

Feuille 2 Inclusion de Requêtes Conjonctives

Dans cette feuille on présente un algorithme qui permet de tester l'inclusion de requêtes conjonctives. Pour toute la feuille, on suppose que le schéma de base de données, \mathbf{R} , le domaine \mathbf{dom} et l'ensemble de variables \mathbf{var} sont fixés. On prouve le théorème suivant :

Théorème 1. *Le problème suivant est décidable :*

INPUT : Deux requêtes $q, r \in \text{CQ}$.

QUESTION : $q \subseteq r$?

Remarque 1. *On peut même prouver que le problème est NP-complet. Voir les exercices.*

1 Rappels

On rappelle les termes importants concernant une requête conjonctive dans la Figure 1 ci-dessous :

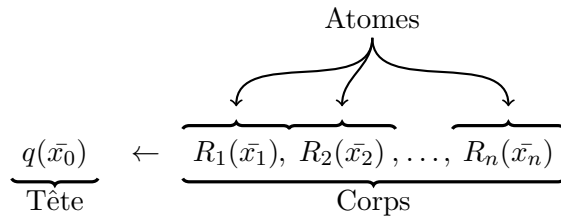


FIGURE 1 – Notations pour une requête conjonctive

De plus on rappelle que pour évaluer une requête conjonctive, on a besoin de la notion d'affectation de variables. Une affectation est une application $f : \mathbf{var} \cup \mathbf{dom} \rightarrow \mathbf{dom}$ telle que $f(d) = d$ pour toute donnée $d \in \mathbf{dom}$. L'évaluation de la requête q sur une instance \mathbf{I} ci-dessus est l'ensemble de tuples suivant :

$$q(\mathbf{I}) = \{f(\bar{x}_0) \mid \text{pour toute affectation } f \text{ telle que } R_i(f(\bar{x}_i)) \in \mathbf{I} \text{ pour tout } i \leq n\}$$

2 Préliminaires

L'algorithme fonctionne selon le principe suivant. On commence par prouver que pour deux requêtes $q, r \in \text{CQ}$, la propriété $q \subseteq r$ est équivalente à une propriété *syntactique* simple de q et r (on appelle ce résultat le *théorème d'homomorphisme*). Il est ensuite facile d'écrire un algorithme qui teste cette propriété.

Homomorphisme entre deux requêtes. On considère deux requêtes q et r que nous représentons ci-dessous,

$$\begin{aligned} q(\bar{x}_0) &\leftarrow R_1(\bar{x}_1), \dots, R_n(\bar{x}_n) \\ r(\bar{y}_0) &\leftarrow S_1(\bar{y}_1), \dots, S_m(\bar{y}_m) \end{aligned}$$

Un homomorphisme de q vers r est une application $h : \mathbf{var} \cup \mathbf{dom} \rightarrow \mathbf{var} \cup \mathbf{dom}$ telle que $h(d) = d$ pour toute donnée $d \in \mathbf{dom}$ (les données ne sont pas modifiées par h) et qui satisfait les deux propriétés suivantes :

1. $h(\bar{x}_0) = \bar{y}_0$ (h envoie le tuple de tête de q sur celui de r).
2. Pour tout $i \leq n$, il existe $j \leq m$ tel que $h(\bar{x}_i) = \bar{y}_j$ et $R_i = S_j$ (h envoie tout atome de q vers un atome de r).

Remarque 2. *Attention, au sens : un homomorphisme va d'une requête vers une autre.*

Exercice 1

Pour chaque paire de requêtes parmi les suivantes, dites si il existe un homomorphisme (et dans ce cas donnez le).

- $q_1(x_1) \leftarrow R(x_1, x_2, x_2), Q('a', x_1, x_1)$.
- $q_2(x_1) \leftarrow R(x_1, x_2, x_2), Q('a', x_3, x_1), R(x_3, x_2, x_4)$.
- $q_3(x_4) \leftarrow R(x_4, x_3, x_3), Q(x_5, x_1, x_2), Q('b', 'c', x_5)$.
- $q_4('c') \leftarrow R('c', 'd', 'd'), Q('a', 'c', 'c')$.

3 Algorithme

On peut maintenant présenter le théorème d'homomorphisme et l'utiliser pour obtenir un algorithme qui teste l'inclusion de requêtes conjonctives.

Théorème 2. *On considère deux requêtes conjonctives r et q . Les deux propriétés suivantes sont équivalentes :*

1. $r \subseteq q$.
2. Il existe un homomorphisme h de q vers r .

