

# Analizador Sintáctico y Semántico para $\lambda^{bn}$

3 de julio de 2017

Teoría de Lenguajes

Integrante	LU	Correo electrónico
Gabriel Eric Thibeault	114/13	gabriel.eric.thibeault@gmail.com
Gonzalo Ciruelos Rodríguez	063/14	Gonzalo.ciruelos@gmail.com
Luis Agustín Nieto	46/01	lnieto@dc.uba.ar



### Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja) Intendente Güiraldes 2160 - C1428EGA Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359 http://www.fcen.uba.ar

### 1. Introducción

En el presente trabajo práctico se crea un analizador sintáctico y semántico para un subconjunto del lenguaje cálculo lambda tipado, sobre booleanos y naturales  $\lambda^{bn}$ , para dicha implementación se utilizó PLY que es una implementación en Python de las clásicas herramientas lex y yacc.

El lexer está implementado en lexer.py, en el mismo definimos las expresiones regulares y tokkes que sirven para definir si cadenas de entrada son o no válidas acorde a la gramática creada. El parser, donde definimos las producciones de la gramática, está implementado en parser.py

### 2. Gramática

La mayor dificultad del trabajo fue armar una gramática que sirviese para para el lenguaje pedido, el mismo es:

```
 \texttt{M} ::= x \mid \texttt{true} \mid \texttt{false} \mid \texttt{if} \; \texttt{M1} \; \texttt{then} \; \texttt{M2} \; \texttt{else} \; \texttt{M3} \mid \x:\texttt{T.M} \mid \texttt{M1} \; \texttt{M2} \mid \texttt{0} \mid \texttt{succ}(\texttt{M}) \\ \mid \texttt{pred}(\texttt{M}) \mid \texttt{iszero}(\texttt{M})
```

Y la gramática que se implementó:

 $G = \langle \{E, S, C, L, T'\}, \{var, true, false, iszero, succ, pred, (,), 0, Bool, Nat, if, then, else, \backslash V, :, .\}, P, E \rangle, \text{ con P:}$ 

$$E \rightarrow SL$$

$$E \rightarrow SL$$

$$E \rightarrow S$$

$$S \rightarrow SC$$

$$S \rightarrow \lambda$$

$$C \rightarrow (E)$$

$$C \rightarrow var$$

$$C \ \rightarrow \ true$$

$$C \rightarrow false$$

$$C \rightarrow 0$$

$$C \rightarrow iszero(E)$$

$$C \rightarrow succ(E)$$

$$C \rightarrow pred(E)$$

$$C \ \to \ if \ E \ else \ E \ then \ C$$

$$L \rightarrow V: T.E$$

$$T \rightarrow T' \rightarrow T$$

$$T \rightarrow T'$$

$$T' \rightarrow (T->T')$$

$$T' \rightarrow Bool$$

$$T' \rightarrow Nat$$

(1)

### 3. Código y Tests

#### 3.1. Uso

Utilizando de base el código del taller de ply, se necesitan instalar las dependencias que estan en requirements.txt.

```
pip install -r --user requirements.txt
./CLambda --debug 'if true then false else true'
```

