

RV COLLEGE OF ENGINEERING[®]
BENGALURU-560059
(Autonomous Institution Affiliated to VTU, Belagavi)



“Intermediate Code Generation”

Report

Compiler Design Project

(18IS54)

Submitted By

Ankit Kumar Singh (1RV18IS007)



Under the Guidance of

B. K. Srinivas

Asst. Professor

in partial fulfillment for the award of degree of
Bachelor of Engineering
in
INFORMATION SCIENCE AND ENGINEERING
2020-21

RV COLLEGE OF ENGINEERING[®], BENGALURU - 560059
(Autonomous Institution Affiliated to VTU, Belagavi)

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING



CERTIFICATE

Certified that the Mini Project work entitled 'Intermediate Code Generation' has been carried out as a part of Compiler Design(18IS54) in partial fulfillment for the award of degree of **Bachelor of Engineering in Information Science and Engineering** of the Visvesvaraya Technological University, Belagavi during the year **2020-2021** by **Ankit Kumar Singh (1RV18IS007)** who is a bonafide student of **RV College of Engineering[®]**, Bengaluru. It is certified that all the corrections/suggestions indicated for the internal assessment have been incorporated in the report deposited in the departmental library. The report has been approved as it satisfies the academic requirements in respect of work prescribed by the institution for the said degree.

B. K. Srinivas

Asst. Professor

Department of ISE,
RVCE, Bengaluru-59

Dr Sagar B M

Head of the Department

Department of ISE,
RVCE, Bengaluru-59

Name of the Examiners

Signature with Date

1. _____

2. _____

Contents:

1. Introduction
 - a. Intermediate Code Generation
 - b. Yacc Script
 - c. C Program
2. Design of Programs
 - a. Code
 - b. Explanation
3. Test Cases
 - a. Without Errors
 - b. With Errors
4. Implementation
5. Readme File

Introduction:

Intermediate Code Generation:

Intermediate Code Generation phase is the glue between the frontend and backend of the compiler design stages. The final goal of a compiler is to get programs written in a high-level language to run on a computer. This means that, eventually, the program will have to be expressed as machine code which can run on the computer. Many compilers use a medium-level language as a stepping-stone between the high-level language and the very low-level machine code. Such stepping-stone languages are called intermediate code. It provides lower abstraction from source level and maintains some high level information.

Intermediate Code can be represented in many different formats depending whether it is language-specific (e.g. Bytecode for Java) or language-independent (three-address-code). We have used three-address-code to make it independent.

Most common independent intermediate representations are:

1. Postfix notation
2. Three Address Code
3. Syntax tree

Three Address Code

Three address code is a type of intermediate code which is easy to generate and can be easily converted to machine code. It makes use of at most three addresses and one operator to represent an expression and the value computed at each instruction is stored in a temporary variable generated by the compiler. The compiler decides the order of operation given by three address codes.

General representation: $x = y \text{ op } z$

An address can be a name, constant or temporary.

- Assignments $x = y \text{ op } z$; $x = \text{op } y$.
- Copy $x = y$.
- Unconditional jump $\text{goto } L$.
- Conditional jumps $\text{if } x \text{ relop } y \text{ goto } L$.
- Parameters $\text{param } x$.
- Function call $y = \text{call } p$

Yacc Script

Yacc provides a general tool for describing the input to a computer program. The Yacc user specifies the structures of his input, together with code to be invoked as each such structure is recognized. Yacc turns such a specification into a subroutine that handles the input process frequently, it is convenient and appropriate to have most of the flow of control in the user's application handled by this subroutine. Lexer can be used to make a simple parser. But it needs making extensive use of the user-defined states.

The input subroutine produced by Yacc calls a user-supplied routine to return the next basic input item. Thus, the user can specify his input in terms of individual input characters, or in terms of higher-level constructs such as names and numbers. The user-supplied routine may also handle idiomatic features such as comment and continuation conventions, which typically defy easy grammatical specification.

Yacc is written in portable C. The class of specifications accepted is a very general one: LALR(1) grammars with disambiguating rules.

The structure of our Yacc script is given below; files are divided into three sections, separated by lines that contain only two percent signs, as follows:

Definition section

%%

Rules section

%%

C code section

The definition section defines macros and imports header files written in C. It is also possible to write any C code here, which will be copied verbatim into the generated source file.

In the rules section, each grammar rule defines a symbol in terms of:

1. Other symbols
2. Tokens (or terminal symbols) which come from the lexer.

Each rule can have an associated action, which is executed *after* all the component symbols of the rule have been parsed. Actions are basically C-program statements surrounded by curly braces.

The C code section contains C statements and functions that are copied verbatim to the generated source file. These statements presumably contain code called by the rules in the rules section. In large programs, it is more convenient to place this code in a separate file linked in at compile time.

C Program

This section describes the input C program which is fed to the yacc script for parsing.

The workflow is explained as under:

1. Compile the script using Yacc tool

```
$ yacc -d c_parser.y
```

2. Compile the flex script using Flex tool

```
$ flex c_lexer.l
```

3. After compiling the lex file, a lex.yy.c file is generated. Also, y.tab.c and y.tab.h files are generated after compiling the yacc script.

4. The three files, lex.yy.c, y.tab.c and y.tab.h are compiled together with the options `-ll` and `-ly`

```
$ gcc -o compiler lex.yy.c y.tab.h y.tab.c -ll
```

5. The executable file is generated, which on running parses the C file given as a command line input

```
$ ./compiler test.c
```

The script also has an option to take standard input instead of taking input from a file.

Design of Programs:

Code:

Updated Lexer Code:

```
%{  
  
    #include <stdio.h>  
    #include <string.h>  
    #include <stdlib.h>  
    #include "y.tab.h"  
  
    #define ANSI_COLOR_RED           "\x1b[31m"  
    #define ANSI_COLOR_GREEN        "\x1b[32m"  
    #define ANSI_COLOR_YELLOW       "\x1b[33m"  
    #define ANSI_COLOR_BLUE         "\x1b[34m"  
    #define ANSI_COLOR_MAGENTA      "\x1b[35m"  
    #define ANSI_COLOR_CYAN         "\x1b[36m"  
    #define ANSI_COLOR_RESET        "\x1b[0m"  
  
    struct symboltable  
    {  
        char name[100];  
        char class[100];  
        char type[100];  
        char value[100];  
        int nestval;  
        int lineno;  
        int length;  
        int params_count;  
    }ST[1001];  
  
    struct constanttable  
    {  
        char name[100];  
        char type[100];  
        int length;  
    }CT[1001];  
  
    int currnest = 0;  
    extern int yylval;  
  
    int hash(char *str)
```

```

{
    int value = 0;
    for(int i = 0 ; i < strlen(str) ; i++)
    {
        value = 10*value + (str[i] - 'A');
        value = value % 1001;
        while(value < 0)
            value = value + 1001;
    }
    return value;
}

int lookupST(char *str)
{
    int value = hash(str);
    if(ST[value].length == 0)
    {
        return 0;
    }
    else if(strcmp(ST[value].name, str)==0)
    {
        return value;
    }
    else
    {
        for(int i = value + 1 ; i!=value ; i = (i+1)%1001)
        {
            if(strcmp(ST[i].name, str)==0)
            {
                return i;
            }
        }
        return 0;
    }
}

int lookupCT(char *str)
{
    int value = hash(str);
    if(CT[value].length == 0)
        return 0;