Remarque 3. *A nouveau, il faut faire attention au sens. Pour prouver que $r \subseteq q$ il faut trouver un homomorphisme de q vers r (c'est-à-dire de la plus "grande" requête vers la plus "petite").*

Preuve. On suppose fixées les deux requêtes conjonctives q et r que nous notons,

$$\begin{aligned} q(\bar{x}_0) &\leftarrow R_1(\bar{x}_1), \dots, R_n(\bar{x}_n) \\ r(\bar{y}_0) &\leftarrow S_1(\bar{y}_1), \dots, S_m(\bar{y}_m) \end{aligned}$$

Notre objectif est de prouver l'équivalence entre les deux propriétés présentées dans le théorème. On commence par la direction 1) \Rightarrow 2). On suppose que $r \subseteq q$. Observons que si on voit les variables de r comme des données on peut voir les corps de r comme une instance,

$$\mathbf{I}_r = \{S_1(\bar{y}_1), \dots, S_m(\bar{y}_m)\}$$

De plus par définition on a $\bar{y}_0 \in r(\mathbf{I}_r)$. Ainsi, puisque $r \subseteq q$, on a $\bar{y}_0 \in q(\mathbf{I}_r)$. Par définition, cela signifie qu'il existe une affectation f telle que $f(\bar{x}_0) = \bar{y}_0$ et pour tout $i \leq n$, $R_i(\bar{x}_i) \in \mathbf{I}_r$.

Puisque $\mathbf{I}_r = \{S_1(\bar{y}_1), \dots, S_m(\bar{y}_m)\}$, c'est exactement la définition d'un homomorphisme de q vers r ce qui termine la preuve de la direction 1) \Rightarrow 2).

On passe maintenant à la direction 2) \Rightarrow 1). On suppose qu'il existe un homomorphisme h de q vers r et on veut montrer que $r \subseteq q$. On prend donc une instance \mathbf{I} quelconque et on prouve que $r(\mathbf{I}) \subseteq q(\mathbf{I})$. Soit $\bar{c} \in r(\mathbf{I})$. On doit montrer que $\bar{c} \in q(\mathbf{I})$ (c'est-à-dire trouver une affectation appropriée).

Puisque $\bar{c} \in r(\mathbf{I})$, on a une affectation f telle que $f(\bar{y}_0) = \bar{c}$ et pour tout $j \leq m$, $S_j(\bar{y}_j) \in \mathbf{I}$. On prouve maintenant que l'affectation $f \circ h$ obtenue en composant notre homomorphisme h avec f , montre que $\bar{c} \in q(\mathbf{I})$. Puisque h est un homomorphisme, on sait que,

- $f \circ h(\bar{y}_0) = f(h(\bar{x}_0)) = f(\bar{y}_0) = \bar{c}$.
- Pour tout $i \leq n$, on sait qu'il existe $j \leq m$ tel que $R_i = S_j$ et $h(\bar{x}_i) = \bar{y}_j$. On a donc,

$$R_i(f \circ h(\bar{x}_i)) = S_j(f(\bar{y}_j))$$

Et puisqu'on sait que $S_j(f(\bar{y}_j)) \in \mathbf{I}$ par définition de f , on a bien $R_i(f \circ h(\bar{x}_i)) \in \mathbf{I}$. On conclut donc que $\bar{c} \in q(\mathbf{I})$ ce qui termine la preuve. \square

Exercice 2

Utilisez le Théorème 2 pour donner un algorithme qui teste l'inclusion entre les requêtes conjonctives. En déduire un second algorithme qui teste l'équivalence entre les requêtes conjonctives.

Exercice 3

Prouver les propriétés suivantes :

- Prouver que l'inclusion de requêtes conjonctives est dans **NP**.
- Prouver que l'inclusion de requêtes conjonctives est **NP**-difficile (on pourra réduire le problème de 3-Colorabilité d'un graphe).

4 Exercices

Exercice 4

On considère les deux requêtes suivantes $q_1, q_2 \in \text{CQ}$. Déterminer si $q_1 \subseteq q_2$, $q_2 \subseteq q_1$ ou $q_1 \cong q_2$.