#### 3.2. lexer.py

```
\#! coding: utf-8
   """ Calculator \ lexer \ example."""
  import ply.lex as lex
  from .ast import TNat, TBool, TArrow, LBool, LZero, LVar, LLambda, LSucc,
      LPred, LApp, LIfThenElse, LIsZero
  from difflib import get_close_matches
   ,, ,, ,,
   Lista de tokens
  El analizador léxico de PLY (al llamar al método lex.lex()) va a buscar
  para cada uno de estos tokens una variable "t_TOKEN" en el módulo actual.
11
  Sí, es súper nigromántico pero es lo que hay.
  t\_TOKEN puede ser:
15
16
  - Una expresión regular
17
  - Una función cuyo docstring sea una expresión regular (bizarro).
  En el segundo caso, podemos hacer algunas cosas "extras", como se
  muestra aquí abajo.
21
22
   """
23
24
  tokens = (
       'TRUE'
26
       'FALSE' ,
27
       'ZERO',
       'IF',
       'THEN',
30
       'ELSE',
31
32
       'VAR',
       'LAM'.
34
       'COLON',
35
       'DOT',
       'LPARENS',
38
       'RPARENS',
39
```

```
40
        'ISZERO' ,
        'SUCC',
42
        'PRED',
43
         'ARROW' ,
         'NAT',
46
         'BOOL' ,
47
48
   )
50
51
   t_{-}IF = r'if'
   t_THEN = r'then'
   t_{-}ELSE = r'else'
54
  t_LAM = r' \setminus '
  t_{-}COLON = r':
   t\_DOT = r' \setminus .'
58
   t\_LPARENS = r' \setminus ('
   t_RPARENS = r,\),
61
62
   t_ISZERO = r'iszero'
  t\_SUCC = r 'succ'
   t\_PRED = r'pred'
  t\_ARROW = r'->'
  t_NAT = r'Nat'
   t\_BOOL = r'Bool'
69
   t_{ignore} = ' \setminus t'
71
  tokenizables = [
73
   'if',
   'then',
   'else',
   77
   , ) ,
81
   'iszero',
   'succ',
   'pred',
84
   \dot{x}_{->}, \dot{x}_{->}, \dot{x}_{+}
85
   'Nat'
86
   'Bool']
   def t_TRUE(t):
89
        r'true'
90
        t.value = LBool(True)
        return t
92
  def t_FALSE(t):
        r'false
        t.value = LBool(False)
```

```
return t
97
    \mathbf{def} \ t_{-} \mathbf{ZERO}(t):
99
         r '0 '
100
         t.value = LZero()
101
         return t
103
   \# Sin i, s, p
104
    \mathbf{def} \ t_{-}VAR(t):
105
         r'[abcjxyz]'
         t.value = LVar(t.value)
107
        return t
108
    \mathbf{def} \ \mathbf{t} = \mathbf{error}(\mathbf{t}):
110
         errStr = t.value.split(', ', 1)[0]
111
         print('Error de sintaxis: la cadena "{}" no puede ser tokenizada.'
112
                . format (errStr))
113
         if get\_close\_matches(errStr, tokenizables, n = 1):
114
              print('Habrás querido usar "{}"?'
115
                      .format(*get\_close\_matches(errStr, tokenizables, n = 1)))
         exit(0)
118
   # Build the lexer
119
    lexer = lex.lex()
120
    def apply_lexer(string):
         """Aplica el lexer al string dado."""
122
         lexer.input(string)
123
124
         return list (lexer)
```

