

Haskell: Introducere

- Responsabil: Vlad Nedelcu [mailto:nedelcu_vlad@yahoo.com]
- Data publicării: 23.03.2019
- Data ultimei modificări: 14.04.2019 (adăugare soluții)

Obiective

Scopul acestui laborator este acomodarea cu primele noțiuni din Haskell și observarea incipientă a diferențelor dintre acesta și Racket.

Aspectele urmărite sunt:

- diferențele principale dintre cele 2 limbaje
- definirea funcțiilor în Haskell
- pattern matching
- list comprehensions
- tipuri
- gărzi

Introducere

În jurul anilor 1990 un comitet de cercetători în limbaje de programare (Simon Marlow, Simon Peyton Jones, Philip Wadler etc) au creat un limbaj nou care a ajuns să fie standardul de-facto în cercetarea din domeniul programării funcționale. Inspirat dintr-o varietate de limbaje – Miranda, ML, Scheme, APL, FP – limbajul a influențat la rândul lui majoritatea limbajelor de programare cunoscute.

Haskell este un limbaj funcțional **pur**. Spre deosebire de limbajele imperative (sau Racket unde există `set!`), în Haskell aproape toate funcțiile sunt pure. Funcțiile impure sunt marcate diferit prin intermediul sistemului de tipuri.

În plus față de Racket, Haskell are **tipare statica** (și tipuri **polimorfice**). Fiecare expresie are un tip și este sarcina programatorului să efectueze conversiile necesare între tipuri dacă este necesar. De cele mai multe ori, informația despre ce face o funcție se găsește integral în tipul acesteia și numele ei. Astfel, se pune accentul pe *ce* face funcția, nu *cum* efectuează ea operațiile cerute.

Deși tipurile există, programatorul nu este obligat să depună efort în a le scrie în program. Haskell deține inferență de tipuri puternică. Exceptând cazurile în care se folosesc concepte avansate, programatorul poate lucra fără a scrie o singură semnătură de funcție (deși nu e recomandat pentru că se pierde o parte din documentația funcției).

În plus, tipurile ajută programatorul în procesul de scriere a codului transformându-l într-un exercițiu de rezolvare a unui puzzle: pur și simplu trebuie să găsești tipurile folosind funcții existente sau scriind alte funcții. Pentru a căuta funcțiile ce respectă o semnătură se poate folosi Hoogle [http://www.haskell.org/hoogle/], un motor de căutare similar Google dar doar pentru funcții Haskell.

Deosebirea fundamentală față de alte limbaje este **evaluarea leneșă**. Funcțiile nu vor fi evaluate și expresiile nu vor fi reduse până în momentul în care valoarea lor este necesară. Programatorul poate astfel lucra cu date infinite, extrăgând din ele strictul necesar pentru a obține soluția. Ca dezavantaj, analiza performanței unui cod Haskell este puțin mai dificilă, dar există instrumente auxiliare (dezvoltate în Haskell).

În cele ce urmează vom parcurge fiecare dintre aceste particularități, accentuând diferențele față de Racket. Pe scurt, o **paralela** între cele două limbaje arată în felul următor:

Limbaaj	Evaluare	Tipare	Efecte laterale
---------	----------	--------	-----------------

Limbaaj	Evaluare	Tipare	Efecte laterale
Racket	Aplicativă	Dinamică	Da
Haskell	Leneșă	Statică	Nu

Laboratoarele de Haskell din cadrul cursului se vor axa pe evidențierea acestor diferențe și a beneficiilor care rezultă din ele. Pentru noțiuni avansate despre limbaj, consultați primele 2 resurse din [bibliografie](#) (sau Reddit, Stack Overflow, Planet Haskell, etc.)

