[Faculty of Science Information and Computing Sciences]

Compiler Construction

Stefan Holdermans

Dept. of Information and Computing Sciences, Utrecht University P.O. Box 80.089, 3508 TB Utrecht, The Netherlands E-mail: stefan@cs.uu.nl

> Web pages: http://people.cs.uu.nl/stefan/ January 19, 2009

10. Type inference

Agenda

§10 Type inference

Universiteit Utrecht

[Faculty of Science Information and Computing Sciences]

Contents

§10

§10.1 Untyped lambda-calculus

§10.2 Simply typed lambda-calculus

§10.3 System F

§10.4 Hindley-Milner typing

§10.5 Algorithm W

Universiteit Utrecht

Faculty of Science Information and Computing Sciences] 10118-121121 2 000

Syntax

§10.1

variables Var Tmterms

 $t ::= x | \lambda x. t_1 | t_1 t_2$

Universiteit Utrecht

Faculty of Science Information and Computing Sciences

(B) (B) (E) (E) (E) (9 000

► Church, Kleene (1930s).

- ▶ Formal system designed to investigate function definition, function application, and recursion.
- ▶ Idealised, minimalistic functional programming language.
- Only three language constructs: variables. lambda-abstraction, function application, (Other constructs can be encoded with these.)

Universiteit Utrecht

Faculty of Science Information and Computing Sciences 10110121212121 2 999

Alpha-equivalence and beta-substitution 810.1

 λ is a binder: $\lambda x. t_1$ binds x in t_1 . Unbound variables are called free.

Alpha-equivalence: terms that only differ in the names of their bound variables are considered equal. For example: λx , x and λy , y are alpha-equivalent.

Alpha-conversion: consistently renaming bound variables while avoiding free variables from being captured. For example: λf , λx , f x z can be alpha-converted into λf , λy , f y z, but not into λf , λz , f z z.

Beta-substitution: capture-avoiding substitution of free variables, performing alpha-conversion where necessary. For example: $[z \mapsto y](\lambda f, \lambda x, f \mid x \mid z) = \lambda f, \lambda x, f \mid x \mid y$ and $[z \mapsto x](\lambda f, \lambda u, f \mid u \mid x),$

 $v \in Val$ values

 $v ::= \lambda x. t$

Big-step operational semantics:

 $t \Downarrow v$ evaluation

 $\frac{}{\lambda x.\,t_1 \Downarrow \lambda x.\,t_1} \, [\text{e-lam}]$

 $\frac{t_1 \Downarrow \lambda x. \, t_{11} \quad t_2 \Downarrow v_2 \quad [x \mapsto v_2] t_{11} \Downarrow v}{t_1 \, t_2 \Downarrow v} \text{ [e-app]}$

For example: $(\lambda x. \lambda y. x) (\lambda x. x) (\lambda x. \lambda y. y) \Downarrow \lambda x. x$.

[Faculty of 9]
Universiteit Utrecht

Information and Computing Sciences]

10.2 Simply typed lambda-calculus

Additional language constructs—such as local definitions, natural numbers, boolean constants, conditionals, arithmetic and relational operators, and even recursion—can be introduced as mere syntactic sugar.

For example:

let $x = t_1$ in t_2 ni $=_{def}$ $(\lambda x. t_2) t_1$

In the sequel, we will just assume that some of these additional constructs are in fact added to the core calculus, so that we can use for example natural numbers in our example.

Universiteit Utrecht

[Faculty of Science Information and Computing Sciences]

Simple types

§10.2

To study typing for the lambda-calculus, we extend the syntax of lambda-terms with mandatory type annotations for lambda-abstractions

For example:

 $\lambda x : Nat. x$

 $(\lambda f: Bool \rightarrow Nat. \lambda x: Bool. f x) (\lambda y: Bool. 42)$

12

Faculty of Science

Ty types Tmterms

 $:= \tau_1 \rightarrow \tau_2$ $::= x \mid \lambda x : \tau \cdot t_1 \mid t_1 t_2$

To render the sets of types inhabited, we add type constants such as Nat and Bool.