```



```

        else if(strcmp(CT[value].name,str)==0)
            return 1;
        else
        {
            for(int i = value + 1 ; i!=value ; i = (i+1)%1001)
            {
                if(strcmp(CT[i].name,str)==0)
                {
                    return 1;
                }
            }
            return 0;
        }
    }

void insertSTline(char *str1, int line)
{
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(strcmp(ST[i].name,str1)==0)
        {
            ST[i].lineno = line;
        }
    }
}

void insertST(char *str1, char *str2)
{
    if(lookupST(str1))
    {
        if(strcmp(ST[lookupST(str1)].class,"Identifier")==0 &&
strcmp(str2,"Array Identifier")==0)
        {
            printf("Error use of array\n");
            exit(0);
        }
        return;
    }
    else
    {
        int value = hash(str1);
        if(ST[value].length == 0)

```

```

    {
        strcpy(ST[value].name, str1);
        strcpy(ST[value].class, str2);
        ST[value].length = strlen(str1);
        ST[value].nestval = 9999;
        ST[value].params_count = -1;
        insertSTline(str1, yylineno);
        return;
    }

    int pos = 0;

    for (int i = value + 1 ; i != value ; i = (i+1)%1001)
    {
        if(ST[i].length == 0)
        {
            pos = i;
            break;
        }
    }

    strcpy(ST[pos].name, str1);
    strcpy(ST[pos].class, str2);
    ST[pos].length = strlen(str1);
    ST[pos].nestval = 9999;
    ST[pos].params_count = -1;
}

}

void insertSTtype(char *str1, char *str2)
{
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(strcmp(ST[i].name, str1) == 0)
        {
            strcpy(ST[i].type, str2);
        }
    }
}

void insertSTvalue(char *str1, char *str2)
{
    for(int i = 0 ; i < 1001 ; i++)

```

```

        {
            if(strcmp(ST[i].name,str1)==0 && ST[i].nestval ==
currnest)
            {
                strcpy(ST[i].value,str2);
            }
        }
    }

void insertSTnest(char *s, int nest)
{
    if(lookupST(s) && ST[lookupST(s)].nestval != 9999)
    {
        int pos = 0;
        int value = hash(s);
        for (int i = value + 1 ; i!=value ; i = (i+1)%1001)
        {
            if(ST[i].length == 0)
            {
                pos = i;
                break;
            }
        }

        strcpy(ST[pos].name,s);
        strcpy(ST[pos].class,"Identifier");
        ST[pos].length = strlen(s);
        ST[pos].nestval = nest;
        ST[pos].params_count = -1;
        ST[pos].lineno = yylineno;
    }
    else
    {
        for(int i = 0 ; i < 1001 ; i++)
        {
            if(strcmp(ST[i].name,s)==0 )
            {
                ST[i].nestval = nest;
            }
        }
    }
}

```

```

void insertSTparamscount(char *s, int count1)
{
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(strcmp(ST[i].name,s)==0 )
        {
            ST[i].params_count = count1;
        }
    }
}

int getSTparamscount(char *s)
{
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(strcmp(ST[i].name,s)==0 )
        {
            return ST[i].params_count;
        }
    }
    return -1;
}

void insertSTF(char *s)
{
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(strcmp(ST[i].name,s)==0 )
        {
            strcpy(ST[i].class,"Function");
            return;
        }
    }
}

void insertCT(char *str1, char *str2)
{
    if(lookupCT(str1))
        return;
    else

```

```

{
    int value = hash(str1);
    if(CT[value].length == 0)
    {
        strcpy(CT[value].name,str1);
        strcpy(CT[value].type,str2);
        CT[value].length = strlen(str1);
        return;
    }

    int pos = 0;

    for (int i = value + 1 ; i!=value ; i = (i+1)%1001)
    {
        if(CT[i].length == 0)
        {
            pos = i;
            break;
        }
    }

    strcpy(CT[pos].name,str1);
    strcpy(CT[pos].type,str2);
    CT[pos].length = strlen(str1);
}

}

void deletedata (int nesting)
{
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(ST[i].nestval == nesting)
        {
            ST[i].nestval = 99999;
        }
    }
}

}

int checkscope(char *s)
{
    int flag = 0;

```

```

    for(int i = 0 ; i < 1000 ; i++)
    {
        if(strcmp(ST[i].name,s)==0)
        {
            if(ST[i].nestval > currnest)
            {
                flag = 1;
            }
            else
            {
                flag = 0;
                break;
            }
        }
    }
    if(!flag)
    {
        return 1;
    }
    else
    {
        return 0;
    }
}

int check_id_is_func(char *s)
{
    for(int i = 0 ; i < 1000 ; i++)
    {
        if(strcmp(ST[i].name,s)==0)
        {
            if(strcmp(ST[i].class,"Function")==0)
                return 1;
        }
    }
    return 0;
}

int checkarray(char *s)
{
    for(int i = 0 ; i < 1000 ; i++)
    {
        if(strcmp(ST[i].name,s)==0)

```

```

        {
            if(strcmp(ST[i].class,"Array Identifier")==0)
            {
                return 0;
            }
        }
    }
    return 1;
}

int duplicate(char *s)
{
    for(int i = 0 ; i < 1000 ; i++)
    {
        if(strcmp(ST[i].name,s)==0)
        {
            if(ST[i].nestval == currnest)
            {
                return 1;
            }
        }
    }

    return 0;
}

int check_duplicate(char* str)
{
    for(int i=0; i<1001; i++)
    {
        if(strcmp(ST[i].name, str) == 0 && strcmp(ST[i].class,
"Function") == 0)
        {
            printf("Function redeclaration not allowed\n");
            exit(0);
        }
    }
}

int check_declaration(char* str, char *check_type)
{
    for(int i=0; i<1001; i++)
    {

```

```

        if(strcmp(ST[i].name, str) == 0 && strcmp(ST[i].class,
"Function") == 0 || strcmp(ST[i].name, "printf")==0 )
        {
            return 1;
        }
    }
    return 0;
}

int check_params(char* type_specifier)
{
    if(!strcmp(type_specifier, "void"))
    {
        printf("Parameters cannot be of type void\n");
        exit(0);
    }
    return 0;
}

char gettype(char *s, int flag)
{
    for(int i = 0 ; i < 1001 ; i++ )
    {
        if(strcmp(ST[i].name,s)==0)
        {
            return ST[i].type[0];
        }
    }
}

void printST()
{
    printf("%10s | %15s | %10s | %10s | %10s | %15s | %10s
|\n", "SYMBOL", "CLASS", "TYPE", "VALUE", "LINE NO", "NESTING", "PARAMS
COUNT");
    for(int i=0;i<100;i++) {
        printf("-");
    }
    printf("\n");
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(ST[i].length == 0)

```



```

        {
            continue;
        }
        printf("%10s | %15s | %10s | %10s | %10d | %15d | %10d
|\n",ST[i].name, ST[i].class, ST[i].type, ST[i].value, ST[i].lineno,
ST[i].nestval, ST[i].params_count);
    }
}

```

```

void printCT()
{
    printf("%10s | %15s\n","NAME", "TYPE");
    for(int i=0;i<81;i++) {
        printf("-");
    }
    printf("\n");
    for(int i = 0 ; i < 1001 ; i++)
    {
        if(CT[i].length == 0)
            continue;

        printf("%10s | %15s\n",CT[i].name, CT[i].type);
    }
}
char curid[20];
char curtype[20];
char curval[20];

```

%}

DE "define"
IN "include"