#### 3.3. parser.py

```
"""Parser LR(1) de lambda."""
 import ply.yacc as yacc
 from .lexer import tokens
  from .ast import TNat, TBool, TArrow, LBool, LZero, LVar, LLambda, LSucc,
    LPred, LApp, LIfThenElse, LIsZero
  precedence = (
     ('right', 'ARROW'),
  #
11
 #
    E \rightarrow S L
12
 #
     \mid S
 #
14
 #
15
    16
  def p_expr_s_lambda(p):
17
     'expr : S lambda
18
     if p[1] is None:
        p[0] = p[2]
     else:
21
        p[0] = LApp(p[1], p[2])
24
  \mathbf{def} \ \mathbf{p}_{-}\mathbf{expr}_{-}\mathbf{s}(\mathbf{p}):
25
     'expr : S'
26
     p[0] = p[1]
27
28
  29
    #
30
    S \rightarrow S cont
  #
31
 #
32
 #
33
 #
34
    35
  def p_s_factor(p):
36
     'S : S cont
37
     if p[1] is None:
        p[0] = p[2]
     else:
40
        p[0] = LApp(p[1], p[2])
41
  \mathbf{def} \ \mathbf{p}_{-}\mathbf{s}_{-}\mathbf{empty}(\mathbf{p}):
43
     'S : '
44
45
  ############################ cont
```

```
C \rightarrow (E)
  #
   #
           var
  #
            true
   #
           false
51
   #
           0
   #
           iszero(E)
   #
           succ(E)
54
   #
           pred(E)
55
         | if E then E else C
   #
  #
  #
58
      def p_cont_expr(p):
60
       'cont : LPARENS expr RPARENS'
61
       p[0] = p[2]
62
63
   def p_cont_var(p):
64
       'cont : VAR'
65
       p[0] = p[1]
67
   def p_cont_true(p):
68
       'cont : TRUE'
69
       p[0] = p[1]
70
71
   def p_cont_false(p):
72
       'cont : FALSE
73
       p[0] = p[1]
75
   def p_cont_zero(p):
76
        'cont : ZERO'
77
       p[0] = p[1]
79
   def p_cont_iszero(p):
80
       'cont : ISZERO LPARENS expr RPARENS'
81
       p[0] = LIsZero(p[3])
83
   \mathbf{def} \ \mathbf{p}_{-}\mathbf{cont}_{-}\mathbf{succ}(\mathbf{p}):
84
       'cont : SUCC LPARENS expr RPARENS'
85
       p[0] = LSucc(p[3])
87
   \mathbf{def} \ \mathbf{p}_{-}\mathbf{cont}_{-}\mathbf{pred}(\mathbf{p}):
88
       'cont : PRED LPARENS expr RPARENS'
       p[0] = LPred(p[3])
90
91
   def p_cont_ifthenelse(p):
92
       'cont : IF expr THEN expr ELSE cont'
93
       p[0] = LIfThenElse(p[2], p[4], p[6])
94
95
   #
      L \rightarrow V : T . E
  #
  #
101
  #
```

```
102
   def p_lambda(p):
103
       'lambda : LAM VAR COLON type DOT expr'
       p[0] = LLambda(p[2], p[4], p[6])
105
106
107
   108
      #
109
  #
      T \rightarrow T' \rightarrow T
110
         \mid T'
   #
111
112
   #
      T' \rightarrow (T \rightarrow T')
   #
113
   #
            Bool
114
   #
            Nat
115
   #
116
   #
117
      118
119
   \mathbf{def} \ \mathbf{p}_{-}\mathbf{type}_{-}\mathbf{arrow}(\mathbf{p}):
120
       'type : typex ARROW type'
       p[0] = TArrow(p[1], p[3])
122
123
   def p_type_typex(p):
124
       'type : typex
       p[0] = p[1]
126
127
   def p_typex_arrow(p):
128
       'typex : LPARENS type ARROW typex RPARENS'
       p[0] = TArrow(p[2], p[4])
130
131
   def p_typex_nat(p):
132
       'typex : NAT'
       p[0] = TNat()
134
135
   def p_typex_bool(p):
       'typex : BOOL'
137
       p[0] = TBool()
138
139
   def p_error(p):
141
       if p is not None:
142
           exit ('Error de sintaxis: "{}" no esperado en la posicion {}.'.
143
               format(p.value, p.lexpos))
       else:
           exit ('Error de sintaxis: fin de linea no esperado ("$").')
145
148
   # Build the parser
149
   parser = yacc.yacc(debug=True)
150
   def apply_parser(str):
```

```
p = parser.parse(str)
153
      p.add_judgement('h4x0r', 'turururu') # Llamo a add judgement del
          padre asi
                                           # propaga todos los judgements
155
                                              hacia
                                           \# abajo.
156
      \# print(repr(p))
157
      while p is not None and not p.is_value():
158
          p = p.value()
159
          \# print(repr(p))
      return p
161
 3.4.
       tests.py
  \#!/usr/bin/python3
  from lambda_calculus import parse, lex
  tests = |
      ('0', '0', 'Nat'),
      ('true', 'true', 'Bool'),
      ('if true then 0 else false', None, None),
      ('\\x:Bool.if x then false else true', '\ x: Bool . if x then false
          else true', 'Bool -> Bool'),
      ('(\x:Bool.succ(x)) true', None, None),
       ('succ(succ(succ(0)))', '3', 'Nat'),
       ('x', None, None),
      ('succ(succ(pred(0)))', '2', 'Nat'),
       '0 0', None, None),
      (')\x: Nat-> Nat.\y: Nat.(\x: Bool. if z then x y else 0)', '\x: Nat->
          Nat . 

 \ y : Nat . 
 \ z : Bool . if z then x y else 0', '(Nat -\!\!> Nat
          ) -> Nat -> Bool -> Nat'),
      ('(\x: Nat->Nat.\y: Nat.(\z: Bool. if z then x y else 0)) (\j: Nat. succ
          , Nat '),
      \# Nuestros tests.
      (\ '(\ \ z : Nat. \ pred(z)) \ ((\ \ \ X : Nat->Nat. \ x \ succ(succ(0))) \ \ \ \ \ \ '1
           , 'Nat'),
       ('(\x: Nat.succ(x)) 0 0', None, None),
       ('(\\x:Nat->Nat. x succ(succ(0))) \y:Nat. y', '2', 'Nat')
      ('(\x: Nat. if iszero(x) then succ(x) else pred(x)) pred(0)', '1', '
          Nat'),
      ('(\x: Nat. if iszero(x) then succ(x) else pred(x)) succ(0)', '0', '
          Nat'),
      ('(\x: Nat. if iszero(x) then succ(x) then pred(x)) succ(0)', None,
24
          None),
25
27
  for test in tests:
      try:
          p = parse(test[0])
31
          if str(p.value()) = test[1] and str(p.type()) = test[2]:
32
              print('PASSED', test[0])
```

```
{f else}:
34
               print('FAILED', test[0])
               if str(p.value()) != test[1]:
                   print('Got', str(p.value()), 'but expected', test[1])
37
               if str(p.type()) != test[2]:
                   print('Got', str(p.type()), 'but expected', test[2])
       except:
40
           if test[1] is None:
41
               print('PASSED', test[0])
42
           else:
               print('FAILED', test[0])
```

