TDDD55 - Compilers and Interpreters
Lesson 3

November 22 2011

Kristian Stavåker (kristian.stavaker@liu.se)

Department of Computer and Information Science
Linköping University

LESSON SCHEDULE

November 1,	13.15 -15:	Formal languages and automata theory
November 11,	8.15 - 10:	Formal languages and automata theory, Flex
November 22,	15.15 - 17:	Bison and intermediate code generation
December 6,	15.15 - 17:	Exam preparation

TODAY

- Introduction to the Bison parser generator tool
- Introduction to quadruples and intermediate code generation
- Hints to laboratory assignments 3 and 4

LABORATORY ASSIGNMENTS

In the laboratory exercises you should get some practical experience in compiler construction.

There are 4 separate assignments to complete in 4x2 laboratory hours. You will also (most likely) have to work during non-scheduled time.

LABORATORY ASSIGNMENTS

Lab 1 Top-Down Parsing
Lab 2 Scanner Specification
Lab 3 Parser Generators
Lab 4 Intermediate Code Generation

1. TOP-DOWN PARSING

- Some grammar rules are given
- Your task:
  - Rewrite the grammar (elimate left recursion, etc.)
  - Implement your grammar in a C++ class named **Parser**. The **Parser** class should contain a method named **Parse** that returns the value of expressions in the language.

2. SCANNER SPECIFICATION

- Finish a scanner specification given in a *scanner.l* flex file, by adding rules for C and C++ style comments, identifiers, integers, and reals.

3. PARSER GENERATORS

- Finish a parser specification given in a *parser.y* bison file, by adding rules for expressions, conditions and function definitions, ... You also need to augment the grammar with error productions.

4. INTERMEDIATE CODE GENERATION

- The purpose of this assignment to learn about how parse trees (abstract syntax trees) can be translated into intermediate code.
- You are to finish a generator for intermediate code by adding rules for some language statements.

LABORATORY SKELETON

- ~TDDD55
  - /lab
    - /doc
      - Documentation for the assignments.
    - /lab1
      - Contains all the necessary files to complete the first assignment
    - /lab2
      - Contains all the necessary files to complete the second assignment
    - /lab3-4
      - Contains all the necessary files to complete assignment three and four

HANDING IN AND DEADLINE

- Demonstrate the working solutions to your lab assistant during scheduled time. Then send the modified files to the same assistant as well as answers to questions if any (put *TDDD55 <Name of the assignment>* in the topic field). One e-mail per group.

- Deadline for all the assignments is: **December 15, 2011**.

- Remember to register yourself in the webreg system, www.ida.liu.se/webreg (closed, e-mail your laboratory assistant)

BISON – PARSER GENERATOR

PURPOSE OF A PARSER

- The parser accepts tokens from the scanner and verifies the syntactic correctness of the program.
  - Syntactic correctness is judged by verification against a formal grammar which specifies the language to be recognized.
- Along the way, it also derives information about the program and builds a fundamental data structure known as parse tree or abstract syntax tree (ast).
- The abstract syntax tree is an internal representation of the program and augments the symbol table.

