Universität Freiburg
Institut für Informatik

Georges-Köhler Allee, Geb. 51
D-79110 Freiburg

Fang Wei-Kleiner

fwei@informatik.uni-freiburg.de

Advanced Databases and Information Systems
Summerterm 2019
Discussion on 27/06/2019

10. Sheet: Conjunctive Queries

Exercise 1 (Evaluation of conjunctive queries)
Consider the following sample instantiation $\mathcal{I}$ of a database.

`Sales`	*PName*	*SName*	*CName*
	Audi A7	Autohaus Wenz	Meier
	Audi A8	Autohaus Klein	Meier
	Audi A8	Autohaus Wenz	Smith
	Suzuki GSX	Motorsport AG	Hofmann

`Part`	*PName*	*Type*
	Audi A8	Auto
	Audi A7	Auto
	Suzuki GSX	Motorrad

`Cust`	*CName*	*CAddr*
	Meier	Freiburg
	Smith	Freiburg
	Hofmann	Mannheim

`Supp`	*SName*	*SAddr*
	Autohaus Wenz	Freiburg
	Autohaus Klein	Mannheim
	Motorsport AG	Mannheim

Compute the evaluation result of the following queries on instance $\mathcal{I}$ and informally describe their meaning. Note that constants inside the queries are distinguished by *italic* font.

a) q_1: `ans`(C) $\leftarrow$ `Sales`(P,S,C), `Cust`(C,*Freiburg*), `Supp`(S,*Freiburg*)

Die Anfrage gibt die Namen aller Kunden aus Freiburg zurück die mindestens ein Produkt bei einem Lieferanten aus Freiburg eingekauft haben. Für die obige Instanz ergibt sich die Lösung $q_1(\mathcal{I})$

`ans`	C
	Meier
	Smith

b) q_2: `ans`(S,P) $\leftarrow$ `Sales`(P,S,*Meier*), `Supp`(S,*Mannheim*), `Part`(P,*Auto*)

Die Anfrage gibt den Zulieferernamen und das zugehörige Produkt aller Auto-Käufe von Herrn Meier von einem Zulieferer aus Mannheim zurück. Für die obige Instanz ergibt sich die Lösung $q_2(\mathcal{I})$

`ans`	S	P
	Authohaus Klein	Audi A8

c) q_3: `ans`(S,P) $\leftarrow$ `Sales`(P,S,*Meier*), `Supp`(S,*Mannheim*), `Part`(P2,*Auto*)

Die Anfrage gibt den Zulieferernamen und das zugehörige Produkt aller Käufe von Herrn Meier von einem Zulieferer aus Mannheim zurück, vorausgesetzt dass mindestens ein Auto in der Datenbank (in der `Part`-Relation) eingetragen ist. Für die obige Instanz ergibt sich die Lösung $q_3(\mathcal{I})$

`ans`	S	P
	Authohaus Klein	Audi A8

d) q_4: ans(C1,C2) $\leftarrow$ Cust(C1,*Freiburg*), Cust(C2,*Freiburg*), Sales(P1,S1,C1), Sales(P2,S2,C2), Supp(S1,X), Supp(S2,X)

Die Anfrage berechnet alle Paare von Freiburger Personen, die Produkte von Zulieferern aus dem selben Ort gekauft haben. Für die obige Instanz ergibt sich die Lösung $q_4(\mathcal{I})$

ans	C1	C2
	Meier	Meier
	Meier	Smith
	Smith	Meier
	Smith	Smith

Exercise 2 (Containment)
Consider the following pairs of Conjunctive Queries and decide for each pair q_i, q_i' if $q_i \sqsubseteq q_i'$, $q_i' \sqsubseteq q_i$, and $q_i \equiv q_i'$ holds. If such relationships hold provide the corresponding containment mappings. Otherwise, show that no such mapping exists.

a) q_1: ans(X,Y) $\leftarrow$ R(X,Z), R(Z,T), S(T,Y) und q_1': ans(X,Z) $\leftarrow$ R(X,X), S(X,Z)

q_1 ist nicht enthalten in q_1', weil es keine Enthaltensein-Abbildung von q_1' nach q_1 gibt. Das Teilziel $R(X,X)$ aus q_1' muss entweder auf $R(X,Z)$ oder auf $R(Z,T)$ abgebildet werden. Da in beiden Fällen X auf zwei verschiedene Variablen abgebildet werden muss, existiert diese Abbildung nicht. Umgekehrt ist q_1' enthalten in q_1 mit der Enthaltensein-Abbildung $\theta = \{X \rightarrow X, Y \rightarrow Z, Z \rightarrow X, T \rightarrow X\}$ von q_1 nach q_1'. Da nur eine der beiden Enthaltensein-Beziehungen gilt sind die Anfragen nicht äquivalent.

b) q_2: ans(X) $\leftarrow$ R(X,Y), S(Y,Z), S(Y',Z') und q_2': ans(Y) $\leftarrow$ S(A,B), R(Y,A), R(Y',A)

Es gilt $q_2 \equiv q_2'$. Das Containment Mapping von q_2 nach q_2' ist $\theta_1 = \{X \mapsto Y, Y \mapsto A, Z \mapsto B, Y' \mapsto A, Z' \mapsto B\}$, folglich gilt $q_2' \sqsubseteq q_2$. In der umgekehrten Richtung existiert das Containment-Mapping $\theta_2 = \{Y \mapsto X, A \mapsto Y, B \mapsto Z, Y' \mapsto X\}$ von q_2' nach q_2, also gilt auch $q_2 \sqsubseteq q_2'$.

c) q_3: ans(U,Z) $\leftarrow$ R(U,V), R(X,Y), S(Y,Z), S(V,X) und q_3': ans(U,V) $\leftarrow$ R(Y,U), R(U,X), S(U,V), S(X,Y)
Es gilt $q_3 \not\sqsubseteq q_3'$: wegen dem Kopfprädikat ans muss U nach U und V nach Z abgebildet werden. Somit bilder auch das Rumpfprädikat S(U,V) aus q_3' nach S(U,Z) ab, welches nicht in q_3 enthalten ist. Umgekehrt gilt $q_3' \sqsubseteq q_3$. Das Containment-Mapping von q_3 nach q_3' ist $\theta = \{U \mapsto U, Z \mapsto V, V \mapsto X, X \mapsto Y, Y \mapsto U\}$.

