Semaine 12: Interprétation de Lisp

Nous présentons maintenant un interpréteur pour Lisp.

C'est utile pour deux raisons.

1. De nombreux systèmes informatiques incluent un petit langage, qui est souvent interprété.
2. L'interpréteur nous dit de manière précise comment les programmes Lisp en particulier, et les programmes fonctionnels en général, sont évalués.

Conception de l'interpréteur

L'interpréteur prend en entrée une expression de type *Data*, qui est représenté par la structure de donnée présentée la semaine dernière :

```
abstract class Data;
case class NUM(x: Int)                     extends Data;
case class STR(x: String)                  extends Data;
case class SYM(name: String)               extends Data;
case class NIL                             extends Data;
case class CONS(car: Data, cdr: Data)      extends Data;
```

L'interpréteur retourne une autre expression *Data* en sortie, qui représente une valeur Lisp.

Par exemple, s'il est appliqué à l'expression d'entrée

CONS(SYM(""), CONS(NUM(2), CONS(NUM(7), NIL)))*

l'interpréteur doit retourner l'expression de sortie

NUM(14)

Mais maintenant, une question se pose : Quand l'entrée de l'interpréteur est un symbole, tel que *SYM("x")*, quelle doit-être la sortie ?

Cela dépend si le nom x est défini au point où le symbole est évalué, et si oui, à quelle valeur il est lié.

L'interpréteur a besoin de mémoriser les noms définis dans un *environnement.*

Environnements

Un environnement est une structure de données qui associe des valeurs Lisp à des noms.

Les deux opérations fondamentales sur un environnement sont :

- *lookup* : Étant donné un nom, renvoie la valeur associée à ce nom.
- *extend* : Étant donné une liaison nom/valeur, étendre l'environnement avec cette liaison en renvoyant un nouvel environnement.

Cela mène au design de classe suivant :

```
class Environment with {
  def lookup(n: String): Data = ...
  def extend(name: String, v: Data) = ...
}
```

On a aussi besoin de définir une valeur *EmptyEnvironment*, qui représente l'environnement vide.

L'évaluateur

On présente maintenant l'évaluateur Lisp.

Il prend en entrée la représentation interne d'une expression Lisp, qui peut être :

- un atome : symboles, nombres, chaînes de caractères.
- une combinaison, qui à son tour peut être une forme spéciale ou une application.

On réduit les formes spéciales Lisp pour lier un nom (ex. *define*, ***let***, *letrec*, *lambda*, ...) à seulement ***val*** (avec la signification de Scala) et *lambda*.

Si bien, qu'au lieu de la définition Lisp

```
(define (add m n) (+ m n)) <rest of code>
```

on écrira

```
(val add (lambda (m n)(+ m n)) <rest of code>)
```


Remarque : l'exemple montre la puissance du filtrage de motif – en une seule expression, on peut spécifier que

- la liste d'arguments doit avoir pour longueur deux, et que
- les deux arguments doivent être des instances de *NUM*,
- et on rend accessibles les éléments des deux arguments à travers les noms *arg1* et *arg2*.

Notez que l'expression ***case*** est elle-même une fonction, de type (*List*[*Data*])*Data*, qui filtre l'argument de *FUN*.

Fonctions prédéfinies

Il doit aussi y avoir un moyen d'interpréter les symboles ¡¡standards¿¿ tels que $*$.

On peut utiliser un environnement initial pour enregistrer les valeurs de tels symboles.

Mais quelle doit être la valeur de $*$?

On a besoin de créer une nouvelle sous-classe de *Data*, qui représente les fonctions sur les expressions *Lisp* :

```
case class FUN(f: (List[Data])Data) extends Data;
```

Par exemple, l'opération $*$ sera associée à la valeur :

```
FUN(case [NUM(arg1), NUM(arg2)] ⇒ NUM(arg1 * arg2))
```


Différentes structures de données peuvent être utilisées pour implanter les environnements avec différents compromis de performance. (Exemples ?)

