Theorem

Let $Q_1 : \; ans(\vec{U}) \leftarrow R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})$ be a conjunctive query. To Q_1 there exists a minimal (in the number of subgoals in the body) equivalent conjunctive query $Q_2 : \; ans(\vec{V}) \leftarrow S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})$, such that $\{S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})\} \subseteq \{R_1(\vec{U_1}), \ldots, R_n(\vec{U_n})\}$.

Proof

Let Q_3 be a minimal equivalent conjunctive query to Q_1.
There exist containment mappings θ from Q_1 to Q_3, resp. λ from Q_3 to Q_1.
W.l.o.g. let $\{S_1(\vec{V_1}), \ldots, S_m(\vec{V_m})\}$ be those subgoals of Q_1, which are image w.r.t. λ and let Q_2 be the corresponding conjunctive query.
(i) $Q_1 \sqsubseteq Q_2$.
(ii) $\theta \circ \lambda$ is a containment mapping $Q_2 \sqsubseteq Q_1$.
(iii) Minimality follows, because λ Q_2 cannot have more subgoals than Q_3.

Minimization algorithm

Consider all containment mappings from the query to itself and choose one of those, for which the image has a minimal number of subgoals.

- Complexity?