Projet du cours « Compilation »
Jalon 2 : Interprétation de HOPIX

version numéro lundi 4 janvier 2016, 13 :21 :48 (UTC+0100)

1 Syntaxe abstraite

Dans les sections suivantes, lorsque l'on utilisera de la syntaxe concrète dans les spécifications, on fera référence implicitement aux arbres de syntaxe abstraites sous-jacents définis par la grammaire suivante :

p	::=	$\overline{\text{d}}$	*Programme*
d	::=	`type` T $\overline{\alpha}$:= d_t	*Définition de type*
	|	`extern` x : τ	*Déclaration d'une valeur externe*
	|	d_v	*Définition de valeur(s)*
d_t	::=	`{` $\overline{\ell : \tau}$ `}`	*Type d'enregistrement*
	|	`{` $\overline{K : \overline{\tau}}$ `}`	*Type somme*
	|	•	*Type abstrait*
d_v	::=	`val` x $\overline{\text{m}}$ $:\tau^?$ `:=` e	*Valeur simple*
	|	`rec` $\overline{x\ \overline{\text{m}} :\tau^? \texttt{:=}\ \text{e}}$	*Valeurs mutuellement récursives*
τ	::=	T`[`$\overline{\tau}$`]`	*Type construit*
	|	α	*Variable de type*
e	::=	n	*Entier*
	|	b	*Booléen*
	|	c	*Caractère*
	|	s	*Chaîne de caractères*
	|	x	*Variable*
	|	K`(`$\overline{\text{e}}$`)`	*Construction d'une donnée étiquetée*
	|	`{`$\overline{\ell$ `:=` e`}`	*Construction d'un enregistrement*
	|	`(`e : τ`)`	*Annotation de type*
	|	d_v `;` e	*Définition locale*
	|	e e	*Application*
	|	`\`m `=>` e	*Fonction anonyme*
	|	e`#`ℓ	*Accès à un champ*
	|	e`#`ℓ `<-` e	*Modification d'un champ*
	|	e `?` $\overline{\text{b}}$	*Analyse de motifs*
	|	`if` e `then` e `else` e `fi`	*Conditionnelle*
b	::=	m `=>` e	*Cas d'analyse*
m	::=	K	*Etiquette*
	|	x	*Motif universel liant*
	|	`_`	*Motif universel non liant*
	|	`(`m : τ`)`	*Annotation de type*
	|	`{`$\overline{\ell$ `=` m`}`	*Enregistrement*
	|	n	*Entier*
	|	b	*Booléen*
	|	c	*Caractère*
	|	s	*Chaîne de caractères*
	|	K`(`$\overline{\text{m}}$`)`	*Valeurs étiquetées*
	|	m `|` m	*Disjonction*
	|	m `&` m	*Conjonction*

Dans cette grammaire, on a utilisé les *métavariables* suivantes :
— T pour représenter un constructeur de type.
— x pour représenter un identificateur de valeur.
— ℓ pour représenter une étiquette de champs d'enregistrement.
— K pour représenter le nom d'un constructeur de données.
— τ pour représenter un type.
— e pour représenter une expression.

— m pour représenter un motif.
— n pour représenter un litéral de type entier.
— b pour représenter un litéral de type booléen.
— c pour représenter un litéral de type caractère.
— s pour représenter un litéral de type chaîne de caractères.
— α pour représenter une variable de type.
— b pour représenter une branche d'analyse de motifs.
Par ailleurs, on a aussi écrit $\overline{X}$ pour représenter une liste potentiellement vide de X.

2 Interprétation

2.1 Valeurs

Les valeurs du langage sont définies par la grammaire suivante :

v	::=	n	*Entier*
	|	b	*Booléen*
	|	c	*Caractère*
	|	s	*Chaîne de caractères*
	|	()	*Unité*
	|	$K(\overline{v})$	*Valeur étiquetée*
	|	a	*Adresse d'un enregistrement*
	|	(m => e)$[E]$	*Fermeture*

Environnement d'évaluation Les identificateurs de programme sont associés à des valeurs à l'aide d'un environnement d'évaluation E. On écrit « $E[x]$ » pour parler de la valeur de x dans l'environnement E. On écrit « $E + x \mapsto v$ » pour parler de l'environnement E étendu par l'association entre l'identificateur x et la valeur v.

