

中国矿业大学计算机学院

2019 级本科生课程设计报告

课程名称 系统软件开发实践

报告时间 2022.4.10

学生姓名 王杰永

学 号 03190886

专 业 计算机科学与技术

任课教师 张博

成绩考核

编号	课程教学目标	占比	得分
1	目标 1: 针对编译器中词法分析器软件要求, 能够分析系统需求, 并采用 FLEX 脚本语言描述单词结构。	15%	
2	目标 2: 针对编译器中语法分析器软件要求, 能够分析系统需求, 并采用 Bison 脚本语言描述语法结构。	15%	
3	目标 3: 针对计算器需求描述, 采用 Flex/Bison 设计实现高级解释器, 进行系统设计, 形成结构化设计方案。	30%	
4	目标 4: 针对编译器软件前端与后端的需求描述, 采用软件工程进行系统分析、设计和实现, 形成工程方案。	30%	
5	目标 5: 培养独立解决问题的能力, 理解并遵守计算机职业道德和规范, 具有良好的法律意识、社会公德和社会责任感。	10%	
总成绩			
指导教师		评阅日期	

目录

一 综合实验 1	1
1.1 实验目的	1
1.2 实验内容	1
1.3 实验步骤	1
1.3.1 Windows 系统下进行实验	1
1.3.2 Linux 系统下进行实验	2
1.4 抽象语法树的构建	3
1.4.1 产生式	3
1.4.2 $1+2-3*2/5$ 的抽象语法树	3
1.4.3 $(1+2)-(2*6)$ 的抽象语法树	4
1.5 实验总结	4
1.5.1 你在编程过程中遇到了哪些难题?	4
1.5.2 你的收获有哪些?	5
二 综合实验 2	5
2.1 实验目的	5
2.2 实验内容	5
2.3 实验步骤	5
2.3.1 Windows 系统下进行实验	6
2.3.2 Linux 系统下进行试验	6
2.4 源码分析	6
2.4.1 fb3-2.y 源码分析	7
2.4.2 fb3-2.l 源码分析	9
2.4.3 fb3-2.h 源码分析	10
2.4.4 fb3-2.funcs.c 源码分析	12
2.5 抽象语法树的建立、遍历与计算	13
2.5.1 基本节点 ast	13
2.5.2 数字节点 numval	14
2.5.3 比较运算节点 ast	14
2.5.4 标识符节点 symref	15
2.5.5 赋值节点 symasgn	15
2.5.6 内置函数节点 fncall	16
2.5.7 自定义函数节点 ufncall	16
2.5.8 创建自定义函数	17
2.5.9 表达式求值	18
2.6 Debug 模式	19
2.7 增加函数 pow 及其重载函数的详细过程分析	19
2.7.1 二操作数 pow 函数代码改动	19
2.7.2 二操作数 pow 的计算逻辑	20
2.7.3 二操作数 pow 函数抽象语法树的建立过程	21
2.7.4 二操作数 pow 的计算结果	22
2.7.5 三操作数 pow 函数	22
2.8 实验总结	24

三 综合实验 3.....	24
3.1 实验目的.....	24
3.2 实验内容.....	25
3.3 实验步骤.....	25
3.3.1 文法修改.....	25
3.3.2 移进/归约冲突的解决	26
3.4 实验结果.....	28
3.4.1 Windows 系统下的实验结果.....	28
3.4.2 Linux 系统下的实验结果	29
3.5 实验总结.....	29
四 综合实验 4.....	29
4.1 实验目的.....	29
4.2 实验内容.....	30
4.3 项目实施思路及技术栈.....	30
4.4 实验步骤.....	31
4.4.1 修改 flex&bison 源码，添加对外计算接口	31
4.4.2 编译，打包动态链接库	32
4.4.3 前端 UI 界面	32
4.4.4 后端逻辑.....	33
4.5 成果展示.....	34
4.6 实验总结.....	35

一 综合实验 1

1.1 实验目的

阅读《flex&Bison》第三章。使用 flex 和 Bison 开发一个具有全部功能的计算器，包括如下功能：

- 1) 支持变量；
- 2) 实现复制功能；
- 3) 实现比较表达式（大于小于等）；
- 4) 实现 if/then/else 和 do/while 流程控制；
- 5) 用户可以自定义函数；
- 6) 简单的错误恢复机制。

1.2 实验内容

- 1) 阅读 flex Python 第三章 P47~60，重点学习抽象语法树。
- 2) 阅读 fb3-1.y、fb3-1.l、fb3-1funcs.c、fb3-1.h。
- 3) 撰写实验报告，结合实验结果，给出移进规约过程，即抽象语法树的构建过程，如 $(1+2)-(2*6)$ 、 $1+2-3*2/5$ 。
- 4) 提交报告和实验代码。

1.3 实验步骤

1.3.1 Windows 系统下进行实验

Windows 环境下，编译 flex 与 bison 文件后，生成了对应的词法/语法分析程序的 c 语言文件。使用 gcc 编译器编译，得到可执行程序，运行结果顺利，结果如下图所示。

```
D:\Here\recently\系统软件开发实践\综合实验\3-1\code>fb3-1.tab.exe
> |-1
=    1
> |123
=   123
> (1 + 2) - (2 * 6)
=    -9
> 1 + 2 - 3 * 2 / 5
=    1.8
> |
```

图 1-1 windows 系统下的运行结果

但是，其他使用 `cl` 编译器的同学，编译生成的 `c` 文件后，运行可执行文件，不论输入的值是多少，计算器输出值均为 0。

为了弄明白问题所在，我查看 `flex` 文件，发现其中匹配数字的模式对应的动作如下图：

```
[0-9]+ "." [0-9]* {EXP}? |
"." ? [0-9]+ {EXP}? { yylval.d = atof(yytext); return NUMBER; }
```

图 1-2 匹配数字模式的动作

我加入了如下 `printf` 语句。目的是为了查看是否成功匹配到数字（输出 `yytext`）、并且是否成功将串转换成浮点数（输出 `atof` 的返回值）。

```
[0-9]+ "." [0-9]* {EXP}? |
"." ? [0-9]+ {EXP}? {
    printf("%s, %.2f\n", yytext, atof(yytext));
    yylval.d = atof(yytext); return NUMBER; }
```

图 1-3 加入打印语句

重新编译，运行，发现数字可以成功匹配到，但是 `atof` 函数返回值始终是 0。于是可以将问题定位至 `atof` 函数的使用上。

查询手册得知，`atof` 函数位于头文件 `<stdlib.h>` 中，而程序没有引用 `stdlib` 库，因此出现了错误。故在 `flex` 文件头部加入该头文件引用，发现可以解决问题。而使用 `gcc` 编译时，自动引入 `stdlib` 库，因此不会出现问题。