Un exemplu pentru diferența de evaluare dintre Racket și Haskell:

```
(define (square x) (* x x))

square x = x * x
```

Evaluarea aplicativă din Racket va forța evaluarea parametrului x și apoi va apela funcția square:

```
(square (+ 1 2))
(square 3)
(* 3 3)
9
```

Evaluarea leneșă din Haskell va evalua parametrul x la cerere în cadrul apelului funcției square:

```
square (1 + 2)
(1 + 2) * (1 + 2)
3 * (1 + 2)
3 * 3
9
```

GHC. GHCi

Există o varietate de compilatoare și interpretoare pentru Haskell. În momentul de față, limbajul și evoluția lui sunt strâns legate de eforturile dezvoltatorilor de la Glasgow. GHC, *The Glorious Glasgow Haskell Compilation System*, este și compilatorul pe care-l vom folosi pentru cursul de PP.

Pentru a avea o experiență bună cu acest limbaj, recomandarea este să vă instalați [Haskell Stack](https://docs.haskellstack.org/en/stable/install_and_upgrade/) [https://docs.haskellstack.org/en/stable/install_and_upgrade/] (cititi instrucțiunile de aici: [Limbaje](#)). Față de compilator și suita minimală de pachete, Haskell Stack aduce în plus o suită de biblioteci utile pentru dezvoltarea unor aplicații reale.

Codul Haskell poate fi atât compilat cât și interpretat. Pentru interpretare vom folosi **ghci**, iar pentru compilare vom folosi **ghc**. Fișierele de cod Haskell au în mod normal extensia `.hs`, dar se poate folosi și `.lhs` pentru variantele de **Literate Haskell** (programare ca o poveste - comentariile ocupă majoritatea textului în timp ce secvențele de cod sunt puține - este formatul preferat pentru publicarea de articole despre Haskell pe bloguri, oricine poate copia textul articolului într-un fișier și îl poate compila și rula apoi).

În cadrul laboratorului, recomandăm folosirea **ghci** în modul următor:

- se salvează definițiile de funcții într-un fișier cu extensia `.hs` (ex.: `lab6-ex.hs`)
- se deschide consola sistemului (ex.: `cmd/powershell` pe Win) în directorul cu fișierul `.hs`
- se rulează `stack exec ghci nume.hs` pentru a-l încărca în interpretor
- se rulează în interpretor apelurile de funcții de test necesare, verificarea de tipuri, etc
- dacă se dorește **editarea fișierului** se poate face într-un terminal separat (recomandat) sau folosind `:e (:edit)` din `ghci`
- ATENȚIE! după editarea fișierului, pentru a reîncărca definițiile se folosește `:re (:reload)`
- dacă se dorește încărcarea altor module în interpretor (pentru testare), se poate folosi `:m +Nume.Modul`
- pentru a verifica tipul unor expresii se folosește `:t expresie`
- pentru a ieși din interpretor se folosește `:q` (sau EOF - `^Z` pe Linux/OSX, `^D` pe Windows)

Modulul `Prelude` conține funcții des folosite și este inclus implicit în orice fișier (deși poate fi exclus la nevoie, consultați [bibliografia](#)).

În cadrul unor programe reale va trebui să compilați sursele. Pentru aceasta va trebui să aveți un modul `Main.hs` care să conțină o funcție `main`. Modulul poate importa alte module prin `import Nume.Modul`. Pentru temă, nu va trebui să efectuați acești pași, scheletul de cod ce va fi oferit va conține partea de interacțiune cu mediul exterior. 😊

Comentarii

În Haskell, avem comentarii pe un singur rând, folosind `--` (2 de - legați) sau comentarii pe mai multe rânduri, încadrate de `{ - și - }`. În plus, există comentarii pentru realizarea de documentație dar nu vom insista pe ele aici.

Funcții

O funcție anonimă în **Racket**

```
(lambda (x y) (+ x y))
```

poate fi tradusă imediat în Haskell utilizând tot o *abstracție lambda*

```
\x y -> x + y
```

Sintaxa Haskell este ceva mai apropiată de cea a **calculului Lambda**: \ anunță parametrii (separați prin spații), iar `→` precizează corpul funcției.

