4. Containment and Minimization of Conjunctive Queries and Their Unions

Instructions: Work in groups of 2 students. You can write up your answers by hand (provided your handwriting is legible) or use a word processing system like Latex or Word. Note that experience has shown that Word is in general difficult to use for this kind of task. If you prefer to write up your solution by hand, submit a scanned electronic version. Please, include name and email address in your submission.

1. Containment of Selfjoin-free Conjunctive Queries with Built-Ins

For Coursework 3, you have studied the containment problem for conjunctive queries without selfjoins and without built-ins. In this question, we want to also admit built-ins in the queries.

Question: How difficult is it to decide containment of arbitrary conjunctive queries that have no selfjoin (but may have built-ins)? Can this problem be solved in polynomial time? Or is it NP-complete?

Hint: The queries that you consider have the form $Q(\bar{x})$:– L, M, where L is a conjunction of relational atoms and M is a conjunction of built-in atoms. You can assume without loss of generality that whenever for two terms s, t we have that $M \models s = t$, then s and t are syntactically equal. In other words, the queries are normalized in such a way that M does not entail any equalities between the terms in Q. Still in other words, you can exclude queries of the kind $Q(x)$:– $r(x,y), s(y,z), x \leq y, y \leq x$, where the built-in atoms entail that x and y are equal.

(8 Points)

2. Injective and Surjective Mappings

For the proofs in Question 3, it will be necessary to make use of some facts about injective and surjective mappings on finite sets. The purpose of this exercise is to review those facts.

Recall that a mapping $f\colon X \to Y$ from a set X to a set Y is *injective* if for all x, $x' \in X$ we have that $f(x) = f(x')$ implies that $x = x'$. In other words, an injective mapping maps any two distinct elements of X to distinct elements of Y.

Recall as well that a mapping $f\colon X \to Y$ is called *surjective* if for every $y \in Y$ there exists some $x \in X$ such that $f(x) = y$. In other words, every element y of Y has a preimage in X with respect to f.

Recall as well that for mappings $f\colon X \to Y$ and $g\colon Y \to Z$ the *composition* $g \circ f$ is a mapping from X to Z defined as $(g \circ f)(x) = g(f(x))$.

1. Prove that if $f\colon X \to Y$ and $g\colon Y \to Z$ are injective, then $g \circ f$ is injective.
2. Prove that if $f\colon X \to Y$ and $g\colon Y \to Z$ are surjective, then $g \circ f$ is surjective.
3. If the composition $g \circ f$ is injective, what can you conclude about g and f? Prove your answer.
4. Similarly, if the composition $g \circ f$ is surjective, what can you conclude about g and f? Prove your answer.
5. Suppose X is a finite set and $f\colon X \to X$ is injective. What can you conclude about f? Prove your answer.
6. Similarly, suppose X is a finite set and $f\colon X \to X$ is surjective. What can you conclude about f? Prove your answer.

(3 Points)

3. Minimisation of Conjunctive Queries

Recall that *simple* or relational *conjunctive queries* (RCQs) 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 RCQ with n atoms and Q' be an equivalent RCQ 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.

(5 Points)

Proposition 2: Let Q and Q' be two equivalent minimal RCQs. Then Q and Q' are identical up to renaming of variables.

(6 Points)

4. Containment of Unions of Conjunctive Queries

We consider again queries without built-in atoms, which we called relational or simple queries.
Let $Q_1, \ldots, Q_n$ be conjunctive queries, each defined as $Q_i(\bar{x})$:– L_i. Then

$$Q := \bigcup_{i=1}^{n} Q_i$$

defines a new query, the union of the Q_i. Over an instance $\mathbf{I}$, the query Q returns the result $Q(\mathbf{I}) = \bigcup_{i=1}^{n} Q_i(\mathbf{I})$. We call such a query a *union of conjunctive queries*. Note that if Q is the union of the conjunctive queries Q_i, then all the Q_i have the same arity, indicated by using the same vector of distinguished variables $\bar{x}$ for all Q_i. Clearly, the arity of the Q_i is also the arity of Q.
Find out how to decide the following variants of the contaiment problem:

1. "UCQ in CQ": Given a union of conjunctive queries $Q = \bigcup_{i=1}^{n} Q_i$ and a conjunctive query Q', is Q contained in Q'?

2. "CQ in UCQ": Given a conjunctive query Q and a union of conjunctive queries $Q' = \bigcup_{j=1}^{m} Q'_j$, is Q contained in Q'?

3. "UCQ in UCQ": Given a union of conjunctive queries $Q = \bigcup_{i=1}^{n} Q_i$ and union of conjunctive queries $Q' = \bigcup_{j=1}^{m} Q'_j$, is Q contained in Q'?

For each case, give a decidable criterion for containment and show that your criterion is correct. Also, assess the complexity of each problem.

Hint: When proving your criterion for Case 2, you may want to take the proof of the Homomorphism Theorem as a starting point for the proof of your new criterion.

(8 Points)

Submission: 30 May, 11:30 pm, by email