

Verificación de Tipos

CI4721 – Lenguajes de Programación II

Ernesto Hernández-Novich
<emhn@usb.ve>

Universidad “Simón Bolívar”

Copyright ©2012-2016

Funciones Polimórficas

- Una **Función Polimórfica** puede ser ejecutada con argumentos de tipos diferentes.
 - No se trata de sobrecarga – es **una** función.
 - No se trata de *templates* (C++, Java) – no hay copia de código.
 - Polimorfismo paramétrico (Haskell, ML).



Funciones Polimórficas

- Una **Función Polimórfica** puede ser ejecutada con argumentos de tipos diferentes.
 - No se trata de sobrecarga – es **una** función.
 - No se trata de *templates* (C++, Java) – no hay copia de código.
 - Polimorfismo paramétrico (Haskell, ML).
- Casi cualquier lenguaje tiene alguna función u operador polimórfico.
 - Acceder a posiciones de arreglos – sirve con arreglos de cualquier tipo.
 - Obtener la dirección de un objeto – la obtiene para cualquier objeto.
 - Seguir apuntadores – opera con apuntadores a cualquier cosa.

Funciones Polimórficas

- Una **Función Polimórfica** puede ser ejecutada con argumentos de tipos diferentes.
 - No se trata de sobrecarga – es **una** función.
 - No se trata de *templates* (C++, Java) – no hay copia de código.
 - Polimorfismo paramétrico (Haskell, ML).
- Casi cualquier lenguaje tiene alguna función u operador polimórfico.
 - Acceder a posiciones de arreglos – sirve con arreglos de cualquier tipo.
 - Obtener la dirección de un objeto – la obtiene para cualquier objeto.
 - Seguir apuntadores – opera con apuntadores a cualquier cosa.

¿Cómo verificar tipos de manera general
si el *usuario* puede definir funciones polimórficas?



¿Por qué son interesantes?

Longitud de una lista – sin polimorfismo

```
type link = *cell;  
  cell = record  
    info : integer;  
    next : link  
  end;  
function length (l : link) : integer;  
var len : integer;  
begin  
  len := 0;  
  while l <> nil do begin  
    len := len + 1;  
    l := *l.next  
  end;  
  return len  
end;
```



¿Por qué son interesantes?

Longitud de una lista – sin polimorfismo

```

type link = *cell;
  cell = record
    info : integer;
    next : link
  end;
function length (l : link) : integer;
var len : integer;
begin
  len := 0;
  while l <> nil do begin
    len := len + 1;
    l := *l.next
  end;
  return len
end;

```

Hace falta una función `length` por *cada* tipo de lista.



¿Por qué son interesantes?

Longitud de una lista – con polimorfismo

```
function length(x) =  
  if null(x) then 0  
    else length(tail(x)) + 1
```

`length` no necesita conocer el tipo contenido por la lista – ¡no lo usa!

¿Por qué son interesantes?

What kind of sorcery is this?

```
> length([42, 17, 69])
3
> length(["foo", "bar", "baz", "qux"])
4
> length([])
0
```

- La *única* función `length` se comporta como si tuviera los tipos

¿Por qué son interesantes?

What kind of sorcery is this?

```
> length([42, 17, 69])
3
> length(["foo", "bar", "baz", "qux"])
4
> length([])
0
```

- La *única* función `length` se comporta como si tuviera los tipos
 - `list(int) → int`

¿Por qué son interesantes?

What kind of sorcery is this?

```
> length([42, 17, 69])
3
> length(["foo", "bar", "baz", "qux"])
4
> length([])
0
```

- La *única* función `length` se comporta como si tuviera los tipos
 - `list(int) → int`
 - `list(string) → int`

¿Por qué son interesantes?

What kind of sorcery is this?

```
> length([42, 17, 69])
3
> length(["foo", "bar", "baz", "qux"])
4
> length([])
0
```

- La *única* función `length` se comporta como si tuviera los tipos
 - `list(int) → int`
 - `list(string) → int`
 - `list(whatever) → int`

What do you mean “whatever”?



Variables de Tipo

Whatever floats your boat. . .

- En la última aplicación de `length` no solamente no está claro sino que *no importa* cuál es el contenido de la lista – cualquiera funcionaría.



Variables de Tipo

Whatever floats your boat. . .

- En la última aplicación de `length` no solamente no está claro sino que *no importa* cuál es el contenido de la lista – cualquiera funcionaría.
- Las **Variables de Tipo** denotan expresiones con tipo *desconocido*.
 - Identificadas con letras griegas minúsculas ($\alpha, \beta, \gamma, \dots$).
 - A veces la variable terminará teniendo un tipo específico.
 - A veces la variable terminará indicando “cualquier tipo” – whatever. . .



Variables de Tipo

Whatever floats your boat. . .