În Racket, puteam da nume funcției folosind legarea dinamică prin intermediul lui `define` sau legarea statică prin intermediul lui `let`. Exemplele următoare ilustrează cele 2 cazuri, împreună cu apelul funcției.

```
(define f (lambda (x y) (+ x y)))  
(f 2 3)  
  
(let ((f (lambda (x y) (+ x y)))) (f 2 3))
```

În Haskell, funcția echivalentă ar fi

```
f = \x y -> x + y
```

sau

```
f x y = x + y
```

Și apelul

```
f 2 3
```

Observați că se renunță la paranteze. Unul dintre principiile dezvoltării limbajului a fost oferirea de construcții cât mai simple pentru lucrurile folosite cel mai des. Fiind vorba de programarea funcțională, aplicarea de funcții trebuia făcută cât mai simplu posibil.

Tot din același considerent, operatorii în Haskell sunt scriși în forma **infixată**. Restul apelurilor de funcții sunt în forma prefixată, exact ca în Racket. Totuși, se pot folosi ambele forme: pentru a folosi un operator prefixat se folosesc paranteze, în timp ce pentru a infixa o funcție se folosesc *back-quotes* (```).

```
2 + 3 == (+) 2 3  
elem 2 [1,2,3] == 2 `elem` [1, 2, 3]
```

Tot pentru simplitate, Haskell permite folosirea unor secțiuni – elemente de zăhărel sintactic care se vor traduce în funcții anonime:

```
(2 +) == \x -> 2 + x  
(+ 2) == \x -> x + 2  
(- 2) == -2  
(2 -) == \x -> 2 - x
```

Atenție: În construcția `(- x)` operatorul `-` este unar, nu binar (este echivalentul funcției negate).

Tipuri de bază

În această secțiune vom prezenta tipurile existente în limbajul Haskell. Veți observa că limbajul este mult mai bogat în tipuri decât Racket. Programatorul își va putea defini alte tipuri proprii dacă dorește.

Pentru a putea vedea tipul unei expresii în ghci folosiți `:t expresie`.

```
> :t 'a'
'a' :: Char
```

Operatorul `::` separă o expresie de tipul acesteia.

```
> :t (42::Int)
(42::Int) :: Int
```

Numere, caractere, siruri, booleani

Urmatorii literali sunt valizi in **Haskell** (dupa operatorul `::` sunt precizate tipurile acestora):

```
5      :: Int
'H'    :: Char
"Hello" :: String -- sau [Char] -- lista de Char
True   :: Bool
False  :: Bool
```

Observați că există tipul caracter și tipul șir de caractere. Tipul `String` este de fapt un sinonim pentru tipul `[Char]` - tipul listă de caractere. Astfel, operațiile pe liste vor funcționa și pe șiruri.

Tipurile numerice sunt puțin diferite față de alte limbaje de programare cunoscute:

```
> :t 42
42 :: Num a => a
```

În exemplul de mai sus, `a` este o variabilă de tip (stă pentru orice fel de tip) restricționată (prin folosirea `=>`) la toate tipurile numerice (`Num a`).

Important de reținut este faptul că există 2 tipuri întregi: `Int` și `Integer`. Primul este finit, determinat de arhitectură, în timp ce al doilea este infinit, putând ajunge oricât de mare.

Liste

O listă în Haskell se construiește similar ca în Racket, prin intermediul celor doi constructori:

```
[] -- lista vidă
(:) -- operatorul de adăugare la începutul listei - cons
```

Pentru simplitate, o construcție de forma

```
1:3:5:[]
```

se poate scrie și

```
[1, 3, 5]
```

sau

```
[1, 3 .. 6]
```

sau

```
[1, 3 .. 5]
```

Echivalente pentru `car` și `cdr` din Racket sunt funcțiile `head` și `tail`:

```
> head [1, 2, 3]
1
> tail [1, 2, 3]
[2, 3]
```

Operatorul de **concatenare** este ++:

```
> [1, 2, 3] ++ [4, 5]
[1, 2, 3, 4, 5]
```

