SetlCup Tutorial

Jonas Eilers

January 7, 2016

Contents

Chapter 1

Functionality

The Setlx-addition SetlCup is a LR-Parser-Generator based on JavaCup. The idea is to use a user given scanner- and parser-definition and create an AST out of a given input using the definitions.

In this document the needed syntax of the definitions is examined and the given output is evaluated.

A sample input file is divided into three Sections:

1. Commentpart
2. Scanner-Part
3. Parser-Part

At first the correct call of the program is discussed.

1.1 Using SetlCup

SetlCup has three different variants in which it can be called:

1. `setlx setlcup.stlx -p parser_scanner_file.stlx file_to_be_read.txt`

   With this call there will be no output for the user.

2. `setlx setlcup.stlx -p parser_scanner_file.stlx file_to_be_read.txt -d`

   With this call debugging is possible. It shows the different tables and states and the whole parsing progress. HINT: It is recommended to pipe the output into a file if you are using the "-d" option.

3. `setlx setlcup.stlx -h`

   With this call a little help will be showed, on how to call SetlCup correctly.

1.2 Comment-Part

In the comment-part everything which is written will not be used by the Program itself. It is adviced to comment your idea behind the parser and scanner structure in this section. The section is ended with the "%%%" symbol.

1.3 Scanner-Part

The scanner is responsible for checking whether the input file consists of the defined tokens. It can be written like this:

```
1 INTEGER := 1-9[0-9]*|0;
2 ASTERISK := \*;
3 WHITESPACE := [ ];
4 SKIP := ASTERISK | INTEGER | WHITESPACE;
```

- In line 1 the Token "INTEGER" is defined. Tokens are defined in the following way:
  token_name := regex ;
- In line 2 it is shown, that predefined tokens in Regular Expressions like "$*, +, ?, |, \{, \}, (,), \cdots$" need to be escaped.
- In Line 3 the "SKIP"-Token is shown. In some contexts tokens like Whitespaces are not needed. They can be skipped by defining the "SKIP"-Token with the tokens, which shall be skipped. Multiple tokens need to be seperated by a pipe "|".

1.4 Parser-Part

In this part the grammar-rules are defined with the following syntax:

```
rule_name ::= rule_element:id {: action_code :}
           | rule_element
           | {: action_code :}
           | ;
```

rule_name The rule_name is the name of the rule. It is possible to reference defined rules via their rule_name

rule_element The element can consist of multiple Tokens (defined in the scanner) and rule_names. Each can have an id, which is possible to be used in the action_code.

action_code The action_code is an optional part in a rule. It needs to be at the end of the rule it self. Each rule_element can have an action_code. In this action_code Setlx Code can be written. By using the variable "result" it is possible to pass values between rules. The id of the elements in the respective rule can be referred to by using its name.

| The pipe seperates the rule_elements.

1.5 Example

The first example shows a simple arithmetic grammar. The second example shows how a simple programming language can be parsed using SetlCup.

1.5.1 Arithmetic grammar

The arithmetic grammar and scanner is the following:

```


%%

SEMICOLON := ; ;
TIMES := \* ;
MINUS := - ;
DIVIDE := \\ ;
INTEGER := 0|[1-9][0-9]* ;
NEWLINE := \n ;
WHITESPACE := [ \t\v\n\r\s] ;
MOD := %;
PLUS := \+ ;
LPAREN := \( ;
RPAREN := \) ;
SKIP := WHITESPACE | NEWLINE ;

%%

expr_list ::= expr_list:l expr_part:part {: result := l + [part]; :}
            | expr_part:epart {: result := [epart]; :}
            ;
expr_part ::= expr:e SEMICOLON {: result := e; :}
            ;
expr ::= expr:e PLUS   prod:p {: result := Plus(e , p); :}
       | expr:e MINUS  prod:p {: result := Minus(e , p); :}
       | prod:p               {: result := p;      :}
       ;
prod ::= prod:p TIMES  fact:f {: result := Times(p , f); :}
       | prod:p DIVIDE fact:f {: result := Div(p , f); :}
       | prod:p MOD    fact:f {: result := Mod(p , f); :}
       | fact:f               {: result := f;      :}
       ;
fact ::= LPAREN expr:e_part RPAREN {: result := e_part ;   :}
       | INTEGER:n                 {: result := Integer(eval(n)); :}
       ;
```

