Alexandria University
Faculty of Engineering
Computer and Systems Engineering Dept.
CS 321 – Programming Languages and
Translators



# **Quafios Compiler Collection Semantic Analysis and Code Generation**

# Teamwork:

• Mostafa Abd El-Aziz (65).

# **Used Data Structures:**

In this phase, I modified the parser generator to allow the generation of "semantic actions" which is processed by the parser in run-time. The parser generator takes context free grammar with semantic actions as input, and produces a C file that contains the code for the parser, along with semantic actions (each semantic block is coded as a single function) and an array of pointers to the functions.

# • Representation of Production Rules:

```
/* if we have a rule A --> a b | c d (for example),
 * each of a, b, c, and d are called tokens, and
 * represented by token t. 'a b' and 'c d' are called
 * sub-productions, and represented by subprod t.
 * The whole production rule of A is represented by prod t.
 * Every production rule has its own FOLLOW as a linked list
 * of token t. Similarly, every sub-production has its own FIRST.
/* tokens */
typedef struct token {
    struct token *next;
#define TOKEN TYPE SYMBOL
                             0 /* terminal */
#define TOKEN_TYPE_EPSILON
                               1
#define TOKEN TYPE UNION
                               2
#define TOKEN TYPE ASSIGN
                              3
#define TOKEN TYPE PRODUCT 4 /* nonterminal */
#define TOKEN TYPE DOLARSIGN 5
    int type;
    char *data;
} token t;
/* sub-production rules */
typedef struct subprod {
    struct subprod *next;
    token t *tok head; /* first token */
    token_t *tok_tail; /* last token */
    token t *first; /* FIRST(subprod)*/
    int count;
} subprod t;
/* production rules */
typedef struct product {
    struct product *next;
    char *name; /* name of the production rule */
    subprod t *sub head; /* sub production rules - head */
    subprod_t *sub_tail; /* sub production rules - tail */
    token_t *follow; /* FOLLOW(product)*/
    int firdone; /* first has been computed */
    int foldone; /* follow has been computed */
    int tmp; /* used by ptable construction */
} product t;
```

```
/* list of all production rules in the program */
typedef struct {
    product_t *head;
    product_t *tail;
} all_products_t;
```

The following diagram illustrates the data structures used to represent production rules, and shows the hierarchy of the structures:

# all products t product t 1 subprod t 1 subprod t 2 subprod t n tokens tokens tokens FIRST() FIRST() FIRST() FOLLOW() product t2 subprod t 1 subprod t 2 subprod t n tokens tokens tokens FIRST() FIRST() FIRST() FOLLOW() product tn subprod t 1 subprod t 2 subprod t n tokens tokens tokens FIRST() FIRST() FIRST() FOLLOW()

In any production:  $P \rightarrow \{action\} \ X \ \{action\} \ Y \ \{action\} \ Z \ \{action\}; \ X, \ Y, \ and \ Z \ are encoded as ptent_t (parse table entry) elements. In this phase, I added code to parser generator so that <math>\{action\}$  blocks are also encoded as ptent\_t elements just like other tokens. In the case of  $\{action\}$  blocks, ptent\_t structure will hold the index of the function (in semantic function pointer array, see above) that performs the actions specified in the block.

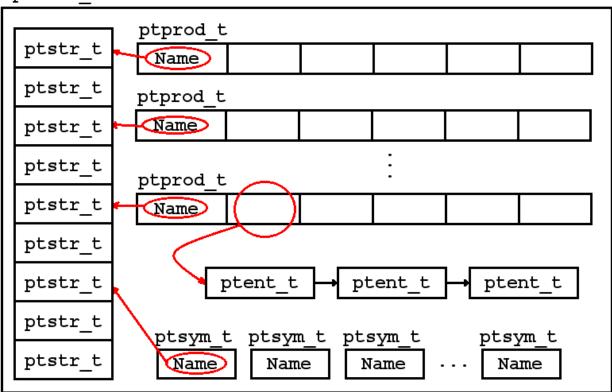
# • Representation of Parse table:

```
/* Parse table is a matrix of M production rules x N symbols.
 * Every row of the table is represented by a ptprod t element.
 * Every entry of the table is represented by a list of ptent t elements.
 * The data structure representing parse table (ptable t)
 * entirely contains all ptprod t instances, all ptent t
 * instances, and a list of terminal symbols of the language
 * (ptsym t). All strings representing the symbols and production
 * names are stored in a string pool (ptstr t).
 */
typedef struct ptsym {
    struct ptsym *next;
    int name ptr; /* pointer to string holding the name in the pool */
    /* these used for comparisons, not stored in binary file */
    int index:
    char *name;
} ptsym t;
typedef struct ptent {
    struct ptent *next;
#define PTENT_TYPE_TERMINATOR 0
#define PTENT_TYPE_PRODUCT 1
                                 1
#define PTENT_TYPE_SYMBOL
                                 2
                                3
#define PTENT TYPE EPSILON
    int type; /* type of the entry (pointer to production rule or symbol) */
    int ptr; /* value of the pointer (index of the prod rule or symbol) */
} ptent t;
typedef struct ptprod {
    /* row of the parse table */
    int name ptr; /* pointer to the string holding production name */
    int *subprods; /* array of entries of the row (N entries)
                      each entry corresponds to symbol
                      in the language */
    /* these are not stored in binary file */
    char *name;
} ptprod t;
typedef struct ptstr {
    struct ptstr *next;
    char *str;
} ptstr t;
typedef struct ptable {
    int symcount; /* N-1 */
```

```
int entcount;
int prodcount; /* M */
int chrcount;
ptsym_t *ptsyms; /* symbols of the language */
ptent_t *ptents; /* parts of production rules */
ptprod_t *ptprods; /* array of parse table rows */
ptstr_t *ptstrs; /* pool of strings */
} ptable t;
```