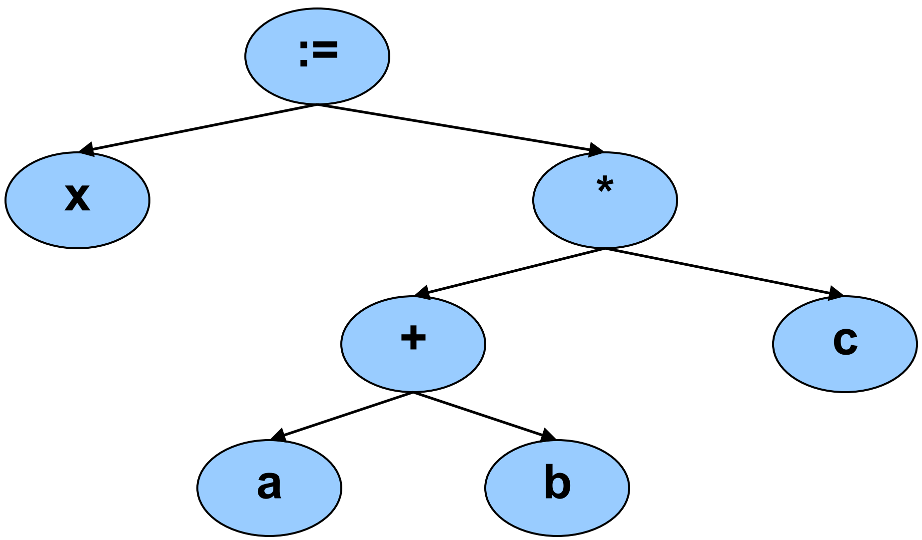
BOTTOM-UP PARSING

- Recognize the components of a program and then combine them to form more complex constructs until a whole program is recognized.

- The parse tree is then built from the bottom and up, hence the name.

BOTTOM-UP PARSING (2)

X := (a + b) * c;



LR PARSING

- A Specific bottom-up technique
  - LR stands for Left->right scan, Rightmost derivation.
  - Probably the most common & popular parsing technique.
  - yacc, bison, and many other parser generation tools utilize LR parsing.
  - Great for machines, not so great for humans …

PROS AND CONS LR PARSING

- Advantages of LR:
  - Accept a wide range of grammars/languages
  - Well suited for automatic parser generation
  - Very fast
  - Generally easy to maintain

- Disadvantages of LR:
  - Error handling can be tricky
  - Difficult to use manually

