Fundamentos de lenguajes de programación Abstracción de datos

Facultad de Ingeniería. Universidad del Valle

Septiembre de 2018





Contenido

- 1 Introducción
- 2 Estrategias para representar tipos de datos
- 3 Definición de tipos de datos
- 4 Sintaxis Abstracta



- Cada vez que se define un conjunto de valores, se está definiendo un nuevo tipo de dato
- En un tipo de dato se define los valores, representaciones y operaciones sobre el mismo
- La representación de los datos es bastante compleja
- Para cambiar la representación de los datos se lleva a cabo una tarea llamada Abstracción de datos



- La abstracción de datos divide en dos los tipos de datos: interface e implementación
- La interfaz de un tipo de datos dice que representa el tipo de dato, sus operaciones y las propiedades de dichas operaciones
- La implementación proporciona una representación específica y un código para las operaciones que hacen uso de esa representación.



- Un tipo de dato que es abstracto se denomina Tipo Abstracto de Dato (TAD).
- El resto del programa fuera del tipo de dato, llamado el cliente del tipo de dato, manipula el nuevo dato solo mediante las operaciones especificadas en la interfaz
- El código del cliente es independiente de la representación si manipula los valores del tipo de dato solo a través de los procedimientos en su interfaz.



- Todo lo relacionado sobre el dato representado debe estar en el código de la implementación
- La parte más importante de la implementación es la especificación de como los datos son representados
- Se utiliza la implementación [v] para la representación de dato v.



Representación de números naturales:

$$\begin{array}{l} (\texttt{zero}) = \lceil 0 \rceil \\ (\texttt{is-zero?} \lceil n \rceil) = \left\{ \begin{array}{l} \texttt{\#t} & n = 0 \\ \texttt{\#f} & n \neq 0 \end{array} \right. \\ (\texttt{successor} \lceil n \rceil) = \lceil n + 1 \rceil & (n \geq 0) \\ (\texttt{predecessor} \lceil n + 1 \rceil) = \lceil n \rceil & (n \geq 0) \end{array}$$



Representación de números naturales:

- En la anterior especificación no se indica cómo se representan los números naturales
- A partir de la especificación se pueden escribir programas para la manipulación de los datos, sin importar su representación



Representación de números naturales:

Satisface (plus
$$[x] [y]$$
) = $[x + y]$ (Succ (succ (succ $[y]$))))



Tres posibles representaciones para los enteros no negativos:

1 Representación Unaria: Donde un entero no negativo n es representado por una lista de n símbolos '#t',

representado por una lista de
$$n$$
 símbolos '#t', $0 = 0$ $1 = (\cos \# t \cdot n)$ (1)

En esta representación, se satisface la especificación escribiendo:

```
(define zero '())
(define iszero? null?)
(define succ (lambda (n) (cons #t n)))
(define pred cdr)
```

Abstracción de datos Ejemplo

2 Representación de Números de Scheme: Se usa la representación interna de números de Scheme.

$$\lceil n \rceil = n$$

Se definen las entidades como:

```
(define zero 0)
(define iszero? zero?)
(define succ (lambda (n) (+ n 1)))
(define pred (lambda (n) (- n 1)))
```



3 Representación Bignum: Los números son representados en base N, para algún entero grande N. Dicha representación es una lista que consiste de números entre 0 y N-1.



Contenido

- 1 Introducción
- 2 Estrategias para representar tipos de datos
- 3 Definición de tipos de datos
- 4 Sintaxis Abstracta



- Cuando es usada la abstracción de datos, los programas tienen la propiedad de independencia de representación.
- Se presentan algunas estrategias para representar tipos de datos. Se ilustran estas estrategias usando el tipo de dato ambiente.
- Un ambiente asocia un valor con cada elemento de un conjunto finito de variables.
- Un ambiente puede ser usado para asociar las variables con sus valores en la implementación de un lenguaje de programación.



- Las variables pueden ser representadas de cualquier manera, siempre y cuando sea posible chequear la igualdad entre dos variables.
- Las variables se pueden representar mediante símbolos, cadenas, referencias en una tabla hash o incluso mediante números.



- Un ambiente es una función cuyo dominio es un conjunto finito de variables y cuyo rango es el conjunto de todos los valores de Scheme.
- De acuerdo a la notación matemática, los ambientes representan todos los conjuntos de la forma $\{(s_1, v_1), \ldots, (s_n, v_n)\}$, donde los s_i son símbolos diferentes y los v_i son valores de Scheme.