- En la última aplicación de `length` no solamente no está claro sino que *no importa* cuál es el contenido de la lista – cualquiera funcionaría.
- Las **Variables de Tipo** denotan expresiones con tipo *desconocido*.
 - Identificadas con letras griegas minúsculas ($\alpha, \beta, \gamma, \dots$).
 - A veces la variable terminará teniendo un tipo específico.
 - A veces la variable terminará indicando “cualquier tipo” – whatever. . .
- Pueden cuantificarse universalmente
 - El tipo *más general* para `length` tendría una variable ligada

$$\forall \alpha. list(\alpha) \rightarrow \mathbf{int}$$

- Se puede *instanciar* la cuantificación con una variable libre particular

$$list(\alpha_1) \rightarrow \mathbf{int}$$

¿Cómo aprovechar las Variables de Tipo?

Se “despejan”, pero se requiere consistencia

- Ayudan a determinar el uso consistente de un símbolo.



¿Cómo aprovechar las Variables de Tipo?

Se “despejan”, pero se requiere consistencia

- Ayudan a determinar el uso consistente de un símbolo.
 - Aparece un símbolo `foo` sin declarar – su tipo será α .



¿Cómo aprovechar las Variables de Tipo?

Se “despejan”, pero se requiere consistencia

- Ayudan a determinar el uso consistente de un símbolo.
 - Aparece un símbolo `foo` sin declarar – su tipo será α .
 - Se usa `foo` como operando en una suma entera – hacemos $\alpha = \mathbf{int}$.



¿Cómo aprovechar las Variables de Tipo?

Se “despejan”, pero se requiere consistencia

- Ayudan a determinar el uso consistente de un símbolo.
 - Aparece un símbolo `foo` sin declarar – su tipo será α .
 - Se usa `foo` como operando en una suma entera – hacemos $\alpha = \mathbf{int}$.
 - Se encuentra `cos(foo)` – establecería α como **float**, pero ya está $\alpha = \mathbf{int}$, así que detectamos un error de tipos



¿Cómo aprovechar las Variables de Tipo?

Se “despejan”, pero se requiere consistencia

- Ayudan a determinar el uso consistente de un símbolo.
 - Aparece un símbolo `foo` sin declarar – su tipo será α .
 - Se usa `foo` como operando en una suma entera – hacemos $\alpha = \mathbf{int}$.
 - Se encuentra `cos(foo)` – establecería α como **float**, pero ya está $\alpha = \mathbf{int}$, así que detectamos un error de tipos
- El proceso que permite determinar las valuaciones para variables de tipo a partir de la construcción del lenguaje en la cual se usan, se denomina **Inferencia de Tipos**.
- Lo frecuente es inferir el tipo de una función a partir de su *cuerpo*.



¿Cómo se usan?

```
type link : *cell;  
procedure process (l : link, procedure p);  
begin  
  while l <> nil do begin  
    p(l);  
    l := *l.next  
  end  
end;
```

- process invoca a p sobre cada elemento de la lista l



¿Cómo se usan?

```
type link : *cell;  
procedure process (l : link, procedure p);  
begin  
  while l <> nil do begin  
    p(l);  
    l := *l.next  
  end  
end;
```

- process invoca a p sobre cada elemento de la lista l
- El parámetro formal indica que p es un procedimiento – tipo $\alpha \rightarrow \beta$

¿Cómo se usan?

```

type link : *cell;
procedure process (l : link, procedure p);
begin
  while l <> nil do begin
    p(l);
    l := *l.next
  end
end;

```

- process invoca a p sobre cada elemento de la lista l
- El parámetro formal indica que p es un procedimiento – tipo $\alpha \rightarrow \beta$
- Se usa en una *instrucción*, no produce resultado – hacemos $\beta = \mathbf{void}$

¿Cómo se usan?

```

type link : *cell;
procedure process (l : link, procedure p);
begin
  while l <> nil do begin
    p(l);
    l := *l.next
  end
end;

```

- process invoca a p sobre cada elemento de la lista l
- El parámetro formal indica que p es un procedimiento – tipo $\alpha \rightarrow \beta$
- Se usa en una *instrucción*, no produce resultado – hacemos $\beta = \mathbf{void}$
- Recibe un argumento de tipo link – hacemos $\alpha = \mathbf{link}$.

Se infiere que todo uso de process debe recibir un **link** \rightarrow **void** como segundo argumento.

Un lenguaje con funciones polimórficas

$$P \rightarrow D ; E$$

$$D \rightarrow D ; D$$

$$D \rightarrow \mathbf{id} : Q$$

$$Q \rightarrow \forall \mathbf{type_variable} . Q$$

$$Q \rightarrow T$$

$$T \rightarrow T \rightarrow T$$

$$T \rightarrow T \times T$$

$$T \rightarrow \mathbf{unary_constructor} (T)$$

$$T \rightarrow \mathbf{basic_type}$$

$$T \rightarrow \mathbf{type_variable}$$

$$T \rightarrow (T)$$

$$E \rightarrow E (E)$$

$$E \rightarrow E , E$$

$$E \rightarrow \mathbf{id}$$

- Funciones curryficadas o no.
- Tuplas.
- **unary_constructor** – para escribir *pointer*, *list*, ...
- **basic_type** – **void**, **bool**, **int**, **float**, ...

¿Cómo hacer la verificación?

