## Monadické parsery

Graham Hutton, Erik Mejer: Functional Pearls: Monadic Parsing in Haskell, 1997 <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/~pszqmh/monparsing.pdf">http://www.cs.nott.ac.uk/~pszqmh/monparsing.pdf</a>

Daan Leijen: *Parsec, a fast combinator parser* <a href="https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/parsec-paper-letter.pdf">https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/parsec-paper-letter.pdf</a>

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

Mnohé analyzátory, ktoré dnes skonštruujeme sa nazývajú trochu inými menami v predošlej prednáške, 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: (zámena poradia argumentov dvojice – rešpektujúc použité zdroje):
         type Parser result = String -> [(result, String)]
      resp.:
         data Parser result = Parser (String -> [(result, String)])
      Primitívne parsery:
      return :: a->Parser a
                                                    -- tento sa volal succeed
      return v = \langle xs - \rangle [(v,xs)]
                                                    -- nečíta zo vstupu, dá výsledok v
      return v xs = [(v,xs)]
     zero :: Parser a
                                                    -- tento sa volal fail
     zero = \xs -> []
                                                    -- neakceptuje nič
     zero xs = []
                                                    -- akceptuje ľubovoľný znak
      item :: Parser Char
                                                    -- tento znak dá do výsledku
               = \xs -> case xs of
      item
                        [] -> []
(v:vs) -> [(v,vs)]
                                                    -- prázdny vstup, neakceptuje
> parse (return 99) ""
                                                    -- akceptuje znak v
> parse (item) "abc"
```

mparser.hs

```
type Parser result = String -> [(result, String)]
```

### 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:



### Kombinátory parserov

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:

#### Typovačka:

- v je výsledok analýzy parsera p, teda v :: a,
- (qf v) je nový analyzátor parametrizovaný výsledkom v, teda (qf v)::Parser b
- (qf v) xs' :: [(b,String)]
- concat :: [[b]] -> [b]

```
type Parser result = String -> [(result, String)]
```

### Ako sa bind používa

#### 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  $x_1 x_2 \dots x_n$ .

# -

## 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

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

ilustrujme na niekoľkých príkladoch použitie bind:

char :: Char -> Parser Char

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

digit :: Parser Char digit = sat isDigit

- -- ten sa volal satisfy
- -- akceptuje symbol, pre ktorý
- -- platí predikát pred
- > parse (sat isLower) "a123ad" [('a',"123ad")]
- -- ten sa volal symbol
- <u>-- akceptuie znak x</u>
- > parse (char 'a') "a123ad" [('a',"123ad")]
- -- akceptuje cifru

> parse (digit) "123ad" [('1',"23ad")]



#### Zjednotenie

zjednotenie/disjunkciu analyzátorov poznáme ako <|>:

```
plus :: Parser a -> Parser a -> Parser a

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

p`plus` q xs = (p xs ++ q xs)

> 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)

```
> 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 (analógia determ, greedy):

## Monády

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

definícia monády v Haskelli (class alias interface):

```
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])
```

bind je >>=

### Monad comprehension

je syntax-sugar asi ako List-comprehension

```
[ f x_1 x_2 .... x_n ]
p_1 >>= \x_1 ->
                                                   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 ->
                                                            X_n < -p_n;
                                  x_n < -p_n
return (f x_1 x_2 \dots x_n)
                                                            return (f x_1 x_2 \dots x_n)
                                                                 -- volal sa token
          :: String -> Parser String
string
                      = return ""
string
                      = char x >>= \ ->
string (x:xs)
                                                                 -- výsledok char x zahoď
                        string xs >>= \setminus ->
                                                                 -- výsledok string xs zahoď
                        return (x:xs)
                                                                 -- výsledok je celé slovo x:xs
                                                                 -- prepíšeme do novej
               :: String -> Parser String
string
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...
```

```
> parse (string "begin") "beginend"
[("begin","end")]
```

## Iterátory {p}\*, {p}+

```
gramatika: many_p -> many1_p / \varepsilon -- l'ubovol'ne vel'akrát p
many0 :: Parser a -> Parser [a]
                                               > parse (many0 digit) "123abc"
many0 p = many1 p +++ return []
                                                [("123","abc")]
• gramatika: many1_p -> p many_p -- aspoň raz p
       :: Parser a -> Parser [a]
many1
                = do {a <- p; as <- many p; return (a:as)}
many1 p

    opakovanie p s oddeľovačom sep: reg.výraz: {p sep}*p / ε

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

#### **data** Parser result = Parser(String -> [(result,String)])

# Zátvorky

Analyzátor pre dobre uzátvorkované () výrazy podľa gramatiky:  $P \rightarrow (P)P/\varepsilon$ 

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

Analyzátor pre dobre uzátvorkované ()[] P -> (P) P | [P] P | ε

# Zátvorky 2

```
naivne
                  :: Parser ()
parenBr
parenBr
                  do {
                           (open `plus` openBr); parenBr;
                           (close `plus` closeBr); parenBr; return () }
                  +++ return ()
   správne
parenBr
                  :: Parser ()
                           do { open; parenBr; close; parenBr; return () }
parenBr
                           +++
                           do { openBr ; parenBr; closeBr ; parenBr; return () }
                           +++
                                              > parse parenBr "([[]])([])"
                           return ()
                                              > parse parenBr "([[(])])([])"
```

## 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 po sebe-idúcich výsledkov iného parsera (f x y)

```
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
```

b op
//
c d

op
//
op d

op

а ор

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) }</pre>
```

```
> 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:

```
> 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

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

- 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 :: Show a => Parser a -> String -> IO()
run p input = case (parse p "" input) of
    Left err -> do { putStr "parse error at" ; print err }
    Right x -> print x
```

```
run (char '!') "!123!"
run (oneOf "!?.") "?123"
run letter "a"
run letter "123"
run digit "123"
run word "abc def"
run word "abc123"
```

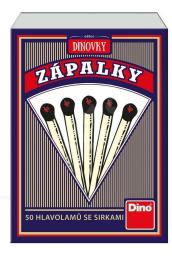
# Zátvorky 3

Analyzátor pre dobre uzátvorkované () výrazy podľa gramatiky:  $P \rightarrow (P)P/\varepsilon$ 

```
-- parser, ktorý nevracia ziadnu hodnotu
paren :: Parser ()
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
```



# 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>)
```



# Minského zápalkový stroj



- A := B
- while (B hrka) do B--; A++; end while
- while (A hrka) do A--; end while; while (B hrka) do B--; A++; end while
- while (A hrka) do A--; end while; while (B hrka) do B--; C++; A++ end while while (C hrka) do C--; B++ end while

A := B + C

A := B\*C

A := B/C

A := B!

A := fib B

- -- presunutie z B do A
- -- vymazanie A
- -- presunutie z B do A
- -- vymazanie A
- -- presunutie z B do A,C
- -- presunutie z C do B

# Lexikálna analýza

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

# 4

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

#### import Text.ParserCombinators.Parsec.Expr

## 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"])
         (Seq [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
   <a href="http://dai.fmph.uniba.sk/courses/FPRO/source/froscal.pdf">http://dai.fmph.uniba.sk/courses/FPRO/source/froscal.pdf</a>
   výsledný analyzátor musí implementovať gramatiku zo zadania KSP
- klasika
   použite Parsec a napíšte analyzátor pre jazyk Brainf\*ck
   <a href="https://en.wikipedia.org/wiki/Brainfuck">https://en.wikipedia.org/wiki/Brainfuck</a>
   v prípade, že parser nebude vaše dielo, budete ho musieť vedieť modifikovať