Nous utilisons ici une approche directe, sans structure de données auxiliaire.

```
abstract class Environment with {
  abstract def lookup(n: String): Data;
  def extend(name: String, v: Data) = new Environment with {
    def lookup(n: String): Data = if (n == name) v else outer.lookup(n);
  }
}
val EmptyEnvironment = new Environment with {
  def lookup(n: String): Data = error("undefined: " + n);
}
```

C'est court et simple, mais la méthode *lookup* prend un temps proportionnel au nombre de liaisons dans l'environnement.

Plus tard, on ajoutera aussi le ***def*** Scala, mais seulement pour les fonctions sans paramètres (vu qu'il y a déjà *lambda* pour exprimer la paramétrisation).

La fonction d'évaluation

```
def eval(x: Data, env: Environment): Data = x match {
  case NUM(_), STR(_), FUN(_) => x
  case SYM(name) =>
    env lookup name
  case CONS(SYM("val"), CONS(SYM(name),
                    CONS(body, CONS(rest, NIL)))) =>
    eval(rest, env.extend(name, eval(body, env)))
  case CONS(SYM("if"), CONS(cond,
                    CONS(thenpart, CONS(elsepart, NIL)))) =>
    if (toBoolean(eval(cond, env))) eval(thenpart, env)
    else eval(elsepart, env)
  case CONS(SYM("quote"), CONS(expr, NIL)) =>
    expr
  case CONS(SYM("lambda"), CONS(params, CONS(body, NIL))) =>
    mkLambda(params, body, env)
  case CONS(operator, operands) =>
    apply(eval(operator, env), toList(operands) map (x => eval(x, env))) }
```


Explications

- Si l'expression d'entrée x est un nombre, une chaîne de caractères ou une fonction, on retourne l'expression elle-même.
- Sinon, si l'expression est un symbole, on retourne le résultat de la recherche du symbole dans l'environnement courant.
- Sinon, si l'expression est une forme spéciale

  (**val** *<name> <body> <rest>*)

  on évalue l'expression *<rest>* dans un environnement étendu par la liaison de *<name>* au résultat de l'évaluation de *<body>*.
- Sinon, si l'expression est une forme spéciale

  (**if** *<cond> <thenpart> <elsepart>*)

  on évalue *<cond>*. Si le résultat est le nombre *0*, on continue avec l'évaluation de *<thenpart>* ; sinon avec l'évaluation de *<elsepart>*. Le test utilise la fonction *toBoolean*, qui est définie ci-dessous.


```
def toBoolean(x: Data) = x match {
  case NUM(0) => False  case _ => True
}
```

- Sinon, si l'expression est une forme spéciale

  (*quote <expr>*)

  on retourne *expr* comme résultat de l'évaluation.
- Sinon, si l'expression est une forme spéciale

  (*lambda <params> expr*)

  on crée une nouvelle fonction en utilisant l'opération *mkLambda* (voir ci-dessous).
- Enfin, si l'expression est une combinaison qui n'est aucune des formes spéciales ci-dessus, ça ne peut être qu'une application de fonction.

  Dans ce cas, on évalue l'opérateur de l'application, ainsi que toutes ses opérandes, et on *applique* ensuite la valeur de l'opérateur aux valeurs des opérandes.

La fonction *apply* vérifie que l'opérateur est bien un noeud *FUN* et applique ensuite l'opération associée à la liste d'arguments.

```
def apply(fn: Data, args: List[Data]): Data = fn match {
  case FUN(f) => f(args);
  case _ => error("application of non-function: " + fn);
}
```

La fonction *toList* convertit une liste Lisp construite à partir de *NIL* et *CONS* en une liste Scala. Elle est définie comme suit.

```
def toList(x: Data): List[Data] = x match {
  case NIL => []
  case CONS(y, ys) => y :: toList(ys)
  case _ => error("malformed list: " + x);
}
```

C'est tout, à part que nous devons encore définir comment une expression *LAMBDA* est convertie en une fonction qui peut être appliquée.

