#### Monadické parsery

Graham Hutton, Erik Mejer: Functional Pearls: Monadic Parsing in Haskell <a href="http://www.cs.nott.ac.uk/~gmh/bib.html#pearl">http://www.cs.nott.ac.uk/~gmh/bib.html#pearl</a>

Graham Hutton, Erik Mejer: Monadic Parser Combinators <a href="http://www.cs.nott.ac.uk/~pszgmh/monparsing.pdf">http://www.cs.nott.ac.uk/~pszgmh/monparsing.pdf</a>

Daan Leijen: *Parsec, a fast combinator parser* <a href="http://research.microsoft.com/en-us/um/people/daan/download/parsec/parsec-letter.pdf">http://research.microsoft.com/en-us/um/people/daan/download/parsec/parsec-letter.pdf</a>

Cieľom tejto prednášky je ukázať na probléme syntaktickej analýzy, ktorý sme začali minule, ideu fungovania monád a monadického štýlu programovania.

Mnohé analyzátory, ktoré skonštruujeme sa historicky nazývajú inými menami, ale ich analógie nájdete aj v predošlej prednáške.

Rôzne typy monád budeme študovať na budúcej prednáške.

## Varovanie

 Namiesto parser.hs z minulej prednášky používame mparser.hs

#### Syntaktický analyzátor

```
Minule sme definovali analyzáror ako nasledujúci typ:
        type Parser symbol result = [symbol] -> [([symbol],result)]
     dnes to bude (zámena poradia – rešpektujúc použité zdroje):
        type Parser result = String -> [(result, String)]
     resp.:
        data Parser result = Parser(String -> [(result, String)])
     Primitívne parsery:
                                                 -- tento sa volal succeed
     return :: a->Parser a
                                                 -- nečíta zo vstupu, dá výsledok v
     return v = \langle xs - \rangle [(v,xs)]
     return v xs = [(v,xs)]
     zero :: Parser a
                                                 -- tento sa volal fail
     -- neakceptuje nič
     zero xs = | |
                                                 -- akceptuje ľubovoľný znak
     item :: Parser Char
                                                 -- tento znak dá do výsledku
              = \xs -> \xs of
     item
                       [] -> []
                                                 -- prázdny vstup, neakceptuje
                      (v:vs) \rightarrow [(v,vs)]
                                                 -- akceptuje znak v
"?: " parse (item) "abc"
[('a',"bc")]
```

#### Kombinátory parserov

- tento kombinátor zbytočne vyrába vnorené výrazy typu \(a,(b,(c,d)))->...
- v tejto časti seq (<\*>) upadne do zabudnutia, nahradí ho bind (>>=),
- nový kombinátor bind (>>=) kombinuje analyzátor p::Parser a s funkciou qf:: a -> Parser b, ktorá ako argument dostane výsledok analýzy analyzátora p a vráti analyzátor q:

```
bind :: Parser a -> (a -> Parser b) -> Parser b 
p `bind` qf = \xs -> concat [ (qf v) xs' | (v,xs')<-p xs])
```

- v je výsledok analýzy p,
- (qf v) je nový analyzátor parametrizovaný výsledkom v.
- concat :: [[a]] -> [a]

#### Ako sa bind používa

```
bind :: Parser a -> (a -> Parser b) -> Parser b p`bind` qf = \xs -> concat [ (qf v) xs' | (v,xs')<-p xs])

spôsob použitia: preprogramujeme seq: p`seq` q = p`bind` (\x_1 -> p`bind` (\x_2 -> p`bind` (\x_2 -> q`bind` (\y -> return (f x_1 x_2 .....x_n)))...)
```

#### spôsob čítania:

najprv pustíme analyzátor  $p_1$ , ktorého výsledok je v premennej  $x_1$ , potom pustíme analyzátor  $p_2$ , ktorého výsledok je v premennej  $x_2$ ,

nakoniec pustíme analyzátor  $p_n$ , ktorého výsledok je v premennej  $x_n$ , výsledok celej analýzy dostaneme ako funkciu f  $x_1 x_2 \dots x_n$ .