Universiteit Utrecht

Faculty of Science Information and Computing Sciences

1011/01/21/21 2 990

Typing

§10.2

Type environments map from variables to types:

TvEnv type environments

 $[] \mid \Gamma_1[x \mapsto \tau]$

As always, we write $\Gamma(x) = \tau$ if the rightmost binding for x in Γ maps x to τ .

The judgements of the typing relation read

 $\Gamma \vdash t : \tau$ typing

> Faculty of Science Information and Computing Sciences1

∈ Val values

 $v ::= \lambda x : \tau, t$

Big-step operational semantics:

evaluation $t \downarrow v$

 $\lambda x : \underline{\tau} \cdot \underline{t_1} \Downarrow \lambda x : \underline{\tau} \cdot \underline{t_1}$ [e-lam]

 $t_1 \Downarrow \lambda x : \tau$. t_{11} $t_2 \Downarrow v_2$ $[x \mapsto v_2]t_{11} \Downarrow v$ [e-app] $t_1 \ t_2 \downarrow v$

Universiteit Utrecht

[Faculty of Science Information and Computing Sciences] 10110121212121 2 999

§10.2

Type rules

 $\frac{\Gamma(x) = \tau}{\Gamma \vdash x : \tau} [t\text{-var}]$

 $\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x : \tau_1 \cdot t_1 : \tau_1 \to \tau_2} [t\text{-lam}]$

 $\frac{\Gamma \vdash t_1 : \textcolor{red}{\tau_2} \rightarrow \textcolor{red}{\tau} \quad \Gamma \vdash t_2 : \textcolor{red}{\tau_2}}{\Gamma \vdash t_1 \ t_2 : \textcolor{red}{\tau}} \ [\textit{t-app}]$

Universiteit Utrecht

Faculty of Science Information and Computing Sciences1 10) 10) 15) 15) 15) 10)

Writing | t | for the untyped lambda-term obtained from erasing all type annotations from the simply typed lambda-term t, we have:

if $t \Downarrow v$, then $|t| \Downarrow |v|$.

That is, types play no rôle at run-time.



Faculty of Science Information and Computing Sciences 10118-121121 2 000

Polymorphism

§10.3

§10.2

Next, we add polymorphism to our language.

The system obtained is known as System F (Girard, 1972) or the second-order polymorphic lambda-calculus (Revnolds, 1974).

The main innovation with respect to the simply typed lambda-calculus is that, in addition to values, functions can also take types as arguments:

 $\Lambda \alpha, \lambda x : \alpha, x$

 $(\Lambda \alpha. \Lambda \beta. \lambda x : \alpha. \lambda y : \beta. x) [Nat] [Bool] 2 false$

Universiteit Utrecht

Faculty of Science Information and Computing Sciences] 10110121121121 2 000

10.3 System F

Universiteit Utrecht

Universiteit Utrecht

Faculty of Science Information and Computing Sciences] 10110121212121 2 999

Syntax

§10.3

TvVar type variables Var term variables

Tvtypes Tmterms

 $::= \alpha \mid \tau_1 \rightarrow \tau_2 \mid \forall \alpha, \tau_1$ $::= x \mid \lambda x : \tau \cdot t_1 \mid t_1 t_2 \mid \Lambda \alpha \cdot t_1 \mid t_1 \tau$

Faculty of Science Information and Computing Sciences1 10118-121121 2 000 Note: functions can take polymorphic functions as arguments.

For example:

 $\lambda f: \forall \alpha. \alpha \rightarrow Nat. f [Nat] 2 + f [Bool] false$

This function takes a "normal" polymorphic function (i.e, of rank 1) as argument and so it has itself a rank-2 type. Its type reads $(v_{\alpha}, \alpha \rightarrow Nat) \rightarrow Nat$. In general, if a function takes a function with a rank-n type as argument, it has itself a rank-(n+1) type.