1.3.2 Linux 系统下进行实验

Linux 系统下由于不存在 `cl` 编译器，默认安装了 `gcc` 编译器，因此在编译后的运行阶段没有任何错误。程序结果如下图。

```
chen@chen: ~/桌面/chen/e5/3-1
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
chen@chen:~/桌面/chen/e5/3-1$ bison -d fb3-1.y
chen@chen:~/桌面/chen/e5/3-1$ flex -ofb3-1.lex.c fb3-1.l
chen@chen:~/桌面/chen/e5/3-1$ gcc fb3-1.tab.c fb3-1.lex.c fb3-1funcs.c -lm -o fb3-1
fb3-1.tab.c: In function 'yyparse':
fb3-1.tab.c:1137:16: warning: implicit declaration of function 'yylex' [-Wimplicit-function-declaration]
    yychar = yylex ();
                  ^
fb3-1funcs.c: In function 'main':
fb3-1funcs.c:104:10: warning: implicit declaration of function 'yyparse' [-Wimplicit-function-declaration]
    return yyparse();
           ^
chen@chen:~/桌面/chen/e5/3-1$ fb3-1
fb3-1: 未找到命令
chen@chen:~/桌面/chen/e5/3-1$ ./fb3-1
> | -1
= -1
> | 123
= 123
> | (1 + 2) - (2 * 6)
= -9
> | 1 + 2 - 3 * 2 / 5
= 1.8
```

图 1-4 Linux 系统下的运行结果

1.4 抽象语法树的构建

在上次实验中就已经发现，Bison 默认采用 LALR(1)的语法分析方法。向后多看一个字符，即可以构建整棵抽象语法树。

1.4.1 产生式

查看 *fb3-1.y* 文件，我们从中可以读出该计算器语法分析所用的产生式。“计算器文法”忽略开始符号 *calclist* 与结束符号 *EOL* 后的产生式集合主要如下：

```

exp → factor
exp → exp + factor
exp → exp - factor
factor → term
factor → factor * term
factor → factor / term
term → NUMBER
term → |term
term → (exp)
term → -term