%%

```

\n    {yylineno++;}
([#][ " "]*({IN})[ ]*([<]?)([A-Za-z]+)[.]?([A-Za-z]*)([>]?))/["\n"|\|/"
"|"\t"]    { }
([#][ " "]*({DE})[ " "]*([A-Za-z]+)(" ")*[0-9]+)/["\n"|\|/" "|\t"]
    { }

```

\\/(.*)

{ }

\\\[^(^*)|[\r\n]|(\^+([^\r\n])))*\^+\\

```

        { }

[ \n\t] ;
";"          { return(';'); }
","          { return(','); }
("{")        { return('{'); }
("}")        { return('}'); }
"("          { return('('); }
")"          { return(')'); }
("[ "|" "<:") { return('['); }
("]" "|" ">") { return(']'); }
":"          { return(':'); }
"."          { return('.'); }

"char"        { strcpy(curtype,yytext); insertST(yytext,
"Keyword");return CHAR;}
"double"      { strcpy(curtype,yytext); insertST(yytext, "Keyword");
return DOUBLE;}
"else"        { insertST(yytext, "Keyword"); return ELSE;}
"float"       { strcpy(curtype,yytext); insertST(yytext,
"Keyword"); return FLOAT;}
"while"       { insertST(yytext, "Keyword"); return WHILE;}
"do"          { insertST(yytext, "Keyword"); return DO;}
"for"         { insertST(yytext, "Keyword"); return FOR;}
"if"          { insertST(yytext, "Keyword"); return IF;}
"int"         { strcpy(curtype,yytext); insertST(yytext, "Keyword");
return INT;}
"long"        { strcpy(curtype,yytext); insertST(yytext,
"Keyword"); return LONG;}
"return"      { insertST(yytext, "Keyword"); return RETURN;}
"short"       { strcpy(curtype,yytext); insertST(yytext,
"Keyword"); return SHORT;}
"signed"      { strcpy(curtype,yytext); insertST(yytext, "Keyword");
return SIGNED;}
"sizeof"      { insertST(yytext, "Keyword"); return SIZEOF;}
"struct"      { strcpy(curtype,yytext); insertST(yytext, "Keyword");
return STRUCT;}
"unsigned"    { insertST(yytext, "Keyword"); return UNSIGNED;}
"void"        { strcpy(curtype,yytext); insertST(yytext,
"Keyword"); return VOID;}
"break"       { insertST(yytext, "Keyword"); return BREAK;}

```

```

"++"      { return increment_operator; }
"--"      { return decrement_operator; }
"<<"      { return leftshift_operator; }
">>"      { return rightshift_operator; }
"<="      { return lessthan_assignment_operator; }
"<"       { return lessthan_operator; }
">="      { return greaterthan_assignment_operator; }
">"       { return greaterthan_operator; }
"=="      { return equality_operator; }
"!="      { return inequality_operator; }
"&&"      { return AND_operator; }
"||"      { return OR_operator; }
"^"       { return caret_operator; }
"*="      { return multiplication_assignment_operator; }
"/="      { return division_assignment_operator; }
"%="      { return modulo_assignment_operator; }
"+="      { return addition_assignment_operator; }
"-="      { return subtraction_assignment_operator; }
"<<="     { return leftshift_assignment_operator; }
">>="     { return rightshift_assignment_operator; }
"&="      { return AND_assignment_operator; }
"^="      { return XOR_assignment_operator; }
"|="      { return OR_assignment_operator; }
"&"       { return amp_operator; }
"!"       { return exclamation_operator; }
"~"       { return tilde_operator; }
"_"       { return subtract_operator; }
"+"       { return add_operator; }
"*"       { return multiplication_operator; }
"/"       { return division_operator; }
%"        { return modulo_operator; }
"|"       { return pipe_operator; }
\=        { return assignment_operator;}

\[^\n]*\[;|,|\)]      {strcpy(curval,yytext);
insertCT(yytext,"String Constant"); return string_constant;}
\[A-Z|a-z]\[;|,|\)]|:]      {strcpy(curval,yytext);
insertCT(yytext,"Character Constant"); return character_constant;}
[a-zA-Z]([a-zA-Z]|[0-9])*\[  {strcpy(curid,yytext); insertST(yytext,
"Array Identifier"); return array_identifier;}
[1-9][0-9]*\[;|,|"
"|\)|<|>|=|!|\||&|\+|\-|\*|\/|\%|~|\||\)|:|\n|\t|\^]
{strcpy(curval,yytext); insertCT(yytext, "Number Constant"); yylval =

```

```

atoi(yytext); return integer_constant;}
([0-9]*)\.[0-9]+)/[;|,|" "|\\|<|>|=|\\!|\\||&|\\+|\\-|\\*|\\/|\\%|~|\\n|\\t|\\^]
    {strcpy(curval,yytext); insertCT(yytext, "Floating Constant");
return float_constant;}
[A-Za-z_][A-Za-z_0-9]* {strcpy(curid,yytext);
insertST(curid,"Identifier"); return identifier;}

(.*?) {
    if(yytext[0]=='#')
    {
        printf("Error in Pre-Processor directive at line no.
%d\\n",yylineno);
    }
    else if(yytext[0]=='/')
    {
        printf("ERR_UNMATCHED_COMMENT at line no.
%d\\n",yylineno);
    }
    else if(yytext[0]=='"')
    {
        printf("ERR_INCOMPLETE_STRING at line no.
%d\\n",yylineno);
    }
    else
    {
        printf("ERROR at line no. %d\\n",yylineno);
    }
    printf("%s\\n", yytext);
    return 0;
}

%%

```

Updated Parser Code:

```
%{  
    #include <stdio.h>  
    #include <string.h>  
    #include <stdlib.h>  
  
    void yyerror(char* s);  
    int yylex();  
    void ins();  
    void insV();  
    int flag=0;  
    #define ANSI_COLOR_RED          "\x1b[31m"  
    #define ANSI_COLOR_GREEN        "\x1b[32m"  
    #define ANSI_COLOR_CYAN         "\x1b[36m"  
    #define ANSI_COLOR_RESET        "\x1b[0m"  
    extern char curid[20];  
    extern char curtype[20];  
    extern char curval[20];  
    extern int currnest;  
    void deletedata (int );  
    int checkscope(char*);  
    int check_id_is_func(char *);  
    void insertST(char*, char*);  
    void insertSTnest(char*, int);  
    void insertSTparamscount(char*, int);  
    int getSTparamscount(char*);  
    int check_duplicate(char*);  
    int check_declaration(char*, char *);  
    int check_params(char*);  
    int duplicate(char *s);  
    int checkarray(char*);  
    char currfunctype[100];  
    char currfunc[100];  
    char currfuncall[100];  
    void insertSTF(char*);  
    char gettype(char*,int);  
    char getfirst(char*);  
    void push(char *s);  
    void codegen();  
    void codeassign();  
    char* itoa(int num, char* str, int base);  
    void reverse(char str[], int length);
```

```

void swap(char*,char*);
void label1();
void label2();
void label3();
void label4();
void label5();
void label6();
void genunary();
void codegencon();
void funcgen();
void funcgenend();
void arggen();
void callgen();

int params_count=0;
int call_params_count=0;
int top = 0,count=0,ltop=0,lno=0;
char temp[3] = "t";
%}

%nonassoc IF
%token INT CHAR FLOAT DOUBLE LONG SHORT SIGNED UNSIGNED STRUCT
%token RETURN MAIN
%token VOID
%token WHILE FOR DO
%token BREAK
%token ENDIF
%expect 1

%token identifier array_identifier func_identifier
%token integer_constant string_constant float_constant character_constant

%nonassoc ELSE

%right leftshift_assignment_operator rightshift_assignment_operator
%right XOR_assignment_operator OR_assignment_operator
%right AND_assignment_operator modulo_assignment_operator
%right multiplication_assignment_operator division_assignment_operator
%right addition_assignment_operator subtraction_assignment_operator
%right assignment_operator

%left OR_operator
%left AND_operator