## Teoría de Lenguajes - Trabajo Práctico Analizador Sintáctico y Semántico para $\lambda^{bn}$

Versión 1.1

1<sup>er</sup> cuatrimestre 2017

Fecha de entrega: miércoles 5 de julio

#### 1. Introducción

Se desea crear un analizador sintáctico y semántico para un subconjunto del cálculo lambda tipado, sobre booleanos y naturales  $(\lambda^{bn})$ . Así, se deberá analizar si una expresión sigue la sintaxis esperada, y, en caso de que sea válida, evaluarla adecuadamente e indicar de qué tipo es la misma. Para la evaluación se puede tomar como referencia la semántica operacional descripta en la 1era clase teórica y práctica de la materia Paradigmas de Lenguajes de Programación[1], aunque no es necesario seguir el proceso de evaluación paso a paso.

### Descripción del lenguaje de entrada

A efectos de este trabajo, el cálculo lambda tipado sobre booleanos y naturales deberá soportar los siguientes términos:

```
M ::= x \mid true \mid false \mid if M1 then M2 else M3 \mid \x:T.M \mid M1 M2 \mid 0 \mid succ(M) \mid pred(M) \mid iszero(M)
```

Siendo,

- true y false: las constantes de verdad,
- if M1 then M2 else M3: el condicional, por lo que si M1 se evalua a true, resulta en M1 y, si no, en M2, y tanto M1 como M2 deben ser del mismo tipo,
- M1 M2: la aplicación de la función denotada por el término M1 al argumento M2, por ejemplo, (\x:Nat.succ(succ(x))) 3, se evalúa a 5, siendo este operador asociativo a izquierda (i.e., M N P se evalúa como (M N) P),
- \x:T.M: una función anónima cuyo parámetro formal es x, de tipo T, y tiene a M como cuerpo, siendo la operación de menor precedencia,
- x: una variable de términos, que puede estar asociada a cualquier término válido,
- succ(M): el término para el cual se evaluará M hasta obtener un número, y se lo incrementará,
- pred(M): el término para el cual se evaluará M hasta obtener un número, y se lo decrementará,
- iszero(M): el término para el cual se evaluará M hasta obtener un número, y evaluará a true o false según sea cero o no.

A su vez, cada expresión bien formada del lenguaje puede tomar alguno de los siguientes tipos:

```
T ::= Bool | Nat | T1 -> T2
```

Siendo.

