Notes on Formal Compiler Construction with the π Framework

Christiano Braga

Instituto de Computação, Universidade Federal Fluminense, Niterói, Brazil

August 30, 2018

http://github.com/ChristianoBraga/BPLC



1. Introduction Example

2. π expressions

3. π commands

4. π^2 : π Framework in Python

5. IMP language

Compiler pipeline

source	lexer	tokens	parser	concrete	AST transformer	abstract	type checker	abstract	code generator	machine	optimizer	optimized
code				syntax		syntax		syntax		code		machine
				tree		tree		tree				code



Compiler pipeline and formal languages

	Regular ContextFree		ContextFree		ContextSensitive			Turing		Turing		
	Grammar		Grammar		Grammar		Grammar		Machine		Machine	
source	lexer	tokens	parser	concrete	AST transformer	abstract	type checker	abstract	code generator	machine	optimizer	optimized
code				syntax		syntax		syntax		code		machine
				tree		tree		tree				code



Compiler pipeline with the π Framework

Automata

- π lib defines a set of constructions common to many programming languages.
- π lib constructions have a formal automata-based semantics in π automata.
- One may execute (or validate) a program in a given language by running its associated π lib program.
- π Framework: http://github.com/ChristianoBraga/BPLC
- Notes on Formal Compiler Construction with the π Framework: https://github.com/ChristianoBraga/BPLC/blob/master/notes/notes.pdf.



A calculator

We wish to compute simple arithmetic expressions such as 5*(3+2).



A calculator: Lexer

```
::= [0..9]
⟨digit⟩
⟨digits⟩
        ::= \langle digit \rangle^+
⟨boolean⟩ ::= 'true' | 'false'
```



A calculator: concrete syntax

```
::= \langle aexp \rangle \mid \langle bexp \rangle
\langle exp \rangle
                           ::= \langle aexp \rangle '+' \langle term \rangle | \langle aexp \rangle '-' \langle term \rangle | \langle term \rangle
\langle aexp \rangle
                           ::= \langle term \rangle '*' \langle factor \rangle \langle term \rangle '/' \langle factor \rangle \langle factor \rangle
⟨term⟩
                          ::= '(' \( aexp\) ')' \ \( digits\)
⟨factor⟩
                           ::= \langle boolean \rangle | '~' \langle bexp \rangle \langle bexp \rangle \langle boolop \langle bexp \rangle
\langle bexp \rangle
                                    ⟨aexp⟩ ⟨iop⟩ ⟨aexp⟩
⟨boolop⟩
                           ::= '=' | '/\' | '\/'
                           ::= '<' | '>' | '<=' | '>='
⟨iop⟩
```



A calculator: abstract syntax

$$\langle exp \rangle \qquad ::= \langle digits \rangle \mid \langle boolean \rangle \mid \langle exp \rangle \langle bop \rangle \langle exp \rangle$$

$$\langle bop \rangle \qquad ::= '+' \mid '-' \mid '*' \mid '/' \mid '=' \mid ' \setminus /' \mid ' \cdot / \mid ' \cdot | '$$



A calculator: π denotations I

Let D in $\langle digits \rangle$, B in $\langle boolean \rangle$ and E_1, E_2 in $\langle exp \rangle$,

$$[D]_{\pi} = Num(D) \tag{1}$$

$$[B]_{\pi} = Boo(B) \tag{2}$$

$$[E_1 + E_2]_{\pi} = Sum([E_1]_{\pi}, [E_2]_{\pi})$$
(3)

$$[E_1 - E_2]_{\pi} = Sub([E_1]_{\pi}, [E_2]_{\pi})$$

$$[E_1 * E_2]_{\pi} = Mul([E_1]_{\pi}, [E_2]_{\pi})$$
 (5)

$$[E_1/E_2]_{\pi} = Div([E_1]_{\pi}, [E_2]_{\pi})$$

$$\tag{6}$$

$$||E_1| < E_2||_{\pi} = Lt(||E_1||_{\pi}, ||E_2||_{\pi})$$
(7)

$$[E_1 <= E_2]_{\pi} = Le([E_1]_{\pi}, [E_2]_{\pi})$$
 (8)

$$[E_1 > E_2]_{\pi} = Gt([E_1]_{\pi}, [E_2]_{\pi})$$
 (9)

$$||E_1\rangle = E_2||_{\pi} = G(||E_1||_{\pi}, ||E_2||_{\pi})$$

$$||E_1\rangle = E_2||_{\pi} = Ge(||E_1||_{\pi}, ||E_2||_{\pi})$$
(10)