[Faculty of Science Information and Computing Sciences]

Type inference

§10.4

For both the simply typed lambda-calculus and System F, implementing a type checker is straightforward. (Exercise: ...)

Although programming languages based on System F are very powerful, writing type annotation on every function parameter is very tedious—especially if types become more involved due to polymorphism.

So, the question is: can we derive an algorithm that takes an erased System-F term as argument and that *infers* all missing type annotations? That way, we can have the full power for System F, without the burden of having to write possibly complex or otherwise tiresome type annotations.

The anwer is **no**, type inference for System F is undecidable (Wells, 1994).



[Faculty of Science Information and Computing Sciences]

10.4 Hindley-Milner typing



[Faculty of Sciences]

The Hindley-Milner system

§10.4

The Hindley-Milner (Hindley, 1969; Milner, 1978) type system is a compromise between the full power of System F and the desire to leave out type annotations.

Hindley-Milner typing comes with two crucial restrictions:

- All types are of at most rank 1, i.e., functions cannot take polymorphic functions as arguments.
- Functions can only have a polymorphic type if they are directly bound in a local definition (let-polymorphism).

The resulting type system is at the heart of languages like Haskell and ML and allows that for each well-typed term a so-called principal (i.e., most polymorphic) type can be inferred

Information and Computing Sciences]

810.4

Semantics

Var Tm

term variables terms

 $t ::= x \mid \lambda x. t_1 \mid t_1 t_2 \mid \mathbf{let} \ x = t_1 \mathbf{in} \ t_2 \mathbf{ni}$

Local definitions play a crucial rôle in typing now and so. rather than syntactic sugar, they form a true language construct now

Universiteit Utrecht

Faculty of Science Information and Computing Sciences]

1011/01/21/21 2 990

Evaluation rules

30 §10.4

— [e-lam] $\lambda x. t_1 \Downarrow \lambda x. t_1$

 $t_1 \Downarrow \lambda x. t_{11}$ $t_2 \Downarrow v_2$ $[x \mapsto v_2 | t_{11} \Downarrow v$ _ [e-app] $t_1 t_2 \downarrow v$

> $t_1 \Downarrow v_1 \quad [x \mapsto v_1]t_2 \Downarrow v$ _ [e-let] $let x = t_1 in t_2 ni \Downarrow v$

> > Faculty of Science

Information and Computing Sciences 10110121121121 2 000 $v \in Val$ values

 $v ::= \lambda x, t$

Big-step operational semantics:

 $t \Downarrow v$ evaluation

Universiteit Utrecht

Faculty of Science Information and Computing Sciences]

10110121212121 2 999

§10.4

Typing

Implementing the rank-1 restriction, the type language is stratified into two levels: types and type schemes.

TyVar type variables \in Ty types

TyScheme type schemes

 $:= \alpha \mid \tau_1 \rightarrow \tau_2$ $\sigma ::= \tau \mid \forall \alpha, \sigma_1$

Type environments map from variables to type schemes:

TvEnv type environments $\Gamma ::= [] | \Gamma_1[x \mapsto \sigma]$

Universiteit Utrecht

Faculty of Science Information and Computing Sciences1 イロンイグライミンイミン を めらの \forall is a binder for type variables: α is bound in $\forall \alpha, \sigma_1$.

We write $ftv(\sigma)$ for the set of type variables that appear free in σ .

Similarly, we write $ftv(\Gamma)$ for the set of type variables that appear free in the codomain of Γ .