Usaremos un Esquema de Traducción Dirigido por Sintaxis para etiquetar el árbol sintáctico – con tres observaciones importantes:



¿Cómo hacer la verificación?

Usaremos un Esquema de Traducción Dirigido por Sintaxis para etiquetar el árbol sintáctico – con tres observaciones importantes:

- 1 Cada ocurrencia de una función polimórfica *puede* tener argumentos de diferente tipo – las variables de tipos se *instancian* con un tipo particular a ese caso (¡qué podría tener variables!).



¿Cómo hacer la verificación?

Usaremos un Esquema de Traducción Dirigido por Sintaxis para etiquetar el árbol sintáctico – con tres observaciones importantes:

- 1 Cada ocurrencia de una función polimórfica *puede* tener argumentos de diferente tipo – las variables de tipos se *instancian* con un tipo particular a ese caso (¡qué podría tener variables!).
- 2 La presencia de variables obliga a operar usando equivalencia estructural – las variables se sustituyen por expresiones particulares para establecer una correspondencia (¡bidireccional!).



¿Cómo hacer la verificación?

Usaremos un Esquema de Traducción Dirigido por Sintaxis para etiquetar el árbol sintáctico – con tres observaciones importantes:

- ❶ Cada ocurrencia de una función polimórfica *puede* tener argumentos de diferente tipo – las variables de tipos se *instancian* con un tipo particular a ese caso (¡qué podría tener variables!).
- ❷ La presencia de variables obliga a operar usando equivalencia estructural – las variables se sustituyen por expresiones particulares para establecer una correspondencia (¡bidireccional!).
- ❸ Como una variable puede aparecer en varias expresiones, las instancias y sustituciones tienen que registrarse para asegurar su consistencia.

Un ejemplo informal antes de la teoría. . .



Un programa en el lenguaje

Varias definiciones seguidas de la expresión a verificar

Para el programa

```
deref :  $\forall \alpha. \text{pointer}(\alpha) \rightarrow \alpha;$   
q :  $\text{pointer}(\text{pointer}(\text{int}));$   
deref(deref(q))
```



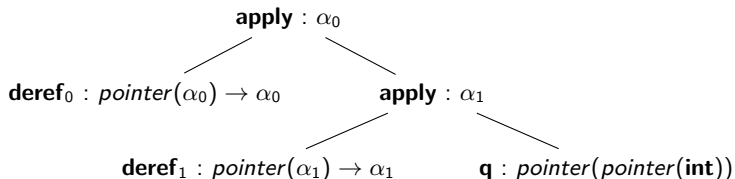
Un programa en el lenguaje

Varias definiciones seguidas de la expresión a verificar

Para el programa

$$\begin{aligned} \mathbf{deref} &: \forall \alpha. \mathit{pointer}(\alpha) \rightarrow \alpha; \\ \mathbf{q} &: \mathit{pointer}(\mathit{pointer}(\mathbf{int})); \\ &\mathbf{deref}(\mathbf{deref}(\mathbf{q})) \end{aligned}$$

se construye el árbol decorado



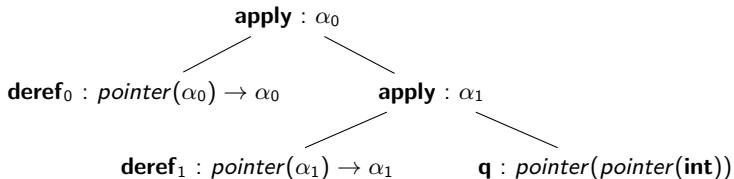
Un programa en el lenguaje

Varias definiciones seguidas de la expresión a verificar

Para el programa

$$\begin{aligned} \mathbf{deref} &: \forall \alpha. \mathit{pointer}(\alpha) \rightarrow \alpha; \\ \mathbf{q} &: \mathit{pointer}(\mathit{pointer}(\mathbf{int})); \\ &\mathbf{deref}(\mathbf{deref}(\mathbf{q})) \end{aligned}$$

se construye el árbol decorado



que permite inferir

$$\alpha_1 = \mathit{pointer}(\mathbf{int})$$

$$\alpha_0 = \mathbf{int}$$

Las variables y su reemplazo

Una **Sustitución** S define un conjunto de asociaciones entre variables de tipos y expresiones de tipos.

- $S(\alpha) = s_\alpha$ indica que en la sustitución S la variable de tipos α debe sustituirse por la expresión de tipos s_α .
- Si la sustitución S no especifica ninguna expresión para una variable particular α , se asume que $S(\alpha) = \alpha$.
- Si t es una expresión de tipos cualquiera, $S(t)$ es la expresión de tipos resultante de aplicar la sustitución S sobre t – se le denomina **instancia**.

Aplicar una sustitución es un procedimiento recursivo.



Aplicación de una sustitución

```

function subst(t : expr) : expr
  if t es un tipo primitivo then
    return t
  else if t es una variable then
    return  $S(t)$ 
  else if  $t = t_1 \rightarrow t_2$  then
    return  $S(t_1) \rightarrow S(t_2)$ 
  end if

```

- *expr* corresponde a una expresión de tipos.
- Otros constructores de tipo – sustituir en cada argumento.
- Complejidad proporcional a la cantidad de nodos de la expresión.

