

Generic HASKELL, Specifically

Dave Clarke and Andres Löh

July 11, 2002

Introduction

This talk is about recent extensions to the Generic HASKELL language.

These extensions have in common that they

- are motivated by examples that we could not do nicely before,
- allow to program a larger class of generic programs with less work,
- extend the application area of Generic HASKELL to generic traversals over a large tree (of values of different datatypes).

Structure

- Plain Generic HASKELL
- Specific Generic HASKELL
 - Default cases
 - Constructor cases
- Conclusions

PART I

Plain Generic HASKELL

Generic HASKELL provides . . .

- a superset of Haskell 98 (which is compiled to Haskell)
- constructs for defining generic (i.e. type-indexed) functions
- generic programming in a style based on work by Hinze

Type-indexed values have kind-indexed types

Example: *gmap*

```
data Conference = MPC | WCGP
data List a     = Nil | Cons a (List a)
data Tree a b  = Leaf a | Node (Tree a b) b (Tree a b)
```

```
gmap{Conference} MPC           ~> MPC
gmap{List} (+1) (Cons 2 (Cons 3 Nil)) ~> Cons 3 (Cons 4 Nil)
gmap{Tree} (*2) (const "bar") (Node (Leaf 21) "foo" (Leaf 5))
                                     ~> Node (Leaf 42) "bar" (Leaf 10)
```

Type-indexed values have kind-indexed types – continued

$$\begin{aligned} gmap\{Conference :: *\} &:: Conference \rightarrow Conference \\ gmap\{List :: * \rightarrow *\} &:: \forall a b . (a \rightarrow b) \rightarrow List\ a \rightarrow List\ b \\ gmap\{Tree :: * \rightarrow * \rightarrow *\} &:: \\ &\forall a b c d . (a \rightarrow c) \rightarrow (b \rightarrow d) \rightarrow Tree\ a\ b \rightarrow Tree\ c\ d \end{aligned}$$

The following shows a valid Generic H_VSKELL definition of a kind-indexed type and the assignment of this type to the generic function *gmap*:

$$\begin{aligned} gmap\{t :: k\} &:: Map\{k\}\ t\ t \\ \mathbf{type}\ Map\{*\}\ s\ t &= s \rightarrow t \\ \mathbf{type}\ Map\{k \rightarrow l\}\ s\ t &= \forall a b . Map\{k\}\ a\ b \rightarrow Map\{l\}\ (s\ a)\ (t\ b) \end{aligned}$$

Type-indexed values have kind-indexed types – continued

The function *gmap* can be defined as

$gmap\{t :: k\}$		$::$	$Map\{k\} t t$
$gmap\{Unit\}$	$Unit$	$=$	$Unit$
$gmap\{:+:\}$	$gmapA\ gmapB\ (Inl\ a)$	$=$	$Inl\ (gmapA\ a)$
$gmap\{:+:\}$	$gmapA\ gmapB\ (Inr\ b)$	$=$	$Inr\ (gmapB\ b)$
$gmap\{*:\}$	$gmapA\ gmapB\ (a*:\ b)$	$=$	$(gmapA\ a)*:\ (gmapB\ b)$
$gmap\{Con\ c\}$	$gmapA\ (Con\ a)$	$=$	$Con\ (gmapA\ a)$
$gmap\{Label\ l\}$	$gmapA\ (Label\ a)$	$=$	$Label\ (gmapA\ a)$
$gmap\{Int\}$	i	$=$	i
$gmap\{Char\}$	c	$=$	c

Observations:

- Type arguments can be type constructors of arbitrary kind.
- There are special constructors (*Con*, *:+:*) to reflect the *structure of types* as well as abstract types (*Int*, *Char*).
- Each case matches the generic function's type.