Universiteit Utrecht

Faculty of Science Information and Computing Sciences

10118-121121 2 000

Typing rules

§10.4

 $\frac{\Gamma(x) = \sigma}{\Gamma \vdash x : \sigma} [t\text{-var}]$

$$\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x, t_1 : \tau_1 \to \tau_2} [t-lam]$$

$$\frac{\Gamma \vdash t_1 : \textcolor{red}{\tau_2} \rightarrow \textcolor{red}{\tau} \quad \Gamma \vdash t_2 : \textcolor{red}{\tau_2}}{\Gamma \vdash t_1 \ t_2 : \textcolor{red}{\tau}} \ [\textit{t-app}]$$

$$\frac{\Gamma \vdash t_1 : \sigma_1 \quad \Gamma[x \mapsto \sigma_1] \vdash t_2 : \tau}{\Gamma \vdash \text{let } x = t_1 \text{ in } t_2 \text{ ni } : \tau} \text{ [t-let]}$$

Universiteit Utrecht

Faculty of Science

Information and Computing Sciences 10110121121121 2 000 The judgements of the typing relation take the form

 $\Gamma \vdash t : \sigma$ typing

The typing relation is defined by a natural deduction system comprised from six rules: one for each of the four term constructors and two for dealing with polymorphism.

Universiteit Utrecht

Typing judgements

Faculty of Science Information and Computing Sciences 10110121212121 2 999

Typing rules (cont'd)

§10.4

 $\frac{\Gamma \vdash t : \sigma_1 \quad \alpha \notin \mathit{ftv}(\Gamma)}{\Gamma \vdash t : \forall \alpha, \sigma_1} \text{ [t-gen]}$

$$\frac{\Gamma \vdash t : \forall \alpha. \sigma_1}{\Gamma \vdash t : [\alpha \mapsto \tau_0] \sigma_1} [t\text{-inst}]$$

The premise $\alpha \notin ftv\{\Gamma\}$ in [t-gen] is needed because we do not have any binders for type variables in our term language.

(Exercise: show that without this premise we can derive $\lambda x, x : \forall \alpha, \forall \beta, \alpha \rightarrow \beta,$

Universiteit Utrecht

Information and Computing Sciences] 10118-121121 2 000

10.5 Algorithm W

[Faculty of Science Information and Computing Sciences]

Challenges

➤ We have to somehow "guess", for every

- lambda-abstraction, what the type of its formal parameter is.
- ► The rules [t-gen] and [t-inst] can be applied to terms of any form. We have to decide when to apply them.

Algorithm W

§10.5

Algorithm W (Damas and Milner, 1982) establishes a procedure for obtaining a principal type, for each well-typed term in the Hindley-Milner system.

Intuitively, a principal type is the most polymorphic type that can be assigned to a given term.

Strategy

Universiteit Utrecht

[Faculty of Science Information and Computing Sciences]

§10.5

< □ > < **⊙**

 Algorithm W proceeds by initially "guessing" a fresh type variable for every parameter type and by incrementally refining theses guesses as more information on the use of parameters becomes available.

 Algorithm W uses a syntax-directed variation of the Hindley-Milner type rules in which generalisation only occurs at let-bindings and instantiation only occurs at the use-sites of variables.

Syntax-directed: for every term, at most one rule applies.

Universiteit Utrecht

The syntax-directed type rules are defined in terms of metaoperations gen and inst:

$$\begin{array}{l} \textit{gen}_{\cdot} \colon \mathbf{TyEnv} \to \mathbf{Ty} \to \mathbf{TyScheme} \\ \textit{gen}_{\Gamma}(\tau) = \\ \textit{let} \left\{ \alpha_{1}, \dots, \alpha_{n} \right\} = \textit{ftv}(\tau) \backslash \textit{ftv}(\Gamma) \\ \textit{in} \ \forall \alpha_{1} \dots \forall \alpha_{n}, \tau \end{array}$$

$$inst : \mathbf{TyScheme} \to \mathbf{Ty}$$

 $inst(\forall \alpha_1, \dots, \forall \alpha_n, \tau_1)$
 $= let \alpha'_1, \dots, \alpha'_n \text{ be fresh}$
 $in [\alpha_1 \mapsto \alpha'_1] \cdots [\alpha_n \mapsto \alpha'_n] \tau_1$

Universiteit Utrecht

Faculty of Science Information and Computing Sciences

10118-121121 2 000 Type substitutions

Algorithm W makes use of type substitutions:

∈ TvSubst type substitutions

$$\theta$$
 ::= $id \mid [\alpha \mapsto \tau] \mid \theta_1 \circ \theta_2$

Universiteit Utrecht

Information and Computing Sciences

 $\Gamma(x) = \frac{\sigma_0}{[t-var]}$

$$\frac{\Gamma[x \mapsto \tau_1] \vdash t_1 : \tau_2}{\Gamma \vdash \lambda x. \ t_1 : \tau_1 \to \tau_2} \ [t\text{-lam}]$$

$$\frac{\Gamma \vdash t_1 : \tau_2 \to \tau \quad \Gamma \vdash t_2 : \tau_2}{\Gamma \vdash t_1 \ t_2 : \tau} \ [t\text{-app}]$$

$$\frac{\Gamma \vdash t_1 : \tau_1 \quad \Gamma[x \mapsto gen_{\Gamma}(\tau_1)] \vdash t_2 : \tau}{\Gamma[t\text{-left}]} \ [t\text{-left}]$$

$$\Gamma \vdash \text{let } x = t_1 \text{ in } t_2 \text{ ni} : \tau$$

 $t:\sigma$). Universiteit Utrecht

Information and Computing Sciences 101181212121 2 990

Information and Computing Sciences]

10118-121121 2 000

§10.5

Applying type substitutions

Applying a type substitution to a type scheme:

$$\begin{split} & id\sigma &= \sigma \\ & [\alpha \mapsto \tau_0]\alpha &= \tau_0 \\ & [\alpha \mapsto \tau_0]\alpha_0 &= \alpha_0 & \text{if } \alpha \not\equiv \alpha_0 \\ & [\alpha \mapsto \tau_0](\gamma_1 \to \tau_2) = [\alpha \mapsto \tau_0]\tau_1 \to [\alpha \mapsto \tau_0]\tau_2 \\ & [\alpha \mapsto \tau_0](\forall \alpha, \sigma_1) &= \forall \alpha, \sigma \\ & [\alpha \mapsto \tau_0](\forall \alpha_0, \sigma_1) &= \forall \alpha_0, [\alpha \mapsto \tau_0]\sigma_1 & \text{if } \alpha \not\equiv \alpha_0 \\ & (\theta_1 \circ \theta_2)\sigma &= \theta_1\theta_2\sigma \end{split}$$

Applying a type substitution to a type environment:

$$\begin{array}{l} \theta[] &= [] \\ \theta(\Gamma_1[x\mapsto\sigma]) = \theta\Gamma_1[x\mapsto\theta\sigma] \end{array}$$
 (Faculty of Science

42

§10.5

Algorithm W makes use of Robinson's unification algorithm (1965).

$$\mathcal{U}: \mathbf{Tv} \times \mathbf{Tv} \rightharpoonup \mathbf{TvSubst}$$

 \mathcal{U} provides a partial function that, for any two types τ_1 and 72, constructs a most general unifier, i.e., a type substitution θ , such that $\theta \tau_1 = \theta \tau_2$ and, for all θ' with $\theta' \tau_1 = \theta' \tau_2$ there is a θ'' with $\theta' \approx \theta'' \circ \theta$. (Where $\theta_1 \approx \theta_2$ iff, $\theta_1 \sigma = \theta_2 \sigma$ for all σ .)

If τ_1 and τ_2 are not unifiable, \mathcal{U} fails.

For example:

$$\mathcal{U}(\alpha \to Bool \to \alpha , Nat \to \beta \to \gamma)$$

= $[\gamma \mapsto Nat] \circ [\beta \mapsto Bool] \circ [\alpha \mapsto Nat]$

Universiteit Utrecht

Universiteit Utrecht

Faculty of Science Information and Computing Sciences]

Algorithm W

§10.5

 $W: \mathbf{TyEnv} \times \mathbf{Tm} \rightharpoonup \mathbf{Ty} \times \mathbf{TySubst}$

Algorithm W takes a type environment Γ and a term t, and produces, if t is well-typed in Γ , a type τ and a type substitution θ , such that $\theta\Gamma \vdash t : gen_{\theta\Gamma}(\tau)$. Moreover, $gen_{\theta\Gamma}(\tau)$ is then a principal type for t in $\theta\Gamma$.