```

```

%left pipe_operator
%left caret_operator
%left amp_operator
%left equality_operator inequality_operator
%left lessthan_assignment_operator lessthan_operator
greaterthan_assignment_operator greaterthan_operator
%left leftshift_operator rightshift_operator
%left add_operator subtract_operator
%left multiplication_operator division_operator modulo_operator

%right SIZEOF
%right tilde_operator exclamation_operator
%left increment_operator decrement_operator


%start program

%%
program
    : declaration_list;

declaration_list
    : declaration D

D
    : declaration_list
    | ;

declaration
    : variable_declaration
    | function_declaration

variable_declaration
    : type_specifier variable_declaration_list ';'

variable_declaration_list
    : variable_declaration_list ','
variable_declaration_identifier | variable_declaration_identifier;

variable_declaration_identifier
    : identifier
{if(duplicate(curid)){printf("Duplicate\n");exit(0);}insertSTnest(curid,c
urrnest); ins(); } vdi

```

```

        | array_identifier
{if(duplicate(curid)){printf("Duplicate\n");exit(0);}insertSTnest(curid,c
urnnest); ins(); } vdi;

```

```

vdi : identifier_array_type | assignment_operator simple_expression ;

```

```

identifier_array_type
    : '[' initialization_params
    | ;

```

```

initialization_params
    : integer_constant ']' initialization {if($$ < 1)
{printf("Wrong array size\n"); exit(0);} }
    | ']' string_initialization;

```

```

initialization
    : string_initialization
    | array_initialization
    | ;

```

```

type_specifier
    : INT | CHAR | FLOAT | DOUBLE
    | LONG long_grammar
    | SHORT short_grammar
    | UNSIGNED unsigned_grammar
    | SIGNED signed_grammar
    | VOID ;

```

```

unsigned_grammar
    : INT | LONG long_grammar | SHORT short_grammar | ;

```

```

signed_grammar
    : INT | LONG long_grammar | SHORT short_grammar | ;

```

```

long_grammar
    : INT | ;

```

```

short_grammar
    : INT | ;

```

```

function_declaration

```



```

        : function_declaration_type
function_declaration_param_statement;

function_declaration_type
        : type_specifier identifier '(' { strcpy(currfunctype,
curtype); strcpy(currfunc, curid); check_duplicate(curid);
insertSTF(curid); ins(); };

function_declaration_param_statement
        : {params_count=0;}params ')' {funcgen();} statement
{funcgenend();};

params
        : parameters_list { insertSTparamscount(currfunc,
params_count); }| { insertSTparamscount(currfunc, params_count); };

parameters_list
        : type_specifier { check_params(curtype);}
parameters_identifier_list ;

parameters_identifier_list
        : param_identifier parameters_identifier_list_breakup;

parameters_identifier_list_breakup
        : ',' parameters_list
| ;

param_identifier
        : identifier { ins();insertSTnest(curid,1);
params_count++; } param_identifier_breakup;

param_identifier_breakup
        : '[' ']'
| ;

statement
        : expression_statement | compound_statement
| conditional_statements | iterative_statements
| return_statement | break_statement
| variable_declaration;

compound_statement
        : {currnest++;} '{' statment_list '}'

```

```

{deletedata(currnest);currnest--;} ;

statement_list
    : statement statement_list
    | ;

expression_statement
    : expression ';'
    | ';' ;

conditional_statements
    : IF '(' simple_expression ')'
{label1();if($3!=1){printf("Condition checking is not of type
int\n");exit(0);}} statement {label2();} conditional_statements_breakup;

conditional_statements_breakup
    : ELSE statement {label3();}
    | {label3();};

iterative_statements
    : WHILE '(' {label4();} simple_expression ')'
{label1();if($4!=1){printf("Condition checking is not of type
int\n");exit(0);}} statement {label5();}
    | FOR '(' expression ';' {label4();} simple_expression
    ';' {label1();if($6!=1){printf("Condition checking is not of type
int\n");exit(0);}} expression ')'statement {label5();}
    | {label4();}DO statement WHILE '(' simple_expression
    ')' {label1();label5();if($6!=1){printf("Condition checking is not of type
int\n");exit(0);}} ';' ;

return_statement
    : RETURN ';' {if(strcmp(currfunctype,"void"))
{printf("Returning void of a non-void function\n"); exit(0);}}
    | RETURN expression ';' { if(!strcmp(currfunctype,
"void"))
{
yyerror("Function is void");
}

if((currfunctype[0]=='i' || currfunctype[0]=='c') && $2!=1)
{

```

```
printf("Expression doesn't match return type of function\n"); exit(0);
    }
```

```
};
```

```
break_statement
```

```
    : BREAK ';' ;
```

```
string_initilization
```

```
    : assignment_operator string_constant {insV();} ;
```

```
array_initialization
```

```
    : assignment_operator '{' array_int_declarations '}';
```

```
array_int_declarations
```

```
    : integer_constant array_int_declarations_breakup;
```

```
array_int_declarations_breakup
```

```
    : ',' array_int_declarations
    | ;
```

```
expression
```

```
    : mutable assignment_operator {push("=");} expression
```

```
{
```

```
    if($1==1 && $4==1)
```

```
    {
```

```
    $$=1;
```

```
}
```

```
else
```

```
{$$=-1; printf("Type mismatch\n"); exit(0);}
```

```
codeassign();
```

```
}
```

```
    | mutable addition_assignment_operator
```

```
{push("+=");}expression {
```

```
    if($1==1 && $4==1)
```

```

$$=1;

else

{$$=-1; printf("Type mismatch\n"); exit(0);}

codeassign();

| mutable subtraction_assignment_operator {push("-=");}
expression {

    if($1==1 && $4==1)

    $$=1;

    else

    {$$=-1; printf("Type mismatch\n"); exit(0);}

    codeassign();

    | mutable multiplication_assignment_operator
    {push("*=");} expression {

        if($1==1 && $4==1)

        $$=1;

        else

        {$$=-1; printf("Type mismatch\n"); exit(0);}

        codeassign();

        | mutable division_assignment_operator
        {push("/=");} expression {

            if($1==1 && $4==1)

            $$=1;

            else

```

```

{ $$=-1; printf("Type mismatch\n"); exit(0);}
}

| mutable modulo_assignment_operator
{push("%=");}expression {

    if($1==1 && $3==1)

    $$=1;

else

{ $$=-1; printf("Type mismatch\n"); exit(0);}

codeassign();

}

| mutable increment_operator
{ push("++");if($1 == 1) $$=1; else $$=-1; genunary();}
| mutable decrement_operator
{push("--");if($1 == 1) $$=1; else $$=-1;}
| simple_expression {if($1 == 1) $$=1; else $$=-1;} ;

simple_expression
: simple_expression OR_operator and_expression
{push("||");} {if($1 == 1 && $3==1) $$=1; else $$=-1; codegen();}
| and_expression {if($1 == 1) $$=1; else $$=-1;}

and_expression
: and_expression AND_operator {push("&&");}
unary_relation_expression {if($1 == 1 && $3==1) $$=1; else $$=-1;
codegen();}
| unary_relation_expression {if($1 == 1) $$=1; else
$$=-1;} ;

unary_relation_expression
: exclamation_operator {push("!");}
unary_relation_expression {if($2==1) $$=1; else $$=-1; codegen();}
| regular_expression {if($1 == 1) $$=1; else $$=-1;} ;

regular_expression

