CSCE 314
Programming Languages

Haskell 101

Dr. Hyunyoung Lee
Contents

1. Historical Background of Haskell
2. Lazy, Pure, and Functional Language
3. Using ghc and ghci
4. Functions
5. Haskell Scripts
6. Exercises
Historical Background

1930s:

Alonzo Church develops the lambda calculus, a simple but powerful theory of functions.
Historical Background

1950s:

John McCarthy develops **Lisp**, the first functional language, with some influences from the lambda calculus, but retaining variable assignments.
1960s:

Peter Landin develops **ISWIM**, the first *pure* functional language, based strongly on the lambda calculus, with no assignments.
Historical Background

1970s:

John Backus develops FP, a functional language that emphasizes higher-order functions and reasoning about programs.
Historical Background

1970s:

Robin Milner and others develop ML, the first modern functional language, which introduced type inference and polymorphic types.
Historical Background

1970s – 1980s:

David Turner develops a number of lazy functional languages, culminating in the Miranda system.
Historical Background

1987:

An international committee of researchers initiates the development of Haskell, a standard lazy pure functional language.
Historical Background

2003:

The committee publishes the Haskell 98 report, defining a stable version of the language. Since then highly influential in language research and fairly widely used in open-source/commercial software.
Haskell is a Lazy Pure Functional Language
“Haskell is a Lazy Pure Functional Language”

Lazy programming language only evaluates arguments when strictly necessary, thus, (1) avoiding unnecessary computation and (2) ensuring that programs terminate whenever possible. For example, given the definitions

\[
\text{omit } x = 0 \\
\text{keep\_going } x = \text{keep\_going } (x+1)
\]

what is the result of the following expression?

\[
\text{omit } (\text{keep\_going } 1)
\]
“Haskell is a Lazy Pure Functional Language”

Pure functional language, as with mathematical functions, prohibits side effects (or at least they are confined):

- Immutable data: Instead of altering existing values, altered copies are created and the original is preserved, thus, there’s no destructive assignment:
  
  ```
  a = 1; a = 2;   -- illegal
  ```

- Referential transparency: Expressions yield the same value each time they are invoked; helps reasoning. Such expression can be replaced with its value without changing the behavior of a program, for example,
  
  ```
  y = f x  and  g = h y y
  ```

  then, replacing the definition of g with  
  
  ```
  g = h (f x) (f x)
  ```

  will get the same result (value).
“Haskell is a Lazy Pure **Functional** Language”

**Functional** language supports the functional programming style where the basic method of computation is application of functions to arguments. For example, in C,

```c
int s = 0;
for (int i=1; i <= 100; ++i)  s = s + i;
```

the computation method is **variable assignment**. In Haskell,

```haskell
sum [1..100]
```

the computation method is **function application**.
Features of Functional Languages

- Higher-order functions are functions that take other functions as their arguments. E.g.,

  ```
  > map reverse ["abc","def"]
  ["cba","fed"]
  ```

- Purity – prohibits side effects
  (Expressions may result in some actions in addition to return values, such as changing state and I/O; these actions are called side effects.)

- Recursion – the canonical way to iterate in functional languages
A Taste of Haskell

\[ f \; [] \; = \; [] \]
\[ f \; (x:xs) \; = \; f \; ys \; ++ \; [x] \; ++ \; f \; zs \]

where

\[ ys \; = \; [a \mid a \leftarrow xs, \; a \leq x] \]
\[ zs \; = \; [b \mid b \leftarrow xs, \; b > x] \]
void f(vector<int> &xs, int first, int last)
{
    int mid;
    if (first < last)
    {
        mid = partition(xs, first, last);
        f(xs, first, mid);
        f(xs, mid+1, last);
    }
    return;
}

int partition(vector<int> &xs, int first, int last)
{
    int k = array[first]; int i = first-1; int j = last+1; int temp;
    do {
        do { j--; } while (k>array[j]);
        do { i++; } while (k<array[i]);
        if (i<j) {
            temp=array[i]; array[i]=array[j]; array[j]=temp;
        }
    } while (i<j);
    return j;
}
Recursive function execution:

\[ f[3,2,4,1,5] \]

\[ f[2,1] \quad ++ \quad [3] \quad ++ \quad f[4,5] \]

\[ f[1] \quad ++ \quad [2] \quad ++ \quad f[\] \]

\[ f[\] \quad ++ \quad [4] \quad ++ \quad f[5] \]

\[ [1] \]

\[ [\] \]

\[ [\] \]

\[ [5] \]
Other Characteristics of Haskell

- Statically typed
- Type inference
- Rich type system
- Succinct, expressive syntax yields short programs
- Indentation matters
- Capitalization of names matters
Using GHC and GHCi

- From a shell window, the compiler is invoked as
  
  > ghc myfile.hs
  
  > ghci (or as > ghc --interactive)
  