# 4

#### Príklady pre bind operátor

```
"?: " parse (item `bind` \x-> return x) "abc"
[('a',"bc")]
"?: " parse (item `bind` \x-> item) "abc"
[('b',"c")]
"?: " parse (item `bind` x-> item `bind` y-> return ((x:[])++(y:[]))) "abc"
[("ab","c")]
"?: " parse (item `bind` \x-> item `bind` \y -> return ([x,y])) "abc"
[("ab","c")]
"?: " parse ( return 4 `bind` \x->return 9 `bind` \y->return (x+y)) "abc"
[(13,"abc")]
```

#### Príklady jednoduchých parserov

Main> parse (item) "abcd" [('a',"bcd")]

ilustrujme na niekoľkých príkladoch použitie bind:

```
sat :: (Char->Bool) -> Parser Char -- ten sa volal satisfy
```

if pred x then return x else zero

-- tell sa voidi satisfy

-- akceptuje symbol, pre ktorý

-- platí predikát pred

Main> parse (sat isLower) "a123ad" [('a',"123ad")]

char :: Char -> Parser Char -- ten sa

char x = sat (y -> x == y)

-- ten sa volal symbol

-- akceptuje znak x

Main> parse (char 'a') "a123ad"

[('a',"123ad")]

digit :: Parser Char -- akceptuje cifru

digit = sat isDigit

Main> parse (digit) "123ad"

[('1',"23ad")]

#### Zjednotenie

zjednotenie/disjunkciu analyzátorov poznáme ako <|>:

```
p 'plus' q = \xs -> (p xs ++ q xs)
p 'plus' q xs = (p xs ++ q xs)

Main> parse (letter 'plus' digit) "123abc"
[('1',"23abc")]
```

definujme analyzátor word, ktorý akceptuje postupnosť písmen, {letter}\*:
 gramatická idea: word -> letter word | ε

word :: Parser String

word = nonEmptyWord `plus` return "" where

nonEmptyWord = letter `bind`  $\x ->$ 

word `bind` \xs ->

return (x:xs)

Main> parse word "abcd"

[("abcd",""),("abc","d"),("ab","cd"),("a","bcd"),("","abcd")]

Problém je, že nechceme dostať toľko riešení, chceme *greedy* verziu...

#### Deterministické plus

predefinujeme zjednotenie analyzátorov:

#### Monády

Monáda je analógia algebraickej štruktúry nazývanej monoid s doménou M a binárnou asociatívnou operáciou MxM->M s (ľavo a pravo)-neutrálnym prvkom.

definícia v Haskelli:

```
class Monad m where
```

```
return :: a -> m a -- neutrálny prvok vzhľadom na >>= >> :: m a -> (a -> m b) -> m b -- asociatívna operácia, nič iné ako bind
```

- chceli by sme vytvorit' Monad Parser t, ale musime predefinovat' typ Parser:
   data Parser result = Parser(String -> [(result, String)])
- to nás stojí trochu syntaktických zmien:

```
parse :: Parser a -> String -> [(a,String)]
parse (Parser p) = p -- inak: parse (Parser p) xs = p xs
```

instance Monad Parser where

```
return v = Parser(\xs -> [(v,xs)])
p >>= f = Parser( -- bind
\xs -> concat [ parse (f v) xs' | (v,xs')<-parse p xs])
```

#### Monad comprehension

```
p_1 >>= \langle x_1 ->
                                [ f x_1 x_2 ... x_n |
                                                    do \{x_1 < -p_1;
p_2 >> = \x_2 ->
                                  x_1 < -p_1
                                                             x_2 < -p_2;
                                   x_2 < -p_2
p_n >> = \langle x_n -> \rangle
                                                             x_n < -p_n;
return (f x_1 x_2 \dots x_n)
                                   x_n < -p_n
                                                             return (f x_1 x_2 \dots x_n)
           :: String -> Parser String
                                                                  -- volal sa token
string
                      = return ""
string
                      = char x >>= \ ->
                                                                  -- výsledok char x zahoď
string (x:xs)
                        string xs >>= \setminus ->
                                                                  -- výsledok string xs zahoď
                                                                  -- výsledok je celé slovo x:xs
                        return (x:xs)
string
               :: String -> Parser String
                                                                  -- prepíšeme do novej
string "" = return ""
                                                                   -- syntaxi
string (c:cs) = do { _ <- char c; _ <- string cs; return (c:cs)} -- explicitné...
string (c:cs) = do {char c; string cs; return (c:cs)}
                                                                  -- čitateľnejšie...
                                                      Main> parse (string "begin") "beginend"
                                                      [("begin","end")]
```