Valeur d'enregistrement Une valeur d'enregistrement $\mathcal{L}$ est écrite « $\{\overline{\ell \ = \ v}\}$ ». On écrit « $\mathcal{L}[\ell$:= $v]$ » pour représenter l'enregistrement similaire en tout champ à $\mathcal{L}$ sauf pour le champ ℓ qui vaut v.

Les valeurs d'enregistrement sont allouées dynamiquement dans une mémoire M. On écrit « $M + a \mapsto \mathcal{L}$ » pour représenter la mémoire qui étend la mémoire M avec une nouvelle adresse a où se trouve stockée la valeur d'enregistrement $\mathcal{L}$. On écrit « $M[a \leftarrow \mathcal{L}]$ » pour représenter la mémoire M modifiée seulement à l'adresse a en y stockant la valeur d'enregistrement $\mathcal{L}$. Enfin, on écrit « $(a \mapsto \mathcal{L}) \in M$ » pour indiquer que la valeur d'enregistrement $\mathcal{L}$ est stockée à l'adresse a de la mémoire M.

2.2 Évaluation des programmes

Un programme p s'évalue à partir d'une mémoire vide et d'un environnement vide en évaluant successivement les définitions de valeurs dans leur ordre d'apparition dans le programme.

Pour spécifier précisément ce processus, il faut donc décrire la façon dont les définitions de valeurs s'évaluent : c'est le rôle de la section 2.3. Les expressions sont les termes qui s'évaluent en des valeurs. Leur évaluation est spécifiée dans la section 2.4. Elle s'appuie sur l'évaluation de l'analyse par cas (section 2.5) et l'analyse de motifs (section 2.6).

2.3 Évaluation des définitions

L'évaluation des définitions s'appuie sur le jugement :

$$E, M \vdash \mathrm{d}_v \Rightarrow E', M'$$

qui se lit « Dans l'environnement E et la mémoire M, la définition d_v étend E et E' et modifie M en M'.

Cette partie de la spécification est omise volontairement. Vous devez réfléchir à une spécification raisonnable.

2.4 Évaluation des expressions

Le jugement d'évaluation des expressions s'écrit :

$$E, M \vdash \mathrm{e} \Downarrow v, M'$$

et se lit « Dans l'environnement d'évaluation E, l'expression e s'évalue en v et change la mémoire M en M'. »

Il est défini par les règles suivantes :

$$\frac{}{E, M \vdash n \Downarrow n, M} \qquad \frac{}{E, M \vdash c \Downarrow c, M} \qquad \frac{}{E, M \vdash b \Downarrow b, M} \qquad \frac{}{E, M \vdash s \Downarrow s, M} \qquad \frac{E(x) = v}{E, M \vdash x \Downarrow v, M}$$

$$\frac{M_0 = M \qquad \forall i \in [1 \dots n], E, M_{i-1} \vdash \mathrm{e}_i \Downarrow v_i, M_i}{E, M \vdash K(\mathrm{e}_1, \dots, \mathrm{e}_n) \Downarrow K(v_1, \dots, v_n), M_n}$$

$$\frac{M_0 = M \qquad \forall i \in [1 \dots n], E, M_{i-1} \vdash \mathrm{e}_i \Downarrow v_i, M_i \qquad M' = M_n + a \mapsto \{\ell_1 \texttt{ := } v_1, \dots, \ell_n \texttt{ := } v_n\}}{E, M \vdash \{\ell_1 \texttt{ := } \mathrm{e}_1, \dots, \ell_n \texttt{ := } \mathrm{e}_n\} \Downarrow a, M'} \qquad \frac{E, M \vdash \mathrm{e} \Downarrow v, M'}{E, M \vdash (\mathrm{e} : \tau) \Downarrow v, M'}$$