```

基于此，我们便可以构建特定表达式的抽象语法树了，如图 1-5、1-6 所示

1.4.2 1+2-3*2/5 的抽象语法树

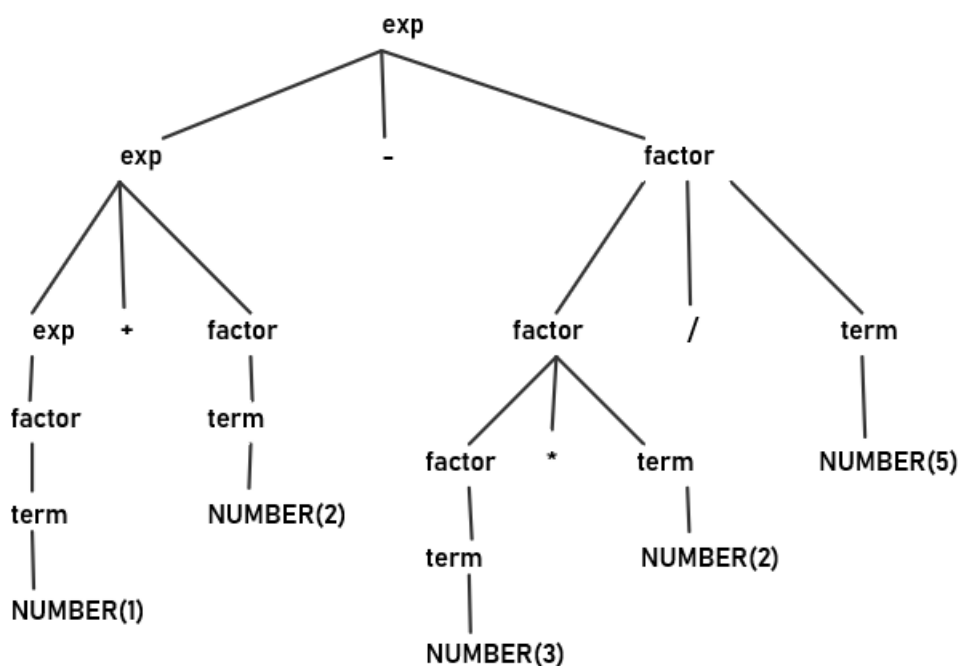


图 1-5 抽象语法树

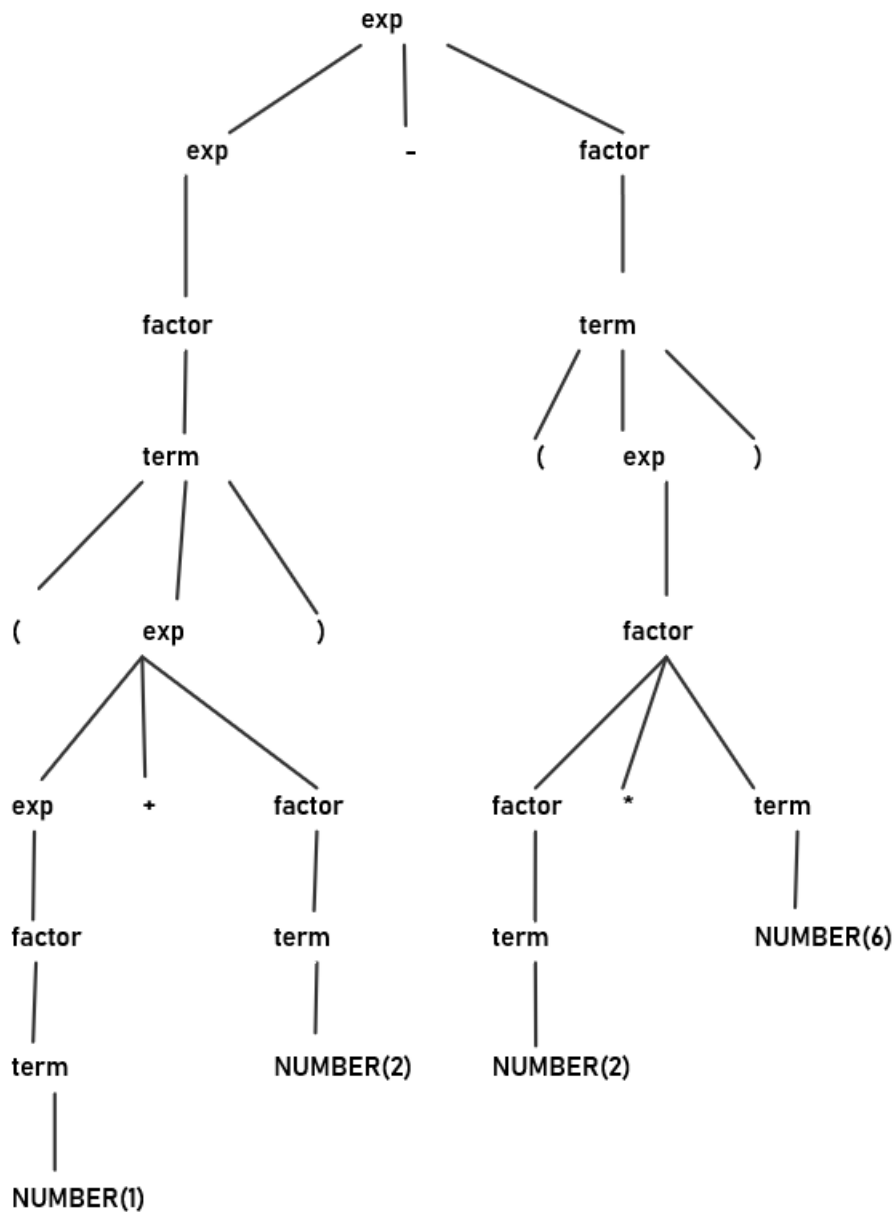
1.4.3 $(1+2)-(2*6)$ 的抽象语法树

图 1-6 抽象语法树

1.5 实验总结

1.5.1 你在编程过程中遇到了哪些难题？

正如在 1.3.1 部分所述，由于在 windows 系统下也使用 gcc 编译器进行编译，因此实验完成的很顺利，没有遇到任何问题。

不过在与其他同学交流后，发现使用 cl 编译器编译后，计算器的输出结果恒为 0。

由此我也动手尝试使用 `cl` 编译器,遇到了同样的问题。解决的步骤已于 1.3.1 部分体现。

1.5.2 你的收获有哪些?

本次实验的源代码相较于前两周更为复杂,我也得以阅读并分析更复杂的词法、语法规则。对计算机语言编译过程中的前端工作有了更深入的理解。

通过解决 `cl` 编译器环境下遇到的问题,我也收获了 `gcc` 编译器的相关知识。

二 综合实验 2

2.1 实验目的

阅读《flex&Bison》第三章。使用 `flex` 和 `Bison` 开发一个具有全部功能的计算器,包括如下功能:

- 1) 支持变量;
- 2) 实现复制功能;
- 3) 实现比较表达式(大于小于等);
- 4) 实现 `if/then/else` 和 `do/while` 流程控制;
- 5) 用户可以自定义函数;
- 6) 简单的错误恢复机制。

2.2 实验内容

- 5)阅读《Flex&Bison》第三章 P60~P79,学习抽象语法树;
- 6)阅读 `fb3-2.y`、`fb3-2.l`、`fb3-2funcs.c`、`fb3-2.h`;
- 7)使用内置函数 `sqrt(n)`、`exp(n)` , `log(n)`
- 8)定义函数 `sq(n)`、`avg(a, b)`,用于计算平方根;
- 9)撰写实验报告,结合实验结果,给出抽象语法树的构建过程。。

2.3 实验步骤

有了上一次对于 `atof` 函数的错误处理策略,本次实验在 Windows 与 linux 系统下都进行的较为顺利。

2.3.1 Windows 系统下进行实验

编译 Flex 与 Bison 文件后，使用 gcc 编译生成的全部 c 文件，得到可执行程序。实验顺利，结果截图如下。

```
D:\Here\recently\系统软件开发实践\综合实验\3-2\code>fb3-2.tab.exe
> sqrt(10)
= 3.162
> exp(2)
= 7.389
> log(7.389)
= 2
> let sq(n) = e = 1; while| ((t = n / e) - e) > .001 do e = avg(e, t);;
Defined sq
> let avg(a, b) = (a + b) / 2;
Defined avg
> sq(10)
= 3.162
> sq(10) - sqrt(10)
= 0.000178
> |
```

图 2-1 Windows 系统下的运行结果

2.3.2 Linux 系统下进行试验

Linux 系统下同理，使用 gcc 编译 flex 与 bison 生成的 c 文件，计算器成功运行。结果如下图。

```
chen@chen:~/桌面/chen/e6$ ^C
chen@chen:~/桌面/chen/e6$ bison -d fb3-2.y
chen@chen:~/桌面/chen/e6$ flex -offb3-2.lex.c fb3-2.l
chen@chen:~/桌面/chen/e6$ gcc fb3-2.tab.c ffb3-2.lex.c fb3-2funcs.c -o fb3-2 -lmfb3-2.tab.c: In function 'yyparse':
fb3-2.tab.c:1184:16: warning: implicit declaration of function 'yylex' [-Wimplicit-function-declaration]
    yychar = yylex ();
                  ^~~~~~
fb3-2funcs.c: In function 'main':
fb3-2funcs.c:438:10: warning: implicit declaration of function 'yyparse' [-Wimplicit-function-declaration]
    return yyparse();
           ^~~~~~
chen@chen:~/桌面/chen/e6$ ./fb3-2
> sqrt(10)
= 3.162
> exp(2)
= 7.389
> log(7.389)
= 2
> let sq(n) = e = 1; while | ((t = n / e) - e) > .001 do e = avg(e, t);;
Defined sq
> let avg(a, b) = (a + b) / 2;
Defined avg
> sq(10)
= 3.162
> sq(10) - sqrt(10)
= 0.000178
> |
```

图 2-2 Linux 系统下的运行结果

2.4 源码分析

本次实验相较于上次，源码在代码量上增加了很多，同时计算器的功能也更为复杂。以下对源码的分析主要围绕如何在已有内置函数 `sqrt`、`exp`、`log`、`print` 的基础上，增加

一个新的内置函数 **pow** 及其重载形式展开。

2.4.1 fb3-2.y 源码分析

首先分析 Yacc 文件。分析器使用 `%union` 来声明语法分析器中符号值的类型。查阅参考书得知，在 **bison** 语法分析器中，每一个语法符号（不论是终结符和非终结符），都可以有一个相应的值。默认情况下这个值是整型，但对于有着更多功能的语法分析器而言，一个整型数字显然不够，因此定义了 `%union` 联合体来记录语法符号的值。

```
1. %union {
2.     struct ast *a;
3.     double d;
4.     struct symbol *s;          /* which symbol */
5.     struct symlist *sl;
6.     int fn;                   /* which function */
7. }
```

当定义了 **union** 类型后，一个很直接的问题是，我们需要告知 **bison** 每个语法符号使用的是 **union** 中的哪种类型的值。查阅参考书得知，声明 **token** 时，在 **token** 名前使用 `<>`，尖括号中写入对应的类型即可。

```
1. /* declare tokens */
2. %token <d> NUMBER
3. %token <s> NAME
4. %token <fn> FUNC
5. %token EOL
```

例如，计算器的 **NUMBER** 语法符号使用 **union** 中的类型 **d**，即 **double** 类型；而 **FUNC** 语法符号使用 **union** 中的 **fn**，即表明内置函数类型的整型数。这里也启发我，如果要新增一个自定义函数的话，需要将为 **fn** 引入新的取值。

