



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Analizador Sintáctico y Semántico para λ^{bn}

2 de julio de 2017

Teoría de Lenguajes

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



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 λ^{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:

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

Y la gramática que se implementó:

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

$$\begin{aligned}
 E &\rightarrow SL \\
 E &\rightarrow SL \\
 E &\rightarrow S \\
 S &\rightarrow SC \\
 S &\rightarrow \lambda \\
 C &\rightarrow (E) \\
 C &\rightarrow var \\
 C &\rightarrow true \\
 C &\rightarrow false \\
 C &\rightarrow 0 \\
 C &\rightarrow iszero(E) \\
 C &\rightarrow succ(E) \\
 C &\rightarrow pred(E) \\
 C &\rightarrow if E else E then C \\
 L &\rightarrow \backslash V : T.E \\
 T &\rightarrow T' \rightarrow T \\
 T &\rightarrow T' \\
 T' &\rightarrow (T - > T') \\
 T' &\rightarrow Bool \\
 T' &\rightarrow Nat
 \end{aligned}$$

(1)

3. Código

3.1. Uso

3.2. lexer.py

```

1  #!/ coding: utf-8
2  """ Calculator lexer example. """
3  import ply.lex as lex
4  from .ast import TNat, TBool, TArrow, LBool, LZero, LVar, LLambda, LSucc,
   LPred, LApp, LIfThenElse, LIsZero
5  from difflib import get_close_matches
6
7  """
8  Lista de tokens
9
10 El analizador léxico de PLY (al llamar al método lex.lex()) va a buscar
11 para cada uno de estos tokens una variable "t_TOKEN" en el módulo actual.
12
13 Sí, es súper nigromántico pero es lo que hay.
14
15 t_TOKEN puede ser:
16
17 – Una expresión regular
18 – Una función cuyo docstring sea una expresión regular (bizarro).
19
20 En el segundo caso, podemos hacer algunas cosas "extras", como se
21 muestra aquí abajo.
22
23 """
24
25 tokens = (
26     'TRUE',
27     'FALSE',
28     'ZERO',
29     'IF ',
30     'THEN',
31     'ELSE',
32
33     'VAR',
34     'LAM',
35     'COLON',
36     'DOT',
37
38     'LPARENS',
39     'RPARENS',
40
41     'ISZERO',
42     'SUCC',
43     'PRED',
44
45     'ARROW',
46     'NAT',
47     'BOOL',
48

```

```

49 )
50
51
52 t_IF = r 'if '
53 t_THEN = r 'then '
54 t_ELSE = r 'else '
55
56 t_LAM = r '\\ '
57 t_COLON = r ':' '
58 t_DOT = r '\\. '
59
60 t_LPARENS = r '\\(' '
61 t_RPARENS = r '\\)' '
62
63 t_ISZERO = r 'iszero '
64 t_SUCC = r 'succ '
65 t_PRED = r 'pred '
66
67 t_ARROW = r '->'
68 t_NAT = r 'Nat '
69 t_BOOL = r 'Bool '
70
71 t_ignore = ' \\t '
72
73 tokenizables = [
74 'if ',
75 'then ',
76 'else ',
77 '\\ ',
78 ': ',
79 '. ',
80 '(',
81 ') ',
82 'iszero ',
83 'succ ',
84 'pred ',
85 '->',
86 'Nat ',
87 'Bool ']
88
89 def t_TRUE(t):
90     r 'true '
91     t.value = LBool(True)
92     return t
93
94 def t_FALSE(t):
95     r 'false '
96     t.value = LBool(False)
97     return t
98
99 def t_ZERO(t):
100     r '0 '
101     t.value = LZero()
102     return t
103
104 # Sin i, s, p
105 def t_VAR(t):

```

```
106     r '[abcjxyz]'
107     t.value = LVar(t.value)
108     return t
109
110 def t_error(t):
111     errStr = t.value.split(' ', 1)[0]
112     print('Error de sintaxis: la cadena "{}" no puede ser tokenizada.'
113           .format(errStr))
114     if get_close_matches(errStr, tokenizables, n = 1):
115         print('Habrás querido usar "{}"?'.
116               .format(*get_close_matches(errStr, tokenizables, n = 1)))
117     exit(0)
118
119 # Build the lexer
120 lexer = lex.lex()
121 def apply_lexer(string):
122     """Aplica el lexer al string dado."""
123     lexer.input(string)
124
125     return list(lexer)
```