Haskell oferă un mod suplimentar de a genera liste: scriem proprietățile pe care ar trebui să le respecte elementele listei într-o sintaxă numită **list comprehension**. Este o sintaxă similară celei din matematică. De exemplu, vrem lista numerelor pare, divizibile cu 3. În matematică, am fi avut ceva de tipul $\{x \mid x \in \mathbb{N}_2, x \equiv 0 \pmod{3}\}$ (pentru \mathbb{N}_2 mulțimea numerelor pare). În Haskell, avem

```
> [x | x <- [0, 2 ..], x `mod` 3 == 0] -- lista numerelor naturale pare, divizibile cu 3
[0,6,12,18,24,30,36,42,48,54,60,66,72,78,84,90,96,102,108, Interrupted.]
```

Observați că se generează elemente la infinit. Pentru fiecare element din lista `[0, 2 ..]` (din expresia `x <- [0, 2 ..]`) se testează condițiile următoare. Dacă toate sunt îndeplinite, se generează elementul din fața `|`.

Pentru a putea vedea o porțiune a fluxului folosim funcțiile `take` și `drop`.

Perechi

În Haskell, elementele unei liste **sunt de același tip**. Dacă dorim construcții cu elemente de tipuri diferite vom folosi tuplurile

```
(3, "Ana") :: (Int, String)
(3, True, "Ana") :: (Int, Bool, String)
```

Haskell oferă funcții pentru extragerea componentelor din perechi

```
> fst (3, "Ana")
3
> snd (3, "Ana")
"Ana"
```

Se ofera de asemenea și funcțiile `zip` și `unzip`, **doar** pentru perechi, dar există și `zip3` și `unzip3` pentru triplete.

Pentru tuplurile cu mai mult de 2 elemente este sarcina programatorului să definească funcțiile folosite.

Definirea funcțiilor

Să considerăm că vrem să scriem o funcție pentru factorialul unui număr. Prima implementare ar folosi sintaxa `if`:

```
factorial_if x = if x < 1 then 1 else x * factorial_if (x - 1)
```

Este obligatoriu ca `if` să conțină **ambele** ramuri și ca acestea să fie **de același tip**.

Mai frumos, putem scrie funcția de mai sus folosind gărzi:

```
factorial_guards x
| x < 1 = 1
| otherwise = x * factorial_guards (x - 1)
```

Observați indentarea: orice linie care face parte din aceeași expresie ca cea de deasupra trebuie să înceapă la exact aceeași indentare. Orice linie care este o subexpresie a expresiei de mai sus (`then` sau `else` în cazul `if`, fiecare gardă în parte, etc.) trebuie să fie indentată mai spre dreapta.

Construcția `otherwise` este echivalentă expresiei `True`. O gardă poate conține orice fel de expresie care se va evalua la o valoare de tip `Bool`.

Fiecare gardă este testată în ordine, prima adevărată este executată.

Pentru a ne apropia de definiția matematică, putem scrie aceeași funcție folosind `case` (echivalentul `switch`-ului din C dar mult mai avansat):

```
factorial_case x = case x < 1 of
  True -> 1
```

```
_ -> x * factorial_case (x - 1)
```

Observați regula indentării aplicată și aici. Expresia `_` semnifică orice valoare, indiferent de valoarea ei. Expresia din `case` poate fi oricare.

Fiecare caz este tratat în ordine, prima potrivire este executată.

În final, putem scrie aceeași funcție folosind **pattern matching**:

```
factorial_pm 0 = 1
factorial_pm x = x * factorial_pm (x - 1)
```

Expresiile din interiorul fiecărei ramuri din `case` sau din interiorul `pattern-match` nu pot fi decât constructorii unui tip (similar cu ce ați făcut la AA). În continuare vom folosi cele 4 stiluri pentru a ilustra funcția care calculează lungimea unei liste

```
length_if l = if l /= [] then 1 + length_if (tail l) else 0
```

```
length_guard l
| l /= [] = 1 + length_guard (tail l)
| otherwise = 0
```

```
length_case l = case l of
  (_ : xs) -> 1 + length_case xs
  _ -> 0
```

```
length_pm [] = 0
length_pm (_:xs) = 1 + length_pm xs
```

Observați în exemplele de mai sus expresivitatea limbajului. Folosirea construcției potrivite duce la definiții scurte și ușor de înțeles.

