

Inclusion de requêtes conjonctives : correction

Exercice 3 Inclusion de requêtes conjonctives

Etant données deux requêtes conjonctives Q_1 et Q_2 , on dit que Q_1 est incluse dans Q_2 (notation $Q_1 \sqsubseteq Q_2$) si, pour toute base de faits F , l'ensemble des réponses à Q_1 dans F est inclus dans l'ensemble des réponses à Q_2 dans F .

1. Montrer la propriété suivante : étant données des requêtes booléennes Q_1 et Q_2 , on a $Q_1 \sqsubseteq Q_2$ (autrement dit toute base de faits qui répond oui à Q_1 répond aussi oui à Q_2) si et seulement si il existe un homomorphisme de Q_2 dans Q_1 .
2. Définir l'homomorphisme de requêtes conjonctives quelconques, dans l'idée d'étendre cette propriété ; puis étendre la propriété à des requêtes conjonctives quelconques en utilisant cette notion.

1. On rappelle que, par définition, la réponse à une requête booléenne conjonctive Q sur une base de faits F est *oui* (plus précisément l'ensemble des réponses à Q sur F est $\{()\}$) s'il existe un homomorphisme de Q dans F .

Soient Q_1 et Q_2 deux requêtes booléennes conjonctives.

\Rightarrow Supposons que $Q_1 \sqsubseteq Q_2$. Considérons la base de faits $F = Q_1$ (ce qu'on peut faire car Q_1 est une formule existentielle conjonctive fermée et a donc la même forme qu'une base de faits). La réponse à Q_1 sur F est oui car il existe un homomorphisme de Q_1 dans F (l'identité). Puisque $Q_1 \sqsubseteq Q_2$, la réponse à Q_2 sur F est également oui. Il existe donc un homomorphisme de Q_2 dans $F = Q_1$.

\Leftarrow Soit h un homomorphisme de Q_2 dans Q_1 . Soit une base de faits quelconque F telle que la réponse à Q_1 sur F est oui. Soit h' un homomorphisme de Q_1 dans F . La composition de ces deux homomorphismes $h' \circ h$ est un homomorphisme de Q_2 dans F . Donc la réponse à Q_2 sur F est oui. F étant une base de faits quelconque, on en déduit que $Q_1 \sqsubseteq Q_2$.

2. On note $Q(x_1 \dots x_k)$ une requête conjonctive où $(x_1 \dots x_k)$ est un ordre sur ses variables libres. Une réponse à $Q(x_1 \dots x_k)$ dans une base de faits F est un tuple de constantes $(h(x_1) \dots h(x_k))$ tel que h est un homomorphisme de Q dans F . De façon équivalente, on peut dire qu'un tuple de constantes $(c_1 \dots c_k)$ est une réponse à $Q(x_1 \dots x_k)$ si la réponse à $Q[x_1/c_1 \dots x_k/c_k]$ est oui, où $Q[x_1/c_1 \dots x_k/c_k]$ est obtenue à partir de Q en remplaçant chaque variable libre x_i par c_i , $1 \leq i \leq k$.

Pour étendre la notion d'homomorphisme à des requêtes conjonctives quelconques il faut prendre en compte les variables libres (variables qui constituent la réponse). Par exemple, considérons les requêtes suivantes :

$$Q_1() = \exists x \exists y p(x, y)$$

$$Q_2(x) = \exists y p(x, y)$$

$$Q_3(y) = \exists x p(x, y)$$

$$Q_4(x, y) = p(x, y).$$

Si l'on considère ces requêtes comme des ensembles d'atomes (des singletons ici) en oubliant la distinction entre variables libres et liées, on a un homomorphisme de chaque requête dans chaque autre. Pourtant, aucune n'est incluse dans une autre au sens de \sqsubseteq . Par exemple, sur la base de faits $F = p(a, b)$, les ensembles de réponses à $Q_1 \dots Q_4$ sont respectivement $\{()\}$, $\{(a)\}$, $\{(b)\}$ et $\{(a, b)\}$.

On définit donc la notion d'homomorphisme "de requête" : un homomorphisme de requête h de $Q_2(x_1 \dots x_k)$ dans $Q_1(y_1 \dots y_k)$ est une application des variables de Q_2 dans les termes de Q_1 telle que :

- (1) $h(Q_2) \subseteq Q_1$ (autrement dit h est un homomorphisme de Q_2 dans Q_1 vues comme des requêtes booléennes), et
- (2) pour tout i de 1 à k , $h(x_i) = y_i$.

Remarque : si les requêtes n'ont pas le même nombre de variables libres, on ne peut pas les comparer car leurs réponses n'ont pas le même format.

On étend la preuve de la question 1 ainsi :

\Rightarrow Supposons que $Q_1(y_1 \dots y_k) \sqsubseteq Q_2(x_1 \dots x_k)$. Considérons la base de faits F obtenue en remplaçant chaque variable libre y_i de Q_1 par une nouvelle constante c_i (qui n'apparaît ni dans F , ni dans Q_1 ni dans Q_2). Il existe un homomorphisme h de Q_1 dans F tel que $h(y_i) = c_i$ pour toute variable libre y_i , et $h(z) = z$ pour toute autre variable. Puisque $Q_1 \sqsubseteq Q_2$, $(c_1 \dots c_k)$ est une réponse à Q_2 dans F . Il existe donc un homomorphisme h' de Q_2 dans F tel que $h'(x_i) = c_i$ pour toute variable libre x_i . A partir de h' on construit un homomorphisme de requête h'' de Q_2 dans Q_1 : $h''(x_i) = y_i$ pour toute variable libre x_i , et $h''(z) = h'(z)$ pour toute autre variable.

\Leftarrow Soit h un homomorphisme de requête de $Q_2(x_1 \dots x_k)$ dans $Q_1(y_1 \dots y_k)$. Soit une base de faits quelconque F et $(h'(y_1) \dots h'(y_k))$ une réponse à Q_1 dans F , où h' est un homomorphisme de Q_1 dans F . La composition de ces deux homomorphismes $h' \circ h$ est un homomorphisme de Q_2 dans F . Cet homomorphisme associe à toute variable libre x_i de Q_2 le terme $h' \circ h(x_i)$: comme $h(x_i) = y_i$ pour tout i , on a $(h' \circ h(x_1) \dots h' \circ h(x_k)) = (h'(y_1) \dots h'(y_k))$, qui est donc une réponse à Q_2 dans F . L'ensemble des réponses à Q_1 dans F est donc inclus dans l'ensemble des réponses à Q_2 dans F . F étant une base de faits quelconque, on en déduit que $Q_1 \sqsubseteq Q_2$.