(4)

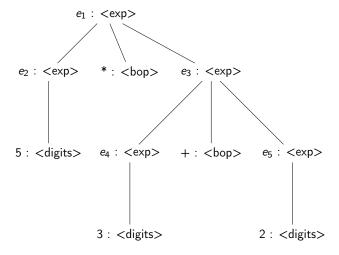


A calculator: π denotations II

- π denotations are functions $[\![\cdot]\!]_{\pi}: AST \to \pi$ lib, where AST denotes the datatype for the abstract syntax tree and π lib denotes the datatype for π lib programs.
- Note that $\llbracket \cdot \rrbracket_{\pi}$ has *trees* as parameters, instances of *AST*. The example expression 5*(3+2) becomes



A calculator: π denotations III





A calculator: π denotations IV



A calculator: executing π lib with π automata

A π automaton is a 5-tuple $\mathscr{A}=(G,Q,\delta,q_0,F)$, where G is a context-free grammar, Q is the set of states, q_0 is the initial state, $F\subseteq Q$ is the set of final states and

$$\delta: L(G)^* \times L(G)^* \times Store \rightarrow Q$$
,

where L(G) is the language generated by G and Store represents the memory. (Elements in a set S^* are represented by terms $[s_1, s_2, ..., s_n]$.)

```
\begin{split} &\delta([Mul(Num(5),Sum(Num(3),Num(2)],\phi,\phi) = \delta([Num(5),Sum(Num(3),Num(2)),\#MUL],\phi,\phi) \\ &\delta([Num(5),Sum(Num(3),Num(2)),\#MUL],\phi,\phi) = \delta([Sum(Num(3),Num(2)),\#MUL],[Num(5)],\phi) \\ &\delta([Sum(Num(3),Num(2)),\#MUL],[Num(5)],\phi) = \delta([Num(3),Num(2),\#SUM,\#MUL],[Num(5)],\phi) \\ &\delta([Num(3),Num(2),\#SUM,\#MUL],[Num(5)],\phi) = \delta([Num(2),\#SUM,\#MUL],[Num(3),Num(5)],\phi) \\ &\delta([Num(2),\#SUM,\#MUL],[Num(3),Num(5)],\phi) = \delta([\#SUM,\#MUL],[Num(2),Num(3),Num(5)],\phi) \\ &\delta([\#SUM,\#MUL],[Num(2),Num(3),Num(5)],\phi) = \delta([\#MUL],[Num(5),Num(5)],\phi) \\ &\delta([\#MUL],[Num(5),Num(5)],\phi) = \delta(\phi,[Num(25)],\phi) \\ &\delta(\phi,[Num(25)],\phi) = Num(25) \end{split}
```



Excerpt of π lib expressions

```
\langle Statement \rangle ::= \langle Exp \rangle

\langle Exp \rangle ::= \langle ArithExp \rangle | \langle BoolExp \rangle

\langle ArithExp \rangle ::= 'Num'(\langle digits \rangle) | 'Sum'(\langle Exp \rangle, \langle Exp \rangle) | 'Sub'(\langle Exp \rangle, \langle Exp \rangle) | 'Mul'(\langle Exp \rangle, \langle Exp \rangle) | 'Mot'(\langle Exp \rangle)
```



π automata semantics for π lib expressions

• Recall that $\delta: L(G)^* \times L(G)^* \times Store \rightarrow Q$, and let $N, N_i \in \mathbb{N}$, $C, V \in L(G)^*$, $S \in Store$,

$$\delta(Num(N) :: C, V, S) = \delta(C, Num(N) :: V, S)$$
(11)

$$\delta(Sum(E_1, E_2) :: C, V, S) = \delta(E_1 :: E_2 :: \#SUM :: C, V, S)$$
 (12)

$$\delta(\#SUM :: C, Num(N_1) :: Num(N_2) :: V, S) = \delta(C, Num(N_1 + N_2) :: V, S)$$
(13)

••

$$\delta(Not(E) :: C, V, S) = \delta(E :: \#NOT :: C, V, S)$$
(14)

$$\delta(\#NOT :: C, Boo(true) :: V, S) = \delta(C, Boo(false) :: V, S)$$

$$\delta(\#NOT :: C, Boo(false) :: V, S) = \delta(C, Boo(true) :: V, S)$$
(16)

- Notation *h*:: *Is* denotes the concatenation of element *h* with the list *Is*.
- *C* represents the *control* stack. *V* represents the *value* stack. *S* denotes the memory store.
- $\delta(\emptyset, V, S)$ denotes an accepting state.



