

# Learn Me a Haskell

Jacob Bishop

2023-07-01

## 1 Introduction

WIP

## 2 Starting Out

### 第一个函数

在 `./test/` 文件夹下创建一个 `baby.hs` 的文件，写入：

```
1 doubleMe x = x + x
```

使用 `ghci` 加载该文件（在本项目根目录时使用 `:l tests/baby`）：

```
1 ghci> :l baby
2 [1 of 1] Compiling Main           ( baby.hs, interpreted )
3 Ok, modules loaded: Main.
4 ghci> doubleMe 9
5 18
6 ghci> doubleMe 8.3
7 16.6
```

一个带有 `if` 的函数：

```
1 doubleSmallNumber x =
2   if x > 100
3   then x
4   else x * 2
```

Haskell 中的 `if` 声明是一个表达式，那么 `else` 是强制性的，因为表达式一定要有所返回。因此加上述函数可以改写为：

```
1 doubleSmallNumber' x = (if x > 100 then x else x * 2) + 1
```

这里的 `'` 符号是 Haskell 中的有效字符，且在 Haskell 中并没有特殊的意义，因此可以用于函数名。通常情况下，使用 `'` 代表着一个函数（非懒加载的函数）的严格版本，或是一个有细微变化的函数或者变量。又因为 `'` 是一个有效字符，那么可以创建以下函数：

```
1 conanO'Brien = "It's a-me, Conan O'Brien!"
```

这里又有两点值得注意的地方。首先，函数名不能以大写开头，稍后会进行说明；其次，该函数并没有任何入参。当一个函数没有入参，我们通常称其为一个定义 *definition*，因为一旦定义了它便不能修改其名称，以及其返回。

### list 的介绍

Haskell 中的 `list` 是 **同质的 homogenous** 数据结构。

#### Note

在 `GHCI` 中可以使用 `let` 关键字定义一个名称。换言之，`GHCI` 中的 `let a = 1` 等同于脚本中的 `a = 1`。

通常使用 `++` 操作符将两个数组进行合并：

```
1 ghci> [1,2,3,4] ++ [9,10,11,12]
2 [1,2,3,4,9,10,11,12]
3 ghci> "hello" ++ " " ++ "world"
4 "hello world"
5 ghci> ['w','o'] ++ ['o','t']
6 "woot"
```

可以使用 `:` 操作符将元素直接添加至数组头部：

```
1 ghci> 'A':" SMALL CAT"
2 "A SMALL CAT"
3 ghci> 5:[1,2,3,4,5]
4 [5,1,2,3,4,5]
```

实际上，`[1, 2, 3]` 是 `1:2:3:[]` 的语法糖，其中 `[]` 为一个空数组。如果头部追加 `3`，`[]` 就变成了 `[3]`，再次进行头部追加 `2`，则变为 `[2, 3]`，以此类推。

如果希望通过索引获取数组中的元素，那么可以使用 `!!` 操作符：

```
1 ghci> "Steve Buscemi" !! 6
2 'B'
3 ghci> [9.4,33.2,96.2,11.2,23.25] !! 1
4 33.2
```

超出索引时则会报错。

数组还可以通过操作符 `<`，`<=`，`==`，`>` 以及 `>=` 操作符来进行比较，而比较的方式则是顺序比较。当进行头部比较元素相等时，再进行下一个元素进行比较。

数组的四种基础操作 `head`，`tail`，`last` 以及 `init`：

```
1 ghci> head [5,4,3,2,1]
2 5
3 ghci> tail [5,4,3,2,1]
4 [4,3,2,1]
5 ghci> last [5,4,3,2,1]
6 1
7 ghci> init [5,4,3,2,1]
8 [5,4,3,2]
```

当使用上述四种操作时，需要注意是否应用于空数组，这样的错误在编译期并不能被发现。其它的操作：

1. `length` 获取数组长度；
2. `null` 检查数组是否为空；
3. `reverse` 翻转数组；
4. `take` 获取数组的头几个元素的数组；