Curry vs Uncurry

Fiind un limbaj funcțional, în Haskell funcțiile sunt valori de prim rang. Astfel, putem afla tipul unei funcții:

```
f x y = x + y

> :t f
f :: Num a => a -> a -> a
```

Fiecare argument este separat prin `→` de următorul sau de rezultat.

Amintindu-ne de discuția despre funcții curry și uncurry, rezultatul următor nu trebuie să ne surprindă

```
> :t f 3
f 3 :: Num a => a -> a
```

Toate funcțiile din Haskell sunt în formă **curry**.

Pentru a face o funcție uncurry putem folosi funcția `uncurry :: (a → b → c) → (a, b) → c` dacă vrem să transformăm o funcție echivalentă sau putem s-o definim în mod direct folosind perechi.

Funcționale uzuale

Haskell oferă un set de funcționale (similar celui din Racket) pentru cazuri de folosire des întâlnite (dacă nu mai știți ce fac, încercați să ghiciți citind semnătura și numele ca o documentație):

```
> :t map
map :: (a -> b) -> [a] -> [b]
> :t filter
filter :: (a -> Bool) -> [a] -> [a]
> :t foldl
foldl :: (a -> b -> a) -> a -> [b] -> a
> :t foldr
foldr :: (a -> b -> b) -> b -> [a] -> b
```

```
> :t zip
zip :: [a] -> [b] -> [(a, b)]
```

Folosirea lor duce la un cod mai ușor de citit și de întreținut.

Point-free programming

În Haskell, compunerea funcțiilor se realizează cu ajutorul operatorului `.`. De interes este și operatorul `$` definit ca:

```
f $ x = f x
```

El are avantajul de a grupa expresiile din dreapta și stânga lui înainte de aplicarea funcției, scăpând astfel de paranteze:

```
length $ 3 : [1, 2] == length (3 : [1, 2])
```

Împreună cu `curry`, `uncurry` și `flip` aceste funcții duc la un stil de programare în care valoarea în care se evaluează funcția nu este prezentă. Urmăriți exemplul următor de transformare

```
square x = x*x
inc x = x+1
f1 x = inc (square x)
f2 x = inc $ square x
f3 x = inc . square $ x
f4 = inc . square
```

Stilul are câteva avantaje în domeniul expresivității și al verificării programului dar poate duce ușor la cod obfuscant.

Resurse

- [Exerciții](#)
- [Cheatsheet](#)
- [Soluții](#)

Referințe

- [Learn You a Haskell \[http://learnyouahaskell.com/chapters\]](http://learnyouahaskell.com/chapters), este utilă pentru toate laboratoarele de Haskell plus ceva extra
- [Hoogle \[http://www.haskell.org/hoogle/\]](http://www.haskell.org/hoogle/), search for Haskell functions starting from types
- [Hackage \[http://hackage.haskell.org/packages/hackage.html\]](http://hackage.haskell.org/packages/hackage.html), repository pentru pachete Haskell
- [Pointfree on Haskell Wiki \[http://www.haskell.org/haskellwiki/Pointfree\]](http://www.haskell.org/haskellwiki/Pointfree)
- [Pointfree Style - Why is it good for \[http://buffered.io/posts/point-free-style-what-is-it-good-for\]](http://buffered.io/posts/point-free-style-what-is-it-good-for)
- [Advantages of point-free on Lambda the Ultimate \[http://lambda-the-ultimate.org/node/3233\]](http://lambda-the-ultimate.org/node/3233)