Algunos ejemplos triviales...

Siendo la sustitución

$$S = \{\alpha = \mathbf{int}, \beta = \mathbf{float}, \gamma = \mathit{pointer}(\beta)\}$$



Algunos ejemplos triviales...

Siendo la sustitución

$$S = \{\alpha = \mathbf{int}, \beta = \mathbf{float}, \gamma = \mathit{pointer}(\beta)\}$$

- $\mathit{pointer}(\mathbf{int})$ es instancia de $\mathit{pointer}(\alpha)$.

Algunos ejemplos triviales...

Siendo la sustitución

$$S = \{\alpha = \mathbf{int}, \beta = \mathbf{float}, \gamma = \mathit{pointer}(\beta)\}$$

- $\mathit{pointer}(\mathbf{int})$ es instancia de $\mathit{pointer}(\alpha)$.
- $\mathit{pointer}(\mathbf{float})$ es instancia de γ .



Algunos ejemplos triviales...

Siendo la sustitución

$$S = \{\alpha = \mathbf{int}, \beta = \mathbf{float}, \gamma = \mathit{pointer}(\beta)\}$$

- $\mathit{pointer}(\mathbf{int})$ es instancia de $\mathit{pointer}(\alpha)$.
- $\mathit{pointer}(\mathbf{float})$ es instancia de γ .
- $\mathbf{int} \rightarrow \mathbf{int}$ es instancia de $\alpha \rightarrow \alpha$



Algunos ejemplos triviales...

Siendo la sustitución

$$S = \{\alpha = \mathbf{int}, \beta = \mathbf{float}, \gamma = \mathit{pointer}(\beta)\}$$

- $\mathit{pointer}(\mathbf{int})$ es instancia de $\mathit{pointer}(\alpha)$.
- $\mathit{pointer}(\mathbf{float})$ es instancia de γ .
- $\mathbf{int} \rightarrow \mathbf{int}$ es instancia de $\alpha \rightarrow \alpha$
- $\mathbf{int} \rightarrow \mathbf{float}$ no es instancia de $\alpha \rightarrow \alpha$
– reemplazo inconsistente de α



Algunos ejemplos triviales...

Siendo la sustitución

$$S = \{\alpha = \mathbf{int}, \beta = \mathbf{float}, \gamma = \mathit{pointer}(\beta)\}$$

- $\mathit{pointer}(\mathbf{int})$ es instancia de $\mathit{pointer}(\alpha)$.
- $\mathit{pointer}(\mathbf{float})$ es instancia de γ .
- $\mathbf{int} \rightarrow \mathbf{int}$ es instancia de $\alpha \rightarrow \alpha$
- $\mathbf{int} \rightarrow \mathbf{float}$ no es instancia de $\alpha \rightarrow \alpha$
 - reemplazo inconsistente de α
- $\mathbf{int} \rightarrow \alpha$ no es instancia de $\alpha \rightarrow \alpha$
 - ocurrencias de α sin reemplazar.



Unificación

- Decimos que dos expresiones de tipos t_1 y t_2 se **unifican** si existe una sustitución S tal que $S(t_1) = S(t_2)$.
- En la práctica nos interesa la **unificación más general** S tal que
 - $S(t_1) = S(t_2)$ – es una sustitución.
 - $\forall S' \neq S : S'(t_1) = S'(t_2) \Rightarrow S'$ es instancia de S
 - S es el unificador que impone menos restricciones.
- De aquí en más al hablar de unificar está implícito que se trata del unificador más general.

Supongamos que sabemos encontrar el unificador más general.

Esquema de Verificación

- Ante una definición – $\mathbf{id}_1(\mathbf{id}_2) = E$
 - Crear variables de tipo α y β .
 - Asociar el tipo $\alpha \rightarrow \beta$ con \mathbf{id}_1 – es la función.
 - Asociar el tipo α con \mathbf{id}_2 – es el argumento en el dominio.
 - Inferir un tipo para E – ha de ser $s \rightarrow t$.
 - Unificar y cuantificar variables que no tengan restricciones.



Esquema de Verificación

- Ante una definición – $\mathbf{id}_1(\mathbf{id}_2) = E$
 - Crear variables de tipo α y β .
 - Asociar el tipo $\alpha \rightarrow \beta$ con \mathbf{id}_1 – es la función.
 - Asociar el tipo α con \mathbf{id}_2 – es el argumento en el dominio.
 - Inferir un tipo para E – ha de ser $s \rightarrow t$.
 - Unificar y cuantificar variables que no tengan restricciones.
- Ante una aplicación – $E_1(E_2)$
 - Inferir tipos para E_1 y E_2 .
 - E_1 debe tener tipo de la forma $s \rightarrow s'$.
 - Sea t el tipo inferido para E_2 .
 - Encontrar el unificador de s y t
 - Si existe, s' es el tipo inferido para $E_1(E_2)$.
 - Si no existe, hay un error de tipos