- For multi-file programs, use --make option

- GHCi operates on an eval-print-loop:

  ```
  > sqrt (3^2 + 4^2)
  5.0
  >
  ```

  - User types in a Haskell expression
  - The interpreter evaluates it and prints out the result
  - Waits for the next expression

- Efficient edit-compile-run cycle, e.g., using Emacs with haskell-mode (https://github.com/serras/emacs-haskell-tutorial/blob/master/tutorial.md) helps indenting, debugging, jumping to an error, etc.
Using GHCi

- Useful basic GHCi commands:
  - `:?`: Help! Show all commands
  - `:load test`: Open file test.hs or test.lhs
  - `:reload`: Reload the previously loaded file
  - `:main a1 a2`: Invoke main with command line args a1 a2
  - `:!`: Execute a shell command
  - `:edit name`: Edit script name
  - `:edit`: Edit current script
  - `:type expr`: Show type of expr
  - `:quit`: Quit GHCi

- Commands can be abbreviated. E.g., `:r` is `:reload`

- At startup, the definitions of the “Standard Prelude” are loaded
The Standard Prelude

Haskell comes with a large number of standard library functions. In addition to the familiar numeric functions such as + and *, the library also provides many useful functions on lists.

-- Select the first element of a list:

> head [1,2,3,4,5]
1

-- Remove the first element from a list:

> tail [1,2,3,4,5]
[2,3,4,5]
-- Select the nth element of a list:

> [1,2,3,4,5] !! 2
3

-- Select the first n elements of a list:

> take 3 [1,2,3,4,5]
[1,2,3]

-- Remove the first n elements from a list:

> drop 3 [1,2,3,4,5]
[4,5]

-- Append two lists:

> [1,2,3] ++ [4,5]
[1,2,3,4,5]
-- Reverse a list:
> reverse [1,2,3,4,5]
[5,4,3,2,1]

-- Calculate the length of a list:
> length [1,2,3,4,5]
5

-- Calculate the sum of a list of numbers:
> sum [1,2,3,4,5]
15

-- Calculate the product of a list of numbers:
> product [1,2,3,4,5]
120
Functions (1)

- Function and parameter names must start with a lower case letter, e.g., myFun1, arg_x, personName, etc. (By convention, list arguments usually have an _ suffix on their name, e.g., xs, ns, nss, etc.)

- Functions are defined as equations:
  $$\text{square } x = x \times x \quad \text{add } x \ y = x + y$$

- Once defined, apply the function to arguments:
  > square 7 > add 2 3
  49 5

  In C, these calls would be square(7); and add(2,3);

- Parentheses are often needed in Haskell too
  > add (square 2) (add 2 3)
  9
Functions (2)

- Function application has the highest precedence
  
  \text{square} \ 2 \ + \ 3 \ \text{means} \ (\text{square} \ 2) \ + \ 3 \ \text{not} \ \text{square} \ (2+3)

- Function call associates to the left and is by pattern matching (first one to match is used)

- Function application operator $ has the lowest precedence and is used to rid of parentheses.
  
  \text{sum} \ ([1..5] \ +\ [+ \ [6..10]]) \ \rightarrow \ \text{sum} \ \$ \ [1..5] \ +\ [+ \ [6..10]]

- Combinations of most symbols are allowed as function
  
  \text{x \ @#$%^&*-+@#$% \ y} = "\text{What on earth?}" \smiley

Another (more reasonable) example:

\text{x \ +/- \ y} = (x+y, \ x-y)

> 10 \ +/- \ 1

(11,9)
Function Application

In mathematics, function application is denoted using parentheses, and multiplication is often denoted using juxtaposition or space.

\[ f(a,b) + c \cdot d \]

Apply the function \( f \) to \( a \) and \( b \), and add the result to the product of \( c \) and \( d \)

In Haskell, function application is denoted using space, and multiplication is denoted using \( \ast \).

\[ f \ a \ b + c \ast d \]

As previously, but in Haskell syntax
Examples

Mathematics

\[ f(x) \]
\[ f(x,y) \]
\[ f(g(x)) \]
\[ f(x,g(y)) \]
\[ f(x)g(y) \]

Haskell

\[ f \ x \]
\[ f \ x \ y \]
\[ f \ (g \ x) \]
\[ f \ x \ (g \ y) \]
\[ f \ x \ * \ g \ y \]
Evaluating Functions (1)

Think of evaluating functions as substitution and reduction

\[
\text{add } x \ y = x + y; \quad \text{square } x = x \times x
\]

\[
\text{add } (\text{square } 2) \ (\text{add } 2 \ 3)
\]

— apply \text{square}

\[
\text{add } (2 \times 2) \ (\text{add } 2 \ 3)
\]

— apply \text{\ast}

\[
\text{add } 4 \ (\text{add } 2 \ 3)
\]

— apply inner \text{add}

\[
\text{add } 4 \ (2 + 3)
\]