最后，计算器的语法规则及相应的语法制导翻译模式如下：

```
1. stmt: IF exp THEN list          { $$ = newflow('I', $2, $4, NULL); }
2.     | IF exp THEN list ELSE list { $$ = newflow('I', $2, $4, $6); }
3.     | WHILE exp DO list         { $$ = newflow('W', $2, $4, NULL); }
4.     | exp
5. ;
6.
7. list: /* nothing */ { $$ = NULL; }
8.     | stmt ';' list { if ($3 == NULL)
```

```

9.         $$ = $1;
10.        else
11.        $$ = newast('L', $1, $3);
12.        }
13.    ;
14.
15. exp: exp CMP exp      { $$ = newcmp($2, $1, $3); }
16.   | exp '+' exp      { $$ = newast('+', $1,$3); }
17.   | exp '-' exp      { $$ = newast('-', $1,$3); }
18.   | exp '*' exp      { $$ = newast('*', $1,$3); }
19.   | exp '/' exp      { $$ = newast('/', $1,$3); }
20.   | '|' exp          { $$ = newast('|', $2, NULL); }
21.   | '(' exp ')'      { $$ = $2; }
22.   | '-' exp %prec UMINUS { $$ = newast('M', $2, NULL); }
23.   | NUMBER          { $$ = newnum($1); }
24.   | FUNC '(' explist ')' { $$ = newfunc($1, $3); }
25.   | NAME            { $$ = newref($1); }
26.   | NAME '=' exp     { $$ = newasgn($1, $3); }
27.   | NAME '(' explist ')' { $$ = newcall($1, $3); }
28. ;
29.
30. explist: exp
31.   | exp ',' explist { $$ = newast('L', $1, $3); }
32. ;
33. symlist: NAME      { $$ = newsymlist($1, NULL); }
34.   | NAME ',' symlist { $$ = newsymlist($1, $3); }
35. ;
36.
37. calclist: /* nothing */
38.   | calclist stmt EOL {
39.       if(debug) dumpast($2, 0);
40.       printf("= %4.4g\n> ", eval($2));
41.       treefree($2);
42.   }
43.   | calclist LET NAME '(' symlist ')' '=' list EOL {
44.       dodef($3, $5, $8);
45.       printf("Defined %s\n> ", $3->name); }
46.
47.   | calclist error EOL { yyerrok; printf("> "); }
48. ;

```

可以看出，语法分析的产生式规则较为简单。值得注意的是，在行 39 出现了关键字"debug"，由此可以看出该计算器具有 debug 功能。

2.4.2 fb3-2.l 源码分析

对于词法分析的 `flex` 文件，通过前面若干次实验，已经比较熟悉了。

值得注意的一些规则如下：

```
1. /* comparison ops */
2. ">"      { yylval.fn = 1; return CMP; }
3. "<"      { yylval.fn = 2; return CMP; }
4. "<>"     { yylval.fn = 3; return CMP; }
5. "=="     { yylval.fn = 4; return CMP; }
6. ">="     { yylval.fn = 5; return CMP; }
7. "<="     { yylval.fn = 6; return CMP; }
8.
9. /* built in functions */
10. "sqrt"   { yylval.fn = B_sqrt; return FUNC; }
11. "exp"    { yylval.fn = B_exp; return FUNC; }
12. "log"    { yylval.fn = B_log; return FUNC; }
13. "print"  { yylval.fn = B_print; return FUNC; }
```

从这里可以看出，无论匹配到的 `token` 是 `CMP` 还是 `FUNC`，都使用 `union` 中的 `fn` 作为区别。可以推断出，`fn` 的值不能唯一确定当前的语法符号，语法制导翻译会通过 `fn` 与 `token` 类型共同确定。

在上述代码段中还可以看到，内置函数 `sqrt`、`exp` 等的词法规则为 `union` 的 `fn` 赋值 `B_sqrt`、`B_exp`。在 `fb3-2.h` 文件中可以找到这些变量的定义。

```
1. enum bifs {          /* built-in functions */
2.   B_sqrt = 1,
3.   B_exp,
4.   B_log,
5.   B_print
6. };
```

这些变量是枚举变量的常数值。这样的编码风格使得程序的可读性强，容易维护。同时，如果我们想增加内置函数的话，需要在 `bifs` 枚举类型下增加新的枚举值。

还有这样几个词法匹配模式值得我们去注意：

```
1. "//".*
2. [\t]  /* ignore white space */
```

这两条词法模式匹配规则，由于其后没有任何动作，因此计算器允许我们输入"/"作为注释或是任意空格、制表符来“美化”输入。

```
1. \\n    printf("c> "); /* ignore line continuation */
2. "\n"   { return EOL; }
```

最后一条规则很简单，当我们输入换行符时，返回一个语法分析的结束符号 EOL 以此结束当前表达式，从而求值。

"\\n"规则则允许我们“换行继续输入”，这里的反斜杠是转义符的含意，当匹配到反斜杠且其后紧跟换行符时，当前表达式的计算不会停止，词法分析器仅输出一个"c>"符号以此表示该行输入紧跟上一行。即，该计算器具有分多行输入的功能。

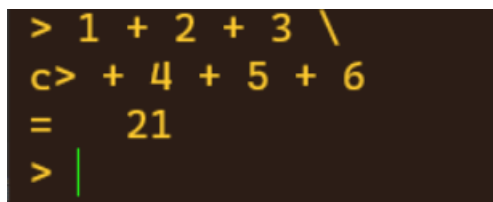


图 2-3 计算器具有多行输入功能

最后，还有着 debug 模式的词法匹配规则。规则如下。

```
1. /* debug hack */
2. "debug"[0-9]+ { debug = atoi(&yytext[5]); printf("debug set to %d\n",
    debug); }
```

从这里也能看出，当输入 debug+ 数字的时候，会进入到 debug 模式。将在 xxxxxxxx 部分演示。

2.4.3 fb3-2.h 源码分析

与上一次实验一样，由于词法/语法规则足够多，因此将全部的辅助 c 代码放于另外一组文件中——fb3-2.h 与 fb3-2funcs.c.

在 fb3-2.h 头文件中，定义了各类全局变量与不同语法符号的结构体（抽象语法树节点）信息。

◆ 符号表

```
1. /* symbol table */
2. struct symbol {          /* a variable name */
3.     char *name;
4.     double value;
```

```

5.  struct ast *func;    /* stmt for the function */
6.  struct symlist *syms; /* list of dummy args */
7.  };
8.
9.  /* simple symtab of fixed size */
10. #define NHASH 9997
11. struct symbol symtab[NHASH];

```

符号表使用结构体定义。每一个符号都可以有一个变量和一个用户自定义函数。
value 域用来保存符号值，**func** 域指向用抽象语法树表示的该函数用户代码。

◆ 各种不同类型的节点

```

1. struct ast {
2.     int nodetype;
3.     struct ast *l;
4.     struct ast *r;
5. };
6. struct fncall {          /* built-in function */
7.     int nodetype;        /* type F */
8.     struct ast *l;
9.     enum bifs functype;
10. };
11. struct ufncall {        /* user function */
12.     int nodetype;        /* type C */
13.     struct ast *l;       /* list of arguments */
14.     struct symbol *s;
15. };
16. struct flow {
17.     int nodetype;        /* type I or W */
18.     struct ast *cond;    /* condition */
19.     struct ast *tl;      /* then or do list */
20.     struct ast *el;      /* optional else list */
21. };
22. struct numval {
23.     int nodetype;        /* type K */
24.     double number;
25. };
26. struct symref {
27.     int nodetype;        /* type N */
28.     struct symbol *s;
29. };
30. struct symasgn {
31.     int nodetype;        /* type = */