Esquema de Verificación

- Ante una definición – $\mathbf{id}_1(\mathbf{id}_2) = E$
 - Crear variables de tipo α y β .
 - Asociar el tipo $\alpha \rightarrow \beta$ con \mathbf{id}_1 – es la función.
 - Asociar el tipo α con \mathbf{id}_2 – es el argumento en el dominio.
 - Inferir un tipo para E – ha de ser $s \rightarrow t$.
 - Unificar y cuantificar variables que no tengan restricciones.
- Ante una aplicación – $E_1(E_2)$
 - Inferir tipos para E_1 y E_2 .
 - E_1 debe tener tipo de la forma $s \rightarrow s'$.
 - Sea t el tipo inferido para E_2 .
 - Encontrar el unificador de s y t
 - Si existe, s' es el tipo inferido para $E_1(E_2)$.
 - Si no existe, hay un error de tipos
- Si ocurre una función polimórfica, eliminar el cuantificador con variables nuevas antes de inferir su tipo.

Esquema de Verificación

- Ante una definición – $\mathbf{id}_1(\mathbf{id}_2) = E$
 - Crear variables de tipo α y β .
 - Asociar el tipo $\alpha \rightarrow \beta$ con \mathbf{id}_1 – es la función.
 - Asociar el tipo α con \mathbf{id}_2 – es el argumento en el dominio.
 - Inferir un tipo para E – ha de ser $s \rightarrow t$.
 - Unificar y cuantificar variables que no tengan restricciones.
- Ante una aplicación – $E_1(E_2)$
 - Inferir tipos para E_1 y E_2 .
 - E_1 debe tener tipo de la forma $s \rightarrow s'$.
 - Sea t el tipo inferido para E_2 .
 - Encontrar el unificador de s y t
 - Si existe, s' es el tipo inferido para $E_1(E_2)$.
 - Si no existe, hay un error de tipos
- Si ocurre una función polimórfica, eliminar el cuantificador con variables nuevas antes de inferir su tipo.
- Cada **id** encontrado por *primera* vez recibe una variable de tipos.

Verificando nuestro primer programa

```
deref :  $\forall \alpha. \text{pointer}(\alpha) \rightarrow \alpha;$   
q :  $\text{pointer}(\text{pointer}(\mathbf{int}));$   
deref(deref(q))
```



Verificando nuestro primer programa

deref : $\forall \alpha. \text{pointer}(\alpha) \rightarrow \alpha;$
q : $\text{pointer}(\text{pointer}(\mathbf{int}));$
deref(deref(q))

Expresión : Tipo

Sustitución

q : $\text{pointer}(\text{pointer}(\mathbf{int}))$



Verificando nuestro primer programa

deref : $\forall \alpha. \text{pointer}(\alpha) \rightarrow \alpha;$
q : $\text{pointer}(\text{pointer}(\mathbf{int}));$
deref(deref(q))

Expresión : Tipo

Sustitución

q : $\text{pointer}(\text{pointer}(\mathbf{int}))$

deref₁ : $\text{pointer}(\alpha_1) \rightarrow \alpha_1$



Verificando nuestro primer programa

$\mathbf{deref} : \forall \alpha. \mathit{pointer}(\alpha) \rightarrow \alpha;$
 $\mathbf{q} : \mathit{pointer}(\mathit{pointer}(\mathbf{int}));$
 $\mathbf{deref}(\mathbf{deref}(\mathbf{q}))$

Expresión : Tipo	Sustitución
$\mathbf{q} : \mathit{pointer}(\mathit{pointer}(\mathbf{int}))$ $\mathbf{deref}_1 : \mathit{pointer}(\alpha_1) \rightarrow \alpha_1$ $\mathbf{deref}_1(\mathbf{q}) : \mathit{pointer}(\mathbf{int})$	$\alpha_1 = \mathit{pointer}(\mathbf{int})$



Verificando nuestro primer programa

$\mathbf{deref} : \forall \alpha. \mathit{pointer}(\alpha) \rightarrow \alpha;$
 $\mathbf{q} : \mathit{pointer}(\mathit{pointer}(\mathbf{int}));$
 $\mathbf{deref}(\mathbf{deref}(\mathbf{q}))$

Expresión : Tipo	Sustitución
$\mathbf{q} : \mathit{pointer}(\mathit{pointer}(\mathbf{int}))$ $\mathbf{deref}_1 : \mathit{pointer}(\alpha_1) \rightarrow \alpha_1$ $\mathbf{deref}_1(\mathbf{q}) : \mathit{pointer}(\mathbf{int})$ $\mathbf{deref}_0 : \mathit{pointer}(\alpha_0) \rightarrow \alpha_0$	$\alpha_1 = \mathit{pointer}(\mathbf{int})$



Verificando nuestro primer programa

$\mathbf{deref} : \forall \alpha. \mathit{pointer}(\alpha) \rightarrow \alpha;$
 $\mathbf{q} : \mathit{pointer}(\mathit{pointer}(\mathbf{int}));$
 $\mathbf{deref}(\mathbf{deref}(\mathbf{q}))$