(15)

π lib commands

 Commands are language constructions that require both an environment and a memory store to be evaluated.

```
⟨Statement⟩ ::= <Cmd>
⟨Exp⟩ ::= 'Id'(<String>)
⟨Cmd⟩ ::= 'Assign'(<Id>, <Exp>)
| 'Loop'(<BoolExp>, <Cmd>)
| 'CSeq'(<Cmd>, <Cmd>)
```

 From a syntactic standpoint, they extend both statements and expressions, as an identifier is an expression.



π automata semantics for π lib commands I

- A location $l \in Loc$ denotes a memory cell.
- Storable and Bindable sets denote the data that may be mapped to by identifiers and locations on the memory and environment respectively.
- Store = Id → Storable, Env = Loc → Bindable, Loc ⊆ Store,
 N ⊆ Loc, Bindable.
- Now the transition function is $\delta: L(G)^* \times L(G)^* \times Env \times Store \rightarrow Q$, and let $W \in String$, $C, V \in L(G)^*$, $S \in Store$, $E \in Env$, $B \in Bindable$, $I \in Loc$, $T \in Storable$, $X \in Exp >$, $M, M_1, M_2 \in Cmd >$, and



π automata semantics for π lib commands II

expression $S' = S/[I \mapsto N]$ means that S' equals to S in all indices but I that is bound to N,

$$\delta(Id(W)::C,V,E,S) = \delta(C,B::V,E,S),$$
where $E[W] = I$ and $S[I] = B$,
$$(17)$$

$$\delta(Assign(W,X)::C,V,E,S) = \delta(X::\#ASSIGN::C,W::V,E,S'), \tag{18}$$

$$\delta(\#ASSIGN::C,T::W::V,E,S) = \delta(C,V,E,S'), \tag{19}$$

where
$$E[W] = I$$
 and $S' = S/[I \mapsto T]$,

$$\delta(Loop(X,M)::C,V,E,S) = \delta(X::\#LOOP::C,Loop(X,M)::V,E,S), \tag{20}$$

$$\delta(\#LOOP :: C, Boo(true) :: Loop(X, M) :: V, E, S) = \delta(M :: Loop(X, M) :: C, V, E, S), \tag{21}$$

$$\delta(\#LOOP :: C, Boo(false) :: Loop(X, M) :: V, E, S) = \delta(C, V, E, S), \tag{22}$$

$$\delta(CSeq(M_1, M_2) :: C, V, E, S) = \delta(M_1 :: M_2 :: C, V, E, S).$$
(23)



π lib expressions in Python I

https://github.com/ChristianoBraga/BPLC/blob/master/python/pi.ipynb

```
class Statement:
    def __init__(self, *args):
        self.opr =args

def __str__(self):
    ret =str(self.__class__.__name__)+"("
    for o in self.opr:
        ret +=str(o)
    ret +=")"
    return ret
class Exp(Statement): pass
class ArithExp(Exp): pass
```



π lib expressions in Python II

```
class Num(ArithExp):
    def __init__(self, f):
        assert(isinstance(f, int))
        ArithExp.__init__(self,f)

class Sum(ArithExp):
    def __init__(self, e1, e2):
        assert(isinstance(e1, Exp) and isinstance(e2, Exp))

ArithExp.__init__(self, e1, e2)

...
```



π lib expressions in Python III

```
class BoolExp(Exp): pass
class Eq(BoolExp):
    def __init__(self, e1, e2):
        assert(isinstance(e1, Exp) and isinstance(e2, Exp))
        BoolExp.__init__(self, e1, e2)
...
```



π lib expressions in Python IV

```
exp =Sum(Num(1), Mul(Num(2), Num(4)))
print(exp)

Sum(Num(1)Mul(Num(2)Num(4)))
```