Tipo de dato Ambiente: Interfaz

La interfaz del tipo de dato ambiente tiene tres procedimientos:

donde
$$g(s') = \begin{cases} v & \text{si } s' = var \\ f(s') & \text{de otra forma} \end{cases}$$



Tipo de dato Ambiente: Interfaz

- El procedimiento empty-env debe producir una representación del ambiente vacío.
- El procedimiento apply-env aplica una representación de un ambiente a un argumento.
- El procedimiento (extend-env var val env) produce un nuevo ambiente que se comporta como env, excepto que su valor en el símbolo var es val.

Tipo de dato Ambiente: Implementación

Representación como estructura de datos:

Cada ambiente puede ser construido mediante una expresión en la siguiente gramática:

 De acuerdo a esta gramática los ambientes pueden ser representados como listas en Scheme.



Tipo de dato Ambiente: Implementación

Representación como estructura de datos:

```
empty-env: () \rightarrow Env
```

```
(define empty-env
  (lambda () (list 'empty-env)))
```

extend-env: $Var \ X \ SchemeValue \ X \ Env \rightarrow Env$



Tipo de dato Ambiente: Implementación

Representación como estructura de datos:

apply-env: Env $X \ Var \rightarrow SchemeValue$



Tipo de dato Ambiente: Implementación

Representación Procedimental:

- La interfaz del tipo de dato ambiente tiene una propiedad importante: ella tiene exactamente una entidad observadora apply-env.
- Esto permite representar un ambiente como un procedimiento que toma una variable y retorna su valor asociado.
- Se define empty-env y extend-env de tal manera que retornan un procedimiento que al ser aplicado devuelve el valor asociado a la variable en el ambiente.



Tipo de dato Ambiente: Implementación

Representación Procedimental:

```
empty-env: () \rightarrow Env
```

```
(define empty-env
(lambda ()
(lambda (search-var)
(report-no-binding-found search-var))))
```

extend-env: $Var \ X \ SchemeValue \ X \ Env \rightarrow Env$

Tipo de dato Ambiente: Implementación

Representación Procedimental:

apply-env: $Env \ X \ Var \rightarrow SchemeValue$

```
(define apply-env
  (lambda (env search-var)
       (env search-var)))
```



Contenido

- 1 Introducción
- 2 Estrategias para representar tipos de datos
- 3 Definición de tipos de datos
- 4 Sintaxis Abstracta



Hasta el momento hemos visto como definir tipos de datos recursivos mediante diferentes métodos. Por ejemplo:

```
<Lc-exp> := <var-exp> (<identifier>)
:= <lambda-exp> (lambda (<identifier>) <Lc-exp>)
::= <app-exp> (<Lc-exp> <Lc-exp>)
```



Por ejemplo para una representación basada en listas:



Así mismo, hemos intentado manipular y operar tipos de datos recursivos. Por ejemplo, para las expresiones del cálculo λ se han definido procedimientos como occurs-free?:

- No obstante, esta definición de occurs-free? es díficil de leer (es díficil decir que (car (cadr exp)) se refiere a la declaración de una variable en una expresión lambda o que (caddr exp) se refiere a su cuerpo).
- Esta definición establece cierta dependencia con la implementación de las expresiones del cálculo λ como listas.



- Se puede mejorar estos aspectos introduciendo interfaces para la creación y manipulación de datos de un cierto tipo.
- Una interfaz para un tipo de dato consta de procedimientos constructores y procedimientos observadores.

 Los procedimientos observadores pueden ser predicados o extractores.



Interfaz para expresiones cálculo λ

Para las expresiones del cálculo λ tenemos la siguiente interfaz:

Constructores:

var-exp : $Var \rightarrow Lc$ -exp

lambda-exp : $Var \times Lc\text{-}exp \rightarrow Lc\text{-}exp$

app-exp : Lc-exp x Lc-exp \rightarrow Lc-exp

Predicados:

var-exp? : Lc-exp \rightarrow Bool lambda-exp? : Lc-exp \rightarrow Bool app-exp? : Lc-exp \rightarrow Bool





Interfaz para expresiones cálculo λ

Extractores:

```
\begin{array}{lll} \text{var-exp-} \text{var} & : \textit{Lc-exp} \rightarrow \textit{Var} \\ \text{lambda-exp-} \text{bound-var} & : \textit{Lc-exp} \rightarrow \textit{Var} \\ \text{lambda-exp-} \text{body} & : \textit{Lc-exp} \rightarrow \textit{Lc-exp} \\ \text{app-exp-} \text{rator} & : \textit{Lc-exp} \rightarrow \textit{Lc-exp} \\ \text{app-exp-} \text{rand} & : \textit{Lc-exp} \rightarrow \textit{Lc-exp} \\ \end{array}
```



Interfaces para tipos de datos recursivos Interfaz para expresiones cálculo λ

Ahora es posible escribir una versión de occurs-free? que depende solo de la interfaz.