Compilation of generic functions

- Each case is converted into one Haskell function.
- The type of each case is determined by the function's type.

$$\begin{aligned} & \vdots \\ \text{case_gmap_Sum} & :: \forall a b c d . (a \rightarrow c) \rightarrow (b \rightarrow d) \rightarrow a::b \rightarrow c::d \\ \text{case_gmap_Sum } \text{gmapA } \text{gmapB } (\text{Inl } a) & = \text{Inl } (\text{gmapA } a) \\ \text{case_gmap_Sum } \text{gmapA } \text{gmapB } (\text{Inr } b) & = \text{Inr } (\text{gmapB } b) \\ \text{case_gmap_Prod} & :: \forall a b c d . (a \rightarrow c) \rightarrow (b \rightarrow d) \rightarrow a::b \rightarrow c::d \\ \text{case_gmap_Prod } \text{gmapA } \text{gmapB } (a::b) & = (\text{gmapA } a)::(\text{gmapB } b) \\ & \vdots \end{aligned}$$

Structure of data types

$$\begin{array}{lcl}
 \mathbf{data} \ A \ arg_1 \ \dots \ arg_k & = & \mathit{Con}_1 \ T_{1,1} \ T_{1,2} \ \dots \ T_{1,n_1} \\
 & | & \mathit{Con}_2 \ T_{2,1} \ T_{2,2} \ \dots \ T_{2,n_2} \\
 & \vdots & \\
 & | & \mathit{Con}_m \ T_{m,1} \ T_{m,2} \ \dots \ T_{m,n_m} \\
 \mathbf{type} \ A^\circ \ arg_1 \ \dots \ arg_k & = & \mathit{Con} \ (T_{1,1} \ :* \ T_{1,2} \ :* \ \dots \ :* \ T_{1,n_1}) \\
 & \text{:}+ & \mathit{Con} \ (T_{2,1} \ :* \ T_{2,2} \ :* \ \dots \ :* \ T_{2,n_2}) \\
 & \vdots & \\
 & \text{:}+ & \mathit{Con} \ (T_{m,1} \ :* \ T_{m,2} \ :* \ \dots \ :* \ T_{m,n_m})
 \end{array}$$

- Both $\text{:}+ \text{:}$ and $\text{:} * \text{:}$ are right associative.
- A constructor without elements is replaced by *Unit*.
- Labelled fields are marked by *Label*.

$$\begin{array}{lcl}
 \mathbf{data} \ List \ a & = & \mathit{Nil} \ | \ \mathit{Cons} \{ \mathit{hd} \ :: \ a, \ \mathit{tl} \ :: \ (List \ a) \} \\
 \mathbf{type} \ List^\circ \ a & = & \mathit{Con} \ \mathit{Unit} \ \text{:}+ \ \mathit{Con} \ (\mathit{Label} \ a \ :* \ (\mathit{Label} \ (List \ a)))
 \end{array}$$

Structure of data types – continued

data $a :+: b$ = $Inl\ a \mid Inr\ b$
data $a **: b$ = $a **: b$
data $Unit$ = $Unit$
data $Con\ a$ = $Con\ a$
data $Label\ a$ = $Label\ a$

- For each datatype in a program, Generic HVSHELL generates an equivalent data type using $:+:$, $**:$, $Unit$, Con and $Label$.
- An embedding-projection pair mapping between both versions of the type is also generated.

data $EP\ a\ b$ = $EP\{from :: a \rightarrow b, to :: b \rightarrow a\}$
 ep_List :: $EP\ (List\ a)\ (List^\circ\ a)$

Structure of datatypes – more examples

```
data Fork a      = Fork a a
type Forko a    = Con (a :: a)
data Perfect a  = ZeroP a | SuccP (Perfect (Fork a))
type Perfecto a = Con a
                :+: Con (Perfect (Fork a))
data Tree a b   = Leaf a
                | Node (Tree a b) b (Tree a b)
type Treeo a b = Con a
                :+: Con (Tree a b :: b :: Tree a b)
```

Specialisation of generic functions – example

```

data Tree a b    = Leaf a
                  | Node (Tree a b) b (Tree a b)
type Treeo a b  = Con a
                  :+: Con (Tree a b :+: b :+: Tree a b)

```

```

gmap {Tree String}  ~> gmap {Tree} (gmap {String})
gmap {Tree}         ~> wrapper around gmap {Treeo}
gmap {Treeo}       ~> λa b → gmap {:+:}
                               (gmap {Con} a)
                               (gmap {:+:}
                                (gmap {Tree} a b)
                                (gmap {:+:}
                                 b (gmap {Tree} a b)
                                )
                               )
gmap {String} ~> gmap {[Char]} ~> gmap{[]} (gmap {Char})
gmap{[]}      ~> wrapper around gmap{[ ]o}
gmap{[ ]o}   ~> ...