- Bool: el tipo para los booleanos (e.g., isZero(0): Bool),
- Nat: el tipo para los naturales, (e.g., succ(pred(0)): Bool),
- T1 -> T2: el tipo para las funciones que van de T1 a T2, por ejemplo, \x:Bool.if x then succ(0) else pred(succ(succ(0))): Bool -> Nat

Una expresión del cálculo lambda está en forma normal si no puede evaluarse más. Así, los valores están en forma normal, aunque no todops los términos que no pueden evaluarse más son valores. Finalmente, si tenemos un término cerrado (i.e., que no tiene variables libres) y que está bien tipado, podremos obtener un valor a partir de él.

Los valores para nuestro lenguaje serán los siguientes:

```
V ::= true | false | \xriantering x:T.M | n
con n como macro de succ^n(0)
```

### 3. Descripción del lenguaje de salida

Dada una cadena de entrada, se deberá evaluar si efectivamente respeta la sintaxis esperada de Cálculo Lambda y, si no hay algún de error de tipado<sup>1</sup>, se deberá evaluar la expresión hasta llegar a un valor, e indicar el tipo del mismo.

### 3.1. Ejemplos

Entrada	Resultado	Salida
0	OK	0:Nat
true	OK	true:Bool
if true then 0 else false	ERROR: Las dos opciones del if deben tener el mismo tipo	
\x:Bool.if x then false else true	ОК	\x:Bool.if x then false else true:Bool->Bool
\x:Nat.succ(0)	OK	\x:Nat.succ(0):Nat->Nat
\z:Nat.z	OK	\z:Nat.z:Nat->Nat
(\x:Bool.succ(x)) true	ERROR: succ espera un valor de tipo Nat	
succ(succ(succ(0)))	OK	succ(succ(succ(0))):Nat
x	ERROR: El término no es cerrado (x está libre)	
succ(succ(pred(0)))	OK	succ(succ(0)):Nat
\x:Nat.succ(x)	OK	\x:Nat.succ(x):Nat->Nat
0 0	ERROR: La parte izquierda de la aplicación (0) no es una función con dominio en Nat	
\x:Nat->Nat.\y:Nat. (\z:Bool.if z then x y else 0)	OK	<pre>\x:Nat-&gt;Nat.\y:Nat.(\z:Bool.if z then x y else 0): (Nat-&gt;Nat)-&gt;(Nat-&gt;(Bool-&gt;Nat))</pre>
<pre>(\x:Nat-&gt;Nat.\y:Nat.  (\z:Bool.if z then x y else 0))  (\j:Nat.succ(j)) 8 true</pre>	OK	9:Nat

<sup>&</sup>lt;sup>1</sup>Esto es, no se corresponde el tipo esperado con el recibido

### 4. Modo de uso

El programa deberá poder ejecutarse como un comando de consola del sistema operativo, con los siguientes requerimientos:

#### Sinopsis:

./CLambda [EXPRESION]

#### Descripción:

El programa deberá recibir una cadena con la expresión a evaluar (EXPRESION), y deberá devolver por *standard output* (stdout) el resultado de la evaluación.

Si no se especificara la cadena en la llamada, se esperará recibirla por *standard input* (stdin). En caso de que hubiera algún inconveniente al ejecutar el programa, se deberá terminar el programa con código de salida de error (esto es, un código mayor a 0), y mostrar los detalles por standard error (stderr). En caso de que la llamada fuera correcta, se deberá devolver el resultado, y, si no fuera correcta la expresión, incluir todos los detalles del error por stderr y no retornar nada por stdout.

### 5. Implementación

Para el lenguaje descripto se solicita crear un analizador léxico, generando los tokens necesarios de acuerdo a la gramática que se haya diseñado, y un analizador sintáctico, que deberá generar el árbol sintáctico para la cadena procesada, en caso de que sea válida, o un mensaje de error adecuado, en caso de que no lo sea. Para este último caso, será deseable indicar el lugar preciso en el cual se encuentra el problema detectado.