```

```

32.  struct symbol *s;
33.  struct ast *v;      /* value */
34. };

```

计算器的功能相较于上一次实验增多了很多，抽象语法树节点类型的增多是最直接的体现。

以上是各类节点的定义。每一个节点都有一个 `nodetype`，遍历树的代码使用这个变量来判断当前访问的节点类型——`nodetype` 的取值及含意在源码的注释中已经全部表述。

```

1.  /* nodes in the Abstract Syntax Tree */
2.  /* all have common initial nodetype */
3.
4.  /* node types
5.   * + - * / |
6.   * 0-7 comparison ops, bit coded 04 equal, 02 less, 01 greater
7.   * M unary minus
8.   * L statement list
9.   * I IF statement
10.  * W WHILE statement
11.  * N symbol ref
12.  * = assignment
13.  * S list of symbols
14.  * F built in function call
15.  * C user function call
16.  */

```

通过分析源码可以发现，内置函数使用 `fncall` 节点，流程控制表达式使用 `flow` 节点，常量使用 `numval` 节点，符号引用使用 `symref` 节点。因此，若想要增加一个自定义的内置函数，在语法制导翻译时，需要使用 `ufncall` 节点类型。

2.4.4 fb3-2.funcs.c 源码分析

该文件中是对 `fb3-2.h` 头文件中各个已经声明的函数的实现。包括抽象语法树的构造函数——`newast()`, `newcmp()`, `newfunc()`等，抽象语法树的求值函数 `eval()`，定义新的内置函数的函数 `dodef()`，释放语法树所占空间的函数 `treefree()`，以及 `debug` 模式下起作用的 `dumpast()`函数。

该部分源码定义了抽象语法树的建立逻辑，代码量较大。在部分 5 抽象语法树的建立中详细描述。

2.5 抽象语法树的建立、遍历与计算

在 4.3 中已经提到，该桌面计算器的源码中定义了多种类型的节点结构，根据归约出的不同语法符号，会创建不同类型的节点。

2.5.1 基本节点 ast

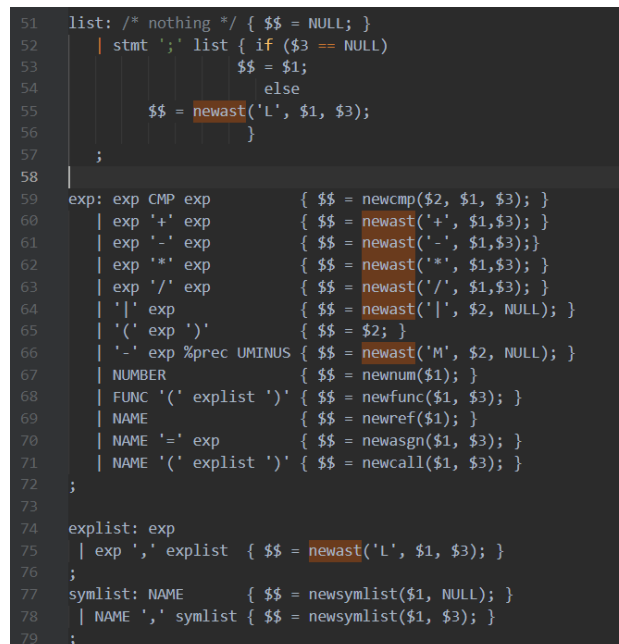
创建基本节点 ast 的函数 `newast()` 的源码如下。

```

1. struct ast *
2. newast(int nodetype, struct ast *l, struct ast *r)
3. {
4.     struct ast *a = malloc(sizeof(struct ast));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = nodetype;
11.    a->l = l;
12.    a->r = r;
13.    return a;
14. }

```

可以看出，ast 节点有左右两棵子树，因此推断二元运算符才会创建 ast 节点。我们在语法分析的.y 文件中查找 `newast` 函数，得到如下结果。



```

51 list: /* nothing */ { $$ = NULL; }
52 | stmt ';' list { if ($3 == NULL)
53     $$ = $1;
54     else
55     $$ = newast('L', $1, $3);
56 }
57 ;
58
59 exp: exp CMP exp { $$ = newcmp($2, $1, $3); }
60 | exp '+' exp { $$ = newast('+', $1,$3); }
61 | exp '-' exp { $$ = newast('-', $1,$3); }
62 | exp '*' exp { $$ = newast('*', $1,$3); }
63 | exp '/' exp { $$ = newast('/', $1,$3); }
64 | '|' exp { $$ = newast('|', $2, NULL); }
65 | '(' exp ')' { $$ = $2; }
66 | '-' exp %prec UMINUS { $$ = newast('M', $2, NULL); }
67 | NUMBER { $$ = newnum($1); }
68 | FUNC '(' explist ')' { $$ = newfunc($1, $3); }
69 | NAME { $$ = newref($1); }
70 | NAME '=' exp { $$ = newasgn($1, $3); }
71 | NAME '(' explist ')' { $$ = newcall($1, $3); }
72 ;
73
74 explist: exp
75 | exp ',' explist { $$ = newast('L', $1, $3); }
76 ;
77 symlist: NAME { $$ = newsymlist($1, NULL); }
78 | NAME ',' symlist { $$ = newsymlist($1, $3); }
79 ;

```

图 2-4 语法分析中查找 `newast` 函数

对于基本运算符"+-*/"以及"|"运算符，均以左右两表达式的值\$1、\$3 来创建 ast 节点。同时，在图 2-4 的行 55 可以看出，整棵抽象语法树的根节点类型也为 ast。

2.5.2 数字节点 numval

```
1. struct ast *
2. newnum(double d)
3. {
4.     struct numval *a = malloc(sizeof(struct numval));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = 'K';
11.    a->number = d;
12.    return (struct ast *)a;
13. }
```

数字类型节点的 **nodetype** 为 'K'。由于数字类型节点一定是抽象语法树的叶子节点，因此 **numval** 结构不含左右子树指针。创建完成后，将数字类型强转为类型 **ast** 返回。

2.5.3 比较运算节点 ast

```
1. struct ast *
2. newcmp(int cmptype, struct ast *l, struct ast *r)
3. {
4.     struct ast *a = malloc(sizeof(struct ast));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = '0' + cmptype;
11.    a->l = l;
12.    a->r = r;
13.    return a;
14. }
```

可以看出，比较运算创建的节点类型仍是 **ast**，之所以不能和普通的加减乘除写在一个函数内的原因是比较运算的 **nodetype** 确定方法与加减乘除不同。

2.5.4 标识符节点 symref

```
1. struct ast *
2. newref(struct symbol *s)
3. {
4.     struct symref *a = malloc(sizeof(struct symref));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = 'N';
11.    a->s = s;
12.    return (struct ast *)a;
13. }
```

结合 symref 的结构定义，对函数 newref 可以更好地理解。

```
1. struct symref {
2.     int nodetype;           /* type N */
3.     struct symbol *s;
4. };
```