A sample input file:

```
1 + 2 * 3 - 4;
1 + 2 + 3 + 4;
1 + ( 2 * 3 ) * 5 % 6;
```

The output AST:

```
ExprList([Minus(Plus(Integer(1), Times(Integer(2), Integer(3))),
    Integer(4)),
Plus(Plus(Plus(Integer(1), Integer(2)), Integer(3)), Integer(4)),
Plus(Integer(1), Mod(Times(Times(Integer(2), Integer(3)), Integer
    (5)), Integer(6)))])
```

1.5.2 Programming language grammar

The programming language grammar and scanner:

```

%%

SEMI := ; ;
TIMES := \* ;
MINUS   := - ;
DIV    := \\ ;
MOD := %;
PLUS   := \+ ;
LPAR := \( ;
RPAR := \) ;
LBRACE := \{ ;
RBRACE := \} ;
COMMA := , ;
ASSIGN := = ;
EQ := == ;
NE := != ;
LT := < ;
GT := > ;
LE := <= ;
GE := >= ;
AND := && ;
OR := \|\| ;
NOT := ! ;
FUNCTION := function ;
RETURN := return ;
IF := if ;
ELSE := else ;
WHILE := while ;
FOR := for ;
PRINT := print ;
QUIT := exit ;
STRING := \"(?:\\.|[^\"])*\" ;
NEWLINE := \n ;
COMMENTS := //[^\n]* ;
WHITESPACE := [ \t\v\n\r\s] ;
SKIP := WHITESPACE | NEWLINE | COMMENTS ;
INTEGER := 0|[1-9][0-9]* ;
DECIMAL := 0\.[0-9]+|[1-9][0-9]*\.[0-9]+ ;
ZID := [a-zA-Z_][a-zA-Z0-9_]* ;


%%
program ::= dfnStmntList:d {: result := Program(d); :};

dfnStmntList
    ::= definition:d dfnStmntList:dl {: result := [d] + dl; :}
     |  statement:stmts  dfnStmntList:dsl {: result := [stmts] +
        dsl; :}
     | {: result := []; :}
     ;

definition ::= FUNCTION ZID:function_name LPAR paramList:param_list
     RPAR LBRACE stmntList:statement_list RBRACE
         {: result := Function(function_name, param_list,
             statement_list);:}
     ;

stmntList
```

```
    ::= statement:s stmntList:sl {: result := [s] + sl ; :}
     |  {: result := []; :}
     ;

statement
    ::= assignment:a SEMI {: result := Ass(a); :}
     |  PRINT LPAR printExprList:printexpr_list RPAR SEMI       {: result := Print(printexpr_list); :}
     |  IF LPAR boolExpr:b RPAR LBRACE stmntList:st_list1 RBRACE {: result := If(b, st_list1); :}
     |  WHILE LPAR boolExpr:b RPAR LBRACE stmntList:st_list2 RBRACE {: result := While(b, st_list2); :}
     |  FOR LPAR assignment:i_a SEMI boolExpr:b SEMI assignment:e_a RPAR LBRACE stmntList:st_list3 RBRACE {: result := For(i_a, b, e_a, st_list3);  :}
     |  RETURN expr:e SEMI {: result := Return(e); :}
     |  RETURN SEMI {: result := Return(); :}
     |  expr:e SEMI {: result := Expr(e); :}
     |  QUIT SEMI {: result := Exit(); :}
     ;

printExprList
    ::= printExpr:p COMMA nePrintExprList:np {: result := [p] + np ; :}
     |  printExpr:p {: result := [p]; :}
     |  {: result := []; :}
     ;

nePrintExprList
    ::= printExpr:p {: result := [p]; :}
     |  printExpr:p COMMA nePrintExprList:np {: result := [p] + np ; :}
     ;

printExpr
    ::= STRING:string  {: result := PrintString(string); :}
     |  expr:e   {: result := e; :}
     ;

assignment
    ::= ZID:id ASSIGN expr:e {: result := Assign(id, e); :}
     ;

paramList
    ::= ZID:id COMMA neIDList:nid {: result := [id] + nid ; :}
     |  ZID:id {: result := [id] ; :}
     |  {: result := []; :}
     ;

neIDList