```

```

: regular_expression relational_operators sum_expression
{if($1 == 1 && $3==1) $$=1; else $$=-1; codegen();}
| sum_expression {if($1 == 1) $$=1; else $$=-1;} ;

relational_operators
: greaterthan_assignment_operator {push(">=");} |
lessthan_assignment_operator {push("<=");} | greaterthan_operator
{push(">");} | lessthan_operator {push("<");} | equality_operator
{push("==");} | inequality_operator {push("!=");} ;

sum_expression
: sum_expression sum_operators term {if($1 == 1 &&
$3==1) $$=1; else $$=-1; codegen();}
| term {if($1 == 1) $$=1; else $$=-1;} ;

sum_operators
: add_operator {push("+");}
| subtract_operator {push("-");} ;

term
: term MULOP factor {if($1 == 1 && $3==1) $$=1; else
$$=-1; codegen();}
| factor {if($1 == 1) $$=1; else $$=-1;} ;

MULOP
: multiplication_operator {push("*");} |
division_operator {push("/");} | modulo_operator {push("%");} ;

factor
: immutable {if($1 == 1) $$=1; else $$=-1;}
| mutable {if($1 == 1) $$=1; else $$=-1;} ;

mutable
: identifier {
push(curid);
if(check_id_is_func(curid))
{printf("Function name used as
Identifier\n"); exit(8);}
if(!checkscope(curid))
{printf("%s\n",curid);printf("Undeclared\n");exit(0);}
if(!checkarray(curid))
{printf("%s\n",curid);printf("Array ID has

```

```

no subscript\n");exit(0);}
                                if(gettype(curid,0)=='i' ||
gettype(curid,1)=='c')
                                $$ = 1;
                                else
                                $$ = -1;
                                }
                                | array_identifier
{if(!checkscope(curid)){printf("%s\n",curid);printf("Undeclared\n");exit(
0);}} '[' expression ']'
                                {if(gettype(curid,0)=='i' ||
gettype(curid,1)=='c')
                                $$ = 1;
                                else
                                $$ = -1;
                                };

immutable
: '(' expression ')' {if($2==1) $$=1; else $$=-1;}
| call {if($1==-1) $$=-1; else $$=1;}
| constant {if($1==1) $$=1; else $$=-1;};

call
: identifier '('{
                                if(!check_declaration(curid, "Function"))
                                { printf("Function not declared");
exit(0);}
                                insertSTF(curid);
                                strcpy(currfunccall,curid);
                                if(gettype(curid,0)=='i' ||
gettype(curid,1)=='c')
                                {
                                $$ = 1;
                                }
                                else
                                $$ = -1;
                                call_params_count=0;
                                }
                                arguments ')'
                                { if(strcmp(currfunccall,"printf"))
                                {

```

```

if(getSTparamscount(currfuncall)!=call_params_count)
    {
        yyerror("Number of
arguments in function call doesn't match number of parameters");
        exit(8);
    }
}
callgen();
};

arguments
    : arguments_list | ;

arguments_list
    : arguments_list ',' exp { call_params_count++; }
    | exp { call_params_count++; };

exp : identifier {arggen(1);} | integer_constant {arggen(2);} |
string_constant {arggen(3);} | float_constant {arggen(4);} |
character_constant {arggen(5);} ;

constant
    : integer_constant { insV(); codegencon(); $$=1; }
    | string_constant { insV(); codegencon(); $$=-1; }
    | float_constant { insV(); codegencon(); }
    | character_constant { insV(); codegencon(); $$=1; };

%%

extern FILE *yyin;
extern int yylineno;
extern char *yytext;
void insertSTtype(char *,char *);
void insertSTvalue(char *, char *);
void incertCT(char *, char *);
void printST();
void printCT();

struct stack
{
    char value[100];
    int labelvalue;
}s[100],label[100];

```



```

void push(char *x)
{
    strcpy(s[++top].value,x);
}

void swap(char *x, char *y)
{
    char temp = *x;
    *x = *y;
    *y = temp;
}

void reverse(char str[], int length)
{
    int start = 0;
    int end = length -1;
    while (start < end)
    {
        swap((str+start), (str+end));
        start++;
        end--;
    }
}

char* itoa(int num, char* str, int base)
{
    int i = 0;
    int isNegative = 0;

    if (num == 0)
    {
        str[i++] = '0';
        str[i] = '\0';
        return str;
    }

    if (num < 0 && base == 10)
    {
        isNegative = 1;
        num = -num;
    }
}

```

```
}
```

```
while (num != 0)
```

```
{
```

```
    int rem = num % base;
```

```
    str[i++] = (rem > 9)? (rem-10) + 'a' : rem + '0';
```

```
    num = num/base;
```

```
}
```

```
if (isNegative)
```

```
    str[i++] = '-';
```

```
str[i] = '\\0';
```

```
reverse(str, i);
```

```
return str;
```

```
}
```

```
void codegen()
```

```
{
```

```
    strcpy(temp, "t");
```

```
    char buffer[100];
```

```
    itoa(count, buffer, 10);
```

```
    strcat(temp, buffer);
```

```
    printf("%s = %s %s
```

```
%s\\n", temp, s[top-2].value, s[top-1].value, s[top].value);
```

```
    top = top - 2;
```

```
    strcpy(s[top].value, temp);
```

```
    count++;
```

```
}
```

```
void codegencon()
```

```
{
```

```
    strcpy(temp, "t");
```

```
    char buffer[100];
```

```
    itoa(count, buffer, 10);
```

```
    strcat(temp, buffer);
```

```
    printf("%s = %s\\n", temp, curval);
```

```
    push(temp);
```

```
    count++;
```

```

}

int isunary(char *s)
{
    if(strcmp(s, "--")==0 || strcmp(s, "++")==0)
    {
        return 1;
    }
    return 0;
}

void genunary()
{
    char temp1[100], temp2[100], temp3[100];
    strcpy(temp1, s[top].value);
    strcpy(temp2, s[top-1].value);

    if(isunary(temp1))
    {
        strcpy(temp3, temp1);
        strcpy(temp1, temp2);
        strcpy(temp2, temp3);
    }
    strcpy(temp, "t");
    char buffer[100];
    itoa(count, buffer, 10);
    strcat(temp, buffer);
    count++;

    if(strcmp(temp2, "--")==0)
    {
        printf("%s = %s - 1\n", temp, temp1);
        printf("%s = %s\n", temp1, temp);
    }

    if(strcmp(temp2, "++")==0)
    {
        printf("%s = %s + 1\n", temp, temp1);
        printf("%s = %s\n", temp1, temp);
    }

    top = top - 2;
}

```

```

}

void codeassign()
{
    printf("%s = %s\n",s[top-2].value,s[top].value);
    top = top - 2;
}

void label1()
{
    strcpy(temp,"L");
    char buffer[100];
    itoa(lno,buffer,10);
    strcat(temp,buffer);
    printf("IF not %s GoTo %s\n",s[top].value,temp);
    label[++ltop].labelvalue = lno++;
}

void label2()
{
    strcpy(temp,"L");
    char buffer[100];
    itoa(lno,buffer,10);
    strcat(temp,buffer);
    printf("GoTo %s\n",temp);
    strcpy(temp,"L");
    itoa(label[ltop].labelvalue,buffer,10);
    strcat(temp,buffer);
    printf("%s:\n",temp);
    ltop--;
    label[++ltop].labelvalue=lno++;
}

void label3()
{
    strcpy(temp,"L");
    char buffer[100];
    itoa(label[ltop].labelvalue,buffer,10);
    strcat(temp,buffer);
    printf("%s:\n",temp);
    ltop--;
}

