

Extensible and Modular Generics for the Masses

Sean Leather

Utrecht University

26 September, 2011

Previously...

You learned about:

- Datatypes and Kinds
- Lightweight Implementation of Generics and Dynamics (LIGD)
- Scrap Your Boilerplate (SYB)

This time...

We're going to talk about the library Extensible and Modular Generics for the Masses (EMGM).

- Define an example generic function
- Introduce the run-time type representation
- Add datatype-generic support
- Demonstrate support for ad-hoc cases
- Change the representation to be extensible and modular
- Define other generic functions
 - ▶ Producer functions
 - ▶ Higher-kinded datatypes
 - ▶ Abstracting over more than one type

Defining an Example: Equality (1)

Defining a generic function in EMGM involves several steps. First, let's decide what the “ideal” type signature should look like.

```
geq :: a → a → Bool
```

Defining an Example: Equality (2)

Next, we need to define a **newtype** for the generic function.

```
newtype Geq a = Geq { selEq :: a → a → Bool }
```

This is similar the use of **newtype** in LIGD.

Defining an Example: Equality (3)

Now, we implement the structural components of our generic function.

```
geq_unit      Unit      Unit      = True
geq_int       i         j         = i  $\equiv$  j
geq_char      c         d         = c  $\equiv$  d
geq_plus r_a r_b (L a1) (L a2) = selEq r_a a1 a2
geq_plus r_a r_b (R b1) (R b2) = selEq r_b b1 b2
geq_plus r_a r_b _       _       = False
geq_prod r_a r_b (a1 : $\times$ : b1) (a2 : $\times$ : b2) = selEq r_a a1 a2  $\wedge$ 
                                                selEq r_b b1 b2
```

Defining an Example: Equality (4)

That should look familiar. Here's `geq` in LIGD.

```
geq (RUnit)      Unit      Unit      = True
geq (RInt  )      i        j        = i  $\equiv$  j
geq (RChar)      c        d        = c  $\equiv$  d
geq (RSum ra rb) (L a1)  (L a2)  = geq ra a1 a2
geq (RSum ra rb) (R b1)  (R b2)  = geq rb b1 b2
geq (RSum ra rb) –        –        = False
geq (RProd ra rb) (a1 :×: b1) (a2 :×: b2) = geq ra a1 a2 ∧
                                                geq rb b1 b2
```

Defining an Example: Equality (5)

Next, we create an instance of the `Generic` type class using our generic functions.

```
instance Generic Geq where  
  runit      = Geq geq_unit  
  rint       = Geq geq_int  
  rchar      = Geq geq_char  
  rsum ra rb = Geq (geq_plus ra rb)  
  rprod ra rb = Geq (geq_prod ra rb)
```

How does this tie the recursive knot with `selEq` ?

Defining an Example: Equality (6)

At this point, our generic function is (partially) usable.

```
selEq (rprod rchar rint) ('Q' :×: 42) ('Q' :×: 42) ≡ True
```

But that's not good enough...

Defining an Example: Equality (7)

We want to hide the **type representation** argument...

```
geq :: (Rep a) => a -> a -> Bool  
geq = selEq rep
```

... to make it **implicit**:

```
geq ('Q' :×: 42) ('Q' :×: 42) ≡ True
```

The Mechanics: Run-time Type Representation (1)

Now, let's talk about the run-time type representation machinery that allows us to define functions such as `geq`.

First, you should recall these structure representation types. They are the same as those in LIGD.

```
data Unit = Unit
```

```
data a :+: b = L a | R b
```

```
data a :×: b = a :×: b
```

The Mechanics: Run-time Type Representation (2)

The `Generic` class has a method for each representation type.

```
class Generic g where  
  runit :: g Unit  
  rint  :: g Int  
  rchar :: g Char  
  rsum  :: g a → g b → g (a :+: b)  
  rprod :: g a → g b → g (a :×: b)
```

An instance of `Generic` defines a type-indexed function.

The Mechanics: Run-time Type Representation (3)

To make the representation value implicit, we use the `Rep` class.

```
class Rep a where  
  rep :: (Generic g)  $\Rightarrow$  g a
```

This allows us to substitute `rep` for any instance of `Generic`.

