Typage et programmation
Examen 2003-2004

François Pottier
Francois.Pottier@inria.fr

18 décembre 2003

La plupart des questions peuvent être traitées sans avoir nécessairement répondu à celles qui précèdent.

Un algorithme d'unification du premier ordre

On se donne un ensemble de *constructeurs* F, chacun étant muni d'une arité entière, positive ou nulle, fixe. On suppose que cet ensemble contient au moins un constructeur F_0 d'arité nulle.

Un *chemin* π est une suite finie d'entiers strictement positifs. On note ϵ le chemin de longueur nulle, i le chemin de longueur 1 constitué de la lettre i, si i est un entier strictement positif, et $\pi.\pi'$ la concaténation des chemins π et π'. Un *arbre* t est une fonction partielle des chemins vers les constructeurs dont le domaine est non vide, clos par préfixe, et cohérent avec l'arité des constructeurs – trois conditions que l'on peut écrire comme suit :

1. $t(\epsilon)$ est défini ;
2. si $t(\pi.\pi')$ est défini, alors $t(\pi)$ l'est également ;
3. $t(\pi.i)$ est défini si et seulement si le constructeur $t(\pi)$ a une arité supérieure ou égale à i.

Un arbre est dit *fini* si son domaine est un ensemble fini, *infini* dans le cas contraire. L'intérêt de cette définition des arbres est d'englober les arbres finis et infinis dans un cadre uniforme.

Question 1. *Supposons donnés deux constructeurs* `int` *et* $\rightarrow$*, d'arités respectives 0 et 2. Quels arbres peut-on naturellement associer aux notations usuelles* $\mathtt{int} \rightarrow \mathtt{int}$ *et* $\mu\mathtt{X}.\mathtt{int} \rightarrow \mathtt{X}$ *? (On en explicitera le domaine et on indiquera, parmi les chemins éléments du domaine, lesquels ont pour image* `int` *et lesquels ont pour image* $\rightarrow$*.) Ces arbres sont-ils finis ou infinis ?* $\diamond$

Si $t(\pi)$ est défini, le *sous-arbre de t à la position π*, noté t/π, est l'arbre qui à π' associe $t(\pi.\pi')$. Un arbre est dit *régulier* s'il admet un nombre fini de sous-arbres distincts. Si F est un constructeur d'arité n et si $t_1, \ldots, t_n$ sont des arbres, on notera $F(t_1, \ldots, t_n)$ l'arbre t défini par $t(\epsilon) = F$ et $t(i.\pi) = t_i(\pi)$ pour $i \in \{1, \ldots, n\}$.

Question 2. *On définit souvent les arbres finis comme les termes engendrés par la définition inductive suivante :*

$$T ::= F\,T_1\,\ldots\,T_n \qquad \text{où } n \text{ est l'arité de } F$$

Indiquer comment tout terme T de cette forme peut être interprété comme un arbre au sens de la définition précédente, noté $[\![T]\!]$, et vérifier que cet arbre est fini. Réciproquement, vérifier que tout arbre fini est l'interprétation d'un tel terme. $\diamond$

Dans la suite, on s'intéresse à la résolution de contraintes d'égalité entre termes avec variables. On se donne un ensemble infini de *variables* `X`. La syntaxe des *termes* `T` est alors définie inductivement :

$$\mathtt{T} ::= \mathtt{X} \mid F\,\mathtt{T}_1\,\ldots\,\mathtt{T}_n \qquad \text{où } n \text{ est l'arité de } F$$

Une *multi-équation* e est un ensemble de termes. La syntaxe des contraintes est alors la suivante :

$$C ::= \texttt{false} \mid \mathtt{T} =^? \mathtt{T} \mid e \mid C \wedge C \mid \exists \mathtt{X}.C$$

On a distingué les équations binaires $\mathtt{T}_1 =^? \mathtt{T}_2$ des multi-équations e, car ces deux formes satisferont des invariants différents et seront représentées de façon différente en mémoire. Néanmoins, d'un point de vue logique, il s'agit, dans les deux cas, de contraintes d'égalité.