If t is ill-typed in Γ . Algorithm W fails.

Faculty of Science Information and Computing Sciences (B) (B) (E) (E) (E) (9 00C

 $\mathcal{U}(\tau_1, \tau_2)$ = fail in all other cases The side conditions $\alpha_1 \notin ftv(\tau_2)$ and $\alpha_2 \notin ftv(\tau_2)$ are

 $\mathcal{U}(\tau_{11} \to \tau_{12}, \tau_{21} \to \tau_{22}) = let \ \theta_1 = \mathcal{U}(\tau_{11}, \tau_{21})$

known as the "occurs check" and prevent the construction of infinite types.

 $= [\alpha_1 \mapsto \tau_2]$ $= [\alpha_2 \mapsto \tau_1]$ if $\alpha_2 \notin ftv(\tau_1)$

in $\theta_2 \circ \theta_1$

Universiteit Utrecht

Unification (cont'd)

 $\mathcal{U}(\alpha,\alpha)$

 $\mathcal{U}(\alpha_1, \tau_2)$

 $\mathcal{U}(\tau_1, \alpha_2)$

Faculty of Science Information and Computing Sciences 10110121212121 2 999

if $\alpha_1 \notin ftv(\tau_2)$

 $\theta_2 = \mathcal{U}(\theta_1 \tau_{12}, \theta_1 \tau_{22})$

Algorithm W: variables

Universiteit Utrecht

§10.5

```
W(\Gamma, x) =
         x \in dom(\Gamma)
   then (inst (\Gamma(x)), id)
         fail
```



```
\begin{split} \mathcal{W}(\Gamma \ , \ \lambda x. \ t_1) &= \\ \textit{let} \ \alpha_1 \ \textit{be} \ \textit{fresh} \\ (\tau_2 \ , \theta_1) &= \mathcal{W}(\Gamma[x \mapsto \alpha_1] \ , \ t_1) \\ \textit{in} \ (\theta_1 \alpha_1 \to \tau_2, \theta_1) \end{split}
```

Universiteit Utrecht

[Faculty of Science Information and Computing Sciences]

Algorithm W: local definitions

§10.5

 $\begin{array}{l} \mathcal{W}(\Gamma \ , \ \mathbf{let} \ x = t_1 \ \mathbf{in} \ t_2 \ \mathbf{ni}) = \\ \mathit{let} \ (\tau_1 \ , \ \theta_1) = \mathcal{W}(\Gamma \ , \ t_1) \\ (\tau \ , \ \theta_2) = \mathcal{W}(\theta_1 \Gamma[x \mapsto \mathit{gen}_{\theta_1 \Gamma}(\tau_1)] \ , \ t_2) \\ \mathit{in} \ (\tau \ , \ \theta_2 \circ \theta_1) \end{array}$

```
[Faculty of Science Information and Computing Sciences]
```

```
\begin{split} \mathcal{W}(\Gamma \ , \ t_1 \ t_2) &= \\ \textit{let} \ \alpha \ \textit{be fresh} \\ (\tau_1 \ , \ \theta_1) &= \mathcal{W}(\Gamma \ , \ t_1) \\ (\tau_2 \ , \ \theta_2) &= \mathcal{W}(\theta_1 \Gamma \ , \ t_2) \\ \theta_3 &= \mathcal{U}(\theta_2 \ \tau_1 \ , \ \tau_2 \rightarrow \alpha) \\ \textit{in} \ (\theta_3 \alpha \ , \ \theta_3 \circ \theta_2 \circ \theta_1) \end{split}
```

Universiteit Utrecht

Algorithm W: applications

[Faculty of Science Information and Computing Sciences]

Universiteit Utrecht