#### Iterátory {p}\*, {p}+

```
many_p \rightarrow many1_p \mid \epsilon
many0
           :: Parser a -> Parser [a]
many0 p
          = many1 p +++ return []

 many1<sub>p</sub> -> p many<sub>p</sub>

                                                Main> parse (many0 digit) "123abc"
        :: Parser a -> Parser [a] [("123","abc")]
many1
                  = do {a <- p; as <- many p; return (a:as)}
many1 p
  opakovanie p s oddeľovačom sep: {p sep}*p | ε
         :: Parser a -> Parser b -> Parser [a]
sepby
p `sepby` sep = (p `sepby1` sep) +++ return []
                                           "?: " parse (digit `sepby` char ',') "1,2,3abc"
p {p sep}*
                                           [("123","abc")]
sepby1
                 :: Parser a -> Parser b -> Parser [a]
p \cdot sepby1 \cdot sep = do \{ a < -p; as < -many(do {_ < -sep; v < -p; return v});
                          return (a:as) }
                  Main> parse ((many0 digit) `sepby` (char '+' `plus` char '*')) "1+2*3abc"
                  [(["1","2","3"],"abc")]
                                                                           mparser.hs
```

### Zátvorky

Analyzátor pre dobre uzátvorkované výrazy podľa gramatiky: P -> ( P ) P | ε

```
:: Parser Char
open
                    = char '('
open
                                                    " Main> parse (paren) "()(())"
                    = char ')'
close
                                                    [((),"")]
verzia 1
                    :: Parser () -- nezaujíma nás výstupná hodnota
paren
                    = do { open; paren; close; paren; return () }
paren
                                                   Main> parse parenBin "()"
                       +++
                                                   [(Node Nil Nil,"")]
                       return ()
                                                   Main> parse parenBin "()()"
                                                   [(Node Nil (Node Nil Nil),"")]
   verzia 2
                                       -- vnútorná reprezentácia
data Bin = Nil | Node Bin Bin
          deriving(Show, Read, Eq)
                                         -- analyzátor zo String do Bin
parenBin :: Parser Bin
parenBin
            = do {
                    open; x<-parenBin; close; y<-parenBin; return (Node x y) }
                +++
                                                   Main> parse parenBin "(())()"
                                                   [(Node (Node Nil Nil) (Node Nil Nil),"")]
               return Nil
                                                   Main> parse parenBin "())()"
                                                   [(Node Nil Nil,")()")]
                                                                                   mparser.hs
```

#### $P -> (P) P | [P] P | \epsilon$

### Zátvorky 2

```
naivne
parenBr
                   :: Parser ()
parenBr
                  do {
                            (open `plus` openBr); parenBr;
                            (close `plus` closeBr); parenBr; return () }
                   +++ return ()
   reálne
parenBr
                   :: Parser ()
parenBr
                            do { open; parenBr; close; parenBr; return () }
                            +++
                            do { openBr ; parenBr; closeBr ; parenBr; return () }
                            +++
                                                    Main> parse parenBr "([[]])([])"
                            return ()
                                                    [((),"")]
                                                    Main> parse parenBr "([[(])])([])"
                                                    [((),"([[(])])([])")]
                                                                                 mparser.hs
```

#### Aritmetické výrazy

parser pre aritmetické výrazy

l'avo-rekurzívna gramatika:

```
<expression> ::=
    <expression> + <expression> |
    <expression> * <expression> |
    <expression> - <expression> |
    <expression> / <expression> |
    identifier |
    number |
    ( <expression> )
```

```
gramatika LL(1):
```

#### expr naivne

Gramatika je ľavo-rekurzívna, preto sa to zacyklí:

- riešenie je zlé, čo ale stojí za zmienku je typ addop :: Parser (Int -> Int -> Int)
- parser, ktorého vnútorná reprezentácia toho, čo zanalyzuje je funkcia,
- táto funkcia sa potom použije na kombináciu dvoch posebe-idúcich výsledkov iného parsera (f x z)

```
chainl
```

```
<expr> ::= <term> | <term> + <expr> | <term> - <expr>
<term> ::= <factor> | <factor> * <term> | <factor> / <term>
<factor> ::= number | ( <expr> )
```