3.3. parser.py

```

1  """Parser LR(1) de lambda."""
2  import ply.yacc as yacc
3  from .lexer import tokens
4  from .ast import TNat,TBool,TArrow, LBool, LZero, LVar, LLambda, LSucc,
    LPred, LApp, LIfThenElse, LIsZero
5
6  precedence = (
7      ('right', 'ARROW'),
8  )
9
10 ##### EXPRESION
11 #####
12 #
13 # E -> S L
14 #   | S
15 #
16 #####
17
18 def p_expr_s_lambda(p):
19     'expr : S lambda'
20     if p[1] is None:
21         p[0] = p[2]
22     else:
23         p[0] = LApp(p[1], p[2])
24
25 def p_expr_s(p):
26     'expr : S'
27     p[0] = p[1]
28
29 ##### S
30 #####
31 #
32 # S -> S cont
33 #   |
34 #
35 #####
36
37 def p_s_factor(p):
38     'S : S cont'
39     if p[1] is None:
40         p[0] = p[2]
41     else:
42         p[0] = LApp(p[1], p[2])
43
44 def p_s_empty(p):
45     'S : '
46
47 ##### cont
48 #####
49 #

```

```

48 # C -> (E)
49 #   | var
50 #   | true
51 #   | false
52 #   | 0
53 #   | iszero(E)
54 #   | succ(E)
55 #   | pred(E)
56 #   | if E then E else C
57 #
58 #
#####

59
60 def p_cont_expr(p):
61     'cont : LPARENS expr RPARENS'
62     p[0] = p[2]
63
64 def p_cont_var(p):
65     'cont : VAR'
66     p[0] = p[1]
67
68 def p_cont_true(p):
69     'cont : TRUE'
70     p[0] = p[1]
71
72 def p_cont_false(p):
73     'cont : FALSE'
74     p[0] = p[1]
75
76 def p_cont_zero(p):
77     'cont : ZERO'
78     p[0] = p[1]
79
80 def p_cont_iszero(p):
81     'cont : ISZERO LPARENS expr RPARENS'
82     p[0] = LIsZero(p[3])
83
84 def p_cont_succ(p):
85     'cont : SUCC LPARENS expr RPARENS'
86     p[0] = LSucc(p[3])
87
88 def p_cont_pred(p):
89     'cont : PRED LPARENS expr RPARENS'
90     p[0] = LPred(p[3])
91
92 def p_cont_ifthenelse(p):
93     'cont : IF expr THEN expr ELSE cont'
94     p[0] = LIfThenElse(p[2], p[4], p[6])
95
96
97 ##### LAMBDA
98 #####
99 #
100 # L -> \ V : T . E
101 #

```



```

#####

102
103 def p_lambda(p):
104     'lambda : LAM VAR COLON type DOT expr'
105     p[0] = LLambda(p[2], p[4], p[6])
106
107
108 ##### TIPOS
109 #####
110 #
111 #   T -> T' -> T
112 #       | T'
113 #
114 #   T' -> (T -> T')
115 #       | Bool
116 #       | Nat
117 #
118 #####

118
119
120 def p_type_arrow(p):
121     'type : typex ARROW type'
122     p[0] = TArrow(p[1], p[3])
123
124 def p_type_typex(p):
125     'type : typex'
126     p[0] = p[1]
127
128 def p_typex_arrow(p):
129     'typex : LPARENS type ARROW typex RPARENS'
130     p[0] = TArrow(p[2], p[4])
131
132 def p_typex_nat(p):
133     'typex : NAT'
134     p[0] = TNat()
135
136 def p_typex_bool(p):
137     'typex : BOOL'
138     p[0] = TBool()
139
140
141 def p_error(p):
142     if p is not None:
143         exit('Error de sintaxis: "{}" no esperado en la posicion {}'.format(p.value, p.lexpos))
144     else:
145         exit('Error de sintaxis: fin de linea no esperado ("$").')
146
147
148
149 # Build the parser
150 parser = yacc.yacc(debug=True)
151
152 def apply_parser(str):

```

```
153     p = parser.parse(str)
154     p.add_judgement('h4x0r', 'turururu') # Llamo a add judgement del
        padre asi
155                                     # propaga todos los judgements
        hacia
156                                     # abajo.
157     # print(repr(p))
158     while p is not None and not p.is_value():
159         p = p.value()
160         # print(repr(p))
161     return p
```

Teoría de Lenguajes – Trabajo Práctico

Analizador Sintáctico y Semántico para λ^{bn}

Versión 1.1

1^{er} 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 (λ^{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 descrita 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.

