COMP4075: Lecture 7

Functional Programming Patterns:
Functor, Foldable, and Friends

Henrik Nilsson

University of Nottingham, UK

Type Classes and Patterns

- In Haskell, many functional programming patterns are captured through specific type classes.
- Additionally, the type class mechanism itself and the fact that overloading is prevalent in Haskell give raise to other programming patterns.

Semigroups and Monoids (1)

Semigroups and monoids are algebraic structures:

- ***Semigroup***: a set (type) S with an ***associative*** binary operation $\cdot : S \times S \rightarrow S$:

$$\forall a, b, c \in S \;:\; (a \cdot b) \cdot c = a \cdot (b \cdot c)$$

- ***Monoid***: a semigroup with an ***identity element***:

$$\exists e \in S, \forall a \in S \;:\; e \cdot a = a \cdot e = a$$

Semigroups and Monoids (2)

- Semigroups and monoids are patterns that appear frequently in everyday programming.
- Being explicit about when such structures are used
  - makes code clearer
  - offer opportunities for reuse
- The standard Haskell libraries provide type classes to capture these notions.

Class *Semigroup*

Class definition (most important methods):

$$\begin{array}{l} \textbf{class}\ \mathit{Semigroup}\ a\ \textbf{where} \\ \quad (\diamond) \qquad :: a \to a \to a \\ \quad \mathit{sconcat} :: \mathit{NonEmpty}\ a \to a \end{array}$$

Minimum complete definition: $(\diamond)$ (ASCII: `<>`)
(There is thus a default definition for $\mathit{sconcat}$.)

$\mathit{NonEmpty}$ is the non-empty list type:

$$\textbf{data}\ \mathit{NonEmpty}\ a = a :|\ [a]$$

Instances of *Semigroup* (1)

A list $[a]$ is a semigroup (for any type a):

$$\begin{array}{l} \textbf{instance}\ \mathit{Semigroup}\ [a]\ \textbf{where} \\ \quad (\diamond) = (+\!\!+) \end{array}$$

$\mathit{Maybe}\ a$ is a semigroup if a is one:

$$\begin{array}{l} \textbf{instance}\ \mathit{Semigroup}\ a \\ \qquad\qquad \Rightarrow \mathit{Semigroup}\ (\mathit{Maybe}\ a)\ \textbf{where} \\ \quad \mathit{Nothing} \diamond y \qquad\quad = y \\ \quad x \qquad\ \diamond \mathit{Nothing} = x \\ \quad \mathit{Just}\ x \ \ \diamond \mathit{Just}\ y \quad = x \diamond y \end{array}$$

Instances of *Semigroup* (2)

Addition and multiplication are associative; a numeric type with either operation forms a semigroup.

But which one to pick? Both are equally useful!

Idea:

- $\mathit{Sum}\ a$: the semigroup $(a, (+))$
- $\mathit{Product}\ a$: the semigroup $(a, (*))$

Instances of *Semigroup* (3)

Semigroup instances for $\mathit{Sum}\ a$ and $\mathit{Product}\ a$:

$$\begin{array}{l} \textbf{instance}\ \mathit{Num}\ a \Rightarrow \mathit{Semigroup}\ (\mathit{Sum}\ a)\ \textbf{where} \\ \quad (\diamond) = (+) \\ \textbf{instance}\ \mathit{Num}\ a \Rightarrow \mathit{Semigroup}\ (\mathit{Product}\ a)\ \textbf{where} \\ \quad (\diamond) = (*) \end{array}$$

Instances of *Semigroup* (4)

Similarly, any type with a total ordering forms a semigroup with maximum or minimum as the associative operation:

- $Max\ a$: the semigroup $(a,\ max)$
- $Min\ a$: the semigroup $(a,\ min)$

Semigroup instances:

$$\textbf{instance}\ Ord\ a \Rightarrow Semigroup\ (Max\ a)\ \textbf{where}$$
$$(\diamond) = max$$
$$\textbf{instance}\ Ord\ a \Rightarrow Semigroup\ (Min\ a)\ \textbf{where}$$
$$(\diamond) = min$$

Instances of *Semigroup* (5)