- La interfaz para el tipo de dato ambiente consta de los constructores empty-env y extended-env y del procedimiento observador apply-env.
- La interfaz para el tipo de dato lista consta de los constructores empty-list y cons, de los procedimientos extractores car y cdr y del predicado list?.

Diseño de interfaces para tipos de datos recursivos

Receta general para el diseño de interfaces de datos recursivos

- Incluir un constructor para cada clase de dato (regla de producción) en el tipo de dato.
- 2 Incluir un predicado para cada clase de dato en el tipo de dato.
- Incluir un extractor para cada pieza de dato pasada a un constructor del tipo de dato.



Interfaz define-datatype

- Para tipos de datos complejos, resulta tedioso construir interfaces rápidamente.
- La interfaz define-datatype es una herramienta de Scheme pra construir e implementar interfaces para tipos de datos.



Interfaces para tipos de datos recursivos Interfaz define-datatype

Las interfaces son especificadas mediante la expresión define-datatype que tiene la forma general:

```
(define-datatype nombre-tipo nombre-predicado-tipo
{(nombre-variante {(nombre-campo predicado )}*)}*)
```

- Esta declaración crea un tipo de dato llamado *nombre-tipo* con algunas variantes.
- Cada variante tiene un nombre (nombre-variante) y cero o más campos, cada uno con un nombre y un predicado asociado.



Interfaces para tipos de datos recursivos Interfaz define-datatype

- Dos tipos no pueden tener el mismo nombre, tampoco dos variantes, aunque pertenezcan a tipos diferentes, pueden tener el mismo nombre.
- Para cada variante, un procedimiento constructor es creado. Si hay n campos en una variante, su constructor recibe n argumentos, prueba sí cada uno de ellos satisface el predicado asociado y retorna un nuevo valor de dicha variante del tipo de dato.
- El nombre nombre-predicado-tipo es ligado a un predicado. Este predicado determina sí su argumento es un valor perteneciente al tipo nombre-tipo.



Interfaz define-datatype: Ejemplos

```
(define-datatype lc-exp lc-exp?
(var-exp (id identifier?))
(lambda-exp (id identifier?))
(body lc-exp?))
(app-exp (rator lc-exp?))
(rand lc-exp?)))
```



Interfaz define-datatype: Ejemplos

Tipo de dato s-list



Interfaz define-datatype: Ejemplos

Tipo de dato s-list



Para determinar a que objeto de un tipo de dato pertenece una variante y extraer sus componentes, se usa la forma cases, la cual tiene la sintaxis general:

```
(cases nombre-tipo expresion
  {(nombre-variante ( {nombre-campo}*) consecuente)}*
  (else por-defecto))
```



- La expresión cases evalua *expresion*. Esto da como resultado un valor *v* de tipo *nombre-tipo*.
- Si v es una variante nombre-variante, cada uno de los campos nombre-campo son asociados al valor del correspondiente campo de v. Luego la expresión consecuente es evaluada y su valor es retornado.
- Si v no es una de las variantes y la cláusula else es específicada, la expresión por-defecto es evaluada y su valor retornado.
- Si no existe una cláusula else, entonces tiene que existir una variante para todos los tipos de dato.

cases: Ejemplo



Interfaces para tipos de datos recursivos Más ejemplos

Tipo de dato bin-tree



Interfaces para tipos de datos recursivos Más ejemplos

Tipo de dato bin-tree

Definir un procedimiento que permita encontrar la suma de los enteros en las hojas de un árbol. Usando cases se tiene:



Interfaces para tipos de datos recursivos Más ejemplos

Ambientes



Ambientes



Contenido

- 1 Introducción
- 2 Estrategias para representar tipos de datos
- 3 Definición de tipos de datos
- 4 Sintaxis Abstracta





Dada la gramática de las expresiones del cálculo lambda:

```
 \begin{array}{ll} \langle \operatorname{expresi\'on} \rangle & ::= \langle \operatorname{identificador} \rangle \\ & ::= \left( \operatorname{lambda} \left( \langle \operatorname{identificador} \rangle \right) \langle \operatorname{expresi\'on} \rangle \right) \\ & ::= \left( \langle \operatorname{expresi\'on} \rangle \langle \operatorname{expresi\'on} \rangle \right) \\ \end{array}
```