5. `drop` 移除数组的头几个元素，并返回剩余元素的数组；
6. `maximum` 获取最大值；
7. `minimum` 获取最小值；
8. `sum` 求和；
9. `product` 求积；
10. `elem` 元素是否存在于数组中。

```
1 ghci> length [5,4,3,2,1]
2 5
3
4 ghci> null [1,2,3]
5 False
6 ghci> null []
7 True
8
9 ghci> reverse [5,4,3,2,1]
10 [1,2,3,4,5]
11
12 ghci> take 3 [5,4,3,2,1]
13 [5,4,3]
14 ghci> take 1 [3,9,3]
15 [3]
16 ghci> take 5 [1,2]
17 [1,2]
18 ghci> take 0 [6,6,6]
19 []
20
21 ghci> drop 3 [8,4,2,1,5,6]
22 [1,5,6]
23 ghci> drop 0 [1,2,3,4]
24 [1,2,3,4]
25 ghci> drop 100 [1,2,3,4]
26 []
27
28 ghci> minimum [8,4,2,1,5,6]
29 1
30 ghci> maximum [1,9,2,3,4]
31 9
32
33 ghci> sum [5,2,1,6,3,2,5,7]
34 31
35 ghci> product [6,2,1,2]
36 24
37 ghci> product [1,2,5,6,7,9,2,0]
```

```

38 0
39
40 ghci> 4 `elem` [3,4,5,6]
41 True
42 ghci> 10 `elem` [3,4,5,6]
43 False

```

## Texas 排列

```

1 ghci> [1..20]
2 [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]
3 ghci> ['a'..'z']
4 "abcdefghijklmnopqrstuvwxyz"
5 ghci> ['K'..'Z']
6 "KLMNOPQRSTUVWXYZ"

```

带有 step 的排列:

```

1 ghci> [2,4..20]
2 [2,4,6,8,10,12,14,16,18,20]
3 ghci> [3,6..20]
4 [3,6,9,12,15,18]

```

而对于浮点数的排列需要注意精度问题:

```

1 ghci> [0.1, 0.3 .. 1]
2 [0.1,0.3,0.5,0.7,0.8999999999999999,1.0999999999999999]

```

以下是若干用于生产无限长度数组的函数:

**cycle** 循环周期:

```

1 ghci> take 10 (cycle [1,2,3])
2 [1,2,3,1,2,3,1,2,3,1]
3 ghci> take 12 (cycle "LOL ")
4 "LOL LOL LOL "

```

**repeat** 重复:

```

1 ghci> take 10 (repeat 5)
2 [5,5,5,5,5,5,5,5,5,5]

```

另外就是 **replicate** 函数可以重复单个元素:

```

1 ghci> replicate 3 10
2 [10,10,10]

```

## 列表表达式

数学里的 集合表达式 *set comprehensions* 例如  $S = 2 \cdot x | x \in \mathbb{N}, x \leq 10$ ; Haskell 中的列表表达式, 例如 1 至 10 数组中每个元素乘以 2:

```
1 ghci> [x*2 | x <- [1..10]]
2 [2,4,6,8,10,12,14,16,18,20]
```

为列表表达式添加条件（或称谓语 predicate）：

```
1 ghci> [x*2 | x <- [1..10], x*2 >= 12]
2 [12,14,16,18,20]
3 ghci> [ x | x <- [50..100], x `mod` 7 == 3]
4 [52,59,66,73,80,87,94]
```

将列表表达式置于一个函数中便于复用：

```
1 ghci> boomBangs xs = [ if x < 10 then "BOOM!" else "BANG!" | x <- xs, odd x]
2 ghci> boomBangs [7..13]
3 ["BOOM!", "BOOM!", "BANG!", "BANG!"]
```

多个谓词也是可以的：

```
1 ghci> [ x | x <- [10..20], x /= 13, x /= 15, x /= 19]
2 [10,11,12,14,16,17,18,20]
```

除此之外，还可以处理若干数组：

```
1 ghci> [ x*y | x <- [2,5,10], y <- [8,10,11]]
2 [16,20,22,40,50,55,80,100,110]
```

当然也可以加上谓词：

```
1 ghci> [ x*y | x <- [2,5,10], y <- [8,10,11], x*y > 50]
2 [55,80,100,110]
```

那么对于字符串也可以使用列表表达式：

```
1 ghci> let nouns = ["hobo", "frog", "pope"]
2 ghci> let adjectives = ["lazy", "grouchy", "scheming"]
3 ghci> [adjective ++ " " ++ noun | adjective <- adjectives, noun <- nouns]
4 ["lazy hobo", "lazy frog", "lazy pope", "grouchy hobo", "grouchy frog",
5 "grouchy pope", "scheming hobo", "scheming frog", "scheming pope"]
```

现在让我们编写一个自己的 `length`，命名 `length'`（这里的 `_` 意为无需使用的变量）：

```
1 length' xs = sum [1 | _ <- xs]
```