The following diagram illustrates the data structures used to represent parsing table, and shows the hierarchy of the structures:





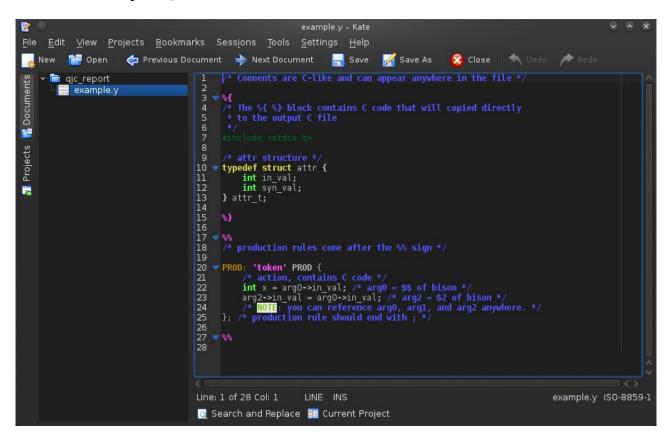
#### Attribute structure:

Each semantic actions has a data structure that holds its attributes. The contents and size of the structure are defined by the language specification itself. Ever production has a linked list that holds the attribute structures of all tokens (and the root production), the data structure is called (attribute frame) and is allocated once production is removed from the simulation stack to be replaced with ptent\_t elements.

# **Algorithms and Techniques:**

# • Parsing of context free grammar and semantic actions:

The parser generator takes CFG as input and produces a C file that contains the parser, parse table, and semantic actions as output. The syntax of the input CFG is like bison syntax, and is illustrated below:



## Processing of semantic actions on run-time:

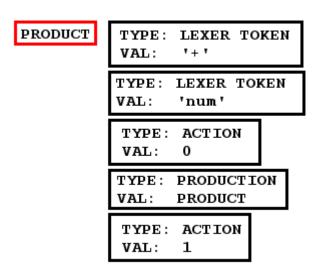
The following example shows how semantic actions are compiled along with the generated parser and the generated parse table.

```
PRODUCT: '+' 'num' {
    /* calculate sum */
    arg3->in_val = arg0->in_val + atoi(curlexeme);
} PRODUCT {
    /* return result */
    arg0->syn_val = arg3->syn_val;
};
```

In the output C file, two functions are generated for the two semantic actions above, and an array of pointers to those functions is created.

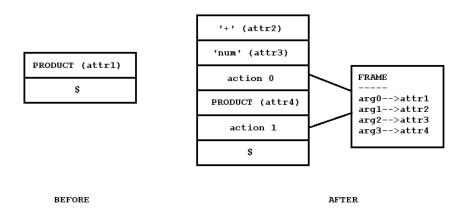
```
void func0() {
attr_ptr_t *frame = get attribute frame of current production;
arg0 = get attribute structure 0 from the frame;
arg1 = get attribute structure 1 from the frame;
arg2 = get attribute structure 2 from the frame;
arg3 = get attribute structure 3 from the frame;
 /* calculate sum */
arg3->in_val = arg0->in_val + atoi(curlexeme);
void func1() {
 attr ptr t *frame = get attribute frame of current production;
 arg0 = get attribute structure 0 from the frame;
arg1 = get attribute structure 1 from the frame;
arg2 = get attribute structure 2 from the frame;
arg3 = get attribute structure 3 from the frame;
 /* return result */
arg0->syn_val = arg3->syn_val;
void (*func calls[2]) () = {
_func0,
 _func1
```