Expresión : Tipo	Sustitución
$\mathbf{q} : \mathit{pointer}(\mathit{pointer}(\mathbf{int}))$	
$\mathbf{deref}_1 : \mathit{pointer}(\alpha_1) \rightarrow \alpha_1$	
$\mathbf{deref}_1(\mathbf{q}) : \mathit{pointer}(\mathbf{int})$	$\alpha_1 = \mathit{pointer}(\mathbf{int})$
$\mathbf{deref}_0 : \mathit{pointer}(\alpha_0) \rightarrow \alpha_0$	
$\mathbf{deref}_0(\mathbf{deref}_1(\mathbf{q})) : \mathbf{int}$	$\alpha_0 = \mathbf{int}$



Un programa más interesante

```

length :  $\beta$ ;
  x :  $\gamma$ ;
  if :  $\forall \alpha. \mathbf{bool} \times \alpha \times \alpha \rightarrow \alpha$ ;
  null :  $\forall \alpha. \mathit{list}(\alpha) \rightarrow \mathbf{bool}$ ;
  tail :  $\forall \alpha. \mathit{list}(\alpha) \rightarrow \mathit{list}(\alpha)$ ;
  0 : int;
  1 : int;
  + : int  $\times$  int  $\rightarrow$  int;
  match :  $\forall \alpha. \alpha \times \alpha \rightarrow \alpha$ ;
    match(length(x), if(null(x), 0, length(tail(x)) + 1)
  
```

La intención de **match** es establecer la equivalencia de los tipos de ambos argumentos – deducir el tipo de **length**.



Un programa más interesante

Expresión : Tipo

$x : \gamma$

Sustitución



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$ $\text{length} : \beta$	



Un programa más interesante

Expresión : Tipo

$x : \gamma$

length : β

length(x) : δ

Sustitución

$\beta = \gamma \rightarrow \delta$



Un programa más interesante

Expresión : Tipo

$x : \gamma$

length : β

length(x) : δ

$x : \gamma$

Sustitución

$\beta = \gamma \rightarrow \delta$



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
length : β	
length (x) : δ	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
null : $list(\alpha_n) \rightarrow \mathbf{bool}$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	
$\mathbf{tail}(x) : \mathit{list}(\alpha_n)$	$\alpha_t = \alpha_n$



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	
$\mathbf{tail}(x) : \mathit{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\mathbf{length} : \mathit{list}(\alpha_n) \rightarrow \delta$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	
$\mathbf{tail}(x) : \mathit{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\mathbf{length} : \mathit{list}(\alpha_n) \rightarrow \delta$	
$\mathbf{length}(\mathbf{tail}(x)) : \delta$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	
$\mathbf{tail}(x) : \mathit{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\mathbf{length} : \mathit{list}(\alpha_n) \rightarrow \delta$	
$\mathbf{length}(\mathbf{tail}(x)) : \delta$	
$\mathbf{1} : \mathbf{int}$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	
$\mathbf{tail}(x) : \mathit{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\mathbf{length} : \mathit{list}(\alpha_n) \rightarrow \delta$	
$\mathbf{length}(\mathbf{tail}(x)) : \delta$	
$\mathbf{1} : \mathbf{int}$	
$\mathbf{+} : \mathbf{int} \times \mathbf{int} \rightarrow \mathbf{int}$	



Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\mathbf{length} : \beta$	
$\mathbf{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\mathbf{null} : \mathit{list}(\alpha_n) \rightarrow \mathbf{bool}$	
$\mathbf{null}(x) : \mathbf{bool}$	$\gamma = \mathit{list}(\alpha_n)$
$\mathbf{0} : \mathbf{int}$	
$x : \mathit{list}(\alpha_n)$	
$\mathbf{tail} : \mathit{list}(\alpha_t) \rightarrow \mathit{list}(\alpha_t)$	
$\mathbf{tail}(x) : \mathit{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\mathbf{length} : \mathit{list}(\alpha_n) \rightarrow \delta$	
$\mathbf{length}(\mathbf{tail}(x)) : \delta$	
$\mathbf{1} : \mathbf{int}$	
$\mathbf{+} : \mathbf{int} \times \mathbf{int} \rightarrow \mathbf{int}$	
$\mathbf{length}(\mathbf{tail}(x)) + \mathbf{1} : \mathbf{int}$	$\delta = \mathbf{int}$

Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\text{length} : \beta$	
$\text{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\text{null} : \text{list}(\alpha_n) \rightarrow \text{bool}$	
$\text{null}(x) : \text{bool}$	$\gamma = \text{list}(\alpha_n)$
$0 : \text{int}$	
$x : \text{list}(\alpha_n)$	
$\text{tail} : \text{list}(\alpha_t) \rightarrow \text{list}(\alpha_t)$	
$\text{tail}(x) : \text{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\text{length} : \text{list}(\alpha_n) \rightarrow \delta$	
$\text{length}(\text{tail}(x)) : \delta$	
$1 : \text{int}$	
$+ : \text{int} \times \text{int} \rightarrow \text{int}$	
$\text{length}(\text{tail}(x)) + 1 : \text{int}$	$\delta = \text{int}$
$\text{if} : \text{bool} \times \alpha_i \times \alpha_j \rightarrow \alpha_j$	

Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\text{length} : \beta$	
$\text{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\text{null} : \text{list}(\alpha_n) \rightarrow \text{bool}$	
$\text{null}(x) : \text{bool}$	$\gamma = \text{list}(\alpha_n)$
$0 : \text{int}$	
$x : \text{list}(\alpha_n)$	
$\text{tail} : \text{list}(\alpha_t) \rightarrow \text{list}(\alpha_t)$	
$\text{tail}(x) : \text{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\text{length} : \text{list}(\alpha_n) \rightarrow \delta$	
$\text{length}(\text{tail}(x)) : \delta$	
$1 : \text{int}$	
$+ : \text{int} \times \text{int} \rightarrow \text{int}$	
$\text{length}(\text{tail}(x)) + 1 : \text{int}$	$\delta = \text{int}$
$\text{if} : \text{bool} \times \alpha_i \times \alpha_j \rightarrow \alpha_j$	
$\text{if}(\text{null}(x), 0, \text{length}(\text{tail}(x)) + 1) : \text{int}$	$\alpha_i = \text{int}$

Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\text{length} : \beta$	
$\text{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\text{null} : \text{list}(\alpha_n) \rightarrow \text{bool}$	
$\text{null}(x) : \text{bool}$	$\gamma = \text{list}(\alpha_n)$
$0 : \text{int}$	
$x : \text{list}(\alpha_n)$	
$\text{tail} : \text{list}(\alpha_t) \rightarrow \text{list}(\alpha_t)$	
$\text{tail}(x) : \text{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\text{length} : \text{list}(\alpha_n) \rightarrow \delta$	
$\text{length}(\text{tail}(x)) : \delta$	
$1 : \text{int}$	
$+$: $\text{int} \times \text{int} \rightarrow \text{int}$	
$\text{length}(\text{tail}(x)) + 1 : \text{int}$	$\delta = \text{int}$
$\text{if} : \text{bool} \times \alpha_i \times \alpha_j \rightarrow \alpha_j$	
$\text{if}(\text{null}(x), 0, \text{length}(\text{tail}(x)) + 1) : \text{int}$	$\alpha_i = \text{int}$
$\text{match} : \alpha_m \times \alpha_m \rightarrow \alpha_m$	

Un programa más interesante

Expresión : Tipo	Sustitución
$x : \gamma$	
$\text{length} : \beta$	
$\text{length}(x) : \delta$	$\beta = \gamma \rightarrow \delta$
$x : \gamma$	
$\text{null} : \text{list}(\alpha_n) \rightarrow \text{bool}$	
$\text{null}(x) : \text{bool}$	$\gamma = \text{list}(\alpha_n)$
$0 : \text{int}$	
$x : \text{list}(\alpha_n)$	
$\text{tail} : \text{list}(\alpha_t) \rightarrow \text{list}(\alpha_t)$	
$\text{tail}(x) : \text{list}(\alpha_n)$	$\alpha_t = \alpha_n$
$\text{length} : \text{list}(\alpha_n) \rightarrow \delta$	
$\text{length}(\text{tail}(x)) : \delta$	
$1 : \text{int}$	
$+ : \text{int} \times \text{int} \rightarrow \text{int}$	
$\text{length}(\text{tail}(x)) + 1 : \text{int}$	$\delta = \text{int}$
$\text{if} : \text{bool} \times \alpha_i \times \alpha_j \rightarrow \alpha_j$	
$\text{if}(\text{null}(x), 0, \text{length}(\text{tail}(x)) + 1) : \text{int}$	$\alpha_i = \text{int}$
$\text{match} : \alpha_m \times \alpha_m \rightarrow \alpha_m$	
$\text{match}(\text{length}(x), \text{if}(\text{null}(x), 0, \text{length}(\text{tail}(x)) + 1)) : \text{int}$	$\alpha_m = \text{int}$

Un programa más interesante

Finalmente...

El resultado del algoritmo es la sustitución

$$S = \left\{ \begin{array}{l} \alpha_j = \mathbf{int} \\ \alpha_m = \mathbf{int} \\ \delta = \mathbf{int} \\ \gamma = \mathit{list}(\alpha_n) \\ \alpha_t = \alpha_n \\ \beta = \gamma \rightarrow \delta \end{array} \right\}$$

como no se hace ninguna suposición sobre α_n , podemos cuantificar y obtener el tipo inferido

$$\mathbf{length} : \forall \alpha_n. \mathit{list}(\alpha_n) \rightarrow \mathbf{int}$$



Deja Vu

¡Sustituir y unificar es lo que hacemos en el sistema formal!

- La regla para inferir una aplicación funcional sería

$$\frac{\rho \vdash e_1 : \alpha \rightarrow \beta \quad \rho \vdash e_2 : \alpha}{\rho \vdash e_1 (e_2) : \beta}$$

Deja Vu

¡Sustituir y unificar es lo que hacemos en el sistema formal!