Hay dos grupos de herramientas que se pueden utilizar para generar los analizadores léxicos y sintácticos:

- uno utiliza expresiones regulares y autómatas finitos para el análisis lexicografico, y la técnica LALR para el análisis sintáctico. Ejemplos de esto son lex y yacc, que generan código C o C++, JLex y CUP, que generan código Java y ply, que genera código Python. flex y bison son implementaciones libres y gratuitas de lex y yacc. En particular, permitiremos utilizar PLY 3.6<sup>2</sup> (o superior).
- el otro grupo utiliza la técnica ELL(k), tanto para el análisis léxico como para el sintáctico, generando parsers descendentes iterativos recursivos. Ejemplos son JavaCC y ANTLR, que están escritos en Java. JavaCC puede generar código Java o C++. ANTLR puede generar Java, C#, Python y JavaScript.
  En particular, para este trabajo se podrá usar ANTLR 4.X³, y se recomienda utilizar el plugin⁴

### 6. Detalles de la entrega

para Eclipse.

Se deberá enviar el código a la dirección de e-mail tptleng@gmail.com, satisfaciendo lo siguiente:

- el **asunto de mail** debe ser [TL-TP] seguido por el nombre del grupo (e.g., "[TL-TP] The Chomsky Boys").
- en el mail deberán estar copiados todos los/as integrantes del grupo.

<sup>&</sup>lt;sup>2</sup>PLY: Python Lex & Yacc. Disponible en https://pypi.python.org/pypi/ply

<sup>&</sup>lt;sup>3</sup>ANTLR (ANother Tool for Language Recognition): http://www.antlr.org

<sup>&</sup>lt;sup>4</sup>ANTLR plugin for Eclipse: http://antlreclipse.sourceforge.net

La entrega debe incluir lo descripto a continuación:

- un programa que cumpla con lo solicitado,
- el código fuente del mismo, adecuadamente documentado,
- informe enviado por e-mail y entregado impreso, con los siguientes contenidos:
  - · carátula con datos de integrantes del grupo y nombre del grupo,
  - breve introducción al problema a resolver,
  - los tokens, con sus expresiones regulares, y la gramática no ambigua definidos a partir del lenguaje propuesto,
  - indicación del tipo de la gramática definida, de acuerdo a los vistos en clase,
  - el código de la solución, y, si se usaron herramientas generadoras de código, imprimir la fuente ingresada a la herramienta, no el código generado,
  - descripción de cómo se implementó la solución, con decisiones que hayan tenido que tomar y justificación de las mismas,
  - información y requerimientos de software para ejecutar y recompilar el tp (versiones de compiladores, herramientas, plataforma, etc), como un pequeño manual del usuario, que además de los requerimientos, contenga instrucciones para compilarlo, ejectuarlo, información de parámetros y lo que consideren necesario,
  - casos de prueba con expresiones sintácticamente correctas e incorrectas (al menos tres para cada caso) y
  - un resumen de los resultados obtenidos, y conclusiones del trabajo.

Es parte de lo que se espera de la resolución del trabajo práctico la detección de puntos no especificados en el enunciado y su resolución. En cualquier caso, siempre pueden realizar consultas al respecto.

### Referencias

- [1] Descargas de la materia Paradigmas de Lenguajes de Programación, 1er cuatrimestre de 2017. Disponible en http://www.dc.uba.ar/materias/plp/cursos/2017/cuat1/descargas/.
- [2] Aho, A.V., Lam, M.S., Sethi, R., Compilers: Principles, Techniques and Tools Second Edition, Pearson Education, 2007.
- [3] Goyvaerts, J., Levithan, S., Regular Expressions Cookbook, O'Reilly, 2009.
- [4] Grune, D., Jacobs, C.J.H, Parsing Techniques: A Practical Guide Second Edition, Springer, 2008.
- [5] Hopcroft, J.E., Motwani, R., Ullman, J.D., Introduction to Automata Theory, Languages, and Computation Third Edition, Addison Wesley, 2007.
- [6] Parr, T., The Definitive ANTLR 4 Reference, The Pragmatic Programmers, 2012.
- [7] PLY (Python Lex-Yacc). Documentación. Disponible en http://www.dabeaz.com/ply/ply.html