# FCW 1x

# Übung zu Formale Sprachen, Compiler- und Werkzeugbau 1

# WS 2016/17, Übung 4

Abgabetermin: in der KW 50

|   | Gr. 1, Dr. H. Dobler<br>Gr. 2, Dr. G. Kronberger | Name   | Aufwand in h |
|---|--|--------|--------------|
| _ | GI. 2, Dr. G. Kronberger                         | Punkte | Übungsleiter |

Im Moodle-Kurs finden Sie unter Flex und Bison zwei ZIP-Dateien mit den GNU-Implementierungen von lex u. yacc, für Windows. Bei UNIX sind beide Werkzeuge üblicherweise bereits enthalten oder können leicht nachgeladen werden.

### 1. MiniC: Scanner und Parser mit lex und yacc

(8 Punkte)

MiniC ist eine kleine Teilmenge von C, angelehnt an MiniPascal. Unten links ist ein einfaches Programm zur Berechnung des Satzes von Pythagoras dargestellt, rechts die Grammatik von MiniC (die Sie auch im Moodle-Kurs in der Datei *MiniC.syn* finden):

```
"void" "main" "(" ")" "{"
void main() {
                              MC =
  int a, b, cs;
                                         [ VarDecl ]
  scanf(a);
                                        StatSeq
  scanf(b);
                                         "}".
                              VarDecl = "int" ident { "," ident } ";" .
 cs = (a * a) + (b * b);
 printf(cs);
                              StatSeq = Stat { Stat } .
                                        [ ident "=" Expr
                              Stat =
                                           "scanf" "(" ident ")"
                                          "printf" "(" Expr ")"
                                         ] ";"
                                        Term { ( "+" | "-" ) Term } .
                              Expr =
                                        Fact { ( "*" | "/" ) Fact } .
                              Term =
                                        ident | number | "(" Expr ")"
                              Fact =
```

Erzeugen Sie mit lex einen lexikalischen Analysator (*scanner*) und mit yacc einen Syntaxanalysator (*parser*) für MiniC und bauen Sie daraus ein Programm für die Analyse von MiniC-Programmen.

### 2. MiniCpp: Scanner und Parser mit lex und yacc

(10 + 6 Punkte)

Wir werden uns intensiver mit der etwas größeren Sprache MiniCpp beschäftigen, mit der man auch etwas anspruchsvollere Programme schreiben kann, z. B. für das Sieb des Erathostenes:

```
void Sieve(int n); // declaration
void main() {
  int n;
  cout << "n > ";
  cin >> n;
  if (n > 2)
    Sieve(n);
} // main
void Sieve(int n) { // definition
  int col, i, j;
  bool *sieve = 0;
  sieve = new bool[n + 1];
  i = 2;
  while (i \le n) {
    sieve[i] = true;
  } // while
```

```
cout << 2 << " ";
  col = 1;
  i = 3;
  while (i \le n) {
    if (sieve[i]) {
      if (col == 10) {
        cout << endl;</pre>
        col = 0;
      } // if
      col++;
      cout << i << " ";
      j = i * i;
      while (j \le n) {
        sieve[j] = false;
         j = j + 2 * i;
      } // while
    } // if
    i = i + 2i
  } // while
  delete[] sieve;
} // Sieve
```

