4. Minimization of Conjunctive Queries

These are sample solutions to exercises that were given as coursework. They are not intended as models but show each one way to approach the problem set in the exercise. These solutions have been produced by a student for the course in 2006.

3. Minimisation of Conjunctive Queries

Recall that simple conjunctive queries (SCQs) are conjunctive queries without equalities and inequalities. Recall as well that a conjunctive query $q_0$ is a subquery of another conjunctive query $q$ if $q_0$ can be obtained from $q$ by dropping some of the atoms in the body of $q$.

Prove the following two propositions that provide the underpinnings for the algorithm of conjunctive query minimization.

**Proposition 1.** Let $q$ be a SCQ with $n$ atoms and $q'$ be an equivalent SCQ with $m$ atoms where $m < n$. Then there exists a subquery $q_0$ of $q$ such that $q_0$ has at most $m$ atoms in the body and $q_0$ is equivalent to $q$.

Sample solution by Evgeny Kharlamov.

**Proof.** Given that $q$ is equivalent to $q'$, the containments $q' \subseteq q$ and $q \subseteq q'$ hold. Suppose $q$ and $q'$ have the form $q(\vec{x}) : - B(q)$ and $q'(\vec{x}) : - B(q')$, respectively. By the Homomorphism Theorem, there exist substitutions $h : \text{term}(q) \rightarrow \text{term}(q')$ and $h' : \text{term}(q') \rightarrow \text{term}(q)$, such that $h'$ is a homomorphism from $q'$ to $q$ and $h$ is a homomorphism from $q$ to $q'$.

Let us consider the composition $\gamma = h' \circ h$ of these homomorphisms. Let $B_0$ be the sequence of atoms $\gamma B(q)$. The function $\hat{\gamma}$ works as follows:

$$\hat{\gamma} : B(q) \xrightarrow{h} B(q') \xrightarrow{h'} B_0.$$ 

Obviously, $B_0$ consists of at most $|B(q')| = m$ atoms. Observe that by construction $\gamma \vec{x} = \vec{\bar{x}}$ and $\gamma c = c$ for any constant $c$ in $\text{term}(q)$. Hence, every variable in $\vec{\bar{x}}$ occurs in $B_0$ and consequently $q_0(\vec{x}) : - B_0$ defines a query. Note that for the body of the query $q_0$ the inclusion $\gamma \text{body}(q) \subseteq \text{body}(q_0)$ hold. We constructed the query $q_0$ in such a way that it is a subquery of $q$ with at most $m$ atoms in the body. Now we will show that $q_0$ is equivalent to $q$. Observe that we constructed $q_0$ in such a way that $\gamma$ maps $\text{term}(q)$ to $\text{term}(q_0)$ and satisfies the conditions from the definition of a homomorphism. Hence, $\gamma$ is a homomorphism from $q$ to $q_0$ and $q_0 \subseteq q$.

![A diagram or explanation is needed here for a clearer understanding of how this inclusion holds.]

The inclusion $q_0 \subseteq q$ holds, since there exist a homomorphism, which is simply the identity function on $\text{term}(q_0)$. Hence, $q_0 \equiv q$ and $q_0$ has at most $m$ atoms in the body. \(\square\)
**Proposition 2.** Let $q$ and $q'$ be two equivalent minimal SCQs. Then $q$ and $q'$ are identical up to renaming of variables.

**Sample solution by Evgeny Kharlamov.**

*Proof.* Given the equivalence of $q$ and $q'$, there exist two homomorphisms $h$ from $q$ to $q'$ and $h'$ from $q'$ to $q$ such that

$$h: \text{term}(q) \rightarrow \text{term}(q') \quad \text{and} \quad h': \text{term}(q') \rightarrow \text{term}(q).$$

Let us consider the composition $\gamma = h' \circ h$ of these homomorphisms. Using the same reasoning as in Proposition 1, one can show that $\gamma$ is a homomorphism from $q$ to its subquery $q_0$ and $q$ is equivalent to $q_0$. Using the minimality of $q$ we obtain that $q_0$ coincides with $q$. Hence, the composition $\tilde{\gamma} = h' \circ \tilde{h}$ of the form

$$\tilde{\gamma}: \text{body}(q) \xrightarrow{\tilde{h}} \text{body}(q') \xrightarrow{h'} \text{body}(q).$$

is surjective. Obviously any surjective mapping of a finite set to itself is injective. The sets $\text{body}(q)$ and $\text{body}(q')$ are finite, hence, the composition $\tilde{\gamma}$ is injective, so it is bijective. From this we conclude that the first component of $\tilde{\gamma}$, namely $\tilde{h}$, is an injective mapping from $\text{body}(q)$ to $\text{body}(q')$. In a similar way one can show that the composition $\tilde{\gamma}' = \tilde{h} \circ h'$ is bijective and its first component, namely $\tilde{h}'$, is an injective mapping from $\text{body}(q')$ to $\text{body}(q)$. We obtained that $\tilde{h}$ and $\tilde{h}'$ are injective mappings from the finite set $\text{body}(q)$ to the finite set $\text{body}(q')$ and back, respectively. One consequence from this fact is that the sets $\text{body}(q)$ and $\text{body}(q')$ have the same cardinality, namely $|\text{body}(q)| = |\text{body}(q')|$, and the queries are over the same relational schemas, i.e. the sets of relational names occur in the bodies of $q$ and $q'$ are the same. Another consequence is that $\tilde{h}$ and $\tilde{h}'$ are surjective, moreover, they are bijective.

Observe that if the extension of a substitution (to sets of atoms) is a surjective mapping from one set of atoms to another, then the substitution itself is a surjective mapping from the set of terms occurring in one set of atoms to the set of terms occurring in another. The extension $\tilde{\gamma} = \tilde{h}' \circ \tilde{h}$ is surjective, hence, the substitution $\gamma = h' \circ h$ is a surjective mapping from the set of terms $\text{term}(q)$ to itself. Similarly, the substitution $\gamma' = h \circ h'$ is a surjective mapping from the set of terms $\text{term}(q')$ to itself. From the surjectivity of $\gamma$ and $\gamma'$ using the same reasons as we used in the previous paragraph we conclude that $\gamma$ and $\gamma'$ are bijective, their components $h$ and $h'$ are also bijective and $|\text{term}(q)| = |\text{term}(q')|$. 

Observe that the homomorphism $h$ is a bijective mapping from $\text{term}(q)$ to $\text{term}(q')$, such that it is the identity mapping on the set of constants and distinguished variables of $q$. Hence the inequality $|\text{var}(q')| \leq |\text{var}(q)|$ holds (in general, $h$ can map some non-distinguished variables of $q$ to constants of $q'$ that do not appear in $q$). Analogously, bijectivity of the homomorphism $h'$ gives us the inequality $|\text{var}(q)| \leq |\text{var}(q')|$. Hence, the sets of variables in the query $q$ and $q'$ have the same cardinality, namely $|\text{var}(q')| = |\text{var}(q)|$. We obtained that the equivalent queries $q$ and $q'$ are over the same relational schemas, have the same active domains, i.e., the same constants occur in the queries, and the same number of non-distinguished variables. Hence, they are identical up to renaming of variables. In particular, the homomorphisms $h$ and $h'$ are such renamings.  

$\square$