$$\frac{E, M \vdash \mathrm{d}_v \Rightarrow E', M' \qquad E', M' \vdash \mathrm{e} \Downarrow v, M''}{E, M \vdash \mathrm{d}_v; \mathrm{e} \Downarrow v, M''}$$

$$\frac{E, M \vdash \mathrm{e}_1 \Downarrow (\mathrm{m} \texttt{ => } \mathrm{e})[E'], M' \qquad E, M' \vdash \mathrm{e}_2 \Downarrow v', M'' \qquad E', M'' \vdash \mathrm{m} \sim v' \Uparrow E'' \qquad E'', M'' \vdash \mathrm{e} \Downarrow v, M'''}{E, M \vdash \mathrm{e}_1\, \mathrm{e}_2 \Downarrow v, M'''}$$

$$\frac{E, M \vdash \mathrm{e} \Downarrow a, M' \qquad (a \mapsto \{\dots \ell \texttt{ := } v \dots\}) \in M'}{E, M \vdash \mathrm{e}\#\ell \Downarrow v, M'} \qquad \frac{E, M \vdash \mathrm{e} \Downarrow a, M' \qquad (a \mapsto \mathcal{L}) \in M' \qquad E, M' \vdash \mathrm{e}' \Downarrow v, M''}{E, M \vdash \mathrm{e}\#\ell \texttt{ <- } \mathrm{e}' \Downarrow (), M''[a \leftarrow \mathcal{L}[\ell \texttt{ := } v]]}$$

$$\frac{E, M \vdash \mathrm{e} \Downarrow v_s, M' \qquad E, M' \vdash v_s \sim \overline{\mathrm{b}} \Downarrow v, M''}{E, M \vdash \mathrm{e}\,\texttt{?}\ \overline{\mathrm{b}} \Downarrow v, M''}$$

2.5 Analyse par cas

L'analyse par cas d'une valeur v consiste à traiter une liste de cas représentés par des branches $\overline{\mathrm{b}}$ de la forme « m `=>` e » en évaluant la première de ces branches dont le motif capture la valeur v. Ce dernier mécanisme est représenté dans la section suivante.

$$\frac{E, M \vdash v \sim \mathrm{m} \Uparrow E' \qquad E', M \vdash \mathrm{e} \Downarrow v', M'}{E, M \vdash v \sim (\mathrm{m} \texttt{ => } \mathrm{e})\overline{\mathrm{b}} \Downarrow v', M'} \qquad \frac{E, M \vdash v \not\sim \mathrm{m} \qquad E, M \vdash v \sim \overline{\mathrm{b}} \Downarrow v', M'}{E, M \vdash v \sim (\mathrm{m} \texttt{ => } \mathrm{e})\overline{\mathrm{b}} \Downarrow v', M'}$$

2.6 Analyse de motifs

L'évaluation des motifs s'appuie sur le jugement :

$$E, M \vdash v \sim \mathrm{m} \Uparrow E'$$

qui se lit « Dans l'environnement E et dans la mémoire M, la valeur v est capturée par le motif m en étendant l'environnement E en E'. »

Cette partie de la spécification est omise volontairement. Vous devez réfléchir à une spécification raisonnable.

3 Travail à effectuer

La seconde partie du projet est l'écriture de l'interprète de la sémantique opérationnelle décrite plus haut.
La compilation s'effectue par la commande« `make` ».
Le projet est à rendre **avant le** :

24 janvier 2015 à 23h59

Pour finir, vous devez vous assurer des points suivants :

— Le projet contenu dans cette archive **doit compiler**.
— Vous devez **être les auteurs** de ce projet.
— Il doit être rendu **à temps**.

Si l'un de ces points n'est pas respecté, la note de 0 vous sera affectée.

4 Log