— apply \text{\textit{+}}

\[
\text{add } 4 \ 5
\]

— apply \text{add}

\[
4 + 5
\]

— apply \text{\textit{+}}

9
Evaluating Functions (2)

- There are many possible orders to evaluate a function
  
  \[
  \text{head} \left(1: (\text{reverse} \ [2,3,4,5])\right) \quad \text{head} \left(1: (\text{reverse} \ [2,3,4,5])\right)
  \]
  
  --- apply reverse
  
  --- apply head
  
  ... many steps omitted here

  \[
  \text{head} \ (1 : [5,4,3,2])
  \]
  
  --- apply head
  
  \[
  1
  \]

- In a **pure** functional language, evaluation order does not affect the *value* of the computation

- It can, however, affect the *amount* of computation and whether the computation *terminates* or not (or fails with a run-time error)

- Haskell evaluates a function’s argument **lazily**

  “Call-by-need” – only apply a function if its value is needed, and “memoize” what’s already been evaluated
Haskell Scripts

A Haskell program consists of one or more scripts. A script is a text file comprising a sequence of definitions, where new functions are defined.

By convention, Haskell scripts usually have a .hs suffix on their filename. This is not mandatory, but is useful for identification purposes.

Loading new script causes new definitions to be in scope:

Prelude> :l test.hs
[1 of 1] Compiling Main     ( test.hs, interpreted )
Ok, modules loaded: Main.
*Main>
My First Script

When developing a Haskell script, it is useful to keep two windows open, one running an editor for the script, and the other running GHCi:

Start an editor, type in the following two function definitions, and save the script as test.hs:

```haskell
double x = x + x
quadruple x = double (double x)
```

In another window start up GHCi with the new script:

```bash
% ghci test.hs
```

Now both the standard library and the file test.hs are loaded, and functions from both can be used:

```haskell
> quadruple 10
40
> take (double 2) [1,2,3,4,5,6]
[1,2,3,4]
```
Leaving GHCi open, return to the editor, add the following definitions, and resave:

```
factorial n = product [1..n]
average ns = sum ns `div` length ns
```

**Note:**
- `div` is enclosed in back quotes, not forward
- `x `f` y` is syntactic sugar for `f x y`
- Any function with two or more arg.s can be used as an infix operator (enclosed in back quotes)
- Any infix operator can be used as a function (enclosed in parentheses), e.g., `(+) 10 20`

GHCi does not automatically detect that the script has been changed, so a `reload` command must be executed before the new definitions can be used:

```
> :r
Reading file "test.hs"
> factorial 10
3628800
> average [1,2,3,4,5]
3
```
The Layout Rule

- Layout of a script determines the structure of definitions
- Commonly use layouts instead of braces and semicolons (which are still allowed and can be mixed with layout)
- Each definition must begin in precisely the same column:

\[
\begin{align*}
\text{a} &= 10 \\
\text{b} &= 20 \\
\text{c} &= 30
\end{align*}
\]

\[
\begin{align*}
\text{a} &= 10 \\
\text{b} &= 20 \\
\text{c} &= 30
\end{align*}
\]

\[
\begin{align*}
\text{a} &= 10 \\
\text{b} &= 20 \\
\text{c} &= 30
\end{align*}
\]

\[
\begin{align*}
\text{a} &= \text{b} + \text{c} \\
\text{where} \\
\text{b} &= 1 \\
\text{c} &= 2 \\
\text{d} &= \text{a} \times 2
\end{align*}
\]

\[
\begin{align*}
\text{a} &= \text{b} + \text{c} \\
\text{where} \\
\{ \text{b} = 1; \\
\text{c} = 2 \} \\
\text{d} &= \text{a} \times 2
\end{align*}
\]

- implicit grouping by layout
- explicit grouping by braces and semicolons
Exercises

(1) Try out the codes in slides 15–24 and 30–32 using GHCi.

(2) Fix the syntax errors in the program below, and test your solution using GHCi.

\[ N = a \text{ `div` length xs} \]
\[ \text{where} \]
\[ a = 10 \]
\[ xs = [1,2,3,4,5] \]

\[ n = a \text{ `div` length xs} \]
\[ \text{where} \]
\[ a = 10 \]
\[ xs = [1,2,3,4,5] \]
(3) Show how the library function `last` that selects the last element of a list can be defined using the functions introduced in this lecture.

\[
\text{last } xs = \text{head } ( \text{reverse } xs )
\]

(4) Can you think of another possible definition?

\[
\text{last } xs = xs !! (\text{length } xs - 1)
\]

(5) Similarly, show how the library function `init` that removes the last element from a list can be defined in two different ways.

\[
\text{init } xs = \text{take } (\text{length } xs - 1) \ xs
\]

\[
\text{init } xs = \text{reverse } (\text{tail } (\text{reverse } xs))
\]