```

Specialisation of generic functions – type expressions

poly{*UltraComplexTree* (*Mostly Harmless*) *ElementType*}

- Complex type expressions and type synonyms in type arguments are simplified: type application is replaced by value application, type abstraction is replaced by type abstraction

poly{*UltraComplexTree*} (*poly*{*Mostly*} (*poly*{*Harmless*}))
(*poly*{*ElementType*})

Specialisation of generic functions – named types

poly{*Mostly*}

poly{*Harmless*}

There are two possibilities how to deal with simple type arguments:

- If there is a case for that type in the generic function, then the specialisation equals that case.

$$\begin{array}{lcl} \textit{poly}\{\textit{Harmless}\} & \rightsquigarrow & \textit{spec_poly_Harmless} \\ \textit{spec_poly_Harmless} & = & \textit{case_poly_Harmless} \end{array}$$

- Otherwise, the default generic behaviour takes place.

$$\begin{array}{lcl} \textit{poly}\{\textit{Mostly}\} & \rightsquigarrow & \textit{spec_poly_Mostly} \\ \textit{spec_poly_Mostly} & \rightsquigarrow & \text{wrapper around } \textit{poly}\{\textit{Mostly}^\circ\} \end{array}$$

PART II

New features of Generic HASKELL

Example datatypes

```
data Var    = V String
data Type  = TVar Var
              | Arrow Type Type
data Expr  = Var Var
              | App Expr Expr
              | Lambda (Var, Type) Expr
              | Let (Var, Type) Expr Expr
```

Default cases

There is a large class of generic functions that do almost nothing (pass on results, preserve structure, . . .) in a large part of the tree.

But in a few places something special happens.

```
-- type VarCollect{*} t           = t → [ Var ]
varcollect {t :: k}              :: VarCollect{k} t
varcollect {Unit}                x           = []
varcollect {:+:} cA cB (Inl a)      = cA a
varcollect {:+:} cA cB (Inr b)      = cB b
varcollect {*:} cA cB (a :* b)      = cA a 'union' cB b
varcollect {Con c} cA (Con a)       = cA a
varcollect {Label l} cA (Label a)   = cA a
varcollect {Char}                x           = []
varcollect {Int}                  x           = []
varcollect {Var}                   v         = [ v ]
```

Only the last line (and maybe the product case) is really application specific. There are many collection functions over other data types that share all the lines except the *Var* line.

Default cases – continued

Default cases allow to abstract out common cases from generic functions.
We start by defining a very general function which is not useful in itself:

```
type Collect {*}      t a          = t → [a]
type Collect {k → l} t a          =
  ∀ u . Collect {k} u a → Collect {l} (t u) a
collect {t :: k}      x          :: ∀ a . Collect {k} t a
collect {Unit}       x          = []
collect {:+:}       cA cB (Inl a) = cA a
collect {:+:}       cA cB (Inr b) = cB b
collect {*:}       cA cB (a *: b) = cA a ++ cB b
collect {Con c}    cA      (Con a) = cA a
collect {Label l} cA      (Label a) = cA a
collect {Char}    x          = []
collect {Int}    x          = []
```


Default cases – translation

$\text{varcollect}\{t :: k\}$		$::$	$\text{Collect}\{k\} t \text{ Var}$
$\text{varcollect}\{ \text{Var} \}$	v	$=$	$[v]$
$\text{varcollect}\{ :: \}$	$cA \ cB \ (a :: b)$	$=$	$cA \ a \ \text{'union'} \ cB \ b$
$\text{varcollect}\{ \text{Unit} \}$		$=$	$\text{collect}\{ \text{Unit} \}$
$\text{varcollect}\{ :: \}$		$=$	$\text{collect}\{ \text{Unit} \}$
$\text{varcollect}\{ \text{Con } c \}$		$=$	$\text{collect}\{ \text{Con } c \}$
$\text{varcollect}\{ \text{Label } l \}$		$=$	$\text{collect}\{ \text{Label } l \}$
$\text{varcollect}\{ \text{Char} \}$		$=$	$\text{collect}\{ \text{Char} \}$
$\text{varcollect}\{ \text{Int} \}$		$=$	$\text{collect}\{ \text{Int} \}$

This is the same as . . .