```

```

void label4()
{
    strcpy(temp,"L");
    char buffer[100];
    itoa(lno,buffer,10);
    strcat(temp,buffer);
    printf("%s:\n",temp);
    label[++ltop].labelvalue = lno++;
}

void label5()
{
    strcpy(temp,"L");
    char buffer[100];
    itoa(label[ltop-1].labelvalue,buffer,10);
    strcat(temp,buffer);
    printf("GoTo %s:\n",temp);
    strcpy(temp,"L");
    itoa(label[ltop].labelvalue,buffer,10);
    strcat(temp,buffer);
    printf("%s:\n",temp);
    ltop = ltop - 2;
}

void funcgen()
{
    printf("func begin %s\n",currfunc);
}

void funcgenend()
{
    printf("func end\n\n");
}

void arggen(int i)
{
    if(i==1)
    {
        printf("refparam %s\n", curid);
    }
}

```

```

    }
    else
    {
        printf("refparam %s\n", curval);
    }
}

void callgen()
{
    printf("refparam result\n");
    push("result");
    printf("call %s, %d\n", currfuncall, call_params_count);
}

int main(int argc , char **argv)
{
    yyin = fopen(argv[1], "r");
    yyparse();

    if(flag == 0)
    {
        printf(ANSI_COLOR_GREEN "Status: Parsing Complete - Valid"
ANSI_COLOR_RESET "\n");
        printf("%30s" ANSI_COLOR_CYAN "SYMBOL TABLE" ANSI_COLOR_RESET
"\n", " ");
        printf("%30s %s\n", " ", "-----");
        printST();

        printf("\n\n%30s" ANSI_COLOR_CYAN "CONSTANT TABLE"
ANSI_COLOR_RESET "\n", " ");
        printf("%30s %s\n", " ", "-----");
        printCT();
    }
}

void yyerror(char *s)
{
    printf(ANSI_COLOR_RED "%d %s %s\n", yylineno, s, yytext);
    flag=1;
    printf(ANSI_COLOR_RED "Status: Parsing Failed - Invalid\n"
ANSI_COLOR_RESET);
}

```

```
        exit(7);
    }

    void ins()
    {
        insertSTtype(curid, curtype);
    }

    void insV()
    {
        insertSTvalue(curid, curval);
    }

    int yywrap()
    {
        return 1;
    }
}
```

Explanation:

The lex code is detecting the tokens from the source code and returning the corresponding token to the parser. In phase 1 it was just printing the token and now it is returning the token so that the parser uses it for further computation. We are using the symbol table and constant table of the previous phase only. We added functions like `insertSTnest()`, `insertSTparamscount()`, `checkscope()`, `deletedata()`, `duplicate()` etc., in order to check the semantics. In the production rules of the grammar semantic actions are written and these are performed by the functions listed above. Along with semantic actions SDT also included a function to generate the 3 address code.

Declaration Section

In this section, all the necessary header files have been included along with function declaration and flag that was needed in the code.

Between declaration and rules section I have listed all the tokens which are returned by the lexer according to the precedence order. We also declared the operators here according to their associativity and precedence. This ensures the grammar we are giving to the parser is unambiguous as LALR(1) parser cannot work with ambiguous grammar.

Rules Section

In this section production rules for the entire C language are written. The grammar production does the syntax analysis of the source code. When a complete statement with proper syntax is matched by the parser. Along with rules semantic actions associated with the rules are also written and corresponding functions are called to do the necessary actions. Then code generation function was also associated with the production so that we can get the desired 3 address code for the given input program.

C-Program Section

In this section the parser links the extern functions, variables declared in the lexer, external files generated by the lexer etc. The main function takes the input source code file and prints the final symbol table and the 3 address code. Apart from these several functions are also written that generate the 3 address code.

Test Cases

Valid test Cases:

```
===== Running TestCase 1 =====
func begin main
L0:
refparam "hi"
refparam result
call printf, 1
t0 = i < n
IF not t0 GoTo L1
GoTo L0:
L1:
func end

Status: Parsing Complete - Valid
      SYMBOL TABLE
-----
SYMBOL | CLASS | TYPE | VALUE | LINE NO | PARAMS COUNT |
-----
i      | Identifier | int |  | 5 | -1 |
n      | Identifier | int |  | 5 | -1 |
do     | Keyword   |  |  | 8 | -1 |
int    | Keyword   |  |  | 5 | -1 |
main   | Function  | void |  | 3 | 0 |
printf | Function  |  |  | 10 | -1 |
while  | Keyword   |  |  | 11 | -1 |
void   | Keyword   |  |  | 3 | -1 |
-----

      CONSTANT TABLE
-----
NAME | TYPE
-----
"hi" | String Constant
```

===== Running TestCase 2 =====

```
func begin myfunc
t0 = a + b
func end
```

```
func begin main
L0:
t1 = 3
t2 = a < t1
IF not t2 GoTo L1
t3 = a + b
a = t3
t4 = 0
i = t4
L2:
t5 = i < b
IF not t5 GoTo L3
t6 = i + 1
i = t6
t7 = b + 1
b = t7
refparam a
refparam b
refparam result
call myfunc, 2
GoTo L2:
L3:
t8 = a + 1
a = t8
GoTo L0:
L1:
func end
```

Status: Parsing Complete - Valid

SYMBOL TABLE

SYMBOL	CLASS	TYPE	VALUE	LINE NO	PARAMS COUNT
a	Identifier	int		3	-1
b	Identifier	int		3	-1
a	Identifier	int	3	10	-1
b	Identifier	int		10	-1
i	Identifier	int		10	-1
for	Keyword			15	-1
return	Keyword			5	-1
int	Keyword			3	-1
main	Function	void		8	0
myfunc	Function	int		3	2
while	Keyword			12	-1
void	Keyword			8	-1

CONSTANT TABLE

NAME	TYPE
0	Number Constant
3	Number Constant

==== Running TestCase 3 =====

```
func begin main
t0 = 5
L0:
t1 = 0
t2 = a > t1
IF not t2 GoTo L1
refparam "Hello world"
refparam result
call printf, 1
GoTo L0:
L1:
t3 = 4
a = t3
L2:
t4 = 0
t5 = a > t4
IF not t5 GoTo L3
refparam "%d"
refparam a
refparam result
call printf, 2
t6 = 4
b = t6
L4:
t7 = 0
t8 = b > t7
IF not t8 GoTo L5
refparam "%d"
refparam b
refparam result
call printf, 2
GoTo L4:
L5:
GoTo L2:
L3:
func end
```

Status: Parsing Complete - Valid

SYMBOL TABLE

SYMBOL	CLASS	TYPE	VALUE	LINE NO	PARAMS COUNT
a	Identifier	int	0	5	-1
b	Identifier	int	0	17	-1
int	Keyword			3	-1
main	Function	int		3	0
printf	Function			8	-1
while	Keyword			6	-1