All products of semigroups are semigroups; e.g.:

$$\textbf{instance}\ (Semigroup\ a, Semigroup\ b)$$
$$\Rightarrow Semigroup\ (a, b)\ \textbf{where}$$
$$(x, y) \diamond (x', y') = (x \diamond x', y \diamond y')$$

$a \to b$ is a semigroup if the range b is a semigroup:

$$\textbf{instance}\ Semigroup\ b$$
$$\Rightarrow Semigroup\ (a \to b)\ \textbf{where}$$
$$f \diamond g = \lambda x \to f\ x \diamond g\ x$$

Exercise: *Semigroup* Instances

What is the value of the following expressions?

$$[1, 3, 7] \diamond [2, 4]$$
$$Sum\ 3 \diamond Sum\ 1 \diamond Sum\ 5$$
$$Just\ (Max\ 42) \diamond Nothing \diamond Just\ (Max\ 3)$$
$$sconcat\ (Product\ 2 :|\ [Product\ 3, Product\ 4])$$
$$([1], Product\ 2) \diamond ([2, 3], Product\ 3)$$
$$((1:) \diamond tail)\ [4, 5, 6]$$

Class *Monoid*

Recall: A monid is a semigroup with an identity element:

$$\textbf{class}\ Semigroup\ a \Rightarrow Monoid\ a\ \textbf{where}$$
$$mempty\ :: a$$
$$mappend :: a \to a \to a$$
$$mappend = (\diamond)$$
$$mconcat\ :: [a] \to a$$
$$mconcat\ = foldr\ mappend\ mempty$$

Minimum complete definition: $mempty$

Instances of $Monoid$ (1)

A list $[a]$ is the archetypical example of a monoid:

$$\textbf{instance}\ Monoid\ [a]\ \textbf{where}$$
$$\quad mempty = [\,]$$

Any semigroup can be turned into a monoid by adjoining an identity element:

$$\textbf{instance}\ Semigroup\ a$$
$$\qquad \Rightarrow Monoid\ (Maybe\ a)\ \textbf{where}$$
$$\quad mempty = Nothing$$

Instances of $Monoid$ (2)

Monoid instances for $Sum\ a$ and $Product\ a$:

$$\textbf{instance}\ Num\ a \Rightarrow Monoid\ (Sum\ a)\ \textbf{where}$$
$$\quad mempty = Sum\ 0$$
$$\textbf{instance}\ Num\ a \Rightarrow Monoid\ (Product\ a)\ \textbf{where}$$
$$\quad mempty = Product\ 1$$

Instances of $Monoid$ (3)

Monoid instances for $Min\ a$ and $Max\ a$:

$$\textbf{instance}\ (Ord\ a, Bounded\ a) \Rightarrow$$
$$\qquad Monoid\ (Min\ a)\ \textbf{where}$$
$$\quad mempty = maxBound$$
$$\textbf{instance}\ (Ord\ a, Bounded\ a) \Rightarrow$$
$$\qquad Monoid\ (Max\ a)\ \textbf{where}$$
$$\quad mempty = minBound$$

Instances of $Monoid$ (4)

All products of monoids are monoids; e.g.:

$$\textbf{instance}\ (Monoid\ a, Monoid\ b)$$
$$\qquad \Rightarrow Monoid\ (a, b)\ \textbf{where}$$
$$\quad mempty = (mempty, mempty)$$

$a \rightarrow b$ is a monoid if the range b is a monoid:

$$\textbf{instance}\ Monoid\ b \Rightarrow Monoid\ (a \rightarrow b)\ \textbf{where}$$
$$\quad mempty\ \_ = mempty$$

Functors (1)

A Functor is a notion that originated in a branch of mathematics called Category Theory.