Default cases – translation, continued

. . . the original definition:

$\text{varcollect}\{t :: k\}$		$::$	$\text{Collect}\{k\} t \text{ Var}$
$\text{varcollect}\{Var\}$	v	$=$	$[v]$
$\text{varcollect}\{::\}$	$cA \ cB \ (a :: b)$	$=$	$cA \ a \ \text{'union'} \ cB \ b$
$\text{varcollect}\{Unit\}$	x	$=$	$[\]$
$\text{varcollect}\{+\}$	$cA \ cB \ (\text{Inl } a)$	$=$	$cA \ a$
$\text{varcollect}\{+\}$	$cA \ cB \ (\text{Inr } b)$	$=$	$cB \ b$
$\text{varcollect}\{Con \ c\}$	$cA \ (\text{Con } a)$	$=$	$cA \ a$
$\text{varcollect}\{Label \ l\}$	$cA \ (\text{Label } a)$	$=$	$cA \ a$
$\text{varcollect}\{Char\}$	x	$=$	$[\]$
$\text{varcollect}\{Int\}$	x	$=$	$[\]$

Note that *varcollect* does not contain any recursive calls to *collect*! All *Var* occurrences, no matter how deep they are in the tree, will be collected by *varcollect*.

Extending a generic function by means of a default case does modify the original behaviour in a deep fashion.

Constructor cases

- Generic HASKELL allows special cases for specific types, but sometimes that is not fine-grained enough!
- Datatypes can have a large number of constructors, and often only one or two constructors need special treatment, whereas the others follow the generic behaviour.
- With a special case for the datatype, the generic behaviour for the uninteresting constructors would need to be reprogrammed explicitly.

⇒ Use constructor cases!

Collecting free variables

Once more the example datatypes:

```
data Var    = V String
data Type  = TVar Var
           | Arrow Type Type
data Expr  = Var Var
           | App Expr Expr
           | Lambda (Var, Type) Expr
           | Let (Var, Type) Expr Expr
```

Let us collect the free variables now:

- In principle the same as *varcollect*.
- For *Lambda* and *Let*, special actions are required.

More about constructor cases

Cases for constructors have the same type as cases for the corresponding type (would) have.

```
-- Lambda :: (Var, Type) -> Expr -> Expr
-- freecollect { case Lambda } (Lambda (v, t) e) :: Collect {*} t Expr
freecollect { case Lambda } (Lambda (v, t) e) = filter (≠ v) (freecollect { Expr } e)
case_freecollect_Case_Lambda :: Expr -> [ Var ]
case_freecollect_Case_Lambda (Lambda (v, t) e) = filter (≠ v) (freecollect { Expr } e)
```

Generic functions containing constructor cases are still *generic*. For instance, the call

```
freecollect { Tree Expr Int }
```

would collect all variables that appear free in one of the expressions in the tree.

Calling constructor cases

For constructor cases to be actually executed, the type translation has to be adapted:

Before:

```
data List a    = Nil | Cons{hd :: a, tl :: (List a)}  
type Listo a  = Con Unit :+: Con (Label a **: (Label (List a)))
```

Now:

```
data Listo a    = Con (Case_Nil a) :+: Con (Case_Cons a)  
type Case_Nil a    = List a  
type Case_Cons a   = List a  
type Case_Nilo a   = Unit  
type Case_Conso a  = Label a **: Label (List a)
```

The specialisation mechanism will do the rest.

Conclusions and future work

- Default and constructor cases allow generic traversals to be conveniently expressed within Generic HASKELL.
- Even generic functions can benefit from specific behaviour.
- The extensions are rather ad-hoc, but useful.
- There are still things we can't do yet.

Try and see yourself!

- Generic HASKELL 1.23 (Beryl) has been released on July 5 and implements the extensions presented here.
- It is available for download at

<http://www.generic-haskell.org>

- There will be a demonstration after lunch.