CONSTANT TABLE

NAME	TYPE
"Hello world"	String Constant
"%d"	String Constant
0	Number Constant
4	Number Constant
5	Number Constant

```

===== Running TestCase 4 =====
func begin main
t0 = 0
t1 = 0
a = t1
L0:
t2 = 10
t3 = a < t2
IF not t3 GoTo L1
t4 = a + 1
a = t4
refparam "H1"
refparam result
call printf, 1
GoTo L0:
L1:
L2:
t5 = 0
t6 = a > t5
IF not t6 GoTo L3
GoTo L2:
L3:
L4:
t7 = 10
t8 = a < t7
IF not t8 GoTo L5
t9 = a + 1
a = t9
GoTo L4:
L5:
func end

```

Status: Parsing Complete - Valid

SYMBOL TABLE

SYMBOL	CLASS	TYPE	VALUE	LINE NO	PARAMS COUNT
a	Identifier	int	10	10	-1
for	Keyword			11	-1
int	Keyword			8	-1
main	Function	int		8	0
printf	Function			13	-1
while	Keyword			16	-1

CONSTANT TABLE

NAME	TYPE
"H1"	String Constant
10	Number Constant
0	Number Constant

Invalid Test Cases:

```

===== Running TestCase 5 =====
func begin main
ERROR at line no. 6
@
6 syntax error @
Status: Parsing Failed - Invalid

```

```

===== Running TestCase 6 =====
func begin main
ERROR at line no. 4
9
4 syntax error 9
Status: Parsing Failed - Invalid

```

```

===== Running TestCase 7 =====
func begin main
Function not declared

```

```

===== Running TestCase 8 =====
func begin myfunc
func end

func begin main
refparam i
refparam n
12 Number of arguments in function call doesn't match number of parameters )
Status: Parsing Failed - Invalid

```

Implementation

The lexer code submitted in the previous phase took care of most of the features of C using regular expressions. Some special corner cases were taken care of using custom regex.

These were:

- A. The Regex for Identifiers
- B. Multiline comments should be supported
- C. Literals
- D. Error Handling for Incomplete String
- E. Error Handling for Nested Comments

The parser code requires exhaustive token recognition and because of this reason, we utilised the lexer code given under the C specifications with the parser. The parser implements C grammar using a number of production rules.

The parser takes tokens from the lexer output, one at a time and applies the corresponding production rules to append to the symbol table with type, value and line of declaration. If the parsing is not successful, the parser outputs the line number with the corresponding error. Along with this semantic actions were also added to each production rule to check if the structure created has some meaning or not. Then we added the function to generate the 3 address code with production so that we can generate the desired intermediate code. In order to generate 3 address code we made use of explicit stack. Whenever we came across an operator, operand or constant we pushed it to stack. Whenever reduction occurred (Since LALR(1) parser is bottom up parser it evaluates SDT when reduction occurs) `codegen()` function generated the 3 address code by creating a new temporary variable and by making use of the entries in the stack, after that it popped those entries from the stack and pushed the temporary variable to the stack so that it gets used in further computation. Similarly functions like labels were used to assign appropriate labels while using conditional statements or iterative statements. All the functions used are described below :

1. `codegen()` : This function is called whenever a reduction of an expression takes place. It creates the temporary variable and displays the desired 3 address code i.e $x = y \text{ op } z$.
2. `codegencon()` : This function is especially written for reductions of expression involving constants since its 3 address code is $x \text{ op } z$.
3. `isunary()` : This function checks if the operator is an unary operator like '++'. If so it returns true else false.
4. `genunary()` : This function is specifically designed to generate 3 address code for unary operations. It makes use of `isunary` function mentioned above. E.g. if $a = i++$ then it converts into $t0 = i + 1, a = t0$.
5. `codeassign()` : This function is specifically designed for assignment operator. It assigns the final temp variable value (after all the evaluation) to the desired variable.
6. `label1()` : It is used while evaluating conditions of loops or if statement. If the condition is not satisfied then it states where to jump to i.e. on which label the control should go.

7. `label2()` : It is used when the statement block pertaining to if statement is over. It tells where the control flow should go once that block is over i.e. it jumps the else statement block.
8. `label3()` : it is used after the whole if else construct is over. It gives label that tells where to jump after the if block is executed.
9. `label4()` : it is used to give labels to starting of loops.
10. `label5()` : it is used after the statement block of the loop. It indicates the label to jump to and also generates the label where the control should go once the loop is terminated.
11. `funcgen()` : it indicates beginning of a function.
12. `funcgenend()` : it indicates the ending of a function.
13. `arggen()` : it displays all the reference parameters that are used in a function call.
14. `callgen()` : it calls the function i.e. displays the appropriate function call according to 3 address code.
15. `ltoa()` : it worked as utility function since we had to name temporary variables and labels it was used to convert int to string and used several functions like reverse , swap to do it.

Github Repository

In order to view the project on Github, scan the QR Code provided on the first page of this report.

README FILE

Intermediate Code Generation

Intermediate Code Generation phase is the glue between the frontend and backend of the compiler design stages. The final goal of a compiler is to get programs written in a high-level language to run on a computer. This means that, eventually, the program will have to be expressed as machine code which can run on the computer. Many compilers use a medium-level language as a stepping-stone between the high-level language and the very low-level machine code. Such stepping-stone languages are called intermediate code. It provides lower abstraction from source level and maintains some high level information.

Intermediate Code can be represented in many different formats depending whether it is language-specific (e.g. Bytecode for Java) or language-independent (three-address-code). We have used three-address-code to make it independent. Most common independent intermediate representations are:

- Postfix notation
- Three Address Code
- Syntax tree

Intermediate Code generation is the 4th phase in the design of a compiler. Before it, a compiler comprises of three other phases:

- Lexical Analysis
- Parsing
- Semantic Analysis

Three Address Code

Three address code is a type of intermediate code which is easy to generate and can be easily converted to machine code. It makes use of at most three addresses and one operator to represent an expression and the value computed at each instruction is stored in a temporary variable generated by the compiler. The compiler decides the order of operation given by three address codes. General representation: $x = y \text{ op } z$ An address can be a name, constant or temporary.

- Assignments $x = y \text{ op } z$; $x = \text{op } y$.
- Copy $x = y$.
- Unconditional jump goto L.
- Conditional jumps if $x \text{ relop } y$ goto L.
- Parameters param x.
- Function call $y = \text{call } p$

Hardware Requirements

Usage of CPU, RAM, and storage space can vary significantly based on user behavior. These hardware recommendations are based on my work on the project.

- A Computer System
- Ethernet connection (LAN) OR a wireless adapter (Wi-Fi)
- Processor: Minimum 1 GHz
- Sufficient memory to install required libraries

Software Requirements

Software requirements for the project are as follows:

- GCC Compiler (for execution of C/C++ scripts)
- Flex Compiler (for execution of lex scripts)
- Bison Compiler (for execution of yacc scripts)
- Bash Shell (To run bash scripts)

File Structure

```
.
├── Intermediate Code Generation
│   ├── parser.y
│   ├── run.sh
│   ├── scanner.l
│   └── tests
├── Lexical Analysis
│   ├── Lexical Analysis.pdf
│   ├── run.sh
│   ├── scanner.l
│   └── tests
├── Parsing
│   ├── parser.y
│   ├── Parsing.pdf
│   ├── run.sh
│   ├── scanner.l
│   └── tests
├── README.md
└── Semantic Analysis
    ├── parser.y
    └── run.sh
```



```

|— scanner.l
|— Semantic Analysis.pdf
|— tests

```

8 directories, 47 files

Clearly, from the above structure, it can be seen that each phase is contained in an individual directory. So, if need be, one can see the working of each individual phase by going in the respective directory. Each directory will contain the following:

- A bash script named run.sh.
- A directory names tests with suitable test cases.
- Lex or/and Yacc code as required.
- Report regarding the working of the phase.