The parse table will be generated as a byte array. The byte array is like a binary file that contains names of tokens, productions, ptent\_t entries, etc... Please refer to the report of last phase to know the format of the parse table file. The ptent\_t structures of PRODUCT production will be as follows:

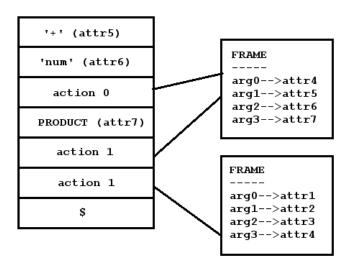


When the parser begins, a "\$" and PRODUCT are pushed into the stack. An attribute structure is allocated for the PRODUCT entry. When product is removed from the stack, it is replaced with the five pten\_t structures specified above.

Each pten\_t, when inserted into stack, is allocated an attribute structure (unless it is an action). Once PRODUCT is removed and replaced, an attribute frame is allocated with arg0 = attribute structure of the removed PRODUCT, and arg1 to arg3 are equal to the new attribute structures of the pten\_t entries. The frame is accessible by all the semantic actions that derived from PROOUCT when it was removed. The following diagram shows what happens to the stack:



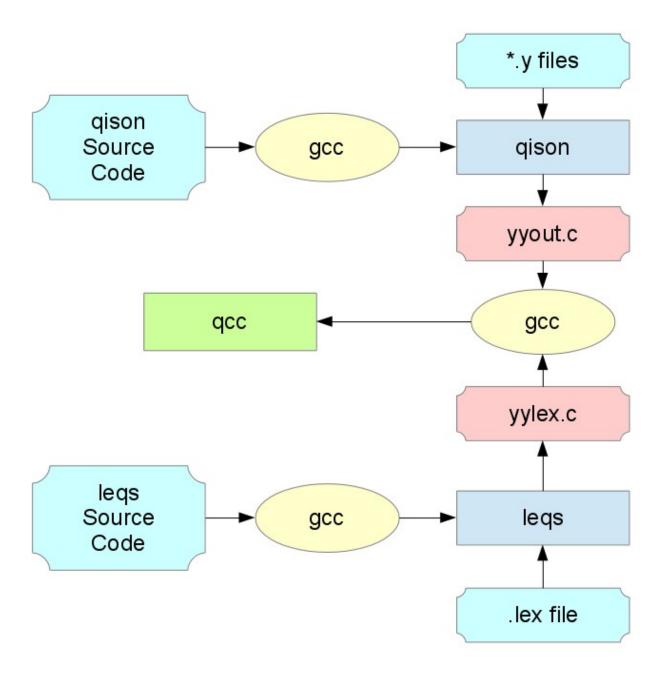
When '+' is matched, the '+' entry is removed from stack, but attr2 is not deallocated beacause it is still referenced by the frame that is associated with action 0 and action 1. The same happens when 'num' is matched. action 0 is then executed: the address of \_func0 is fetched from the pointer array "func\_calls" and a far call is performed. When PRODUCT is removed from stack, it is replaced with the same 5 tokens, but a new frame is allocated this time, and attr4 is inherited to the new frame as its arg0.



The parser keeps executing the actions that way till it terminates.

# Makefile:

The following figure shows the stages of building a compiler using Quafios compiler compiler. The compiler compiler consists of a lexer generator (leqs) and a parser generator (qison).



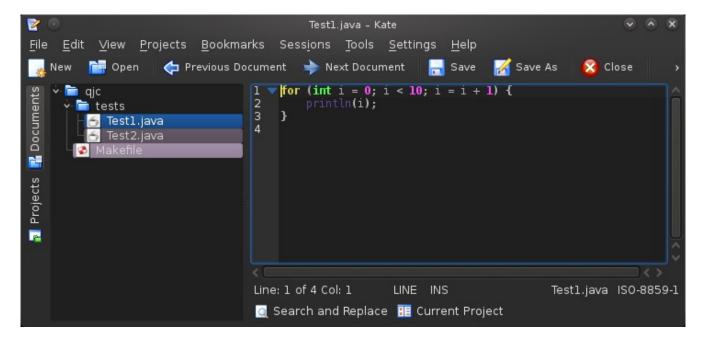
# **Quafios Java Compiler:**

Using Quafios compiler generator, I managed to make a compiler for subset of Java programming language that supports the following:

- Primitive Types (int, float, boolean)
- Arithmetic operations (+, -, \*, /, %)
- Logical operations (&&, ||, !)
- Relational operations (>, >=, <, <=, ==, !=)</li>
- If-else statements
- while loops
- Bitwise operations (&, |, ~) (bonus)
- boolean expressions (bonus)
- for loops (bonus)