π lib expressions in Python V

```
_{1} \exp 2 = Mul(2, 1)
3 AssertionError Traceback (most recent call last)
4 <ipython-input-7-00fd40a79a54> in <module>()
5 \longrightarrow 1 \exp 2 = Mul(2, 1)
7 <ipython-input-5-42a82e58862f> in __init__(self, e1, e2)
       28 class Mul(ArithExp):
8
       29 def __init__(self, e1, e2):
10 --->30 assert(isinstance(e1, Exp) and isinstance(e2, Exp))
       31 ArithExp.__init__(self, e1, e2)
11
       32 class BoolExp(Exp): pass
12
13
4 AssertionError:
```



π automaton for π lib expressions I

```
1 ## Expressions
2 class ValueStack(list): pass
3 class ControlStack(list): pass
4 class ExpKW:
5 SUM ="#SUM"
6 SUB ="#SUB"
7 MUL = "#MUL"
8 EQ = "#EQ"
9 NOT = "#NOT"
```



π automaton for π lib expressions II

```
1 class ExpPiAut(dict):
     def __init__(self):
         self["val"] =ValueStack()
3
         self["cnt"] =ControlStack()
     def __evalSum(self, e):
5
         e1 =e.opr[0]
6
         e2 =e.opr[1]
         self.pushCnt(ExpKW.SUM)
         self.pushCnt(e1)
         self.pushCnt(e2)
     def pushCnt(self, e):
         cnt =self.cnt()
         cnt.append(e)
```



π automaton for π lib expressions III

```
1 ea =ExpPiAut()
2 print(exp)
3 ea.pushCnt(exp)
4 while not ea.emptyCnt():
5     ea.eval()
6     print(ea)
```



π automaton for π lib expressions IV

```
1 Sum(Num(1)Mul(Num(2)Num(4)))
2 {'val': [], 'cnt': ['#SUM', <__main__.Num object at 0x111851470>, <
                                      __main__.Mul object at 0x1118516d8>]
3 {'val': [], 'cnt': ['#SUM', <__main__.Num object at 0x111851470>, '#MUL'
                                      , <__main__.Num object at
                                      0x111851630>, <__main__.Num object
                                      at 0x1118516a0>]}
4 {'val': [4], 'cnt': ['#SUM', <__main__.Num object at 0x111851470>, '#MUL
                                      ', <__main__.Num object at
                                      0x111851630>]}
5 {'val': [4, 2], 'cnt': ['#SUM', <__main__.Num object at 0x111851470>, '#
                                      MUL'1}
6 {'val': [8], 'cnt': ['#SUM', <__main__.Num object at 0x111851470>]}
7 {'val': [8, 1], 'cnt': ['#SUM']}
8 {'val': [9], 'cnt': []}
```



π lib commands I

```
1 class Cmd(Statement): pass
2 class Id(Exp):
    def __init__(self, s):
3
         assert(isinstance(s, str))
         Exp.__init__(self, s)
6 class Assign(Cmd):
     def __init__(self, i, e):
7
         assert(isinstance(i, Id) and isinstance(e, Exp))
         Cmd.__init__(self, i, e)
10 class Loop(Cmd):
     def __init__(self, be, c):
11
12
         assert(isinstance(be, BoolExp) and isinstance(c, Cmd))
         Cmd.__init__(self, be, c)
13
14 class CSeq(Cmd):
     def __init__(self, c1, c2):
15
         assert(isinstance(c1, Cmd) and isinstance(c2, Cmd))
16
         Cmd.__init__(self, c1, c2)
17
```



π lib commands II

```
cmd =Assign(Id("x"), Num(1))
print(type(cmd))
print(cmd)
<class '__main__.Assign'>
Assign(Id(x)Num(1))
```



π automaton for π lib commands I

Environment, Location, Store and commands opcodes.

```
1 ## Commands
2 class Env(dict): pass
3 class Loc(int): pass
4 class Sto(dict): pass
5 class CmdKW:
    ASSIGN = "#ASSIGN"
    LOOP = "#LOOP"
```



π automaton for π lib commands II

 π automaton for commands extends the π automaton for expressions.

```
1 class CmdPiAut(ExpPiAut):
     def __init__(self):
         self["env"] =Env()
         self["sto"] =Sto()
         ExpPiAut.__init__(self)
     def env(self):
         return self["env"]
     def getLoc(self, i):
         en =self.env()
         return en[i]
LO
     def sto(self):
11
         return self["sto"]
12
     def updateStore(self, 1, v):
         st =self.sto()
         st[1] =v
```