The Mechanics: Run-time Type Representation (4)

The instances of `Rep` include all representable types. We start with the universe of base and structure types.

instance Rep Unit **where**

rep = runit

instance Rep Int **where**

rep = rint

instance Rep Char **where**

rep = rchar

instance (Rep a, Rep b) \Rightarrow Rep (a :+: b) **where**

rep = rsum rep rep

instance (Rep a, Rep b) \Rightarrow Rep (a : \times : b) **where**

rep = rprod rep rep

Expanding the Universe (1)

To make our functions truly generic, we need to expand our universe to include user-defined datatypes.

```
class Generic g where
```

```
...
```

```
  rtype :: EP b a  $\rightarrow$  g a  $\rightarrow$  g b
```

Recall the analogous LIGD constructor:

```
RType :: EP b a  $\rightarrow$  Rep a  $\rightarrow$  Rep b
```

Recall the embedding-projection pair datatype.

```
data EP d r = EP { from :: (d  $\rightarrow$  r), to :: (r  $\rightarrow$  d) }
```

Expanding the Universe (2)

The representation for `List` is:

```
rList :: (Generic g) => g a -> g (List a)
rList r_a = rtype (EP fromList toList)
              (rsum runit (rprod r_a (rList r_a)))
```

Again, notice the similarity to LIGD:

```
rList r_a = RType (EP fromList toList)
                (RSum RUnit (RProd r_a (rList r_a)))
```


Expanding the Universe (3)

To add `rList` as another implicit representation, we define an instance of `Rep` for `List`.

```
instance (Rep a)  $\Rightarrow$  Rep (List a) where  
  rep = rList rep
```

Expanding the Universe (4)

To make `geq` a generic function that supports user-defined datatypes, we add another case.

```
geq_dt ep r_a a_1 a_2 = selEq r_a (from ep a_1) (from ep a_2)
```

```
instance Generic Geq where
```

```
...
```

```
rtype ep r_a = Geq (geq_dt ep r_a)
```

Overloaded and Ad-hoc (1)

Let's write a generic `show` function. Think: **deriving** `Show`.

```
gshow :: a → String
```

But we don't have access to the constructor names.

For that, we can add another case to our generic function signature.

```
class Generic g where
```

```
...
```

```
rcon :: String → g a → g a
```

`rcon` is a wrapper around other structure types.

Overloaded and Ad-hoc (2)

We then add `rcon` to wrap each alternative in `rsum` with the name of the constructor.

```
rList :: (Generic g) => g a -> g (List a)
rList r_a = rtype (EP fromList toList)
              (rsum (rcon "Nil" runit)
                  (rcon "Cons" (rprod r_a (rList r_a))))
```

Overloaded and Ad-hoc (3)

Now, we can implement the cases of `Gshow`. Most of the entries are exactly as you would expect (see lecture notes).

```
newtype Gshow a = Gshow { selShow :: a → String }
```

```
gshow_unit      Unit = ""
```

```
...
```

```
gshow_dt      ep ra a    = selShow ra (from ep a)
```

```
gshow_constr s ra a    = "(" ++ s ++  
                          " " ++ selShow ra a ++  
                          ")"
```

```
instance Generic Gshow where
```

```
  runit = Gshow gshow_unit
```

```
...
```

Overloaded and Ad-hoc (4)

The final generic show function looks like this:

```
gshow :: (Rep a) => a -> String  
gshow = selShow rep
```

And it works like this:

```
gshow (Cons 4 (Cons 2 Nil)) ≡ "(Cons 4 (Cons 2 (Nil )))"
```

But the output is ugly! We need to fix it...

Overloaded and Ad-hoc (5)

We want something specific for `List`. Instead of the general `rList` representation based on `rtype`, we can add a special list case to `Generic`.

```
class Generic g where
```

```
...
```

```
list :: g a → g (List a)
```

We also need to register `list` as a representable type.

```
instance (Rep a) ⇒ Rep (List a) where
```

```
rep = list rep
```

Overloaded and Ad-hoc (6)

We extend `gshow` for lists...

```
gshow_list ra Nil          = "[]"  
gshow_list ra (Cons a as) = selShow ra a ++ ":" ++  
                             selShow (list ra) as
```

instance Generic Gshow **where**

```
...  
list ra = Gshow (gshow_list ra)
```

... arriving at a more concise output:

```
gshow (Cons 4 (Cons 2 Nil)) ≡ "4:2: []"
```


Becoming Modular and Extensible (1)

Modifying the `Generic` class for every type is bad. The process is not modular and reduces the reusability of a library. (Just like LIGD.) We can change this with *EMGM*.

Let's try a hierarchy of classes.

```
class (Generic g)  $\Rightarrow$  GenericList g where
```

```
  rlist :: g a  $\rightarrow$  g (List a)
```

```
  rlist = rList
```

```
instance GenericList Gshow where
```

```
  rlist ra = Gshow (gshow_list ra)
```

We can now use `selShow`.

```
selShow (rlist rint) (Cons 2 Nil)  $\equiv$  "2: []"
```

Becoming Modular and Extensible (2)

What happens when we define the following instance?

```
instance (Rep a)  $\Rightarrow$  Rep (List a) where  
  rep = rlist rep
```

GHC complains:

```
Could not deduce (GenericList g)  
  from the context (Rep (List a), Rep a, Generic g)  
  arising from a use of 'rlist' at ...
```

Possible fix:

```
add (GenericList g) to the context of  
  the type signature for 'rep' ...
```

The current type signature for `rep`:

```
rep :: (Generic g, Rep a) => g a
```

What happens if we follow GHC's advise?