2. Descripción del lenguaje de entrada

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

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

Siendo,

- `true` y `false`: las constantes de verdad,
- `if M1 then M2 else M3`: el condicional, por lo que si `M1` se evalúa 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, $(\lambda x:\text{Nat}.\text{succ}(\text{succ}(x))) \ 3$, se evalúa a 5, siendo este operador asociativo a izquierda (i.e., $M \ N \ P$ se evalúa como $(M \ N) \ P$),
- $\lambda x:T.M$: 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:

$T ::= \text{Bool} \mid \text{Nat} \mid T1 \rightarrow T2$

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, $\lambda x:\text{Bool}.\text{if } x \text{ then } \text{succ}(0) \text{ else } \text{pred}(\text{succ}(\text{succ}(0))) : \text{Bool} \rightarrow \text{Nat}$

Una expresión del cálculo lambda está en forma normal si no puede evaluarse más. Así, los valores están en forma normal, aunque no todos 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:

$V ::= \text{true} \mid \text{false} \mid \backslash x:T.M \mid n$

con n como macro de $\text{succ}^n(0)$

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 tipado¹, 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	
$\backslash x:\text{Bool}.\text{if } x \text{ then false else true}$	OK	$\backslash x:\text{Bool}.\text{if } x \text{ then false else true:Bool} \rightarrow \text{Bool}$
$\backslash x:\text{Nat}.\text{succ}(0)$	OK	$\backslash x:\text{Nat}.\text{succ}(0):\text{Nat} \rightarrow \text{Nat}$
$\backslash z:\text{Nat}.z$	OK	$\backslash z:\text{Nat}.z:\text{Nat} \rightarrow \text{Nat}$
$(\backslash x:\text{Bool}.\text{succ}(x)) \text{ true}$	ERROR: succ espera un valor de tipo Nat	
$\text{succ}(\text{succ}(\text{succ}(0)))$	OK	$\text{succ}(\text{succ}(\text{succ}(0))):\text{Nat}$
x	ERROR: El término no es cerrado (x está libre)	
$\text{succ}(\text{succ}(\text{pred}(0)))$	OK	$\text{succ}(\text{succ}(0)):\text{Nat}$
$\backslash x:\text{Nat}.\text{succ}(x)$	OK	$\backslash x:\text{Nat}.\text{succ}(x):\text{Nat} \rightarrow \text{Nat}$
0 0	ERROR: La parte izquierda de la aplicación (0) no es una función con dominio en Nat	
$\backslash x:\text{Nat} \rightarrow \text{Nat}.\backslash y:\text{Nat}.$ $(\backslash z:\text{Bool}.\text{if } z \text{ then } x \text{ y else } 0)$	OK	$\backslash x:\text{Nat} \rightarrow \text{Nat}.\backslash y:\text{Nat}.$ $(\backslash z:\text{Bool}.\text{if } z \text{ then } x \text{ y else } 0):$ $(\text{Nat} \rightarrow \text{Nat}) \rightarrow (\text{Nat} \rightarrow (\text{Bool} \rightarrow \text{Nat}))$
$(\backslash x:\text{Nat} \rightarrow \text{Nat}.\backslash y:\text{Nat}.$ $(\backslash z:\text{Bool}.\text{if } z \text{ then } x \text{ y else } 0))$ $(\backslash j:\text{Nat}.\text{succ}(j)) \text{ 8 true}$	OK	9:Nat

¹Esto 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 especifica 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 lexicográfico, 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 3.6² (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 para Eclipse.

6. Detalles de la entrega

Se deberá enviar el código a la dirección de e-mail tpleng@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.

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

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

⁴ANTLR plugin for Eclipse: <http://antlrclipse.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, ejecutarlo, 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>