NP-compleetheid van conjunctive query containment

In deze oefening tonen we aan dat containment van conjunctive queries NP-compleet is. Voor de eenvoud beperken we ons tot *booleaanse* conjunctive queries: deze hebben een lege head. Een booleaanse conjunctive query Q, toegepast op een database D, geeft als enige antwoordtupel het lege tupel (indien de body van Q matcht in D), ofwel helemaal niets (indien de body niet matcht).

Uit wat we leerden over conjunctive query containment weten we dat voor twee booleaanse conjunctive queries Q_1 en Q_2 geldt: $Q_1 \subseteq Q_2$ als en slechts als er een homomorfisme is van de body van Q_2 naar de body van Q_1.

Definiëer nu formeel het probleem CONTAINMENT = $\{\langle Q_1, Q_2\rangle \mid Q_1$ en Q_2 zijn booleaanse conjunctive queries zodat $Q_1 \subseteq Q_2\}$.

CONTAINMENT in NP: Een certificaat voor lidmaatschap van een input $\langle Q_1, Q_2\rangle$ is een homomorfisme van Q_2 naar Q_1. Het is inderdaad eenvoudig voor een verifier om na te gaan dat een gegeven mapping van de variabelen van Q_2 naar de variabelen van Q_1 inderdaad een homomorfisme is.

NP-compleet: We reduceren 3SAT naar CONTAINMENT. Zij

$$\phi = (a_1 \vee b_1 \vee c_1) \wedge \cdots \wedge (a_\ell \vee b_\ell \vee c_\ell)$$

een booleaanse 3CNF formule over de variabelen $x_1, \ldots, x_k$. Beschouw nu volgende twee conjunctive queries Q_ϕ en Q_{true} over het database schema bestaande uit een ternaire relatie C en een binaire relatie L:

$$Q_\phi() \leftarrow C(a_1, b_1, c_1), \ldots, C(a_\ell, b_\ell, c_\ell), L(x_1, \overline{x_1}), \ldots, L(x_k, \overline{x_k})$$
$$Q_{\text{true}}() \leftarrow C(0,0,1), C(0,1,0), C(0,1,1), \ldots, C(1,1,1), L(1,0), L(0,1)$$

Merk op dat Q_{true} de subgoal $C(0,0,0)$ niet bevat, maar alle andere zeven mogelijke C-subgoals over de elementen 0 en 1 wel. Merk ook op dat Q_ϕ en Q_{true} in polynomiale tijd kunnen geconstrueerd worden vanuit ϕ (Q_{true} hangt tussen haakjes helemaal niet af van ϕ). We tonen nu aan:

$$\phi \text{ satisfiable} \quad \Leftrightarrow \quad Q_{\text{true}} \subseteq Q_\phi$$

$\boxed{\Rightarrow}$ Zij $\alpha : \{x_1, \ldots, x_k\} \to \{0,1\}$ een waarheidstoekenning die ϕ waarmaakt. Breid α uit naar genegeerde literals door $\alpha(\overline{x}) := 1 - \alpha(x)$. Dan is deze α duidelijk een homomorfisme van Q_ϕ naar Q_{true}.

$\boxed{\Leftarrow}$ Zij $h : \{x_1, \overline{x_1}, \ldots, x_k, \overline{x_k}\} \to \{0,1\}$ een homomorfisme van Q_ϕ naar Q_{true}. Omdat elke subgoal $L(x, \overline{x})$ in Q_ϕ wordt afgebeeld op een subgoal $L(1,0)$ of $L(0,1)$ in Q_{true}, is ofwel $h(x) = 1$ en $h(\overline{x}) = 0$, of $h(x) = 0$ en

$h(\overline{x}) = 1$. We kunnen h dus beschouwen als een consistente waarheidstoekenning op $x_1, \ldots, x_k$. Omdat h elke subgoal $C(a_1, b_1, c_\ell)$ afbeeldt op een C-subgoal in Q_{true} die altijd minstens één 1 bevat, is elke clause van ϕ voldaan onder h, en is ϕ dus satisfiable.