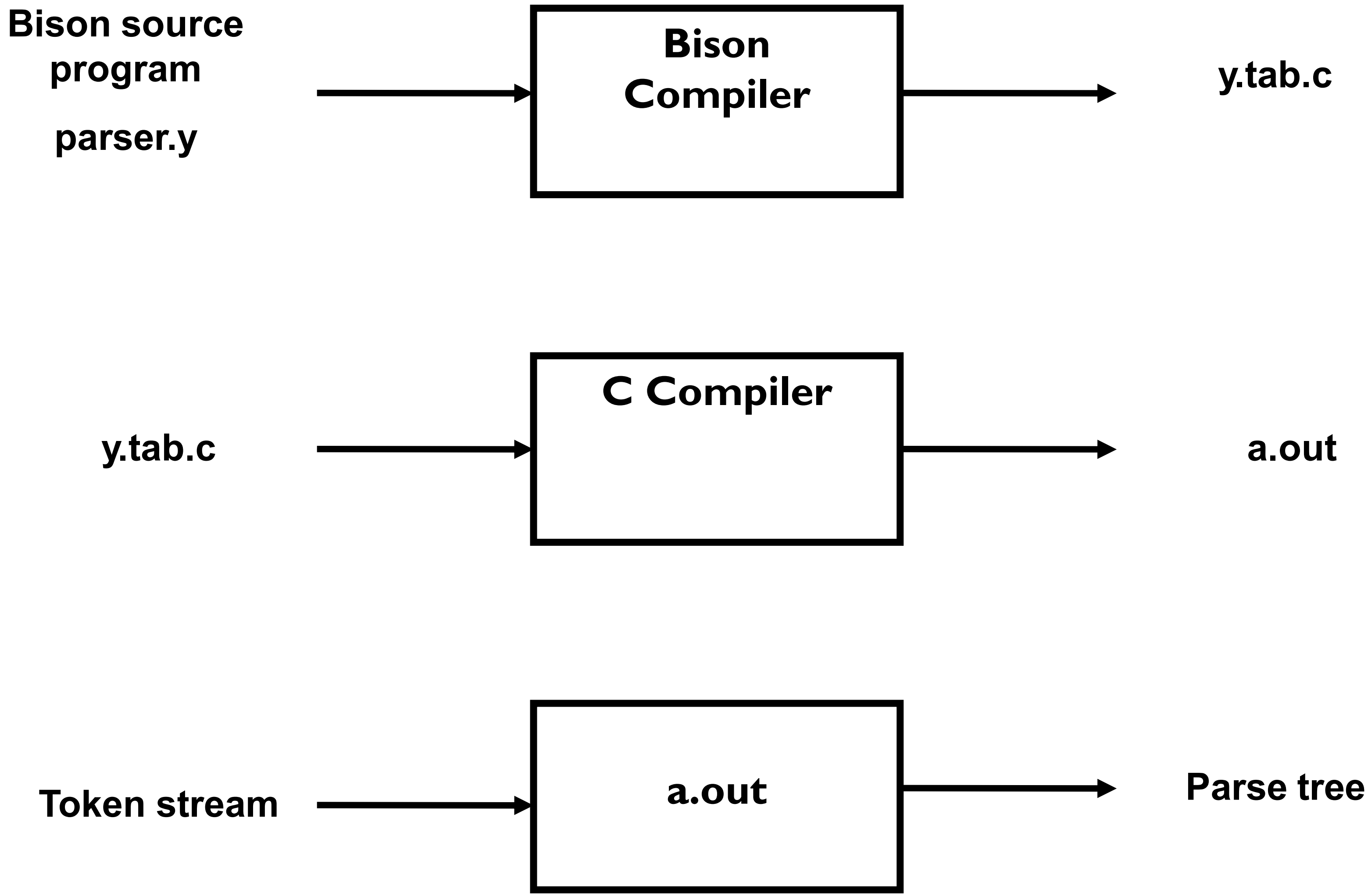
BISON

- **Bison** is a general-purpose parser generator that converts a grammar description of a context-free grammar into a **C** program to parse that grammar

BISON (2)

- Input: a specification file containing mainly the grammar definition
- Output: a C source file containing the parser
- The entry point is the function int yyparse();
  - yyparse reads tokens by calling yylex and parses until
    - end of file to be parsed, or
    - unrecoverable syntax error occurs
  - returns 0 for success and 1 for failure

BISON USAGE


Bison source program
parser.y
Bison Compiler
y.tab.c
y.tab.c
C Compiler
a.out
Token stream
a.out
Parse tree

BISON SPECIFICATION FILE

- A Bison specification is composed of 4 parts.

```
%{
    /* C declarations */
%}
    /* Bison declarations */

%%

    /* Grammar rules */

%%

    /* Additional C code */
```

C DECLARATIONS

- Contains macro definitions and declarations of functions and variables that are used in the actions in the grammar rules
- Copied to the beginning of the parser file so that they precede the definition of yyparse
- Use #include to get the declarations from a header file. If C declarations isn't needed, then the %{ and %} delimiters that bracket this section can be omitted

BISON DECLERATIONS

- Contains declarations that define terminal and non-terminal symbols, and specify precedence

GRAMMAR RULES

- Contains one or more Bison grammar rule, and nothing else.
- Example:
  - expression : expression '+' term { $$ = $1 + $3; } ;
- There must always be at least one grammar rule, and the first %% (which precedes the grammar rules) may never be omitted even if it is the first thing in the file.

ADDITIONAL C CODE

- Copied verbatim to the end of the parser file, just as the C declarations section is copied to the beginning.
- This is the most convenient place to put anything that should be in the parser file but isn't needed before the definition of yyparse.
- The definitions of yylex and yyerror often go here.

SYNTAX ERRORS

- Error productions can be added to the specification
- They help the compiler to recover from syntax errors and to continue to parse
- In order for the error productions to work we need at least one valid token after the error symbol

- Example 1:
  - functionCall : ID ‘(‘ paramList ‘)’
    | ID ‘(‘ error ‘)’

USING BISON WITH FLEX

- Bison and flex are obviously designed to work together
- Bison produces a driver program called yylex() (actually its included in the lex library -ll)
  - #include "lex.yy.c" in the last part of bison specification
  - this gives the program yylex access to bisons' token names

USING BISON WITH FLEX (2)

- Thus do the following:
  - % flex scanner.l
  - % bison parser.y
  - % cc y.tab.c -ly -ll
- This will produce an a.out which is a parser with an integrated scanner included

BISON EXAMPLE 1 (1/2)

```
%{
#include <ctype.h> /* standard C declarations here */
// extern int yylex();
}%
%token DIGIT /* bison declarations */
%%
/* Grammar rules */
line : expr '\n'        {  printf { "%d\n", $1 }; }   ;
expr : expr '+' term    {  $$ = $1 + $3;  }
      | term                                ;
term : term '*' factor      {  $$ = $1 * $3;  }
      | factor                              ;
```

