COSE212: Programming Languages

Lecture 11 — Automatic Type Inference (2)

Hakjoo Oh
2016 Fall

Finding a Solution of Type Equations

Find values for the type variables that make all the equations true.

$$\underbrace{\texttt{proc}\ (\underbrace{f}_{t_f})\ \underbrace{\texttt{proc}\ (\underbrace{x}_{t_x})\ \underbrace{((\underbrace{f\ 3}_{t_3}) - (\underbrace{f\ x}_{t_4}))}_{t_2}}_{t_1}}_{t_0}$$

Equations			Solution		
t_0	$=$	$t_f \rightarrow t_1$	t_0	$=$	(int $\rightarrow$ int) $\rightarrow$ (int $\rightarrow$ int)
t_1	$=$	$t_x \rightarrow t_2$	t_1	$=$	int $\rightarrow$ int
t_3	$=$	int	t_2	$=$	int
t_4	$=$	int	t_3	$=$	int
t_2	$=$	int	t_4	$=$	int
t_f	$=$	int $\rightarrow t_3$	t_f	$=$	int $\rightarrow$ int
t_f	$=$	$t_x \rightarrow t_4$	t_x	$=$	int

Such a solution can be found by the *unification algorithm*.

Unification Algorithm: Example 1

The calculation is split into equations to be solved and substitution found so far. Initially, the substitution is empty:

Equations	Substitution
$t_0 = t_f \to t_1$	
$t_1 = t_x \to t_2$	
$t_3 = \text{int}$	
$t_4 = \text{int}$	
$t_2 = \text{int}$	
$t_f = \text{int} \to t_3$	
$t_f = t_x \to t_4$	

Unification Algorithm: Example 1

Consider each equation in turn. If the equation's left-hand side is a variable, move it to the substitution:

Equations	Substitution
$t_1 = t_x \rightarrow t_2$	$t_0 = t_f \rightarrow t_1$
$t_3 = \text{int}$	
$t_4 = \text{int}$	
$t_2 = \text{int}$	
$t_f = \text{int} \rightarrow t_3$	
$t_f = t_x \rightarrow t_4$	

Unification Algorithm: Example 1

Move the next equation to the substitution and propagate the information to the existing substitution (i.e., substitute the right-hand side for each occurrence of t_1):

Equations			Substitution		
t_3	$=$	int	t_0	$=$	$t_f \rightarrow (t_x \rightarrow t_2)$
t_4	$=$	int	t_1	$=$	$t_x \rightarrow t_2$
t_2	$=$	int			
t_f	$=$	int $\rightarrow t_3$			
t_f	$=$	$t_x \rightarrow t_4$			

Unification Algorithm: Example 1

Same for the next three equations:

Equations	Substitution
$t_4 = \text{int}$	$t_0 = t_f \rightarrow (t_x \rightarrow t_2)$
$t_2 = \text{int}$	$t_1 = t_x \rightarrow t_2$
$t_f = \text{int} \rightarrow t_3$	$t_3 = \text{int}$
$t_f = t_x \rightarrow t_4$	

Equations	Substitution
$t_2 = \text{int}$	$t_0 = t_f \rightarrow (t_x \rightarrow t_2)$
$t_f = \text{int} \rightarrow t_3$	$t_1 = t_x \rightarrow t_2$
$t_f = t_x \rightarrow t_4$	$t_3 = \text{int}$
	$t_4 = \text{int}$

Equations	Substitution
$t_f = \text{int} \rightarrow t_3$	$t_0 = t_f \rightarrow (t_x \rightarrow \text{int})$
$t_f = t_x \rightarrow t_4$	$t_1 = t_x \rightarrow \text{int}$
	$t_3 = \text{int}$
	$t_4 = \text{int}$
	$t_2 = \text{int}$

Unification Algorithm: Example 1

Consider the next equation $t_f = \text{int} \to t_3$. The equation contains t_3, which is already bound to int in the substitution. Substitute int for t_3 in the equation. This is called *applying* the substitution to the equation.

Equations	Substitution
$t_f = \text{int} \to \text{int}$	$t_0 = t_f \to (t_x \to \text{int})$
$t_f = t_x \to t_4$	$t_1 = t_x \to \text{int}$
	$t_3 = \text{int}$
	$t_4 = \text{int}$
	$t_2 = \text{int}$

Move the resulting equation to the substitution and update it.

