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 04/07/2019

9. Sheet: Conjunctive Query Minimization

Exercise 1 (Containment & Minimization)
Consider the following four Conjunctive Queries, where c denotes a constant.

- $q_1 : ans(X, Y) \leftarrow R(X, A), R(A, B), R(B, Y)$
- $q_2 : ans(X, Y) \leftarrow R(X, A), R(A, B), R(B, C), R(C, Y)$
- $q_3 : ans(X, Y) \leftarrow R(X, A), R(B, C), R(D, Y), R(X, B), R(A, C), R(C, Y)$
- $q_4 : ans(X, Y) \leftarrow R(X, A), R(A, c), R(c, B), R(B, Y)$

a) Find all equivalences and containment relationships between the above queries.
Zwar sieht q_4 aus wie q_2, jedoch gibt es keine Konstante c in q_2. Eine Konstante kann niemals auf eine Variable abgebildet werden, d.h. $q_2 \not\sqsubseteq q_4$. Umgekehrt gilt jedoch $q_4 \sqsubseteq q_2$ (das Containment-Mapping bildet B auf Konstante c, C nach B und alle sonstigen Variablen auf sich selbst ab).
Weiterhin gilt $q_1 \not\sqsubseteq q_2$, $q_2 \not\sqsubseteq q_1$ und $q_1 \equiv q_3$ (letztgenannte Äquivalenz wird im zweiten Teil der Aufgabe gezeigt).

b) Minimize all queries.
q_1, q_2 und q_4 sind bereits minimiert. Minimierung von q_3:
ans(X,Y) ← R(X,A), R(B,C), R(D,Y), R(X,B), R(A,C), R(C,Y);
ans(U,V) ← R(U,W), R(P,L), R(N,V), R(U,P), R(W,L), R(L,V);
Abbildung:
θ : {U → X, V → Y, W → A, P → A, L → C, N → D}
Nach dem Entfernen von R(B,C) und R(X,B) erhalten wir q_3':
ans(X,Y) ← R(X,A), R(A,C), R(D,Y), R(C,Y), das umgeschrieben werden kann:
ans(U,V) ← R(U,W), R(W,T), R(P,V), R(T,V) und minimiert:
θ: {U → X, V → Y, W → A, T → C, P → C}
zur Anfrage ans(X,Y) ← R(X,A), R(A,C), R(C,Y) die äquivalent ist zu q_1.
$q_3 \equiv q_1$
Alternative Abbildung:
θ : {U → X, V → Y, W → A, P → A, L → C, N → C}
ans(X, Y) ← R(X,A), R(A,C), R(C,Y) ⇒ $q_3 \equiv q_1$.

Exercise 2 (CQ Minimization)
Instead of eliminating subgoals, query minimization can also be achieved by eliminating variables. Write an algorithm which minimizes queries by eliminating each time at least one variable. Prove that your algorithm generates a minimal query.

The minimization algorithm by removing the variables works analogously to the one by removing subgoals. It involves a stepwise picking a variable v from the current CQ Q, and find a containment mapping ρ from $\overrightarrow{V}$ to $\overrightarrow{V} \setminus v$, such that for each subgoal $R_i(\overrightarrow{V_i})$, $\rho(R_i(\overrightarrow{V_i}))$ can be found in the body of Q. If this is the case, then we can remove all the subgoals containing v. Otherwise the algorithm stops.

To show the removing of variables is equivalent to removing of subgoals, we have to prove: 1. If one variable can be removed, then there is at least one subgoal which can be removed. This is obvious from the algorithm. 2. If one subgoal can be removed, then there is at least one variable can be removed as well. To show this, assume we have found a subgoal $R_i(\overrightarrow{V_i})$, which can be removed. This means there is a containment mapping ρ from the variables from the CQ Q to the variables to Q, such that $\rho(R_i(\overrightarrow{V_i}))$ is mapped to another subgoal other than itself. This means, ρ maps at least one variable $v \in \overrightarrow{V_i}$ to some other variable v' in Q. If ρ maps some other variable v'' to v, then we can simply change the mapping from $v'' \to v$ to $v'' \to v'$, so that the containment holds as well. So we have constructed a new containment mapping from $\overrightarrow{V}$ to $\overrightarrow{V} \setminus v$. Thus we could remove v accordingly.

Exercise 3 (Acyclic CQ)
Apply GYO Algorithm to the following queries and decide whether they are acyclic.

$q(X, T) \leftarrow R1(X, Y, Z), R2(Y, V), R3(Y, Z, U), R4(Z, U, W), R5(U, W, T).$

$q(X, W) \leftarrow R1(X, Y, Z), R3(Y, Z, U), R4(Z, U, W), R5(U, W, X).$