BISON EXAMPLE 1 (2/2)

```
factor : '(' expr ')'     {  $$ = $2;  }
            | DIGIT ;
%%
/* Additional C code */

void yylex () {
  /* A really simple lexical analyzer */
  int c;
  c = getchar ();
  if ( isdigit (c) ) {
    yylval = c - '0' ;
    return DIGIT;
  }
  return c;
}
```

BISON EXAMPLE 2 – MID-RULES

```
thing: A { printf("seen an A"); } B ;

The same as:

thing: A fakename B ;
fakename: /* empty */ { printf("seen an A"); } ;
```

BISON EXAMPLE 3 (1/2)

```
/* Infix notation calculator--calc */

%{
#define YYSTYPE double
#include <math.h>
%}

/* BISON Declarations */
%token NUM
%left '-' '+'
%left '*' '/'
%left NEG     /* negation--unary minus */
%right '^'    /* exponentiation        */

/* Grammar follows */
%%
```

BISON EXAMPLE 3 (2/2)

```
input:    /* empty string */
        | input line
;
line:    '\n'
        | exp '\n'  { printf ("\t%.10g\n", $1); }
;
exp:     NUM              { $$ = $1;         }
        | exp '+' exp      { $$ = $1 + $3;    }
        | exp '-' exp      { $$ = $1 - $3;    }
        | exp '*' exp      { $$ = $1 * $3;    }
        | exp '/' exp      { $$ = $1 / $3;    }
        | '-' exp  %prec NEG { $$ = -$2;       }
        | exp '^' exp      { $$ = pow ($1, $3); }
        | '(' exp ')'      { $$ = $2;         }
;
%%
```

INTERMEDIATE CODE GENERATION

INTERMEDIATE LANGUAGE

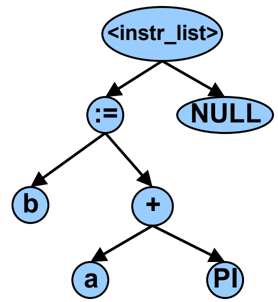
- Is closer to machine code without being machine dependent.
- Can handle temporary variables.
- Means higher portability, intermediate code can easier be expanded to assembly code.
- Offers the possibility of performing code optimizations such as register allocation.

INTERMEDIATE LANGUAGE (2)

- Why use intermediate languages?
- Retargeting - build a compiler for a new machine by attaching a new code generator to an existing front-end and middle-part
- Optimization - reuse intermediate code optimizers in compilers for different languages and different machines
- Code generation - for different source languages can be combined

GENERATION OF INTERMEDIATE CODE

```
program example;
const
  PI = 3.14159;
var
  a : real;
  b : real;
begin
  b := a + PI;
end.
```




```
q_rplus    A    PI   $1
q_rassign  $1   -    B
q_labl     4    -    -
```

INTERMEDIATE LANGUAGES

- Various types of intermediate code are:
  - Infix notation
  - Postfix notation
  - Three address code
    - Triples
    - Quadruples

QUADRUPLES

- You will use quadruples as intermediate language where an instruction has four fields:

operator	operand1	operand2	result

QUADRUPLES

(A + B) * (C + D) - E

operator	operand1	operand2	result
+	A	B	T1
+	C	D	T2
*	T1	T2	T3
-	T3	E	T4

HINTS LABORATORY ASSIGNMENT 3

PARSER GENERATORS

- Finnish a parser specification given in a *parser.y* bison file, by adding rules for expressions, conditions, function definitions, ...

FUNCTIONS

- One suggestion, outline:

```
function : funcnamedecl parameters ':' type variables functions block ';'
{
    // Set the return type of the function
    // Set the function body
    // Set current function to point to the parent again
} ;

funcnamedecl : FUNCTION id
{
    // Check if the function is already defined, report error if so
    // Create a new function information and set its parent to current function
    // Link the newly created function information to the current function
    // Set the new function information to be current function
} ;
```

EXPRESSIONS

- For precedence and associativity you can factorize the rules for expressions …

  or

- you can specify precedence and associativy at the top of the Bison specification file, in the *Bison Declarations* section. Read more about this in the Bison reference(s).