    ::= ZID:id COMMA neIDList:nid {: result := [id] + nid ; :}
     |  ZID:id   {: result := [id] ; :}
     ;


boolExpr
    ::= expr:lhs EQ expr:rhs   {: result := Equation(lhs,rhs); :}
     |  expr:lhs NE expr:rhs   {: result := Inequation(lhs,rhs); :}
     |  disjunction:lhs EQ disjunction:rhs   {: result := Equation(lhs,rhs); :}
     |  disjunction:lhs NE disjunction:rhs   {: result := Inequation(lhs,rhs); :}
```

```
       |   expr:lhs LE expr:rhs   {: result := LessOrEqual(lhs,rhs); :}
       |   expr:lhs GE expr:rhs   {: result := GreaterOrEqual(lhs,rhs); :}
       |   expr:lhs LT expr:rhs   {: result := LessThan(lhs,rhs); :}
       |   expr:lhs GT expr:rhs   {: result := GreaterThan(lhs,rhs); :}
       |   disjunction:d {: result := d; :}
       ;
disjunction
    ::= disjunction:d OR conjunction:c {: result := Disjunction(d,c); :}
       |   conjunction:c {: result := c; :}
       ;
conjunction
    ::= conjunction:c AND boolFactor:f {:result := Conjunction(c,f); :}
       | boolFactor:f {: result := f; :}
       ;
boolFactor
    ::= LPAR boolExpr:be_par RPAR {:   result := be_par; :}
       | NOT boolExpr:e {: result := Negation(e); :}
       ;


expr ::= expr:e PLUS   prod:p {: result := Sum(e,p); :}
        |  expr:e MINUS  prod:p {: result := Difference(e,p); :}
        |  prod:p               {: result := p;     :}
        ;
prod ::= prod:p TIMES  fact:f {: result := Product(p,f); :}
        |  prod:p DIV fact:f {: result := Quotient(p,f); :}
        |  prod:p MOD    fact:f {: result := Mod(p,f); :}
        |  fact:f               {: result := f;     :}
        ;
fact ::= LPAR expr:e_par RPAR {: result := e_par;   :}
        |  INTEGER:n              {: result := Integer(eval(n));   :}
        |  DECIMAL:d                {: result := Decimal(eval(d)); :}
        |  ZID:id_1 LPAR exprList:el RPAR {: result := FunctionCall(id_1,el); :}
        | ZID:id_2 {: result := Variable(id_2); :}
        ;

exprList
    ::= expr:e COMMA neExprList:el {: result := [e] + el; :}
       |  expr:e {: result := [e]; :}
       |  {: result := []; :}
       ;

neExprList
    ::= expr:e COMMA neExprList:el {: result := [e] + el; :}
       |  expr:e {: result := [e]; :}
       ;
```

A sample input file:

```
function factorial(n) {
    if (n == 0) {
               return 1;
    }
    return n * factorial(n - 1);
}

print("Calculation of factorial for i = 1 to 9");
for (i = 0; i < 10; i = i + 1) {
    print(i, "! = ", factorial(i));
```

```
}
print();
```

The output AST:

```
Program([Function("factorial", ["n"], [If(Equation(Variable("n"),
    Integer(0)), [Return(Integer(1))]), Return(Product(Variable("n
    "), FunctionCall("factorial", [Difference(Variable("n"),
    Integer(1))])))]), Print([PrintString("Calculation of factorial
     for i = 1 to 9")]), For(Assign("i", Integer(0)), LessThan(
    Variable("i"), Integer(10)), Assign("i", Sum(Variable("i"),
    Integer(1))), [Print([Variable("i"), PrintString("! = "),
    FunctionCall("factorial", [Variable("i")])])]), Print([])])
```