Possible fix:

```
add (GenericList g) to the context of  
the type signature for 'rep' ...
```

Becoming Modular and Extensible (3)

Instead, let's not assume that `g` is always an instance of `Generic`. We abstract over the type constructor in `Rep`.

```
class Rep g a where  
  rep :: g a
```

Becoming Modular and Extensible (4)

We rewrite the instances from before:

```
instance (Generic g)  $\Rightarrow$  Rep g Unit where
```

```
  rep = runit
```

```
instance (Generic g, Rep g a, Rep g b)  $\Rightarrow$  Rep g (a :+: b)
```

```
  where rep = rsum rep rep
```

```
...
```

And use `GenericList` in the context for the `List` instance instead of `Generic`.

```
instance (GenericList g, Rep g a)  $\Rightarrow$  Rep g (List a) where
```

```
  rep = rlist rep
```

Becoming Modular and Extensible (5)

Lastly, we rewrite the generic show...

```
gshow :: (Rep Gshow a) => a -> String  
gshow = selShow rep
```

... by explicitly filling in the **newtype** Gshow for the **g** parameter. Let's move on to some other examples. Some of them challenge the approaches we've shown so far.

A Simple Producer: Empty

Here is a simple generic producer function in its entirety.

```
newtype Gempty a = Gempty { selEmpty :: a }  
instance Generic Gempty where  
  runit          = Gempty Unit  
  rint           = Gempty 0  
  rchar          = Gempty '\NUL'  
  rsum    ra rb = Gempty (L (selEmpty ra))  
  rprod    ra rb = Gempty (selEmpty ra :×: selEmpty rb)  
  rtype ep ra   = Gempty (to ep (selEmpty ra))  
  rcon s    ra   = Gempty (selEmpty ra)  
  
gempty :: (Rep Gempty a) ⇒ a  
gempty = selEmpty rep
```

Higher Kinds: Crush (1)

We have dealt with types of kind $*$ up to this point. How do we deal with kind $* \rightarrow *$? These include the “container” datatypes: `List a`, `Tree a`, etc.

We use a generic `crush` function as an example.

Higher Kinds: Crush (2)

Recall the standard `foldr` function:

$$\text{foldr} :: (a \rightarrow b \rightarrow b) \rightarrow b \rightarrow [a] \rightarrow b$$

It generalizes to `crushr`:

$$\text{crushr} :: (a \rightarrow b \rightarrow b) \rightarrow b \rightarrow f\ a \rightarrow b$$

- $(a \rightarrow b \rightarrow b)$ — A “combining” function
- b — A “zero” value
- $f\ a$ — A container

Higher Kinds: Crush (3)

The type-indexed function is straightforward.

```
newtype Crush b a = Crush { selCrush :: a → b → b }  
crushr_unit      _      e = e  
...  
crushr_plus  ra rb (L a)  e = selCrush ra a e  
crushr_plus  ra rb (R b)  e = selCrush rb b e  
crushr_prod  ra rb (a × b) e = selCrush ra a (selCrush rb b e)  
crushr_dt ep ra  a      e = selCrush ra (from ep a) e  
instance Generic (Crush b) where  
  runit = Crush crushr_unit  
  ...
```

Higher Kinds: Crush (4)

We have `selCrush`, so how do we write `crushr`? Recall `rep` again.

```
class Rep g a where  
  rep :: g a
```

The type variable `a` has kind `*`. We want to abstract over container types of the form `f a` where `f` has kind `* → *`.

Key: The type of the representation function reflects the kind of the represented type.

```
class FRep g f where  
  frep :: g a → g (f a)
```

Higher Kinds: Crush (5)

Translating an instance from `Rep`

```
instance (Generic g, Rep g a)  $\Rightarrow$  Rep g (List a) where  
  rep = rList rep
```

to `FRep`

```
instance (Generic g)  $\Rightarrow$  FRep g List where  
  frep = rList
```

requires removing all references to the type variables for the container's element.