EXPRESSIONS (2)

- Example with factoring:

```
expression : expression '+' term
{
 // If any of the sub-expressions is NULL, set $$ to NULL
// Create a new Plus node but IntegerToReal casting might be needed
}
|
...
```

CONDITIONS

- For precedence and associativity you can factorize the rules for conditions …

  or

- you can specify precedence and associativy at the top of the Bison specification file, in the *Bison Declarations* section. Read more about this in the Bison reference(s).

HINTS LABORATORY ASSIGNMENT 4

INTERMEDIATE CODE GENERATION

- The purpose of this assignment is to learn how abstract syntax trees can be translated into intermediate code.
- You are to finish a generator for intermediate code (quadruples) by adding rules for some language constructs.
- You will work in the file *codegen.cc*.

BINARY OPERATIONS

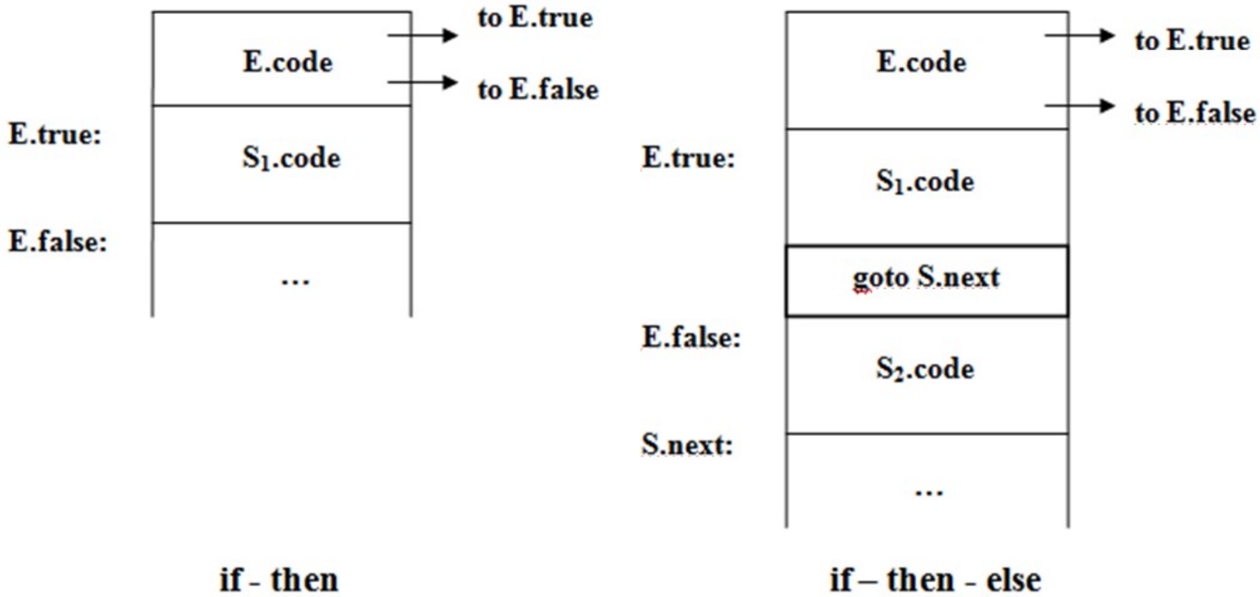
- In *BinaryGenerateCode*:
  - Generate code for left expression and right expression.
  - Generate either a *realop* or *intop* quadruple
    - For relations the type of the result is always integer
    - Otherwise the type of the result is the same as the type of the operands
    - You can use *currentFunction->TemporaryVariable*

ARRAY REFERENCES

- The absolute address is computed as follows:
  - *absAdr = baseAdr + arrayTypeSize * index*

IF STATEMENTS

- $S \rightarrow$ **if E then** S_1
- $S \rightarrow$ **if E then** S_1 **else** S_2



WHILE STATEMENT

- **S → while E do S_1**