π automaton for π lib commands III π semantics for assignment.

```
\begin{split} \delta(Assign(W,X) :: C,V,E,S) &= \delta(X :: \#ASSIGN :: C,W :: V,E,S'), \\ \delta(\#ASSIGN :: C,T :: W :: V,E,S) &= \delta(C,V,E,S'), \\ \text{where } E[W] &= I \text{ and } S' &= S/[I \mapsto T]. \end{split}
```

```
def __evalAssign(self, c):
    i =c.opr[0]
    e = c.opr[1]
    self.pushVal(i.opr[0])
    self.pushCnt(CmdKW.ASSIGN)
    self.pushCnt(e)

def __evalAssignKW(self):
    v =self.popVal()
    i =self.popVal()
    l =self.getLoc(i)
    self.updateStore(l, v)
```

π automaton for π lib commands IV

 π semantics for identifiers.

$$\delta(Id(W)::C,V,E,S) = \delta(C,B::V,E,S),$$
 where $E[W] = I$ and $S[I] = B$.

```
def __evalId(self, i):
    s =self.sto()
    l =self.getLoc(i)
    self.pushVal(s[1])
```



π automaton for π lib commands V

 π semantics for loop: recursive step.

```
\delta(Loop(X,M)::C,V,E,S) = \delta(X::\#LOOP::C,Loop(X,M)::V,E,S).
```

```
def __evalLoop(self, c):
    be =c.opr[0]
    bl =c.opr[1]
    self.pushVal(Loop(be, bl))
    self.pushVal(bl)
    self.pushCnt(CmdKW.LOOP)
    self.pushCnt(be)
```



π automaton for π lib commands VI

 π semantics for loop: basic steps.

```
\delta(\#LOOP :: C, Boo(true) :: Loop(X, M) :: V, E, S) = \delta(M :: Loop(X, M) :: C, V, E, S),\delta(\#LOOP :: C, Boo(false) :: Loop(X, M) :: V, E, S) = \delta(C, V, E, S).
```

computação

π automaton for π lib commands VII

 π semantics for command composition.

```
\delta(CSeq(M_1, M_2) :: C, V, E, S) = \delta(M_1 :: M_2 :: C, V, E, S).
```

```
def __evalCSeq(self, c):
    c1 =c.opr[0]
    c2 =c.opr[1]
    self.pushCnt(c2)
    self.pushCnt(c1)
```



π automaton for π lib commands VIII

Commands are now on the top of the food chain.

```
def eval(self):
1
          c =self.popCnt()
          if isinstance(c, Assign):
              self.__evalAssign(c)
4
          elif c ==CmdKW.ASSIGN:
5
              self.__evalAssignKW()
6
          elif isinstance(c, Id):
7
              self.__evalId(c.opr[0])
          elif isinstance(c, Loop):
              self.__evalLoop(c)
LO
          elif c == CmdKW.I.OOP:
11
              self.__evalLoopKW()
12
          elif isinstance(c, CSeq):
              self.__evalCSeq(c)
          else:
              self.pushCnt(c)
۱6
              ExpPiAut.eval(self)
```

IMP concrete syntax for expressions and commands

```
::= \langle cmd \rangle^*
⟨prog⟩
                            ::= 'while' '('\( bexp\)')' 'do' '\( cmd\) '\}'
⟨cmd⟩
                                    \langle id \rangle ':=' \langle exp \rangle | \langle cmd \rangle ';' \langle cmd \rangle
                            ::= \langle aexp \rangle \mid \langle bexp \rangle
\langle exp \rangle
\langle aexp \rangle
                            ::= \langle aexp \rangle '+' \langle term \rangle | \langle aexp \rangle '-' \langle term \rangle | \langle term \rangle
⟨term⟩
                           ::= \langle term \rangle '*' \langle factor \rangle \langle term \rangle '/' \langle factor \rangle \langle factor \rangle
⟨factor⟩
                           ::= '(' \( aexp\) ')' | \( digits\) | \( id\)
                            ::= \langle id \rangle \mid \langle boolean \rangle \mid ``` \langle bexp \rangle \mid \langle bexp \rangle \langle boolop \rangle \langle bexp \rangle
\langle bexp \rangle
                                    ⟨aexp⟩ ⟨iop⟩ ⟨aexp⟩
                          ::= '=' | '/\' | '\/'
⟨boolop⟩
                           ::= '<' | '>' | '<=' | '>='
\langle iop \rangle
```

Example: iterative factorial

```
while (~ (x = 0))
do {
    y := y * x;
    x := x - 1
};
y
```