Ak tú myšlienku zovšeobecníme dostaneme nasledujúci kód:

```
chainl1 :: Parser a -> Parser (a -> a -> a) -> Parser a

p `chainl1` op = do {a <- p; rest a}

where

rest a = do {f <- op; b <- p; rest (f a b)}

`plus`

return a
```

Aritmetické výrazy:

```
expr = term `chainl1` addop
term = factor `chainl1` mulop
factor = nat `plus` (do {open; n <- expr; close; return n})
mulop = do { char '*'; return (*) } `plus` do { char '/'; return (div) }

Main> parse expr "1+2*3+10"
[(17,""),(8,"0"),(7,"+10"),(3,"*3+10"),(1,"+2*3+10")]
```

#### Cvičenia

parser pre λ-calcul, t.j. Parser LExp:

```
Main> parse lambda "(\\x.(x x) \\x.(x x))"
[(APL (LAMBDA "x" (APL (ID "x") (ID "x"))) (LAMBDA "x" (APL (ID "x") (ID "x"))),"")]
```

parser binárnej/hexa konštanty

```
Main> parse binConst "1101"
[13,"")]
Main> parse hexaConst "FF"
[255,"")]
    parser palindromov,
```

- parser palindromov,t.j. Parser ()
- parser morseovej abecedy,t.j. Parser String

#### Ako na Parsec

#### Daan Leijen autor knižnice

https://hackage.haskell.org/package/parsec

wiki:

http://www.haskell.org/haskellwiki/Parsec

nie na 1.čítanie

http://book.realworldhaskell.org/read/using-parsec.html

module Main where

import Text.ParserCombinators.Parsec

```
run (char '!') "!123!"
run :: Show a => Parser a -> String -> IO()
run p input = case (parse p "" input) of
Left err -> do { putStr "parse error at" ; print err } run letter "123"
Right x -> print x
run (char '!') "!123!"
run (oneOf "!?.") "?123"
run letter "a"
run digit "123"
run digit "123"
run word "abc def"
run word "abc123"
```

### Zátvorky 3

```
paren :: Parser ()
                                   -- parser, ktorý nevracia ziadnu hodnotu
paren = do{ char '('; paren ; char ')' ; paren }
         <|> return ()
data Bin = Nil | Node Bin Bin -- vnútorná reprezentácia
                 deriving(Show, Read, Eq)
parenBin :: Parser Bin
                                    -- analyzátor zo String do Bin
parenBin = do { char '('; x<-parenBin; char ')'; y<-parenBin; return (Node x y) }</pre>
             <|>
             return Nil
                                   -- parser, ktorý vracia hľbku výrazu
nesting :: Parser Int
nesting = do{ char '('; n<-nesting; char ')'; m<-nesting; return (max (n+1) m) }</pre>
         <|> return 0
```

# 1

#### Minského zápalkový stroj

```
stmt ::=
        while <expr> do <stmt> end while
        if <expr> then <stmt> [else <stmt>] end if
        < id > + +
        <id>---
expr ::=
        <id> hrka
        <id> nehrka
        true
        false
        <expr> && <expr>
        <expr> || <expr>
        !<expr>
        (<expr>)
```

### Lexikálna analýza

import Text.ParserCombinators.Parsec.Token import Text.ParserCombinators.Parsec.Language

## 1

#### Haskell & Java Style

import Text.ParserCombinators.Parsec.Language(haskellStyle) import Text.ParserCombinators.Parsec.Language(javaStyle)

```
haskellStyle :: LanguageDef st
                                            javaStyle :: LanguageDef st
haskellStyle= emptyDef
                                            javaStyle = emptyDef
  { commentStart = "{-"
                                               { commentStart = "/*"
  , commentEnd = "-}"
  , commentLine
                                               , commentEnd = "*/"
  , nestedComments = True
                                               , commentLine = "//"
  , identStart = letter
                                               , nestedComments = True
  , identLetter = alphaNum <|> oneOf "_"
                                               , identStart = letter
  , opStart = opLetter haskellStyle
                                               , identLetter = alphaNum <|> oneOf "_"
  , opLetter =
         oneOf ":!#$%&*+./<=>?@\\^|-~"
                                                , reservedNames = []
  , reservedOpNames= []
                                                , reservedOpNames= []
  , reservedNames = []
                                                , caseSensitive = False
  , caseSensitive = True
                                            }
```