由于字符串是数组，因此我们可以使用列表表达式处理并生产字符串。以下是一个移除所有字符但保留大写字符的函数：

```
1 removeNonUppercase st = [ c | c <- st, c `elem` ['A'..'Z']]
2 removeUppercase st = [ c | c <- st, c `notElem` ['A'..'Z']]
```

## 元组

在某种程度上，元组类似于数组 – 存储若干值至单个变量上。然而有一些基础的差异：数组长度可以无限，元组长度固定；数组中元素类型是同质的，而元组则可以是异质的 *heterogenous*。

对于对元组（当且仅当包含两个元素）有以下操作：

**fst** 获取对元组的第一个元素：

```
1 ghci> fst (8,11)
2 8
3 ghci> fst ("Wow", False)
4 "Wow"
```

**snd** 获取对元组的第二个元素：

```
1 ghci> snd (8,11)
2 11
3 ghci> snd ("Wow", False)
4 False
```

另外一个有意思的函数则是 **zip**，它可以将两个数组按对拼接成对元组的数组

```
1 ghci> zip [1,2,3,4,5] [5,5,5,5,5]
2 [(1,5),(2,5),(3,5),(4,5),(5,5)]
3 ghci> zip [1..5] ["one", "two", "three", "four", "five"]
4 [(1,"one"),(2,"two"),(3,"three"),(4,"four"),(5,"five")]
```

当两个数组的长度不一时，**zip** 则按最短的那个进行对齐，长的数组剩余部分则被丢弃，这是因为 Haskell 是懒加载的缘故。

```
1 ghci> zip [5,3,2,6,2,7,2,5,4,6,6] ["im", "a", "turtle"]
2 [(5,"im"),(3,"a"),(2,"turtle")]
3 ghci> zip [1..] ["apple", "orange", "cherry", "mango"]
4 [(1,"apple"),(2,"orange"),(3,"cherry"),(4,"mango")]
```



## 3 Types and Typeclasses

### 相信类型

通过 `:t` 命令可以得知类型：

```
1 ghci> :t 'a'
2 'a' :: Char
3 ghci> :t True
4 True :: Bool
5 ghci> :t "HELLO!"
6 "HELLO!" :: [Char]
7 ghci> :t (True, 'a')
8 (True, 'a') :: (Bool, Char)
9 ghci> :t 4 == 5
10 4 == 5 :: Bool
```

函数同样拥有类型：

```
1 ghci> :t doubleSmallNumber
2 doubleSmallNumber :: (Ord a, Num a) => a -> a
```

而对于有多个入参的函数而言：

```
1 ghci> addThree x y z = x + y + z
2 ghci> :t addThree
3 addThree :: Num a => a -> a -> a -> a
```

参数由 `->` 符分开，并且入参与返回的类型并无差异，之后我们讨论到为什么是由 `->` 分割而不是 `Int, Int, Int -> Int` 或者其他样式的类型。

接下来是一些常规的类型：

**Int**：对于 32 位的机器而言最大值大概是 2147483647 而最小值则是 -2147483647。

**Integer**：同样也是整数，只不过范围会大很多，而 **Int** 则更高效。

**Float**：单精度。

**Double**：双精度。

**Bool**：布尔值。

**Char**：字符。

### 类型变量

那么 `head` 函数的类型是什么呢？

```
1 ghci> :t head
2 head :: [a] -> a
```

这里的 `a` 则是一个 **类型变量 type variable**，这意味着 `a` 可以是任意类型。这非常像其他语言的泛型，但唯有在 Haskell 中它更为强大，因为它允许我们可以轻易的编写通用的函

数，且不使用任何特定的行为的类型。带有类型变量的函数也被称为 **多态函数 polymorphic functions**。

## Typeclasses 101

`typeclass` 类似于一个接口用于定义一些行为。如果一个类型是 `typeclass` 的一部分，这就意味着它支持并且实现了 `typeclass` 中所描述的行为。

那么 `==` 函数的类型签名是什么呢？

```
1 ghci> :t (==)
2 (==) :: Eq a => a -> a -> Bool
```

### Note

`==` 操作符是一个函数，`+`，`*`，`-`，`/` 以及其它的操作符也都是。如果一个函数只包含特殊字符，那么默认情况下它被认做是一个中缀函数。如果想要检查它的类型，将其传递给另一个函数或作为前缀函数调用它，那么则需要用括号将其包围。

这里有趣的是 `=>` 符号，在该符号之前的被称为一个 **类约束 class constraint**。那么上述的类型声明可以被这么理解：等式函数接受任意两个相同类型的值，并返回一个 `Bool`，而这两个值必须是 `Eq` 类的成员（即类约束）。