 Se puede representar cada expresión del cálculo lambda usando el tipo de dato lc-exp definido anteriormente mediante define-datatype



Programadar

- Una BNF especifica una representación particular de un tipo de dato que usa los valores generados por la gramática
- Esta representación es llamada sintaxis concreta o representación externa
- Para procesar dichos datos, se requiere convertirlos a una representación interna o sintaxis abstracta, en la cual los símbolos terminales (como paréntesis) no necesiten ser almacenados ya que no llevan información



- Para crear una sintaxis abstracta a partir de una sintaxis concreta, se debe nombrar cada regla de producción de la sintaxis concreta y cada ocurrencia de un símbolo no terminal
- Para la gramática de las expresiones del cálculo λ , se puede resumir las opciones (sintaxis concreta y abstracta) usando la siguiente notación: •

```
⟨expresión⟩ ::= ⟨identificador⟩
var-exp (id) ← nombre

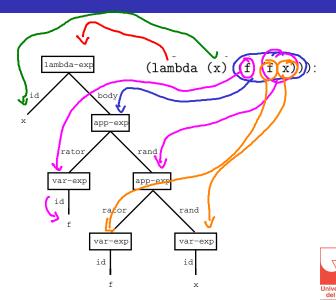
::= (lambda (⟨identificador⟩) ⟨expresión⟩)
lambda-exp (id body)

::= (⟨expresión⟩ ⟨expresión⟩)
app-exp (rator rand)
```

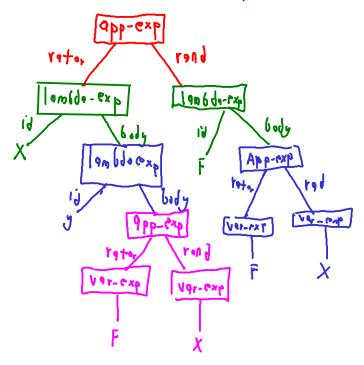


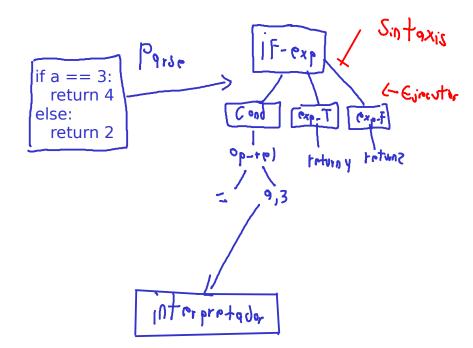
- La sintaxis abstracta de una expresión es más fácil de comprender visualizándola como un árbol de sintaxis abstracta
- El siguiente ejemplo muestra el árbol para la expresión (lambda (x) (f (f x))):





```
(lambda (x) (lambda (y) (f x))
(lambda (f) (f x))
```





- Los árboles de sintaxis son útiles en lenguajes de programación de procesamiento de sistemas ya que los programas que procesan otros programas (como los interpretadores o compiladores) son casi siempre dirigidos por sintaxis
- Esto es que cada parte de un programa es guiado por la regla gramatical asociada con dicha parte, y cualquier subparte correspondiente a un símbolo no terminal puede ser accedido con facilidad



Cuando se procesa la expresión (lambda (x) (f (f x))), primero se debe reconocer como una expresión del cálculo lambda, correspondiente a la regla:

```
\langle \mathrm{expresi\'{o}n} \rangle ::= \texttt{(lambda (} \langle \mathrm{identificador} \rangle \texttt{)} \ \langle \mathrm{expresi\'{o}n} \rangle \texttt{)}
```

■ El parámetro formal es x y el cuerpo es (f (f x)). El cuerpo debe ser reconocido como una app-exp, y así sucesivamente.



El problema de convertir un árbol de sintaxis abstracta a una representación lista-y-símbolo (con lo cual Scheme mostraría las expresiones en su sintaxis concreta), se resuelve con el procedimiento:

- La tarea de derivar el árbol de sintaxis abstracta a partir de una cadena de caracteres es denominado parsing, y es llevado a cabo por un programa llamado parser (analizador sintáctico)
- El siguiente procedimiento deriva la representación en sintaxis concreta a árboles de sintaxis abstracta:





Preguntas

?



Próxima sesión

- Semántica de los conceptos fundamentales de los lenguajes de programación.
- Primer interpretador.