However, for our purposes, we can think of functors as type constructors T (of arity 1) for which a function map can be defined:

$$map :: (a \to b) \to T a \to T b$$

that satisfies the following laws:

$$\begin{aligned} map\ id &= id \\ map (f \circ g) &= map\ f \circ map\ g \end{aligned}$$

Functors (2)

Common examples of functors include (but are not limited to) ***container types*** like lists:

$$\begin{aligned} &mapList :: (a \to b) \to [a] \to [b] \\ &mapList\ \_\ [\,] \qquad\quad = [\,] \\ &mapList\ f\ (x : xs) = f\ x : mapList\ f\ xs \end{aligned}$$

Functors (3)

And trees; e.g.:

$$\begin{aligned} &\mathbf{data}\ Tree\ a = Leaf\ a \\ &\qquad\qquad\quad \mid Node\ (Tree\ a)\ a\ (Tree\ a) \\ &mapTree :: (a \to b) \to Tree\ a \to Tree\ b \\ &mapTree\ f\ (Leaf\ x) \qquad = Leaf\ (f\ x) \\ &mapTree\ f\ (Node\ l\ x\ r) = Node\ (mapTree\ f\ l) \\ &\qquad\qquad\qquad\qquad\qquad\quad (f\ x) \\ &\qquad\qquad\qquad\qquad\qquad\quad (mapTree\ f\ r) \end{aligned}$$

Class *Functor* (1)

Of course, the notion of a functor is captured by a type class in Haskell:

$$\begin{aligned} &\mathbf{class}\ Functor\ f\ \mathbf{where} \\ &\quad fmap :: (a \to b) \to f\ a \to f\ b \\ &\quad (<\$) :: a \to f\ b \to f\ a \\ &\quad (<\$) = fmap \circ const \end{aligned}$$

Class $Functor$ (2)

There is also an infix version that can be viewed as function application lifted over a functor:

$$(<\$>) :: (a \to b) \to f\ a \to f\ b$$
$$(<\$>) = fmap$$

Compare the standard infix function application operator:

$$(\$) :: (a \to b) \to a \to b$$

Class $Functor$ (3)

However, Haskell's type system is not powerful enough to enforce the functor laws.

In general, the programmer is responsible for ensuring that an instance respects all laws associated with a type class.

Note that the type of $fmap$ can be read:

$$(a \to b) \to (f\ a \to f\ b)$$

That is, we can see $fmap$ as promoting a function to work in a different context.

Instances of $Functor$ (1)

As noted, list is a functor:

$$\textbf{instance}\ Functor\ [\,]\ \textbf{where}$$
$$\quad fmap = listMap$$

$Maybe$ is also a functor:

$$\textbf{instance}\ Functor\ Maybe\ \textbf{where}$$
$$\quad fmap\ \_\ Nothing = Nothing$$
$$\quad fmap\ f\ (Just\ x) = Just\ (f\ x)$$

Instances of $Functor$ (2)

Container types are in general instances of functor, including $Array$:

$$\textbf{instance}\ Functor\ (Array\ i)\ \textbf{where} \ldots$$

E.g, given a matrix $m :: Array\ (Int, Int)\ Double$, we can double all elements:

$$fmap\ (*2)\ m$$

Instances of *Functor* (3)

As functors are so common, there is a GHC extension for deriving *Functor* instances in standard cases.

For example, the functor instance for our tree type can be derived:

$$
\begin{aligned}
&\mathbf{data}\ \mathit{Tree}\ a = \mathit{Leaf}\ a \\
&\qquad\qquad\qquad \mid\ \mathit{Node}\ (\mathit{Tree}\ a)\ a\ (\mathit{Tree}\ a) \\
&\qquad\qquad\ \mathbf{deriving}\ \mathit{Functor}
\end{aligned}
$$

Instances of *Functor* (4)

The type of functions from a given domain is an example of a functor that is ***not a container*** type. Map is just function composition:

$$
\begin{aligned}
&\mathbf{instance}\ \mathit{Functor}\ ((\rightarrow)\ a)\ \mathbf{where} \\
&\quad \mathit{fmap} = (\circ)
\end{aligned}
$$

Note that a ***curried*** function type, like

$$a \rightarrow b \rightarrow c = a \rightarrow (b \rightarrow c)$$

thus is a ***nesting*** or ***composition*** of functors:

$$(((\rightarrow)\ a)\ (((\rightarrow)\ b)\ c)) = (((\rightarrow)\ a) \circ ((\rightarrow)\ b))\ c$$

Nesting functors (1)

In practice, functors often appear nested inside other functors, e.g.

$$\mathit{mxs} :: [\mathit{Maybe}\ \mathit{Double}]$$

Such a structure can of course be processed by repeated mapping, e.g.:

$$\mathit{fmap}\ (\mathit{fmap}\ (*2))\ \mathit{mxs}$$

One reading of this is “use fmap to lift $(*2)$ to work on Maybe, and then map that over the list”.