`Eq` `typeclass` 提供了一个用于测试是否相等的接口。

而 `elem` 函数则拥有 `(Eq a) => a -> [a] -> Bool` 这样的类型，因为其在数组中使用了 `==` 用于检查元素是否为期望的值。

一些基础的 `typeclasses`：

`Eq` 如上所述。

`Ord` 覆盖了所有标准的比较函数例如 `>`，`<`，`>=` 以及 `<=`。`compare` 函数接受两个同类型的 `Ord` 成员，并返回一个 `ordering`。`Ordering` 是一个可作为 `GT`，`LT` 或是 `EQ` 的类型，分别意为 大于，小于以及等于。

`Show` 可以表示为字符串。

`Read` 有点类似于 `Show` 相反的 `typeclass`，`read` 函数接受一个字符串并返回一个 `Read` 的成员。

```
1 ghci> read "True" || False
2 True
3 ghci> read "8.2" + 3.8
4 12.0
5 ghci> read "5" - 2
6 3
7 ghci> read "[1,2,3,4]" ++ [3]
8 [1,2,3,4,3]
```

但是如果尝试一下 `read "4"` 呢?

```
1 ghci> read "4"
2 <interactive>:1:0:
3   Ambiguous type variable `a' in the constraint:
4     `Read a' arising from a use of `read' at <interactive>:1:0-7
5   Probable fix: add a type signature that fixes these type variable(s)
```

这里 GHCi 告知它并不知道想要返回什么, 通过检查 `read` 的类型签名:

```
1 ghci> :t read
2 read :: Read a => String -> a
```

也就是说其返回的是 `Read` 所约束的类型, 那么如果在之后没有使用到它, 则没有办法知晓其类型。我们可以显式的使用类型注解, 即在表达式后面加上 `::` 与指定的一个类型:

```
1 ghci> read "5" :: Int
2 5
3 ghci> read "5" :: Float
4 5.0
5 ghci> (read "5" :: Float) * 4
6 20.0
7 ghci> read "[1,2,3,4]" :: [Int]
8 [1,2,3,4]
9 ghci> read "(3, 'a')" :: (Int, Char)
10 (3, 'a')
```

大多数表达式可以被编译器推导出其类型, 但是有时编译器并不知道返回值的类型, 例如 `read "5"` 时该为 `Int` 还是 `Float`。那么为了知道其类型, Haskell 会解析 `read "5"`。然而 Haskell 是一个静态类型语言, 因此它需要在代码编译前知道所有类型。

`Enum` 成员是序列化的有序类型 – 它们可被枚举。`Enum` typeclass 的主要优势是可以使用在列表区间; 它们也同样定义了 `successors` 与 `predecessors`, 即可使用 `succ` 以及 `pred` 函数。在这个类中的类型有: `()`, `Bool`, `Char`, `Ordering`, `Int`, `Integer`, `Float` 以及 `Double`。

`Bounded` 成员拥有上下界:

```
1 ghci> minBound :: Int
2 -2147483648
3 ghci> maxBound :: Char
4 '\1114111'
5 ghci> maxBound :: Bool
6 True
7 ghci> minBound :: Bool
8 False
```

元组的成员如果为 `Bounded` 的一部分, 那么元组也是:

```
1 ghci> maxBound :: (Bool, Int, Char)
2 (True, 2147483647, '\1114111')
```

`Num` 是一个数值类:

```
1 ghci> :t 20
2 20 :: (Num t) => t
3 ghci> 20 :: Int
4 20
5 ghci> 20 :: Integer
6 20
7 ghci> 20 :: Float
8 20.0
9 ghci> 20 :: Double
10 20.0
```

这些类型都在 `Num` typeclass 内。如果检查 `*` 的类型，将会看到：

```
1 ghci> :t (*)
2 (*) :: (Num a) => a -> a -> a
```

`Integral` 同样也是数值 typeclass。

`Floating` 包含 `Float` 与 `Double`。

`fromIntegral` 是一个处理数值的常用函数，而其类型为 `fromIntegral :: (Num b, Integral a) => a -> b`。

## 4 Syntax in Functions

WIP

## 5 Recursion

WIP

## 6 Higher Order Functions

WIP

## 7 Modules

WIP



## **8 Making Our Own Types and Typeclasses**

WIP

## 9 Input and Output

WIP

## 10 Functionally Solving Problems

WIP

## **11 Functors, Applicative Functors and Monoids**

WIP

## 12 A Fistful of Monads

WIP

## 13 For a Few Monads More

WIP

## 14 Zippers

WIP