Exercise 3 (Containment)
Consider the following pairs of Conjunctive Queries and decide if $q_i \sqsubseteq q_i'$, $q_i' \sqsubseteq q_i$, and $q_i \equiv q_i'$ hold using the method of the canonical instance.

a) q_1: ans(X) $\leftarrow$ R(X,Y,X), R(X,Z,Y), S(Y,X) und q_1': ans(X) $\leftarrow$ R(X,Y,Z), S(Y,Z)

Es gilt $q_1 \sqsubseteq q_1'$, $q_1' \not\sqsubseteq q_1$ und folglich $q_1 \not\equiv q_1'$. Um $q_1 \sqsubseteq q_1'$ zu zeigen verwenden wir die Methode der kanonischen Instanz. Die kanonische DB von q_1 ist $D = \{R(0,1,0), R(0,2,1), S(1,0)\}$. Da der Kopf von q_1, ans(0), in $q_1'(D) = \{$ ans(0) $\}$ enthalten ist, ist q_1 in q_1' enthalten. Die kanonische Substitution, die dies zeigt, ist $\theta = \{X \mapsto 0, Y \mapsto 1, Z \mapsto 0\}$. Um $q_1' \not\sqsubseteq q_1$ zu zeigen betrachten wir das erste Literal R(X,Y,X). Dieses Literal müsste auf R(X,Y,Z) in q_1' abgebildet werden, was offensichtlich nicht möglich ist, da X nur auf einen Wert abgebildet werden kann.

b) q_2: ans(X) $\leftarrow$ R(X,Y), R(Y,Z), R(Z,X) und q_2': ans(X) $\leftarrow$ R(X,Y), R(Y,Z), R(Z,U), R(U,V)
Es gilt $q_2' \not\sqsubseteq q_2$, wie man beispielsweise an der Datenbankinstanz $\mathcal{I} = \{R(0,1), R(1,2), R(2,3), R(3,4)\}$: auf dieser Instanz liefer q_2' eine Lösung, während q_2 keine Lösung liefert. Die umgekehrte Richtung lösen wir mit Hilfe der kanonischen Instanz. Im ersten Schritt weisen wir Konstanten für q_2 zu und erhalten q_2: ans(a) :- R(a,b), R(b,c), R(c,a) und $D(q_2)$ - { R(a,b), R(b,c), R(c,a) }. Die gesuchte Abbildung ist $X \mapsto a$, $Y \mapsto b$, $Z \mapsto b$, $U \mapsto a$, $V \mapsto b$; folglich gilt $q_2 \sqsubseteq q_2'$.

Exercise 4 (NP-Completeness)
Prove the Conjunctive Query Containment problem is NP-Complete. To show NP-Hardness, you need to find an NP-Complete problem and make a reduction to the containment problem.

The 3-Colorability Problem

Definition (3-Colorability of Graphs)

Instance: A graph $G = (V, E)$

Question: Can G be colored with the three colors $\{r, g, b\}$ in such a way that two adjacent vertices have a distinct color?

The 3-colorability problem is NP-complete

A graph G is 3-colourable if and only if
there is a graph homomorphism from G to the simplex S_3,
which consists of three vertices that are connected to each other

Reducing 3-Colorability to Evaluation

Theorem (Reduction)

There is a database instance $\mathbf{I}_{3col}$ such that for every finite graph G one can compute in linear time a relational conjunctive query $Q_G() \coloneq L$ such that

$$G \text{ is 3-colorable} \quad \text{if and only if} \quad Q_G(\mathbf{I}_{3col}) = \{()\}$$

Remark (Boolean Queries)

- A query without distinguished variables is called a *boolean* query
- Over an instance, a boolean query returns the empty tuple (), or nothing

This shows NP-hardness of the combined complexity of conjunctive query evaluation

The Reduction

Given graph $G = (V, E)$, where

$V = \{v_1, \ldots, v_n\}$ and

$E = \{(v_{i_l}, v_{j_l}) \mid i_l < j_l,\ 1 \leq l \leq m\}$

We construct $\mathbf{I}_{3col}$ and Q_G as follows

$$\mathbf{I}_{3col} = \{\mathsf{e}(r, b),\ \mathsf{e}(b, r),\ \mathsf{e}(r, g),\ \mathsf{e}(g, r),\ \mathsf{e}(b, g),\ \mathsf{e}(g, b)\}$$

$$Q_G() \coloneq \mathsf{e}(y_{i_1}, y_{j_1}), \ldots, \mathsf{e}(y_{i_m}, y_{j_m})$$

where $y_1, \ldots, y_n$ are new variables and
there is one atom $\mathsf{e}(y_{i_l}, y_{j_l})$ for each edge $(v_{i_l}, v_{j_l}) \in E$

Clearly, there is an $\alpha\colon \{y_1, \ldots, y_n\} \to \{r,\ g,\ b\}$ satisfying Q_G over $\mathbf{I}_{3col}$ iff there is a graph homomorphism from G to S_3