- $q_1(x, y) \leftarrow R(x, z), R(z, t), S(t, y)$
- $q_2(x, z) \leftarrow R(x, x), S(x, z)$

Exercice 5

Même question pour les deux requêtes suivantes :

- $q_1(x) \leftarrow R(x, y, x), R(x, z, y), S(y, x)$
- $q_2(x) \leftarrow R(x, y, z), S(y, z)$

Exercice 6

Parmi les requêtes conjonctives suivantes donnez toutes les inclusions et les équivalences et donnez les homomorphismes correspondants. Donnez également une requête *minimale* équivalente à chaque requête :

$$\begin{aligned}
q_1(x_1, x_2) &\leftarrow R(x_1, y_1), R(y_1, y_2), R(y_2, x_2) \\
q_2(x_1, x_2) &\leftarrow R(x_1, y_1), R(y_1, y_2), R(y_2, y_3), R(y_3, x_2) \\
q_3(x_1, x_2) &\leftarrow R(x_1, y_1), R(y_2, y_3), R(y_4, x_2), R(x_1, y_2), R(y_1, y_3), R(y_3, x_2) \\
q_4(x_1, x_2) &\leftarrow R(x_1, y_1), R(y_1, 'cst'), R('cst', y_2), R(y_2, x_2)
\end{aligned}$$

Exercice 7

Donnez les inclusions et équivalences existantes entre les requêtes conjonctives suivantes. Chaque inclusion devra être justifiée en donnant un homomorphisme. (dans les requêtes, "u, v, x, y, z" représentent des variables et "a, b" des constantes).

$$\begin{aligned}
&\text{— } q_1(x, y) \leftarrow T(u, z, x), P(x, z) \\
&\text{— } q_2(x, y) \leftarrow S(a, z), R(x, y), T(z, y, x) \\
&\text{— } q_3(y, z) \leftarrow Q(z, z), P(y, x), T(b, x, y) \\
&\text{— } q_4(x, y) \leftarrow R(x, y) \\
&\text{— } q_5(z, y) \leftarrow P(z, x), T(b, x, z), Q(y, y), P(z, u), T(b, u, z) \\
&\text{— } q_6(x, a) \leftarrow R(x, a), S(a, b), T(b, a, x), Q(a, a), P(x, a)
\end{aligned}$$

Exercice 8

Soient q, q' deux requêtes, on note $q \subset q'$ pour $q \subseteq q'$ et $q' \not\subset q$. On définit $q \supset q'$ de manière similaire. Pour chaque question on cherche un ensemble de requêtes du calcul conjonctif : (indice : dans chaque cas il suffira de considérer une relation d'arité 2)

1. Donnez un ensemble infini $\{q_1, \dots, q_n, \dots\}$ de requêtes n'utilisant qu'une seule relation et aucune constante et telles que : $q_1 \supset \dots \supset q_n \supset \dots$
2. Donnez un ensemble infini $\{q_1, \dots, q_n, \dots\}$ de requêtes n'utilisant qu'une seule relation et aucune constante et telles que : $q_1 \subset \dots \subset q_n \subset \dots$
3. Donnez un ensemble infini $\{q_1, \dots, q_n, \dots\}$ de requêtes n'utilisant qu'une seule relation et aucune constante et telles que : $q_i \not\subset q_j$ et $q_j \not\supset q_i$ pour tout $i, j \geq 1$.

Exercice 9

Parmi les requêtes conjonctives suivantes, donnez les inclusions existantes. On demande de justifier chaque réponse en donnant l'homomorphisme associé.

$$\begin{aligned}
&\text{— } q_1(x, y) \leftarrow S(y, x'), R(x, z'), R(z, z'), S(u, x'), T(z, v, u) \\
&\text{— } q_2(x, y) \leftarrow R(x, y), S(y, z) \\
&\text{— } q_3(x, y) \leftarrow R(x, y), S(y, x), T(x, 'a', y) \\
&\text{— } q_4(x, y) \leftarrow R(x, 'a'), S(y, 'a') \\
&\text{— } q_5(x, y) \leftarrow R(x, z), S(y, z) \\
&\text{— } q_6(x, y) \leftarrow R(x, v), S(y, u), T(x, z, y), T(u, 'a', v)
\end{aligned}$$

Exercice 10

On considère le fragment **UNAIRE** des requêtes conjonctives où toutes les relations utilisées dans le corps de la requête doivent être unaires (i.e. sont d'arité 1). Par exemple la requête $q_1(x, y) \leftarrow R(x), S(y), R(y)$ est **UNAIRE** mais la requête $q_2(x, y) \leftarrow R(x), S(y), T(x, z, y)$ ne l'est pas. On considère le problème d'évaluation des requêtes **UNAIRE** :

INPUT : $q \in \mathbf{UNAIRE}$, I une instance
QUESTION : Calculer $q(I)$

Quelle est la complexité de ce problème ? Est-il réalisable en temps polynomial ou est-il NP-complet ? Dans le premier cas expliquer comment fonctionnerait un algorithme polynomial. Dans le second cas, donner une réduction prouvant la complétude¹.

1. On rappelle que le même problème pour les requêtes conjonctives est NP-complet