Hier die Grammatik für MiniCpp, die Sie im Moodle-Kurs auch in der Datei MiniCpp.syn finden:

```
{ ConstDecl | VarDef | FuncDecl | FuncDef } .
MiniCpp =
ConstDecl =
                  'const' Type ident Init ';' .
                 '=' (false | true | number ) .
Init =
                 Type [ '*' ] ident [ Init ]
VarDef =
                 { ',' [ '*' ] ident [ Init ] } ';' .
FuncDecl = FuncHead ';' .
FuncDef =
                FuncHead Block .
FuncHead = Type [ '*' ] ident '(' [ FormParList ] ')' .
FormParList = ( 'void' |
                     Type [ '*' ] ident [ '[' ']' ]
                     { ',' Type [ '*' ] ident [ '[' ']' ] } ) .
                  'void' | 'bool' | 'int'
Type =
                  '{' { ConstDecl | VarDef | Stat } '}' .
Block =
                  ( IncStat | DecStat | AssignStat
Stat =
                    CallStat | IfStat
                    WhileStat | BreakStat
                    InputStat | OutputStat | DeleteStat | ReturnStat
                    Block
                  | ';'
IncStat =
DecStat =
                 ident '++' ';' .
                 ident '--' ';'
AssignStat = ident [ '[' Expr ']' ] '=' Expr ';' .

CallStat = ident '(' [ ActParList ] ')' ';' .

ActParList = Expr { ',' Expr } .

IfStat = 'if' '(' Expr ')' Stat [ 'olgo' Stat
IfStat = 'if' '(' Expr ')' Stat [ 'else' Stat ] .
WhileStat = 'while' '(' Expr ')' Stat .
BreakStat = 'break' ';' .
InputStat = 'cin' '>>' ident ';' .
OutputStat = 'cout' '<<' ( Expr | string | 'endl' )</pre>
{ '<<' ( Expr | string
DeleteStat = 'delete' '[' ']' ident ';' .</pre>
                   { '<<' ( Expr | string | 'endl' ) } ';' .
ReturnStat = 'return' [ Expr ] ';' .
Expr = OrExpr .

OrExpr = AndExpr { '|| 'AndExpr } .

AndExpr RelExpr { '&&' RelExpr } .

RelExpr = SimpleExpr
                 [ ( '==' | '!=' | '<' | '<=' | '>' | '>=' )
                    SimpleExpr ] .
SimpleExpr = [ '+' | '-' ]
                 Term { ( '+' | '-' )
                                                          Term
Term = NotFact =
                 NotFact { ( '*' | '/' | '%' ) NotFact } .
                 [ '!' ] Fact .
Fact =
                    'false' | 'true'
                  number
                  | ident [ ( '[' Expr
                                                         ']')
                            | ( '(' [ ActParList ] ')' )
                    'new' Type '[' Expr ']'
                  | '(' Expr ')' .
```

- a) Erzeugen Sie mit lex/flex einen lexikalischen Analysator (*scanner*) und mit yacc/bison einen Syntaxanalysator (*parser*) für MiniCpp und bauen Sie daraus ein Programm für die lexikalische und syntaktische Analyse von MiniCpp-Programmen.
- b) Erweitern Sie Ihre Grammatik aus a) zu einer ATG, sodass der (statische) Funktionsaufrufgraph des analysierten Programms erstellt wird. Gehen Sie so vor, dass von der ATG eine Textdatei (.gv) für *GraphViz* (www.graphviz.org) erzeugt wird, die mit *GVEdit* (oder mit *dot.exe* direkt oder über www.webgraphviz.com) in eine Abbildung umgesetzt werden kann.

# 1 MiniC: Scanner und Parser mit lex und yacc

### Lösungsidee

Als Erstes wird der Scanner mit Hilfe von *lex* erstellt, wobei hier alle Token abgebildet werden müssen.

Im zweiten Schritt wird dann der Parser mit Hilfe der gegeben Grammatik erstellt.

#### Sourcecode

```
_____ ../Source/MiniC.l __
1 /*MiniC.l:
  Lex/Flex description for MiniC.
6 %{
   #include "MiniC.tab.h" /*generated with "bison -d MiniC.y"
   extern int yylval; /*lexical attribute for current token
   extern int yylineno; /*current line number, initalized with 1
11 %}
13 %%
14
15 [ \t]+
          {;}
                        /*ignore white space: blanks and tabs
                                                                   */
16
                        /*skip C comments
               int prevCh = 0, ch = input();
18
               for(;;) {
                 if (ch == EOF)
20
                  break; /*error: end of file within comment
                                                                  */
21
                 if (ch == ' \n')
22
23
                   yylineno++;
                 else if (prevCh == '*' && ch == '/')
24
                  break; /*ok: correct end of comment
                                                                  */
25
                 prevCh = ch;
                 ch = input();
               } /*for*/
28
             { return INT;
31 int
             { return MAIN;
                             }
32 main
            { return PRINTF; }
33 printf
            { return SCANF; }
34 scanf
35 void
             { return VOID;
37 [a-zA-Z][a-zA-Z0-9]* { return IDENT; }
             { yylval = atoi(yytext); return NUMBER; }
39 [0-9]+
```