S 是指向符号表的指针。标识符节点会通过该指针连接到符号表中的某一符号，使用目标符号的值代替该标识符的值。

2.5.5 赋值节点 symasgn

```
1. struct ast *
2. newasgn(struct symbol *s, struct ast *v)
3. {
4.     struct symasgn *a = malloc(sizeof(struct symasgn));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = '=';
11.    a->s = s;
12.    a->v = v;
13.    return (struct ast *)a;
14. }
```

赋值符号的规则为标识符=表达式，因此在 `symasgn` 节点中，使用指针 `s` 指向符号表中的标识符，使用指针 `v` 指向一颗抽象语法树 `ast`，使用 `ast` 的值赋值给标识符。

2.5.6 内置函数节点 `fncall`

```
1. struct ast *
2. newfunc(int functype, struct ast *l)
3. {
4.     struct fncall *a = malloc(sizeof(struct fncall));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = 'F';
11.    a->l = l;
12.    a->functype = functype;
13.    return (struct ast *)a;
14. }
```

`nodetype` 为 'F' 的内置函数节点仅有一棵子树，但增加了 `functype` 域，值为枚举变量 `bifs`。从此处再次看出，如果要增加内置函数的话，需要为枚举类型 `bifs` 增加新值。

2.5.7 自定义函数节点 `ufncall`

```
1. struct ast *
2. newcall(struct symbol *s, struct ast *l)
3. {
4.     struct ufncall *a = malloc(sizeof(struct ufncall));
5.
6.     if(!a) {
7.         yyerror("out of space");
8.         exit(0);
9.     }
10.    a->nodetype = 'C';
11.    a->l = l;
12.    a->s = s;
13.    return (struct ast *)a;
14. }
```

当我们使用 `let` 保留字创建好了自定义函数后，再次使用函数时会创建 `ufncall` 节点。

在计算器的语法分析部分中，我们找到了如图所示的两条文法规则。`symlist` 定义了函数参数列表的匹配规则，调用 `newsymlist` 方法为 `symlist` 中添加参数。`calclist` 定义了标识符 `Let` 的使用规则：

`let 函数名(参数列表) = 表达式`

图 2-5 创建自定义函数的文法规则

创建自定义函数时，调用了 `dodef` 函数，参数为 `$3,$5,$8` 即（标识符，参数列表，表达式）。

```
1.  /* define a function */
2.  void
3.  dodef(struct symbol *name, struct symlist *syms, struct ast *func)
4.  {
5.      if(name->syms) symlistfree(name->syms);
6.      if(name->func) treefree(name->func);
7.      name->syms = syms;
8.      name->func = func;
9.  }
```

结合源码与实验参考书可以得知，当函数被定义时，参数列表和抽象语法树将被简单的保存到符号表中函数名对应的条目中，同时替换了任意可能的旧版本。

2.5.9 表达式求值

对于抽象语法树的求值方法 `eval`，根据节点 `nodetype` 的不同，选用不同的方法递归的调用 `eval` 求值。

部分选择与递归计算的源码如下。

```

1.  /* expressions */
2.  case '+': v = eval(a->l) + eval(a->r); break;
3.  case '-': v = eval(a->l) - eval(a->r); break;
4.  case '*': v = eval(a->l) * eval(a->r); break;
5.  case '/': v = eval(a->l) / eval(a->r); break;
6.  case '|': v = fabs(eval(a->l)); break;
7.  case 'M': v = -eval(a->l); break;
8.
9.  /* comparisons */
10. case '1': v = (eval(a->l) > eval(a->r)) ? 1 : 0; break;
11. case '2': v = (eval(a->l) < eval(a->r)) ? 1 : 0; break;
12. case '3': v = (eval(a->l) != eval(a->r)) ? 1 : 0; break;
13. case '4': v = (eval(a->l) == eval(a->r)) ? 1 : 0; break;
14. case '5': v = (eval(a->l) >= eval(a->r)) ? 1 : 0; break;
15. case '6': v = (eval(a->l) <= eval(a->r)) ? 1 : 0; break;

```

当 `nodetype` 为内置函数时，调用 `callbuiltin` 方法，根据 `functype` 枚举值的不同，进而得知当前需要按照哪一个内置函数去计算。由此，如果要新增一个内置函数的话，需要在 `callbuiltin` 方法中增加一个对应枚举值的 `case`。

```

1. static double
2. callbuiltin(struct fncall *f)
3. {
4.     enum bifs functype = f->functype;
5.     double v = eval(f->l);
6.
7.     switch(functype) {
8.     case B_sqrt:
9.         return sqrt(v);
10.    case B_exp:
11.        return exp(v);
12.    case B_log:
13.        return log(v);

```

```

14. case B_print:
15.     printf("= %4.4g\n", v);
16.     return v;
17. default:
18.     yyerror("Unknown built-in function %d", functype);
19.     return 0.0;
20. }
21. }

```

2.6 Debug 模式

当我们输入 debug1 的时候，计算器会进入 debug 的模式 1。

此时，输入任何表达式，会在求值前打印其所建立的抽象语法树的信息。

```

> debug1 a = 1 * 2
debug set to 1
= a
  binop *
    number 1
    number 2
= 2
> 5 + (3 - 4)
binop +
  number 5
  binop -
    number 3
    number 4
= 4
>

```

图 2-6 debug 模式

2.7 增加函数 pow 及其重载函数的详细过程分析

2.7.1 二操作数 pow 函数代码改动

通过前面的分析，增加内置函数需要改动源码中的三处。

- 1) 词法分析 fb3-2.1 中，新增保留字匹配模式 pow.
- 2) fb3-2.h 头文件中，为枚举类型 bifs 增加取值 B_pow.

<pre> 48 /* built in functions */ 49 "sqrt" { yyval.fn = B_sqrt; return FUNC; } 50 "exp" { yyval.fn = B_exp; return FUNC; } 51 "log" { yyval.fn = B_log; return FUNC; } 52 "print" { yyval.fn = B_print; return FUNC; } 53 "pow" { yyval.fn = B_pow; return FUNC; } 54 </pre>	<pre> 48 enum bifs { /* built-in functions */ 49 B_sqrt = 1, 50 B_exp, 51 B_log, 52 B_print, 53 B_pow 54 }; 55 </pre>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------

图 2-7 前两处代码改动

3) fb3-2funcs.c 中，为函数 `callbuiltin` 中增加 `case B_pow`，并设置相应的计算方法

```

287 static double
288 callbuiltin(struct fncall *f)
289 {
290     enum bifs functype = f->functype;
291     double v = eval(f->l);
292
293     switch(functype) {
294     case B_sqrt:
295         return sqrt(v);
296     case B_exp:
297         return exp(v);
298     case B_log:
299         return log(v);
300     case B_print:
301         printf("= %4.4g\n", v);
302         return v;
303     case B_pow:
304         return pow(eval(f->l->l), v);
305     default:
306         yyerror("Unknown built-in function %d", functype);
307         return 0.0;
308     }
309 }
310