On peut se donner (au moins) trois interprétations naturelles de ces contraintes.

Question 3. *Définir une interprétation des termes et des contraintes dans l'ensemble de tous les arbres (finis ou infinis). (On définira la notion de* valuation ϕ *et les prédicats de* satisfaction $\vdash$*, d'implication* $\Vdash$ *et d'*équivalence $\equiv$ *de contraintes.) Indiquer comment la modifier pour obtenir une interprétation dans l'ensemble des arbres réguliers, ou encore une interprétation dans l'ensemble des arbres finis.* ◇

On note $\bar{\mathtt{X}}$ une multi-équation dont tous les membres sont des variables. On note $\mathtt{T} = e$ ou $e = \mathtt{T}$ la multi-équation $\{\mathtt{T}\} \cup e$, c'est-à-dire la multi-équation dont les membres sont $\mathtt{T}$ et les membres de e. On pourra noter $\mathtt{X}$ la multi-équation $\{\mathtt{X}\}$, dont l'unique membre est $\mathtt{X}$. On note $\mathbb{T}$ un *contexte*, c'est-à-dire un terme doté d'un trou :

$$\mathbb{T} ::= [] \mid F\,\mathtt{T}_1 \ldots \mathbb{T} \ldots \mathtt{T}_n$$

L'algorithme d'unification est alors défini par le système de réécriture suivant :

$\mathtt{T} =^? \mathbb{T}[F\,\mathtt{X}_1 \ldots \mathtt{X}_n]$	$\rightarrow$	$\exists \mathtt{X}.(\mathtt{T} =^? \mathbb{T}[\mathtt{X}] \wedge \mathtt{X} = F\,\mathtt{X}_1 \ldots \mathtt{X}_n)$ si $\mathtt{X} \notin \mathrm{ftv}(\mathtt{T}, \mathbb{T}, \mathtt{X}_1, \ldots, \mathtt{X}_n)$	DÉCOUPAGE
$(\exists \mathtt{X}.C_1) \wedge C_2$	$\rightarrow$	$\exists \mathtt{X}.(C_1 \wedge C_2)$ si $\mathtt{X} \notin \mathrm{ftv}(C_2)$	EXTRUSION
$\mathtt{X}_1 =^? \mathtt{X}_2 \wedge \mathtt{X}_1 = \mathtt{X}_2 = e$	$\rightarrow$	$\mathtt{X}_1 = \mathtt{X}_2 = e$	TAUTOLOGIE
$\mathtt{X}_1 =^? \mathtt{X}_2 \wedge \mathtt{X}_1 = \bar{\mathtt{X}}_1 \wedge \mathtt{X}_2 = e$	$\rightarrow$	$\mathtt{X}_1 = \bar{\mathtt{X}}_1 = \mathtt{X}_2 = e$	FUSION
$\mathtt{X}_1 =^? \mathtt{X}_2 \wedge \mathtt{X}_1 = \bar{\mathtt{X}}_1 = F\,\mathtt{X}_1^1 \ldots \mathtt{X}_1^n \wedge \mathtt{X}_2 = \bar{\mathtt{X}}_2 = F\,\mathtt{X}_2^1 \ldots \mathtt{X}_2^n$	$\rightarrow$	$\mathtt{X}_1 = \bar{\mathtt{X}}_1 = \mathtt{X}_2 = \bar{\mathtt{X}}_2 = F\,\mathtt{X}_1^1 \ldots \mathtt{X}_1^n \wedge \wedge_{i=1}^n \mathtt{X}_1^i =^? \mathtt{X}_2^i$	PROPAGATION
$\mathtt{X}_1 =^? \mathtt{X}_2 \wedge \mathtt{X}_1 = \bar{\mathtt{X}}_1 = F_1\,\mathtt{X}_1^1 \ldots \mathtt{X}_1^{n_1} \wedge \mathtt{X}_2 = \bar{\mathtt{X}}_2 = F_2\,\mathtt{X}_2^1 \ldots \mathtt{X}_2^{n_2}$	$\rightarrow$	$\texttt{false}$ si $F_1 \neq F_2$	ÉCHEC
$C \wedge \texttt{false}$	$\rightarrow$	$\texttt{false}$	ÉCHEC-$\wedge$
$\exists \mathtt{X}.\texttt{false}$	$\rightarrow$	$\texttt{false}$	ÉCHEC-$\exists$

La réécriture se fait modulo commutativité et associativité de la conjonction, modulo commutativité du symbole $=^?$, et modulo un contexte arbitraire.