#### Vnútorná reprezentácia výrazu

```
data Expr =
                                  run hrkExpr "(A nehrka) && (B hrka)"
   Var String HrkaNehrka |
                                  BinOp And (Var "A" Nehrka)
   Con Bool |
                                            (Var "B" Hrka)
   UnOp UnOp Expr |
   BinOp BinOp Expr Expr
                                  run hrkExpr "((A nehrka) && (B hrka))"
   deriving Show
                                  BinOp And (Var "A" Nehrka)
                                            (Var "B" Hrka)
data HrkaNehrka =
   Hrka I
                                  run hrkExpr "!((A nehrka) && (B hrka))"
   Nehrka
                                  UnOp Not
   deriving Show
                                           (BinOp And
                                                    (Var "A" Nehrka)
data UnOp = Not deriving Show
                                                    (Var "B" Hrka))
data BinOp = And | Or deriving Show
```

#### Parser výrazov

```
hrkExpr :: Parser Expr
hrkExpr = buildExpressionParser table factor <?> "hrk-expression"
table = \Gamma
     [Prefix ( do { _ reservedOp "!"; return (UnOp Not)})]
   , [Infix (do { _reservedOp "&&" ; return (BinOp And) }) AssocLeft]
   , [Infix (do { _reservedOp "||" ; return (BinOp Or) }) AssocLeft]
factor = _parens hrkExpr
     <|> (do { id <- identifier ;
                kind <- do { _reserved "hrka"; return Hrka }
                        do { _reserved "nehrka"; return Nehrka };
          return (Var id kind)})
     <|> (do { _reserved "true" ; return (Con True)})
     <|> (do { _reserved "false" ; return (Con False)})
```

# Vnútorná reprezentácia príkazu

```
data Stmt =
        Inc String |
         Dec String |
         If Expr Stmt Stmt |
        While Expr Stmt |
        Seq [Stmt]
         deriving Show
run hrkStat "while A hrka && B nehrka do A++;B++;C-- end while"
Seq [While (BinOp And (Var "A" Hrka) (Var "B" Nehrka))
         (Seq [Inc "A",Inc "B",Dec "C"])]
run hrkStat "if! A hrka then A++ else A-- end if"
Seq [If (UnOp Not (Var "A" Hrka))
        (Seq [Inc "A"])
         (Seg [Dec "A"])]
```

hrkStat :: Parser Stmt stmtparser :: Parser Stmt



#### Parser pre hrkStat

```
hrkStat = do { _whiteSpace; s<-stmtparser; eof; return s}</pre>
stmtparser = fmap Seq (_semiSep1 stmt1)
stmt1 = do { v <- _identifier
                 ; do { _reservedOp "++"; return (Inc v) }
                   <|>
                   do { _reservedOp "--"; return (Dec v) }
          <|>
         do {
                reserved "if"
                  ; b <- hrkExpr
                  ; _reserved "then"
                  ; p <- stmtparser
                  ; _reserved "else"
                  ; q <- stmtparser
                  ; _reserved "end" ; _reserved "if"
                  ; return (If b p q)
             } <|> ...
```

#### Domáca úloha

- ľahšie použite Parsec a napíšte analyzátor pre KSP-Turingov stroj http://dai.fmph.uniba.sk/courses/FPRO/source/tstroj.pdf výsledný analyzátor musí zvládnuť príklady zo zadania KSP
- ťažšie použite Parsec a napíšte analyzátor pre KSP-Frontový Pascal http://dai.fmph.uniba.sk/courses/FPRO/source/froscal.pdf výsledný analyzátor musí implementovať gramatiku zo zadania KSP
- klasika použite Parsec a napíšte analyzátor pre jazyk Brainf\*ck https://en.wikipedia.org/wiki/Brainfuck v prípade, že parser nebude vaše dielo, budete ho musieť vedieť modifikovať