```

图 2-8 增加计算逻辑

其中，前两处改动较为容易理解，但第三处的改动则需要一些特殊的处理。在 7.2 部分中详细叙述。

2.7.2 二操作数 pow 的计算逻辑

`pow` 函数与其它内置函数如 `sqrt`、`log` 的不同之处在于，`pow` 的有两个操作数，这使得其计算逻辑与其余单操作数函数有很大的变化。

在图 2-8 中可以看出，`callbuiltin` 方法先调用 `eval` 计算 `fncall` 节点的子树，得到子树的值后，根据 `functype` 的不同或对结果开方，或对结果取对数等。

由于 `sqrt` 等方法是单操作数的，因此才可以使用 `eval` 的返回值去直接计算。

在双操作数的方法中，我们使用 `eval` 计算子树的值并返回赋值给变量 `v`，那么，计算后 `v` 的值是什么就成了很关键的一个问题。

下式是内置函数计算的语法规则，`explist` 是归约出的参数列表。

$$exp \rightarrow FUNC (explist)$$

观察 `explist` 的语法规则，发现对应的语法制导翻译动作调用 `newast` 方法，创建了 `nodetype` 为 `L` 的 `ast` 节点。

查看 `eval` 方法对 `nodetype` 为 `L` 的节点的计算逻辑，问题得以解决。


```

1. double v;
2. switch(a->nodetype){
3. ...
4. case 'L': eval(a->l); v = eval(a->r); break;
5. ...
6. }
7. return v;

```

可以看出，分别计算左右子树的值后，**返回了右子树的值**。返回右子树的值的原因也就很清晰了：函数的参数列表是**逗号表达式**，逗号表达式的值在一般的程序设计语言中都被定义为最右侧表达式的值。

因此，在图 2-8 中，首先计算了 `pow(x, y)` 的 `y` 的值赋值给 `v`，我们需要重新遍历一遍语法树，计算 `x` 的值（即调用 `eval(f->l->l)`），由此可以得到 `pow` 函数的计算逻辑。

2.7.3 二操作数 `pow` 函数抽象语法树的建立过程

- 词法分析匹配到保留字"`pow`"，为 `yyval` 赋值 `B_pow`，返回语法分析标识符 `FUNC`
- 标识 `FUNC` 压入语法分析符号栈，当前句型的句柄未出现，继续读入下一个单词 `token`
- 词法分析匹配到`'('`，返回语法分析标识符`'('`
- 标识`'('`压入语法分析符号栈，当前句型的句柄未出现，继续读入下一个单词 `token`
- 词法分析分别匹配到 `pow` 函数的两个参数`(x,y)`，交给语法分析
- 语法分析将参数 `x, y` 规约为 `explist`，执行语法制导翻译动作

$$\{\$ \$ = \text{newast}('L', \$1, \$3); \}$$

- 调用 `newast` 方法，创建新的抽象语法树节点，左孩子为 `x`，右孩子为 `y`。

- 将归约出的标识 **explist** 压入语法分析符号栈，当前句型句柄未出现，继续读入下一个单词 **token**
- 词法分析匹配到')'，返回语法分析标识符')

标识')'压入语法分析符号栈，句柄出现，使用下表达式归约。

$$exp \rightarrow FUNC (explist)$$

- 规约后，执行该产生式对应语法制导翻译的动作{ \$\$ = newfunc(\$1, \$3); }, 进而为抽象语法树创建 **pow** 函数的节点。

2.7.4 二操作数 pow 的计算结果

计算结果如下图所示：

```
D:\Here\recently\系统软件开发实践\综合实验\3-2\code>fb3-2.exe
> pow(2, 3)
=      8
> pow(3, 2)
=      9
> pow(1 + 2, 2 * 1)
=      9
> |
```

图 2-9 pow 计算结果

2.7.5 三操作数 pow 函数

实现了二操作数的 **pow** 函数，对于其重载函数——三操作数的 **pow** 函数的添加就很简单了。

三操作数的 **pow** 函数的表达式如下：

$$pow(x, y, c) = x^y \% c$$

由于是重载函数，对于 6.1 中的前两处修改不再需要，只需要将第三处即 `pow` 的计算逻辑处增加额外的修改即可。

```

303 case B_pow:
304     if(f->l->r->nodetype == 'L'){ //说明是带取余的pow，三操作数
305         /*
306             v: 取余
307             eval(f->l->l): x
308             eval(f->l->r->l): y
309         */
310         double x = eval(f->l->l);
311         double y = eval(f->l->r->l);
312         double ans = pow(x, y);
313         ans = ans - (int) (ans / v) * v;
314         return ans;
315     } else {
316         return pow(eval(f->l->l), v);
317     }
318 default:
319     yyerror("Unknown built-in function %d", functype);
320     return 0.0;
321 }
322 }

```

图 2-10 三操作数 `pow` 的代码修改

我们使用函数节点的右孩子节点的 `nodetype` 来判断当前是二操作数还是三操作数的 `pow` 函数。其抽象语法树的节点类型如图 2-11 所示。节点上的字母为 `nodetype` 类型。

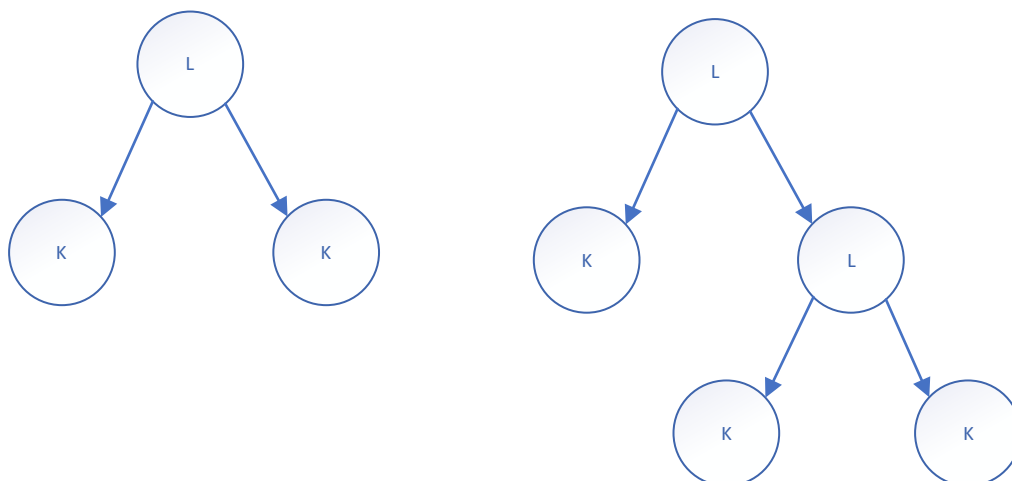
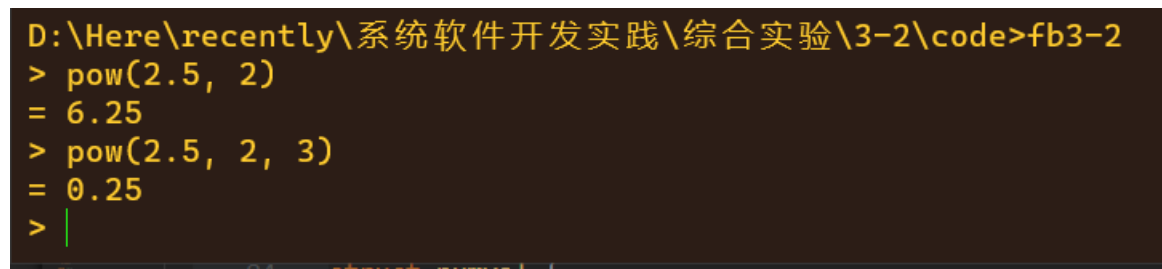