Construction des fonctions

La conversion d'une expression Lisp en une fonction Scala utilise la fonction auxiliaire *mkLambda*.

L'appel à *mkLambda(params, body, env)* renvoie un objet *FUN* avec une fonction (Scala).

Cette fonction associe à une liste d'arguments, qui correspond à la liste de paramètres formels *params*, le résultat de l'évaluation de *body* dans *env*.

Voici comment ça marche :

```
def mkLambda(params: Data, body: Data, env: Environment): Data = {
  val ps: List[String] = toList(params) map {
    case SYM(name) => name
    case _ => error("illegal parameter list");
  }
  FUN(args => eval(body, extendEnv(env, ps, args)))
}
```


- Ainsi, une application de la fonction provoque l'évaluation de *body* dans une environnement étendu par les liaisons qui associent aux noms des paramètres formels les valeurs des arguments effectifs correspondants.
- Notez l'analogie avec le modèle par substitution, où les paramètres formels étaient *remplacés* dans *body* par les arguments effectifs *args*.
- On peut penser un environnement comme une substitution différée : plutôt que faire le remplacement immédiatement, on le réalise quand un symbole est recherché dans l'environnement.
- Il serait aussi possible d'écrire un interpréteur avec substitutions plutôt qu'avec environnements ; mais cet interpréteur serait plus compliqué et moins efficace.

Il reste à définir la fonction *extendEnv* qui étend l'environnement par une liste de liaisons entre les paramètres formels et les valeurs des arguments.

Cette fonction est définie comme suit.

```
def extendEnv(env: Environment,
              ps: List[String], args: List[Data]): Environment =
  (ps, args) match {
    case ([], []) =>
      env
    case (p :: ps1, arg :: args1) =>
      extendEnv(env.extend(p, arg), ps1, args1)
    case _ =>
      error("wrong number of arguments")
  }
```

La fonction vérifie que les deux listes, celle des noms des paramètres et celle des valeurs des arguments, ont la même longueur.

Pour chaque couple nom/valeur, elle applique la méthode *extend* de l'environnement.

Un autre problème avec le schéma du Lisp original est qu'il est très dangereux de retourner des fonctions. Considérons :

$$\textbf{\textit{def}}\ \mathit{incrementer}(x\colon \mathit{Int}) = y \Rightarrow y + x$$

Question : Quel effet a $\mathit{incrementer}(2)(3)$?

Le schéma d'implantation du Lisp original est appelé portée dynamique.

Cela signifie que l'association entre un identificateur et sa valeur est déterminée dynamiquement – elle dépend de la forme de l'environnement au moment où la valeur est référencée.

La portée dynamique offre des possibilités intéressantes, mais elle nuit sévèrement à l'utilisation des fonctions d'ordre supérieur.

Si bien que la plupart des versions modernes de Lisp telles que Scheme ont une portée statique, exactement comme Scala.

Les versions plus anciennes, elisp inclus, ont encore la portée dynamique.

Common Lisp, bien sûr, a les deux.

Par exemple, après renommage :

$$\begin{array}{l}\textbf{\textit{def}}\ f(x\colon \mathit{Int}) = g(y \Rightarrow x + y)\\ \textbf{\textit{def}}\ g(z\colon (\mathit{Int})\mathit{Int}) = z(2);\\ f(1)\end{array}$$

le programme retourne 3, comme attendu.

Mais le programme suivant montre qu'on ne peut pas toujours éviter les problèmes par le renommage.

$$\begin{array}{l}\textbf{\textit{def}}\ \mathit{iterate}(n\colon \mathit{Int},\ f\colon (\mathit{Int})\mathit{Int}) =\\ \quad \textbf{\textit{if}}\ (n == 0)\ (x \Rightarrow x)\\ \quad \textbf{\textit{else}}\ \mathit{nTimes}(n-1,\ (x \Rightarrow f(x)))\end{array}$$

Cette fonction est censée composer son paramètre f n-fois avec lui-même.