- La regla para inferir una aplicación funcional sería

$$\frac{\rho \vdash e_1 : \alpha \rightarrow \beta \quad \rho \vdash e_2 : \alpha}{\rho \vdash e_1 (e_2) : \beta}$$

- Y puedo inferir el tipo para una llamada

$$\frac{\rho \vdash \text{length} : \forall \alpha. \text{list}(\alpha) \rightarrow \mathbf{int} \quad \frac{\dots}{\rho \vdash [42, 17, 69] : \text{list}(\mathbf{int})}}{\rho \vdash \text{length} ([42, 69, 17]) : \mathbf{int}}$$

Algoritmo de Unificación

- ¿Cómo encontrar (si existe) una sustitución S que permita hacer idénticas las expresiones s y t ?
- El caso trivial de unificación es la comparación de igualdad – si s es idéntica a t , la sustitución vacía les unifica.
- Para generalizar a expresiones de tipos, debemos considerar la comparación estructural entre grafos incluyendo ciclos.
 - Las variables de tipos estarán en las hojas – *singleton*.
 - Los tipos primitivos estarán en las hojas – *singleton*.
 - Los constructores de tipos estarán en los nodos interiores.
- Construiremos clases de equivalencia para los nodos.

Un ejemplo

Consideremos las expresiones

$$((\alpha_1 \rightarrow \alpha_2) \times \text{list}(\alpha_3)) \rightarrow \text{list}(\alpha_2)$$

$$((\alpha_3 \rightarrow \alpha_4) \times \text{list}(\alpha_3)) \rightarrow \alpha_5$$



Un ejemplo

Consideremos las expresiones

$$((\alpha_1 \rightarrow \alpha_2) \times \text{list}(\alpha_3)) \rightarrow \text{list}(\alpha_2)$$

$$((\alpha_3 \rightarrow \alpha_4) \times \text{list}(\alpha_3)) \rightarrow \alpha_5$$

Pueden unificarse con la sustitución

$$S = \left\{ \begin{array}{l} \alpha_1 = \alpha_1 \\ \alpha_2 = \alpha_2 \\ \alpha_3 = \alpha_1 \\ \alpha_4 = \alpha_2 \\ \alpha_5 = \text{list}(\alpha_2) \end{array} \right\}$$



Un ejemplo

Consideremos las expresiones

$$((\alpha_1 \rightarrow \alpha_2) \times \text{list}(\alpha_3)) \rightarrow \text{list}(\alpha_2)$$

$$((\alpha_3 \rightarrow \alpha_4) \times \text{list}(\alpha_3)) \rightarrow \alpha_5$$

Pueden unificarse con la sustitución

$$S = \left\{ \begin{array}{l} \alpha_1 = \alpha_1 \\ \alpha_2 = \alpha_2 \\ \alpha_3 = \alpha_1 \\ \alpha_4 = \alpha_2 \\ \alpha_5 = \text{list}(\alpha_2) \end{array} \right\}$$

obteniendo

$$((\alpha_1 \rightarrow \alpha_2) \times \text{list}(\alpha_1)) \rightarrow \text{list}(\alpha_2)$$

Algoritmo de Unificación

Infraestructura

- *Node* – registro con campos para un operador binario, hijos izquierdo y derecho, y apuntador a la clase de equivalencia.
 - Cada clase de equivalencia tiene un nodo representante – se identifica porque tiene el apuntador nulo.
 - Resto de los nodos en la clase de equivalencia apuntan a su representante – directa o indirectamente.
 - Inicialmente, todos los nodos están en su propia clase de equivalencia.
- *find(node)* – encuentra el nodo representante para la clase de equivalencia a la cual pertenece *node*.
- *union(m, n)* – combina las clases de equivalencia de *m* y *n*.
 - Si uno de los dos **no** es una variable, ese será el nuevo representante.
 - Basta hacer que el representante de uno apunte al representante del otro para completar la combinación.



Algoritmo de Unificación

El algoritmo

```
function unify(m, n : Node) : bool
  s ← find(m)
  t ← find(n)
  if s = t then
    return true
  else if s y t son el mismo tipo básico then
    return true
  else if s = op(s1, s2) and t = op(t1, t2) then
    union(s, t)
    return unify(s1, t1) and unify(s2, t2)
  else if var(s) or var(t) then
    union(s, t)
    return true
  else
    return false
  end if
```



Algoritmo de Unificación

¿Cómo encuentro la sustitución?

- Supongamos que m o n es una hoja – corresponde a una variable.
- Supongamos que ha sido incluida en una clase de equivalencia con un nodo que representa una expresión.
- Entonces, para cada variable α
 - $find(\alpha)$ obtiene el nodo representante r para su clase de equivalencia.
 - La expresión representada por r corresponde a $S(\alpha)$.

Bibliografía

- [*Aho*] (Primera Edición)
 - Capítulo 6.6
 - Ejercicios 6.20 a 6.31
- [*Aho*] (Segunda Edición)
 - Secciones 6.3.1, 6.3.2, 6.5.1, 6.5.2 y 6.5.3
 - Ejercicios 6.5.1 y 6.5.2