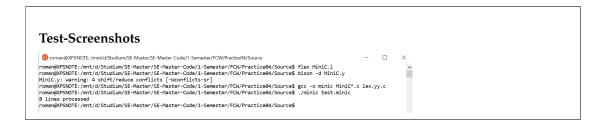
```
{ yylineno++; }
42
            { return yytext[0]; } /*return all other chars
                                  as tokens to parser: '+', ... */
46 %%
48 int yywrap() {
49 return 1; /*on end of input: no further files to scan
                                                                */
50 } /*yywrap*/
52 /* End of MiniC.l
53 ===========*/
                          ____ ../Source/MiniC.y __
1 %{ #include <stdio.h>
   #include <stdlib.h>
   extern int yylineno;
   extern int yylval;
5
6 %}
8 %token INT
9 %token MAIN
10 %token PRINTF
11 %token SCANF
12 %token VOID
14 %token IDENT
15 %token NUMBER
16 %%
18 VOID MAIN '(' ')' Block
19
20 Block:
'{' OptVarDecl StatSeq '}'
22 ;
23 OptVarDecl:
24 /* EPS */
INT Idlist ';'
  ;
28 Idlist:
  IDENT
30 | Idlist ',' IDENT
33 StatSeq:
```

```
/*EPS*/
   Stat
  | StatSeq Stat
36
37
39 Stat:
40
  | IDENT '=' Expr ';'
41
  | SCANF '(' IDENT ')' ';'
  | PRINTF '(' IDENT ')' ';'
44
45
46 Expr:
                  \{ \$\$ = \$1; /*default*/ \}
    Term
47
    | Expr '+' Term \{ \$\$ = \$1 + \$3;
                                            }
  50
51
52 Term:
              \{ \$\$ = \$1; /*default*/ \}
  | Term '*' Fact \{ \$\$ = \$1 * \$3;
  | Term '/' Fact \{ \$\$ = \$1 / \$3; \}
                                            }
  ;
57
58 Fact:
    IDENT
                 \{ \$\$ = \$1; /*default*/ \}
  NUMBER
  | '(' Expr ')' { $$ = $2;
62 ;
63 %%
64 extern FILE * yyin;
65 extern int yylineno;
67 int yyerror(char *message) {
printf("ERROR %s in line %d\n", message, yylineno);
  return 0;
70 }
71 int main(int argc, char ** argv) {
  if(argc > 1) {
    FILE *fin = fopen(argv[1],"r");
73
     if(fin != NULL){
74
      yyin = fin;
75
     } else {
       printf("ERROR: file not found");
77
       exit(-1);
78
     }
79
    }
    yyparse();
```

```
printf("%d lines processed\n",yylineno);
return 0;
}

.../Source/test.minic

void main() {
   int a, b, cs;
   scanf(a);
   scanf(b);
   cs = (a * a) + (b * b);
   printf(cs);
}
```



# 2 MiniCpp: Scanner und Parser mit lex und yacc

### a) Lexikalischen Analysator und Syntaxanalysator

### Lösungsidee

Analog zu Aufgabe-1, wobei hier die erweiterte Grammatik verwendet wird.