Higher Kinds: Crush (6)

How do we come up with ...

```
crushr :: (...)  $\Rightarrow$  (a  $\rightarrow$  b  $\rightarrow$  b)  $\rightarrow$  b  $\rightarrow$  f a  $\rightarrow$  b
```

... given this, ...

```
selCrush :: Crush b a  $\rightarrow$  a  $\rightarrow$  b  $\rightarrow$  b
```

... this, ...

```
frep :: (FRep g f)  $\Rightarrow$  g a  $\rightarrow$  g (f a)
```

... and this?

```
Crush :: (a  $\rightarrow$  b  $\rightarrow$  b)  $\rightarrow$  Crush b a
```

Higher Kinds: Crush (7)

Let's assemble this type jigsaw puzzle:

```
selCrush :: Crush b a → a → b → b
```

```
frep :: (FRep g f) ⇒ g a → g (f a)
```

```
Crush :: (a → b → b) → Crush b a
```

First, `frep ∘ Crush` :

```
frep ∘ Crush ::
```

```
(FRep (Crush b) f) ⇒ (a → b → b) → Crush b (f a)
```

Higher Kinds: Crush (8)

Let's assemble this type jigsaw puzzle:

```
selCrush :: Crush b a → a → b → b
```

```
frep :: (FRep g f) ⇒ g a → g (f a)
```

```
Crush :: (a → b → b) → Crush b a
```

Then, `selCrush ∘ frep ∘ Crush` :

```
selCrush ∘ frep ∘ Crush ::
```

```
(FRep (Crush b) f) ⇒ (a → b → b) → f a → b → b
```

Higher Kinds: Crush (9)

Finally, we can define `crushr`.

```
crushr :: (FRep (Crush b) f) => (a -> b -> b) -> b -> f a -> b  
crushr f z x = selCrush (frep (Crush f)) x z
```

And we can use it, too.

```
gflatten :: (FRep (Crush [a]) f) => f a -> [a]  
gflatten = crushr (:) []  
gflatten (Cons 4 (Cons 2 Nil)) ≡ [4, 2]
```


Higher Abstraction: Map (1)

The standard `map` function is a very handy function. We often want to apply a function to all elements in a list.

```
map :: (a → b) → [a] → [b]
```

Why don't we generalise this to other datatypes as we generalised `foldr` to `crushr`?

```
gmap :: (a → b) → f a → f b
```

Higher Abstraction: Map (2)

The type-indexed function.

newtype Gmap a b = Gmap { selMap :: a → b }

gmap_unit x = x

...

gmap_plus r_a r_b (L a) = L (selMap r_a a)

gmap_plus r_a r_b (R b) = R (selMap r_b b)

gmap_prod r_a r_b (a :×: b) = selMap r_a a :×: selMap r_b b)

gmap_dt ep₁ ep₂ r_a a = (to ep₂ ∘ selMap r_a ∘ from ep₁) a

Higher Abstraction: Map (3)

`gmap` is both a generic consumer and generic producer, so we must abstract over two types, input and output.

```
class Generic2 g where
```

```
  runit2 :: g Unit Unit
```

```
  rint2  :: g Int Int
```

```
  rchar2 :: g Char Char
```

```
  rsum2  :: g a1 a2 → g b1 b2 → g (a1 :+: b1) (a2 :+: b2)
```

```
  rprod2 :: g a1 a2 → g b1 b2 → g (a1 :×: b1) (a2 :×: b2)
```

```
  rtype2 :: EP a2 a1 → EP b2 b1 → g a1 b1 → g a2 b2
```

Higher Abstraction: Map (4)

We define our instance of `Generic2`.

```
instance Generic2 Gmap where  
  runit2 = Gmap gmap_unit  
  ...  
  rtype2 ep1 ep2 ra = Gmap (gmap_dt ep1 ep2 ra)
```

Since we have this new `rtype2` method (rather than `rtype`), we need to redefine our list representation.

```
rList2 :: (Generic2 g) => g a b -> g (List a) (List b)  
rList2 ra = rtype2 (EP fromList toList)  
                  (EP fromList toList)  
                  (rsum2 runit2 (rprod2 ra (rList2 ra)))
```

Higher Abstraction: Map (5)

We can immediately use the list representation to implement the standard `map` on `List` containers.

```
mapList :: (a → b) → List a → List b  
mapList f = selMap (rList2 (Gmap f))
```

But our ultimate goal (as always) is to generalise...

Higher Abstraction: Map (6)

We can't use the `FRep` class. Why?

```
class FRep g f where  
  frep :: g a → g (f a)
```

We must extend it to support the higher-kinded `g (* → * → *)`, i.e. abstraction over two types.

```
class FRep2 g f where  
  frep2 :: g a b → g (f a) (f b)  
instance (Generic2 g) ⇒ FRep2 g List where  
  frep2 = rList2
```

Our instance for `List` is similar to the instance for `FRep`.

Higher Abstraction: Map (7)

We can now define `gmap` with a method similar to how we defined `crushr`.

```
gmap :: (FRep2 Gmap f)  $\Rightarrow$  (a  $\rightarrow$  b)  $\rightarrow$  f a  $\rightarrow$  f b  
gmap f = selMap (frep2 (Gmap f))
```

Conclusions

We have covered the following concepts of using generic functions in EMGM:

- Equality: basic
- Show: ad-hoc, extensible, and modular
- Empty: producer
- Crush: higher-kinded datatypes
- Map: abstraction over more than one type

Next time...

- Projects
- Deriving Generics