Nesting functors (2)

However, in general $f\ (g\ a) = (f \circ g)\ a$, meaning

$$\mathit{fmap}\ (\mathit{fmap}\ (*2)) = (\mathit{fmap} \circ \mathit{fmap})\ (*2)$$

suggesting the following combinator:

$$
\begin{aligned}
&(<\$\$>) :: (\mathit{Functor}\ f, \mathit{Functor}\ g) \Rightarrow \\
&\qquad\qquad (a \rightarrow b) \rightarrow f\ (g\ a) \rightarrow f\ (g\ b) \\
&(<\$\$>) = \mathit{fmap} \circ \mathit{fmap}
\end{aligned}
$$

This allows us to simplify $\mathit{fmap}\ (\mathit{fmap}\ (*2))\ \mathit{mxs}$ to

$$(*2)\ <\$\$>\ \mathit{mxs}$$

Nesting functors (3)

Note that the composition of $fmap$s is mirrored by composition of functors at the type level:

$$[Maybe\ (a)] = [\,]\ (Maybe\ a) = ([\,] \circ Maybe)\ a$$

This can be generalized to any number levels; e.g.

$$(<\$\$\$>) = fmap \circ fmap \circ fmap$$
$$(*2)\ <\$\$\$>\ [[[1,2],[3]],[[4]],[[5]]]$$
$$\Rightarrow [[[2,4],[6]],[[8]],[[10]]]$$

$Data.Functor.Syntax$ defines $<\$\$>$, $<\$\$\$>$ ...

Nesting functors (4)

Note that we also could have defined:

$$(<\$\$>) = fmap\ fmap\ fmap$$

Why?

Exploiting that curried function types are composed functors, $<\$\$>$, $<\$\$\$>$ ... can compose functions where the second function has arity 2, 3, ...:

$$f :: Bool \to Double \to Int \to Double$$
$$(>0)\ <\$\$\$>\ f :: Bool \to Double \to Int \to Bool$$

This is often quite handy in practice.

Class *Foldable* (1)

Class of data structures that can be folded to a summary value.

Many methods; minimal instance $foldMap$, $foldr$:

$$
\begin{array}{l}
\textbf{class}\ Foldable\ t\ \textbf{where} \\
\quad fold \quad :: Monoid\ m \Rightarrow t\ m \to m \\
\quad foldMap :: Monoid\ m \Rightarrow (a \to m) \to t\ a \to m \\
\quad foldr \quad :: (a \to b \to b) \to b \to t\ a \to b \\
\quad foldr' \quad :: (a \to b \to b) \to b \to t\ a \to b \\
\quad foldl \quad :: (b \to a \to b) \to b \to t\ a \to b \\
\quad foldl' \quad :: (b \to a \to b) \to b \to t\ a \to b
\end{array}
$$

Class *Foldable* (2)

(continued)

$$
\begin{array}{l}
foldr1 :: (a \to a \to a) \to t\ a \to a \\
foldl1 :: (a \to a \to a) \to t\ a \to a \\
toList :: t\ a \to [a] \\
null \quad :: t\ a \to Bool \\
length :: t\ a \to Int \\
elem \quad :: Eq\ a \Rightarrow a \to t\ a \to Bool
\end{array}
$$

(Note that $length$ should be understood as $size$.)

Class *Foldable* (3)

(continued)

$$
\begin{aligned}
\mathit{maximum} &:: \mathit{Ord}\ a \Rightarrow t\ a \rightarrow a \\
\mathit{minimum} &:: \mathit{Ord}\ a \Rightarrow t\ a \rightarrow a \\
\mathit{sum} &:: \mathit{Num}\ a \Rightarrow t\ a \rightarrow a \\
\mathit{product} &:: \mathit{Num}\ a \Rightarrow t\ a \rightarrow a
\end{aligned}
$$

Note: foldl typically incurs a large space overhead due to laziness. The version with strict applictaion of the operator, foldl' is typically preferable.