#### Sourcecode

```
17 %}
19 %%
20
21 [ \t]+
            { ; }
                           /*ignore white space: blanks and tabs
                                                                            */
23 "/*"
                          /*skip C comments
                 int prevCh = 0, ch = input();
24
                 for(;;) {
                   if (ch == EOF)
                     break; /*error: end of file within comment
                                                                            */
27
                   if (ch == ' \n')
                     yylineno++;
29
                   else if (prevCh == '*' && ch == '/')
30
                     break; /*ok: correct end of comment
                                                                           */
31
                   prevCh = ch;
                   ch = input();
                 } /*for*/
35
37 11 / / 11
                           /*skip C comments
                 int ch = input();
38
                 for(;;) {
39
                   if (ch == EOF)
                     break; /*error: end of file within comment
42
                   if (ch == ' \n') {
43
                     yylineno++;
                     break;
45
                   }
46
                   ch = input();
47
                 } /*for*/
49
             { return VOID;
51 Void
             { return INT;
                               }
52 int
             { return BOOL;
53 bool
54 string
             { return STRING; }
55
             { return COUT;
56 cout
57 cin
             { return CIN;
                               }
             { return ENDL;
58 endl
             { return IDENT; }
60 ident
61 number
             { return NUMBER; }
             { return TRUE;
62 true
             { return FALSE; }
_{63} false
65 const
             { return CONST; }
```

```
{ return IF;
67 if
           { return ELSE;
68 else
           { return WHILE; }
69 While
           { return BREAK; }
70 break
           { return RETURN; }
71 return
73 new
           { return NEW;
74 delete { return DELETE; }
77 " | | "
           { return OR;
78 "&&"
           { return AND;
79 ">>"
           { return SHIFTR;
           { return SHIFTL;
80 " << "
81 "!="
           { return NOTEQUAL;
82 "=="
           { return EQUAL;
83 "<="
           { return SMALLEREQ; }
84 ">="
           { return BIGGEREQ;
85 "++"
           { return PLUSPLUS;
86 "--"
           { return MINUSMINUS; }
88 [a-zA-Z][a-zA-Z0-9]* {yylval = strdup(yytext); return IDENT; }
90 \\"[^\"]*\" { ; return STRING; }
             { ; return NUMBER; }
92 [0-9]+
93
94 \n
             { yylineno++; }
             { return yytext[0]; } /*return all other chars
                                    as tokens to parser: '+', ... */
99 %%
100
int yywrap() {
  return 1; /*on end of input: no further files to scan
103 } /*yywrap*/
104
105 /* End of MiniC.l
106
                         _{-} ../Source/minicpp/MiniCpp.y _{-}
 1 %{ #include <stdio.h>
     #include <stdlib.h>
     //change output type to char *
     #ifndef YYSTYPE
       # define YYSTYPE char*
```

```
#endif
     extern int yylineno;
     extern char* yylval;
10
12
     //variable definitions
     //=========
13
     FILE *outputFile = NULL;
14
     char *methods[255]; //array of
                          // method calls per method
16
                          //NOTE: should be a dynamic array
17
     int cur = 0; //current index in arry
18
     int i=0; //variable used for loops
20 %}
21
_{22} %token VOID
23 %token INT
24 %token BOOL
25 %token STRING
27 %token COUT
28 %token CIN
29 %token ENDL
31 %token IDENT
32 %token NUMBER
33 %token TRUE
_{34} %token FALSE
36 %token CONST
38 %token IF
39 %token ELSE
40 %token WHILE
41 %token BREAK
42 %token RETURN
44 %token NEW
45 %token DELETE
47 %token OR
48 %token AND
49 %token SHIFTR
50 %token SHIFTL
51 %token NOTEQUAL
52 %token EQUAL
53 %token SMALLEREQ
54 %token BIGGEREQ
55 %token PLUSPLUS
```

```
56 %token MINUSMINUS
59 %%
60 MiniCpp:
61 /*EPS*/
62 | MiniCpp ConstDecl
63 | MiniCpp VarDef
64 | MiniCpp FuncDecl
  MiniCpp FuncDef
65
66
68 ConstDecl:
69 CONST Type IDENT Init ';'
70 ;
72 Init:
73 '=' TypeValue
74 ;
76 TypeValue:
77
    FALSE
   TRUE
78
79 NUMBER
  ;
81
82 VarDef:
83      Type OptStar IDENT OptInit VarDefList ';'
85
86 OptStar:
87 /*EPS*/
   * * *
   ;
89
91 OptInit:
92 /*EPS*/
93 | Init
  ;
94
96 VarDefList:
97 /*EPS*/
   | VarDefList ',' OptStar IDENT OptInit
100
101 FuncDecl:
FuncHead ';'
103
   ;
104
```

```
105 FuncDef:
       FuncHead Block {
                             //loop through all saved methods
107
                             for(i=0;i<cur;i++){</pre>
108
                                  //write to ouput file using dot syntax
                                  //eg. main -> test
110
                                  fprintf(outputFile,"%s -> %s\n",$1,methods[i]);
111
                                  free(methods[i]); //free string
112