Execution Steps

To execute intermediate code generation, run the bash script placed in the directory. The bash script looks like this:

```

#!/bin/bash

YELLOW='\033[1;33m'
NOCOLOR='\033[0m'

function run() {
    flex scanner.l && yacc -d parser.y && gcc y.tab.c lex.yy.c -w
    local total_testcases="$1"
    echo "Running: $total_testcases"
    local start=1
    while [ $start -le $total_testcases ]
    do
        printf "\n\n"
        for i in {1..35}
        do
            echo -ne "="
        done
        echo -ne "  ${YELLOW}Running TestCase $start${NOCOLOR}  "
        for i in {1..35}
        do
            echo -ne "="
        done
        printf "\n"
        local filename="tests/test"$start".c"
        ./a.out $filename
        ((start++))
    done
}

number_of_files=`ls -l ./tests/ | egrep -c '^-'`
run $number_of_files

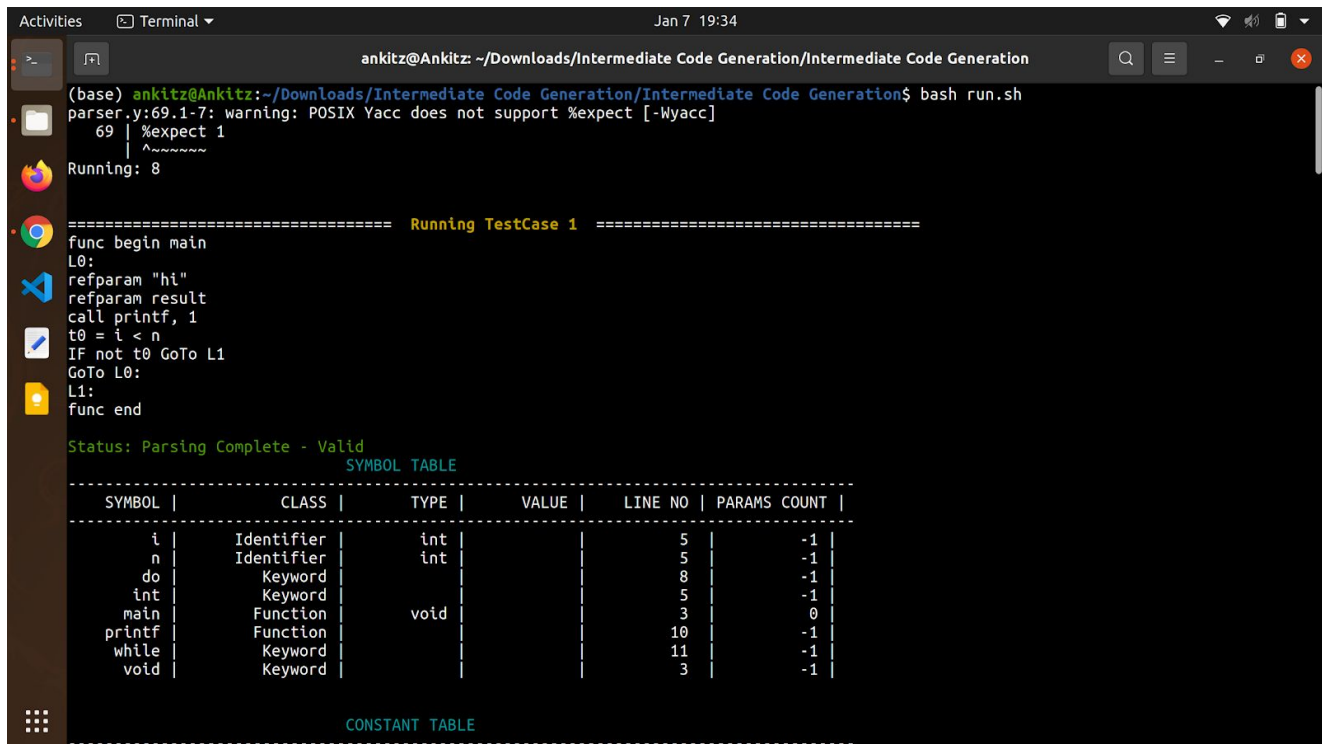
```

The workflow for a lex and yacc file **in general** is explained as under:

- Compile the script using Yacc tool
\$ yacc -d c_parser.y
- Compile the flex script using Flex tool
\$ flex c_lexer.l
- After compiling the lex file, a lex.yy.c file is generated. Also, y.tab.c and y.tab.h files are generated after compiling the yacc script.
- The two files, lex.yy.c and y.tab.c are compiled together
\$ gcc -o compiler lex.yy.c y.tab.h y.tab.c -ll

Usage Examples

Below are the screenshots of the execution of code.



```
(base) ankitz@Ankitz:~/Downloads/Intermediate Code Generation/Intermediate Code Generation$ bash run.sh
parser.y:69.1-7: warning: POSIX Yacc does not support %expect [-Wyacc]
69 | %expect 1
    | ~~~~~
Running: 8

===== Running TestCase 1 =====
func begin main
L0:
refparam "hi"
refparam result
call printf, 1
t0 = i < n
IF not t0 GoTo L1
GoTo L0:
L1:
func end

Status: Parsing Complete - Valid

SYMBOL TABLE
-----
SYMBOL | CLASS | TYPE | VALUE | LINE NO | PARAMS COUNT |
-----
i | Identifier | int | | 5 | -1 |
n | Identifier | int | | 5 | -1 |
do | Keyword | | | 8 | -1 |
int | Keyword | | | 5 | -1 |
main | Function | void | | 3 | 0 |
printf | Function | | | 10 | -1 |
while | Keyword | | | 11 | -1 |
void | Keyword | | | 3 | -1 |

CONSTANT TABLE
-----
```

```
Activities Terminal Jan 7 19:35
ankitz@Ankitz: ~/Downloads/Intermediate Code Generation/Intermediate Code Generation

===== Running TestCase 2 =====
func begin myfunc
t0 = a + b
func end

func begin main
L0:
t1 = 3
t2 = a < t1
IF not t2 GoTo L1
t3 = a + b
a = t3
t4 = 0
i = t4
L2:
t5 = i < b
IF not t5 GoTo L3
t6 = i + 1
i = t6
t7 = b + 1
b = t7
refparam a
refparam b
refparam result
call myfunc, 2
GoTo L2:
L3:
t8 = a + 1
a = t8
GoTo L0:
L1:
func end

Status: Parsing Complete - Valid
SYMBOL TABLE
```

```
Activities Terminal Jan 7 19:35
ankitz@Ankitz: ~/Downloads/Intermediate Code Generation/Intermediate Code Generation

===== Running TestCase 4 =====
func begin main
t0 = 0
t1 = 0
a = t1
L0:
t2 = 10
t3 = a < t2
IF not t3 GoTo L1
t4 = a + 1
a = t4
refparam "H1"
refparam result
call printf, 1
GoTo L0:
L1:
L2:
t5 = 0
t6 = a > t5
IF not t6 GoTo L3
GoTo L2:
L3:
L4:
t7 = 10
t8 = a < t7
IF not t8 GoTo L5
t9 = a + 1
a = t9
GoTo L4:
L5:
func end

Status: Parsing Complete - Valid
SYMBOL TABLE
```

Developer Information

Name: Ankit Kumar Singh

USN: 1RV18IS007

Institute: R. V. College of Engineering

Instructor: Prof B. K. Srinivas