Instances of *Foldable* (1)

All expected instances, e.g.:

- $\mathbf{instance}\ \mathit{Foldable}\ [\,]\ \mathbf{where} \ldots$
- $\mathbf{instance}\ \mathit{Foldable}\ \mathit{Maybe}\ \mathbf{where} \ldots$

And GHC extension allows deriving instances in many cases; e.g.

$$\mathbf{data}\ \mathit{Tree}\ a = \ldots \mathbf{deriving}\ \mathit{Foldable}$$

Instances of *Foldable* (2)

But there are also some instances that are less expected, e.g.:

- $\mathbf{instance}\ \mathit{Foldable}\ (\mathit{Either}\ a)\ \mathbf{where} \ldots$
- $\mathbf{instance}\ \mathit{Foldable}\ ((,)\ a)\ \mathbf{where} \ldots$

This has some arguably odd consequences:

$$
\begin{aligned}
&\mathit{length}\ (1, 2) && \Rightarrow 1 \\
&\mathit{sum}\ (1, 2) && \Rightarrow 2 \\
&\mathit{length}\ (\mathit{Left}\ 1) && \Rightarrow 0 \\
&\mathit{length}\ (\mathit{Right}\ 2) && \Rightarrow 1
\end{aligned}
$$

Example: Folding Over a Tree (1)

Consider:

$$
\begin{aligned}
&\mathbf{data}\ \mathit{Tree}\ a = \mathit{Empty} \\
&\quad\quad \mid \mathit{Node}\ (\mathit{Tree}\ a)\ a\ (\mathit{Tree}\ a) \\
&\quad \mathbf{deriving}\ (\mathit{Show}, \mathit{Eq})
\end{aligned}
$$

Let us make it an instance of $\mathit{Foldable}$:

$$
\begin{aligned}
&\mathbf{instance}\ \mathit{Foldable}\ \mathit{Tree}\ \mathbf{where} \\
&\quad \mathit{foldMap}\ f\ \mathit{Empty} = \mathit{mempty} \\
&\quad \mathit{foldMap}\ f\ (\mathit{Node}\ l\ a\ r) = \\
&\quad\quad \mathit{foldMap}\ f\ l \diamond f\ a \diamond \mathit{foldMap}\ f\ r
\end{aligned}
$$

Example: Folding Over a Tree (2)

We wish to compute the sum and max over a tree of Int. One way:

$$sumMax :: Tree\ Int \rightarrow (Int, Int)$$
$$sumMax\ t = (foldl\ (+)\ 0\ t, foldl\ max\ minBound\ t)$$

Another way, with a single traversal:

$$sumMax :: Tree\ Int \rightarrow (Int, Int)$$
$$sumMax\ t = (sm, mx)$$
$$\quad \textbf{where}$$
$$\qquad (Sum\ sm, Max\ mx) =$$
$$\qquad\quad foldMap\ (\lambda n \rightarrow (Sum\ n, Max\ n))\ t$$

Example: Folding Over a Tree (3)

The latter can be generalized to e.g. computing the sum, product, min, and max in a single traversal:

$$foldMap$$
$$\quad (\lambda n \rightarrow (Sum\ n, Product\ n, Min\ n, Max\ n))$$
$$\quad t$$

Aside: Foldable?

Note that the kind of "folding" captured by the class $Foldable$ in general makes it impossible to recover the structure over which the "folding" takes place.

Such an operation is also known as "reduce" or "crush", and some authors prefer to reserve the term "fold" for ***catamorphisms***, where a separate combining function is given for each constructor, making it possible to recover the structure.

One might thus argue that $Reducible$ or $Crushable$ would have been a more precise name.

MapReduce

Functional mapping and folding (reducing) inspired the MapReduce programming model; e.g.

- Google's original MapReduce framework
- Apache Hadoop

Functional mapping and folding with ***associative*** operator (semigroup) is amenable to parallelization and distribution.

However, achieving scalability in practice required both careful engineering of the frameworks as such, and a good understanding of how to use them on part of the user.