Equations	Substitution
$t_f = t_x \to t_4$	$t_0 = (\text{int} \to \text{int}) \to (t_x \to \text{int})$
	$t_1 = t_x \to \text{int}$
	$t_3 = \text{int}$
	$t_4 = \text{int}$
	$t_2 = \text{int}$
	$t_f = \text{int} \to \text{int}$

Unification Algorithm: Example 1

Apply the substitution to the equation:

Equations	Substitution
$\text{int} \to \text{int} = t_x \to \text{int}$	$t_0 = (\text{int} \to \text{int}) \to (t_x \to \text{int})$
	$t_1 = t_x \to \text{int}$
	$t_3 = \text{int}$
	$t_4 = \text{int}$
	$t_2 = \text{int}$
	$t_f = \text{int} \to \text{int}$

If neither side of the equation is a variable, simplify the equation by yielding two new equations:

Equations	Substitution
$\text{int} = t_x$	$t_0 = (\text{int} \to \text{int}) \to (t_x \to \text{int})$
$\text{int} = \text{int}$	$t_1 = t_x \to \text{int}$
	$t_3 = \text{int}$
	$t_4 = \text{int}$
	$t_2 = \text{int}$
	$t_f = \text{int} \to \text{int}$

Unification Algorithm: Example 1

Switch the sides of the first equation and move it to the substitution:

Equations	Substitution
int $=$ int	$t_0 =$ (int $\rightarrow$ int) $\rightarrow$ (int $\rightarrow$ int)
	$t_1 =$ int $\rightarrow$ int
	$t_3 =$ int
	$t_4 =$ int
	$t_2 =$ int
	$t_f =$ int $\rightarrow$ int
	$t_x =$ int

The final substitution is the solution of the original equations.

Unification Algorithm

For each equation in turn,

1. Apply the current substitituion to the equation.
2. If the left-hand side is a variable, move it to the substitution and substitute the right-hand side for each occurrence of the variable in the substitution. (If the right-hand side is a variable, switch the sides and do the same thing).

Basically, the algorithm follows the two steps. Two execptions: If neither side is a variable, simplify the equation, which eventually generates an equation whose left- or right-hand side is a variable. If the equation is always true, discard it.

Unification Algorithm: Example 2

$$\underbrace{\texttt{proc}\ (\underbrace{f}_{t_f})\ (\underbrace{f\ 11}_{t_1})}_{t_0}$$

$$\begin{array}{rcl} t_0 & = & t_f \to t_1 \\ t_f & = & \mathsf{int} \to t_1 \end{array}$$

Unification Algorithm: Example 2

1.

Equations	Substitution
$t_0 = t_f \to t_1$	
$t_f = \text{int} \to t_1$	

2.

Equations	Substitution
$t_f = \text{int} \to t_1$	$t_0 = t_f \to t_1$

3.

Equations	Substitution
	$t_0 = (\text{int} \to t_1) \to t_1$
	$t_f = \text{int} \to t_1$

The type is *polymorphic* in t_1.

Unification Algorithm: Example 3

$$\underbrace{\texttt{if } \underbrace{x}_{t_x} \texttt{ then } \underbrace{(x - 1)}_{t_1} \texttt{ else } 0}_{t_0}$$

$$\begin{array}{rcl} t_x & = & \text{bool} \\ t_1 & = & t_0 \\ \text{int} & = & t_0 \\ t_x & = & \text{int} \\ t_1 & = & \text{int} \end{array}$$

Unification Algorithm: Example 4

$$\underbrace{\texttt{proc}\ (\underbrace{f}_{t_f})\ (\underbrace{\texttt{iszero}\ \underbrace{(f\ f)}_{t_2}}_{t_1})}_{t_0}$$

$$\begin{aligned} t_0 &= t_f \to t_1 \\ t_1 &= \text{bool} \\ t_2 &= \text{int} \\ t_f &= t_f \to t_2 \end{aligned}$$

Exercises

For each following expression, perform the type inference and find its type, or determine that no such type exists.

1. $\texttt{let}\ x = 4\ \texttt{in}\ (x\ 3)$
2. $\texttt{let}\ f\ = \texttt{proc}\ (z)\ z\ \texttt{in}\ \texttt{proc}\ (x)\ ((f\ x) - 1)$
3. $\texttt{let}\ p\ = \texttt{iszero}\ 1\ \texttt{in}\ \texttt{if}\ p\ \texttt{then}\ 88\ \texttt{else}\ 99$
4. $\texttt{let}\ f\ = \texttt{proc}\ (x)\ x\ \texttt{in}\ \texttt{if}\ (f\ (\texttt{iszero} 0))\ \texttt{then}\ (f\ 11)\ \texttt{else}\ (f\ 22)$

Summary

Automatic type inference:

- derive type equations from the program text, and
- solve the equations by unification.