                                                      //which was produced by strdup
                             }
114
                             //reset method inex
115
                             cur=0;
116
                         }
117
118
119
120 FuncHead:
       Type OptStar IDENT '(' FormParList ')' { $$ = $3;}
122
123
124 FormParList:
    /*EPS*/
     VOID
126
    | Type OptStar IDENT OptBrackets FormParListList
127
129
130 OptBrackets:
    /*EPS*/
131
    1 '[' ']'
132
133
134
135 FormParListList:
    /*EPS*/
    | FormParListList ',' Type OptStar IDENT OptBrackets
137
138
139
140 Type:
141
       VOID
    BOOL
142
    INT
143
145
146 Block:
     '{' BlockList '}'
147
148
149
150 BlockList:
    /*EPS*/
     | BlockList ConstDecl
    | BlockList VarDef
153
```

```
| BlockList Stat
   ;
156
157 Stat:
      IncStat
159
    DecStat
   AssignStat
160
   | CallStat
161
   IfStat
   WhileStat
163
   BreakStat
164
   InputStat
165
    | OutputStat
    DeleteStat
167
   ReturnStat
168
   Block
   | | ; |
171
172
173 IncStat:
   IDENT PLUSPLUS ';'
175
176
177 DecStat:
   IDENT MINUSMINUS ';'
179
180
181 AssignStat:
   IDENT OptExpr '=' Expr ';'
182
183
184
185 OptExpr:
   /*EPS*/
186
   | '[' Expr ']'
187
188
189 CallStat:
    IDENT '(' ')' ';'
                                            { methods[cur]=$1;cur++; /*save method name*/ }
   | IDENT '(' ActParList ')' ';'
                                            { methods[cur]=$1;cur++; /*save method name*/ }
192
193
194 ActParList:
   Expr ActParListList
195
196
197
198 ActParListList:
   /*EPS*/
   | ActParListList ',' Expr
200
201
```

```
203 IfStat:
  IF '(' Expr ')' Stat
   | IF '(' Expr ')' Stat ELSE Stat
208 WhileStat:
   WHILE '(' Expr ')' Stat
209
210
212 BreakStat:
   BREAK ';'
213
214
216 InputStat:
   CIN SHIFTR IDENT ';'
220 OutputStat:
    COUT SHIFTL Expr OutputStatList ';'
   | COUT SHIFTL STRING OutputStatList ';'
   COUT SHIFTL ENDL OutputStatList ';'
224
225
226 OutputStatList:
   /*EPS*/
   OutputStatList SHIFTL Expr
228
   OutputStatList SHIFTL STRING
   | OutputStatList SHIFTL ENDL
231
232
233 DeleteStat:
   DELETE '[' ']' IDENT ';'
235
236
237 ReturnStat:
   RETURN ';'
   RETURN Expr ';'
239
240
241 Expr:
242 OrExpr
243
245 OrExpr:
   AndExpr OrExprList
247
248
249 OrExprList:
   /*EPS*/
    | OrExprList OR AndExpr
```

```
252
254 AndExpr:
   RelExpr AndExprList
257
258 AndExprList:
   /*EPS*/
   AndExprList AND RelExpr
261
262
263 RelExpr:
    SimpleExpr
    | SimpleExpr ExprOp SimpleExpr
265
266
268 ExprOp:
    EQUAL
   NOTEQUAL
270
   1 < 1
271
   | '>'
272
   SMALLEREQ
273
   BIGGEREQ
274
275
276 SimpleExpr:
    OptSign Term
    OptSign Term SimpleExprList
278
279
280
281 OptSign:
   /*EPS*/
282
   1 + 1
   1 1 - 1
284
285
286 SimpleExprList:
   OptSign Term
288
289
290 Term:
NotFact TermList
   ;
292
293
294 TermList:
295 /*EPS*/
    | '*' NotFact
   | '/' NotFact
   | '%' NotFact
299
   ;
300
```

```
301 NotFact:
       Fact
     | '!' Fact
303
304
306 Fact:
       FALSE
307
    TRUE
308
   NUMBER
   IDENT
310
   | IDENT '[' Expr ']'
311
    | IDENT '(' ')
312
    | IDENT '(' ActParList ')'
    | NEW Type '[' Expr ']'
314
    | '(' Expr ')'
315
316
317
318 %%
319 extern FILE * yyin;
320 extern int yylineno;
322 int yyerror(char *message) {
    printf("ERROR %s in line %d\n", message, yylineno);
    return 0;
325 }
326 int main(int argc, char ** argv) {
327
       if(argc > 1) {
328
       FILE *fin = fopen(argv[1],"r");
329
       if(fin != NULL){
330
         yyin = fin;
331
       } else {
         printf("ERROR: file not found");
333
         exit(-1);
334
       }
335
     }
337
       //open grahviz output file
338
       outputFile = fopen("output.gv","w");
339
       if(outputFile == NULL){
           printf("ERROR: could not open output file output.gv");
341
           exit(-1);
342
       }
343
       //graphviz header
345
       fprintf(outputFile,"digraph G {");
346
     yyparse();
347
348
       //grphviz end
       fprintf(outputFile,"}");
349
```

```
fclose(outputFile);
      printf("%d lines processed\n",yylineno);
   return 0;
352
353 }
                          \_ ../Source/minicpp/test.minicpp \_
 void Sieve(int n); // declaration
 void main(void) {
    int n;
    cout << "n > ";
    cin >> n;
    if(n == 0)
      Sieve();
9 }
void Sieve(int n) { // definition
    int col, i, j;
    bool *sieve = 0;
12
    sieve = new bool[n + 1];
    i = 2;
    while (i \leq n) {
    sieve[i] = true;
16
     i++;
17
    } // while
    cout << 2 << " ";
    col = 1;
20
    i = 3;
21
    while (i <= n) {
22
     if (sieve[i]) {
23
        if (col == 10) {
24
         cout << endl;</pre>
25
          col = 0;
        } // if
27
        col++;
28
        cout << i << " ";
29
        j = i * i;
        while (j <= n) {
31
        sieve[j] = false;
32
         j = j + 2 * i;
        } // while
      } // if
35
     i = i + 2;
    } // while
   delete[] sieve;
39 } // Sieve
```