The compiler maintains a symbol table to hold the names and types of variables in the source program. The compiler contains routines for emitting Java bytecode and routines for handling semantic errors and error reporting. Here are some code examples:

# Example 1:



## Example 2:

```
Test2.java - Kate
                                                                                                        V A
    <u>E</u>dit <u>V</u>iew <u>P</u>rojects <u>B</u>ookmarks Sess<u>i</u>ons <u>T</u>ools <u>S</u>ettings <u>H</u>elp
          Open
                       Previous Document
                                              Next Document
                                                                      - Save
                                                                                 🚜 Save As
                                                                                                 🔀 Close
    √- 🛅 qjc
                                             println(" num\tfact");
Documents
                                             println("----");

↓ lests

                                             for (int i = 0; i \le 10; i = i + 1) {
             Test2.java
                                                 int fact = 1;
        Makefile
                                                 int j = 1;
while (j <= i) {
                                        8
Projects
                                        9
                                       10
                                                 }
                                                 print(" ");
                                       11
                                       12
                                                 if (i > 5) {
                                       13
                                                     print("l");
                                       14
                                                 } else {
                                                     print("m");
                                       16
                                                 print("\t");
println(fact);
                                       17
                                             }
                                       19
                                       20
                                       Line: 10 of 20 Col: 6
                                                                LINE INS
                                                                                           Test2.java ISO-8859-1
                                       🔾 Search and Replace 🔢 Current Project
```

# Quafios C Compiler (bonus):

Using Quafios compiler compiler, I managed to write a compiler for subset of C programming language that supports the following:

- Arithmetic operations (+, -, \*, /, %)
- Logical operations (&&, ||, !)
- Relational operations (>, >=, <, <=, ==, !=)</li>
- If-else statements
- while loops
- Bitwise operations (&, |, ~)
- boolean expressions
- for loops
- Variable scope
- Global variables
- Function (and parameters) definition
- Function calls
- Parameter passing (by value and by reference)
- C arrays

The C compiler generates i386 assembly code for GNU/Linux operating system. The generated assembly code is assembled by GNU assembler (as), then linked with glibc and other necessary libraries using GNU linker/loader (ld). The generated binary is a 32-bit binary ELF file that can run on any GNU/Linux system on x86.

References used: ANSI standard specification for C language (1989/1990).

Following are screen-shots for 5 example programs compiled by qcc.

#### Example 1:

```
W
                                                                  test1.c - Kate
File Edit View Projects Bookmarks Sessions Tools Settings Help
📈 New 📄 Open
                        👉 Previous Document \Rightarrow Next Document 📙 Save
                                                                                    Maye As
                                                                                                     Close
                                          l ▼ int fact(int f) {
      tests
P Documents
                                                   int res = 1;
while (f != 0) {
         test1.c
         test2.c
                                                        res = res * f;
f = f - 1;
                                          4
           test4.c
                                                    return res;
Projects
                                            ▼ int main() {
                                                   int x;
for (x = 0; x < 10; x = x+1) {
    printf("fact(%d) = %d\n", x, fact(x));</pre>
0
                                                    return 0;
                                         16
17
                                         Line: 1 of 17 Col: 1
                                                                  LINE INS
                                                                                                                              test1.c ISO-8859-1
                                         🧕 Search and Replace 🔢 Current Project
```

#### Example 2:

```
W
                                                            test2.c - Kate
File Edit View Projects Bookmarks Sessions Tools Settings Help
                      👉 Previous Document 🗼 Next Document
                                                                 - Save
                                                                            Maye As
                                                                                           Close
   v 🛅 tests
P Documents
                                      2
                                        void conserve(int init) {
                                              int y;
if (init) {
          test2.c
                                              y = 5;
} else {
Projects
                                               printf("Value of y: %d\n", y);
                                     12
13
14
                                        ▼int main() {
0
                                               conserve(0);
                                               conserve(1);
                                              conserve(0);
                                               conserve(0);
                                               return 0;
                                     Line: 1 of 19 Col: 1
                                                            LINE INS
                                                                                                                  test2.c ISO-8859-1
                                     🧕 Search and Replace 🔢 Current Project
```

