Logic, Data, and Incomplete Information

The Second Tsinghua Logic Summer School

Take-Home Examination Problems
(due Sunday, July 3, 2022 - before midnight Beijing time)

Problem 1: Let R and S be two binary relation schemas with the same attributes A and B. Give a relational algebra expression for the intersection $R \cap S$ so that the difference operation does not occur in that expression.

Problem 2: By definition, a *homomorphism* from a database D_1 to a database D_2 over the same database schema $\mathbf{S}$ is a function $h : adom(D_1) \to adom(D_2)$ such that for every relation schema R of $\mathbf{S}$ and for all $a_1, \ldots, a_n$ in $adom(D_1)$ such that $(a_1, \ldots, a_n) \in R^{D_1}$, we have that $(h(a_1), \ldots, h(a_n)) \in R^{D_2}$.

Here, where *adom* denotes the active domain, and R^{D_1} and R^{D_2} are the relations in D_1 and D_2, respectively, that interpret the relation schema R.

- Show that homomorphisms *compose*, that is, if h_1 is a homomorphism from D_1 to D_2 and if h_2 is a homomorphism from D_2 to D_3, then the composition $h_2 \circ h_1$ is a homomorphism from D_1 to D_3.
- Show that conjunctive queries are *preserved* under homomorphism, that is, if q is a k-ary conjunctive query for some k, if h is a homomorphism from D_1 to D_2, and if $(a_1, \ldots, a_k) \in q(D_1)$, then $(h(a_1), \ldots, h(a_k)) \in q(D_2)$.
- Is the following statement true or false?
  The query $\exists z(E(x, z) \land (z \neq y) \land E(z, y))$ *is preserved under homomorphisms*
  If this statement is true, give a proof of it; if it is false, give a counterexample.

Problem 3: This problem has several parts

1. Show that every s-t tgd is *preserved under target homomorphisms*, which means the following: Assume that σ is a s-t tgd $\forall \mathbf{x}(\varphi(\mathbf{x}) \to \exists \mathbf{y} \psi(\mathbf{x}, \mathbf{y}))$, I is a source instance, and J, J' are target instances such that there is a homomorphism from J to J' (here, homomorphisms are the identity on elements in the active domain of I). If $(I, J) \models \sigma$, then also $(I, J') \models \sigma$.
2. Show that target tgds are not, in general, preserved under homomorphisms, that is, give an example of a target tgd σ and two target instances J and J' such that $J \models \sigma$, there is a homomorphism from J to J', but $J' \not\models \sigma$.
3. By definition, a *retraction* is a homomorphism $h : J \to J'$ such that J' is a sub-instance of J and h is the identity on the active domain of J'. Prove or disprove the following statement: "target tgds are preserved under retractions".

Note: The notion of homomorphism in this exercise is the notion used in the definition of a *universal solution*, that is, the homomorphism is the identity on constants (elements of the active domain of a source instance).

Also, J' being a sub-instance of J means that if a tuple belongs to one of the relations of J', then this tuple also belongs to the corresponding relation of J; in other words, when J' and J are viewed as sets of "facts", we have that J' is a subset of J.

Problem 4: Let $\mathbf{M} = (\mathbf{S}, \mathbf{T}, \Sigma_{st}, \Sigma_t)$ be a schema mappings such that Σ_{st} is a finite set of s-t tgds and Σ_t is a finite set of target tgds and target egds. Let I_1, I_2 be two source instances and let J_1, J_2 be two target instances such that J_i is a universal solution for I_i, $i = 1, 2$. Show that the following statements are equivalent:

1. J_1 and J_2 are homomorphically equivalent.
2. The setof solutions of I_1 w.r.t. $\mathbf{M}$ coincides with the set of solutions of I_2 w.r.t. $\mathbf{M}$.

Note: The notion of homomorphism in this exercise is the notion used in the definition of a *universal solution*, that is, the homomorphism is the identity on constants (elements of the active domain of a source instance).

Optional Challenge Problem This problem is about the minimization problem for conjunctive queries. First a definition: two queries q and q' are *equivalent* if $q(D) = q(D')$ for every database D. Also, we say that a conjunctive query q' is a *minimal equivalent* query to a conjunctive query q if q' is equivalent to q and has as few conjuncts as any conjunctive query that is equivalent to q.

1. Show that for every conjunctive query q, there is a minimal equivalent conjunctive query q' that is formed by deleting zero or more conjuncts of q.
2. Show that the following problem is NP-hard (i.e., some NP-complete problem can be reduced to it): Given two Boolean conjunctive queries q and q', is q' a *minimal* equivalent conjunctive query to q?
3. Design an algorithm that takes as input a Boolean conjunctive query q and returns as output a minimal conjunctive query q' equivalent to q. What is the running time of your algorithm?