### b) ATG - Erstellung des Funktionsaufrufgraph

#### Lösungsidee

Um den Funktionsaufrufgraph muss die Grammatik um Aktionen erweitert werden. Da der erzeugte Analysator bottom-up Syntaxanalyse betreibt, müssen alle Methodenaufrufe, die sich innerhalb einer Methode befinden temporär gespeichert werden. Nachdem der Analysator dann bei **FuncDef** angekommen ist, steht die umgebende Methode fest und die gespeicherten Methoden können in die Ausgabedatei geschrieben werden.

#### Sourcecode

Die Änderungen am Quellcode dieser Aufgabe ist bereits in Aufgabe 2a enthalten.

```
../Source/minicpp/test2.minicpp _
void testA();
void testB();
3 void testC();
4 void testD();
5 void testE(int i);
6 void test();
7 void testA(){
    testB();
    testD();
10 }
11
void testB() {
      testC();
14 }
16 void test() {
17
    testA();
    testD();
18
19 }
21 void main() {
22 test();
23 testE(5);
24 }
26 //recursion and after main method
27 void testE(int i) {
```

```
28    if(i == 0)
29    return;
30    testE(i-1);
31 }
```

```
Test-Screenshots

oroman@XPSNOTE:/mmt/d/Studium/SE-Master/Code/1-Semester/FCW/Practice04/Source/minicpp

roman@XPSNOTE:/mmt/d/Studium/SE-Master/SE-Master-Code/1-Semester/FCW/Practice04/Source/minicpps flex MiniCpp.1

roman@XPSNOTE:/mmt/d/Studium/SE-Master/SE-Master-Code/1-Semester/FCW/Practice04/Source/minicpps bison -d MiniCpp.y

MiniCpp.y: warning: 2 shift/reduce conflicts [-Nconflicts-sr]

roman@XPSNOTE:/mmt/d/Studium/SE-Master/SE-Master-Code/1-Semester/FCW/Practice04/Source/minicpps gcc *.c -o minicpp

roman@XPSNOTE:/mmt/d/Studium/SE-Master/SE-Master-Code/1-Semester/FCW/Practice04/Source/minicpps ./minicpp test2.minicpp

31 lines processed

roman@XPSNOTE:/mnt/d/Studium/SE-Master/SE-Master-Code/1-Semester/FCW/Practice04/Source/minicpps
```

```
../Source/minicpp/output.gv _
```

```
1 digraph G {testA -> testB
2 testA -> testD
3 testB -> testC
4 test -> testA
5 test -> testD
6 main -> test
7 main -> testE
8 testE -> testE
9 }
```