#### Example 3:

```
W
                                                                  test3.c - Kate
File Edit View Projects Bookmarks Sessions Tools Settings Help
📈 New 📄 Open
                        👉 Previous Document \Rightarrow Next Document
                                                                         - Save
                                                                                     Maye As
                                                                                                      Close
      i tests
P Documents
                                            ▼int main() {
          testl.c
                                                  int y;
printf("Enter integer: ");
scanf("%d", &y);
printf("Read value: %d\n", y);
                                                   return 0:
                                              }
Projects
0
                                         Line: 9 of 9 Col: 1
                                                                  LINE INS
                                                                                                                              test3.c ISO-8859-1
                                         🎑 Search and Replace 🔢 Current Project
```

```
File Edit View Bookmarks Settings Help

make[1]: Leaving directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/leqs'
make -C qison
make -C qison
makell: Entering directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/qison'
./bytearr.sh sim.c > sim arr.c
cc -pedantic -Werror -include config.h -o qison ./product.c ./sim.c ./output.c ./ptable.c ./code.c ./token.c ./cfgfile.c ./bon
us.c ./main.c ./fillow.c ./sim arr.c
makell: Leaving directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/qison'
make -C qcc
makell: Leaving directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/qcc'
...leqs/leqs -n .
.../qison/qison -n . yyout.c ./expr1l.y ./expr14.y ./if.y ./main.y ./expr12.y ./symtab.y ./expr13.y ./expr04.y ./expr01.y ./exp
10.y ./for.y ./expr02.y ./emit.y ./while.y ./expr06.y ./declr.y ./expr2.y ./expr08.y ./expr15.y ./start.y ./expr00.y ./er
ror.y ./expr03.y ./expr05.y ./expr09.y ./expr0
```

## Example 4:

```
test4.c - Kate
W
<u>File Edit View Projects Bookmarks Sessions Tools Settings Help</u>
📈 New 📴 Open
                                    👉 Previous Document \Rightarrow Next Document 📙 Save
                                                                                                                              📝 Save As
                                                                                                                                                       🔀 Close
          i tests
P Documents
                                                                   void set color(int color) {
    printf("%c[1;3%dm", 27, color);
            🌏 testl.c
              test2.c
                  test4.c
                                                                  ▼void rand_color() {
    set_color(rand()%8);
Projects
                                                                  ▼ int main() {
                                                                            main() {
  int i, j;
  int size, stars, spaces;
  printf("Enter size: ");
  scanf("%d", &size);
  srand(time(0));
  for (i = 0; i < size*2-1; i = i + 1) {
     spaces = size - (stars = i + (size-i-1)*2*(i >= size));
     for (j = 1; j < spaces; j = j + 1) {
          printf(" ");
     }
}</pre>
                                                             12
13
2
                                                            16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
                                                                                    for (j = 0; j < stars*2+1; j = j + 1) {
    rand_color();
    printf("*");</pre>
                                                                                    printf("\n");
                                                                             return 0;
                                                             Line: 1 of 30 Col: 1
                                                                                                   LINE INS
                                                                                                                                                                                            test4.c ISO-8859-1
                                                              🧕 Search and Replace 🔢 Current Project
```

```
$≥
  File Edit View Bookmarks Settings Help
make -C tests TEST=test4
make[2]: Entering directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/qcc/tests'
as -32 test4.s -o test4.o
ld -m elf_i386 -L/usr/lib32 -L/usr/lib/gcc/x86_64-unknown-linux-gnu/4.9.2/32/ -o test4 --dynamic-linker /usr/lib32/ld-linux.so
.2 --start-group /usr/lib32/crt1.o /usr/lib32/crti.o /usr/lib32/crtn.o test4.o -lc -lgcc -lgcc_eh
 ./test4
Enter size: 10
            ****
           *****
         *****
       ******
     *********
    ******
   *******
   ****************
   ******
           ******
            *****
              ***
make[2]: Leaving directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/qcc/tests' make[1]: Leaving directory '/home/Files/Documents/foe/BSc in CSE/Semester 08/CS 321/labs/qcc'iocoder ~/Desktop/School/CS 321/labs $ |
  5>
                                                   labs : bash
```

#### Example 5:

```
W
<u>File Edit View Projects Bookmarks Sessions Tools Settings Help</u>
                                             Next Document
                                                                              Save As
                                                                                             🔀 Close
                      revious Document
                                     - Save
P Documents
      i tests
       test.c
                                               for (i = 0; i < 5; i = i + 1) {
   for (j = 0; j < 2; j = j + 1) {
      printf("addr[%d][%d]: %p (%02d)\n", i, j, &x[i][j], x[i][j]);
}</pre>
Projects
                                      12
0
                                                return 0:
                                     14
15
                                     Line: 1 of 15 Col: 1
                                                             LINE INS
                                                                                                                     test.c ISO-8859-1
                                      🞑 Search and Replace 🔢 Current Project
```

#### **THANK YOU**