Question : Quel effet a l'évaluation de cette fonction avec l'environnement de pile du Lisp original ?

Portée dynamique dans le Lisp original

La version originale de Lisp utilisait à la place une pile globale pour les environnements.

A l'entrée d'une fonction, les liaisons pour les paramètres et les variables locales étaient empilées.

A la sortie d'une fonction, la pile était effacée.

Mais les fonctions d'ordre supérieur se comportent alors très étrangement !

Par exemple, l'évaluation de $f(1)$ dans le programme ci-dessus résulterait en une erreur à l'exécution, car la fonction anonyme $y \Rightarrow x + y$ accède à la dernière valeur de x sur la pile, qui est une fonction, pas une valeur entière.

Ce problème avec ce programme particulier pourrait être résolu en renommant les noms de paramètres pour les rendre uniques.

Environnements pour les Lambdas

Notez que l'opération $\mathit{mkLambda}$ garde une référence vers l'environnement qui était l'environnement courant lorsque la fonction a été construite.

Elle utilise alors cet environnement dans une application.

Voici un exemple d'évolution des environnements durant une exécution du programme

$$\begin{array}{l}\textbf{\textit{def}}\ f(x\colon \mathit{Int}) = g(y \Rightarrow x + y)\\ \textbf{\textit{def}}\ g(x\colon (\mathit{Int})\mathit{Int}) = x(2);\\ f(1)\end{array}$$

(voir le tableau).

Prise en compte de la récursivité

Nous avons vu que le schéma d'environnement du Lisp original conduisait à des problèmes.

Mais le nôtre aussi, quand on ajoute la récursivité !

La récursivité est introduite en ajoutant à notre interpréteur une forme spéciale (**def** <name> <expr> <rest>).

Un **def** est comme **val** mise à part qu'il définit une fonction <name> qui peut s'appeler elle-même récursivement.

La méthode pour traiter **val** dans l'interpréteur,

```
case CONS(SYM("val"), CONS(SYM(name),
                   CONS(body, CONS(rest, NIL)))) =>
  eval(rest, env.extend(name, eval(body, env)))
```

ne marche pas pour **def**, pour deux raisons :

- Le corps est évalué trop tôt ; il ne devrait être évalué que lorsqu'on accède à *name*.
- *name* ne fait pas partie de l'environnement visible par *body*, la récursivité est donc impossible.

On résout ces problèmes en ajoutant une nouvelle méthode *extendRec* qui étend un environnement avec une liaison contenant un calcul possiblement récursif.

Évaluation de *def*

Voici comment **def** peut être traité :

```
case CONS(SYM("def"), CONS(SYM(name),
                   CONS(body, CONS(rest, NIL)))) =>
  eval(rest, env.extendRec(name, (env1 => eval(body, env1))))
```

- Cela résout le problème de l'évaluation prématurée, car on étend maintenant l'environnement avec une fonction qui évalue *body*, pas avec le résultat de l'évaluation de *body*.
- De plus, on rend la récursivité possible en changeant la recherche dans un environnement.

Nouveaux environnements

Voici la nouvelle définition des environnements.

```
abstract class Environment with {
  abstract def lookup(n: String): Data;
  def extendRec(name: String, expr: (Environment)Data) =
    new Environment with {
      def lookup(n: String): Data =
        if (n == name) expr(this) else outer.lookup(n);
    }
  def extend(name: String, v: Data) = extendRec(name, (env1 => v));
}
```

Notez que *extendRec* est maintenant l'opération principale pour étendre un environnement ; *extend* est définie en termes de *extendRec*.