On peut vérifier que la réécriture préserve l'invariant suivant : soit la contrainte courante contient `false`, soit (i) toute variable (qu'elle soit libre ou existentiellement quantifiée) est membre d'exactement une multi-équation ; et (ii) toute multi-équation est de la forme soit $\bar{\mathtt{X}}$ soit $\bar{\mathtt{X}} = F\,\mathtt{X}_1 \ldots \mathtt{X}_n$. (On ne demande pas de démontrer cela.) Dans la suite, on ne s'intéressera qu'à des contraintes satisfaisant cet invariant. Les propriétés (i) et (ii) indiquent que les multi-équations constituent une *partition* des variables du problème, et qu'à chaque classe d'équivalence peut être associé un terme de la forme $F\,\mathtt{X}_1 \ldots \mathtt{X}_n$.

Question 4. *À titre d'exemple, supposons qu'on dispose des constructeurs* `int` *et* `bool`*, d'arité* 0*, et des constructeurs* $\rightarrow$ *et* $\times$*, d'arité* 2*. On considère la contrainte suivante, laquelle possède les propriétés (i) et (ii) définies plus haut :*

$$\mathtt{X} \wedge \mathtt{Y} \wedge \mathtt{Z} \wedge \mathtt{X} =^? \mathtt{int} \rightarrow (\mathtt{Y} \times \mathtt{Z}) \wedge \exists \mathtt{Y}.(\mathtt{Y} \wedge \mathtt{W} = \mathtt{bool} \wedge \mathtt{X} =^? \mathtt{Y} \rightarrow (\mathtt{W} \times \mathtt{W}))$$

Quelles en sont les variables libres ? Donner une série de réductions permettant d'amener cette contrainte en forme normale, en indiquant le nom de la (des) règle(s) employée(s) à chaque étape. Indiquer, sans justification, si cette contrainte est satisfiable et quelles sont les valuations qui la satisfont. ◇

Question 5. *Démontrer que toute forme normale (c'est-à-dire toute contrainte irréductible) vis-à-vis de* $\rightarrow$ *est soit* `false`*, soit de la forme* $\exists \bar{\mathtt{X}}.C$*, où* C *est une conjonction de multi-équations satisfaisant les propriétés (i) et (ii) définies plus haut.* ◇

Question 6. *Démontrer que, si termes et contraintes sont interprétés dans l'ensemble de tous les arbres (finis ou infinis), alors toute forme normale autre que* `false` *est satisfiable. Pour cela, on s'intéressera à une conjonction C de multi-équations satisfaisant les propriétés (i) et (ii) définies plus haut, et on exhibera une valuation ϕ qui la satisfait.* ◇

Question 7. *D'après la question précédente, toute forme normale pour $\rightarrow$ est soit* `false`*, soit satisfiable. Quelles autres propriétés de $\rightarrow$ faut-il démontrer pour pouvoir affirmer que le problème de la satisfiabilité d'une contrainte (dans le modèle constitué de tous les arbres, finis ou infinis) est décidable ?* ◇

Question 8. *Vérifier que la valuation ϕ exhibée par la question 6 est en fait régulière, c'est-à-dire que, pour tout* `X`*, ϕ(*`X`*) est un arbre régulier. En déduire qu'une contrainte est satisfiable dans le modèle constitué de tous les arbres (finis ou infinis) si et seulement si elle est satisfiable dans le modèle constitué des arbres réguliers uniquement.* ◇

Question 9. *De quelle façon l'algorithme doit-il être modifié pour décider le problème de la satisfiabilité dans le modèle des arbres finis ? De quelle façon le développement précédent, et en particulier la preuve qui répond à la question 6, est-il alors modifié ?* ◇