Notez aussi que *lookup* permet la récursivité en passant l'environnement courant à la fonction trouvée.

```
val globalEnv = EmptyEnvironment
    .extend ("=",      FUN (case [NUM (arg1), NUM (arg2)] ⇒
                                 NUM (if (arg1 == arg2) 1 else 0)
                             case [STR (arg1), STR (arg2)] ⇒
                                 NUM (if (arg1 == arg2) 1 else 0)))
    .extend ("+",      FUN (case [NUM (arg1), NUM (arg2)] ⇒
                                 NUM (arg1 + arg2)
                             case [STR (arg1), STR (arg2)] ⇒
                                 STR (arg1 + arg2)))
    .extend ("−",      FUN (case [NUM (arg1), NUM (arg2)] ⇒
                                 NUM (arg1 − arg2)))
    .extend ("∗",      FUN (case [NUM (arg1), NUM (arg2)] ⇒
                                 NUM (arg1 ∗ arg2)))
    .extend ("/",      FUN (case [NUM (arg1), NUM (arg2)] ⇒
                                 NUM (arg1 / arg2)))
    .extend ("car",    FUN (case [CONS (x, xs)] ⇒ x))
    .extend ("cdr",    FUN (case [CONS (x, xs)] ⇒ xs))
    .extend ("null?",  FUN (case [NIL] ⇒ NUM (1) case _ ⇒ NUM (0)))
    .extend ("cons",   FUN (case [x, y] ⇒ CONS (x, y)));
```


L'environnement global

On évalue les expressions Lisp dans un environnement initial (global), qui contient les définitions pour les opérations couramment utilisées telles que + ou *cons*.

Voici une version minimalement utile d'un tel environnement.


```
(lambda (n)
   ((lambda (fact)
        (fact fact n))
     (lambda (ft k)
        (if (= k 1)
              1
              (∗ k (ft ft (− k 1)))))))
```


Récursivité par auto-application (self application)

Cette technique met en évidence une connexion profonde en programmation : on peut modéliser la récursivité par l'auto-application.

En fait, toute récursivité est au bout du compte traitée par l'auto-application dans le lambda-calcul, la théorie sous-jacente à la programmation fonctionnelle.

Pour le prouver, sans plus d'explications, voici une version de *faculty* qui n'utilise ni la récursivité, ni les boucles !

Résumé

Nous avons vu comment les programmes fonctionnels sont évalués en écrivant un interpréteur pour Lisp.

La structure de donnée auxiliaire centrale était l'environnement, qui représente toutes les liaisons connues au moment du calcul.

Les environnements remplacent les substitutions dans le modèle formel d'évaluation.

Exercice étendu

Augmenter l'interpréteur Lisp de telle manière qu'il ait sa propre boucle d'interprétation, qui accepte des définitions et des expressions individuelles.

Étant donnée une définition en entrée, l'interpréteur ajoute la liaison définie à l'environnement global.

Étant donnée une expression en entrée, il l'évalue et affiche le résultat.

Utiliser Lisp à partir de Scala

Voici un scénario d'utilisation de l'interpréteur Scala pour évaluer du Lisp :

Premièrement, on définit une fonction Lisp comme une chaîne de caractères :

```
> def facultyDef =
      "def faculty (lambda (n)" +
      "  (if (= n 0)" +
      "      1" +
      "      (* n (faculty (- n 1)))))"
```

Cette fonction peut alors être appliquée comme suit :

```
> evaluate("(" + facultyDef + "(faculty 4))")
24
```


La fonction d'interprétation principale

Voici la fonction principale de l'interpréteur :

```
def evaluate(x: Data): Data = eval(x, globalEnv);
```

Elle évalue un programme Lisp dans l'environnement global et retourne l'objet résultant.

Pour faciliter la vie du programmeur qui tape des expressions Lisp sur la ligne de commande, on inclut aussi une autre version de *evaluate* qui prend et renvoie des expressions Lisp sous forme de chaînes de caractères.

```
def evaluate(s: String): String = evaluate(string2lisp(s)).asString;
```