图 2-11 二操作数（左）与三操作数（右）`pow` 函数的抽象语法树

可见，三操作数的 `pow` 函数与二操作数 `pow` 函数的区别在于根节点的右孩子的节点类型——三操作数为逗号表达式（L），而二操作数为数字节点（K）。因此，在图 2-10

的代码中，行 304 的 if 语句可以实现对两个重载函数的区分。

运行结果如图所示。



```
D:\Here\recently\系统软件开发实践\综合实验\3-2\code>fb3-2
> pow(2.5, 2)
= 6.25
> pow(2.5, 2, 3)
= 0.25
> |
```

图 2-12 pow 重载的运行结果

2.8 实验总结

本次实验收获颇丰。

结合参考书，理解了相当一部分的源代码，并在源码中发掘出了很多实验要求中没有涉及到的计算器模式——多行输入（图 2-3）、debug 模式（图 2-6）等。在仔细理解源码的基础上，修改源码，增加了内置函数 `pow(x,y)` 及其重载的三操作数形式 `pow(x,y,c)`（图 2-12），再次体会到了代码阅读能力与实践编码能力的重要性。

最后，对整个桌面计算器的源码架构与工作逻辑有了更深刻的理解。

三 综合实验 3

3.1 实验目的

阅读《flex&Bison》第三章。使用 flex 和 Bison 开发一个具有全部功能的计算器，包括如下功能：

- 1) 支持变量；
- 2) 实现复制功能；
- 3) 实现比较表达式（大于小于等）；
- 4) 实现 if/then/else 和 do/while 流程控制；
- 5) 用户可以自定义函数；
- 6) 简单的错误恢复机制。

3.2 实验内容

- 10) 阅读《Flex&Bison》第三章 P79 习题 1，学习抽象语法树
- 11) 修改 fb3-2 的相关代码，实现特定函数，并保存为 fb3-3
- 12) 撰写实验报告，结合实验结果，给出抽象语法树的构建过程

3.3 实验步骤

本次实验的重点与前两次综合实验略有不同，前两次实验考察词法、语法分析过程并给出源代码；而本次实验则侧重于编译前端的最后一部分——语法制导翻译。

实验中，我们需要修改特定文法或是语法制导翻译模式，使计算器支持特定的语法。

3.3.1 文法修改

实验要求我们修改计算器自定义函数的语句 `let`，使之支持**大括号**作为函数体作用范围的区域；同时在 `if/else` 等选择判断分支语句处，也要支持**大括号**的使用。

首先观察 `let` 的语法分析段，如下：

```

1. calclist: /* nothing */
2. | calclist stmt EOL {
3.     if(debug) dumpast($2, 0);
4.     printf("= %4.4g\n> ", eval($2));
5.     treefree($2);
6. }
7. | calclist LET NAME '(' symlist ')' '=' list EOL {
8.     dodef($3, $5, $8);
9.     printf("Defined %s\n> ", $3->name); }
10.
11. | calclist error EOL { yyerrok; printf("> "); }
12. ;

```

行 7 给出了 `let` 的产生式。为了使 `let` 后支持大括号，我们去掉行 7 中的 `'='`，并修改 `dodef($3, $5, $7)`。接下来对文法的修改思路就很清晰了——修改非终结符 `list`，使之支持大括号结构。

参考资料中的《ANSI C grammar (Yacc).pdf》，我们对 `list` 产生式修改为如下结构：

```

1. list: '{' list '}' {$$ = $2;}
2. | '{' list stmt '}' {$$ = newast('L', $2, $3);}
3. | stmt {$$ = $1;}
4. | exp {$$ = $1;}

```

最后，由于计算器本身不支持{}这样的终结符，我们在词法分析中，增加左右大括号的词法匹配模式：

```
1. "{" |
2. "}" |
3. { return yytext[0]; }
```

至此，我们的计算器在文法上已经支持大括号结构了。编译结果如下：



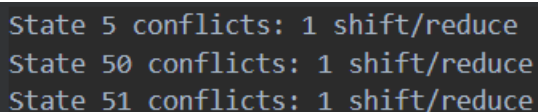
```
D:\Here\recently\系统软件开发实践\综合实验\3-3\test>bison -d fb3-2.y
fb3-2.y: conflicts: 3 shift/reduce
```

图 3-1 第一次编译结果

经过我们修改的文法，存在 3 个移进/归约冲突。我们使用 `bison -v` 命令，生成日志文件，查看具体的移进/归约冲突发生的代码段。

3.3.2 移进/归约冲突的解决

在使用 `bison -v` 指令后，得到 `fb3-2.output` 日志文件。从中可以看出，识别活前缀的 DFA 在状态 5、50、51 发成了移进/归约冲突。



```
State 5 conflicts: 1 shift/reduce
State 50 conflicts: 1 shift/reduce
State 51 conflicts: 1 shift/reduce
```

图 3-2 移进/归约冲突的产生位置

◆ 状态 5 冲突的解决

日志文件中，对状态 5 的描述如下。

```
1. state 5
2.
3. 19 exp: NAME .
4. 20   | NAME . '=' exp
5. 21   | NAME . '(' explist ')'
6.
7.  '=' shift, and go to state 16
8.  '(' shift, and go to state 17
9.
10.  '(' [reduce using rule 19 (exp)]
11. $default reduce using rule 19 (exp)
```

当自底向上的语法分析归约出文法符号 `NAME` 时，如果输入串的下一个符号是 `'(`，

此时识别活前缀的 DFA 既可以移进 '(' 转为状态 17, 又可以使用产生式 $exp \rightarrow NAME$ 归约。即移进/归约冲突。

我们将 NAME 的优先级降低, 优先移进归约左括号后的表达式, 从而解决冲突。

在 Yacc 文件的声明区, 加入 `%nonassoc LOWER`, 并将左括号的改为左结合性 `%left '('` 后, 将产生式 $exp \rightarrow NAME$ 更改为 $exp \rightarrow NAME \%prec LOWER$, 即可解决冲突。

◆ 状态 50 冲突的解决

日志文件中, 对状态 50 的描述如下:

```
1. state 50
2.
3.     1 stmt: IF exp THEN list .
4.     2       | IF exp THEN list . ELSE list
5.
6.     ELSE shift, and go to state 57
7.
8.     ELSE [reduce using rule 1 (stmt)]
9.     $default reduce using rule 1 (stmt)
```

当符号栈中已经存在了归约出的 IF exp THEN list, 且输入串的下一位是 ELSE 时, 此时 DFA 既可以移进 ELSE 转为状态 57, 又可以使用产生式 $stmt \rightarrow IF exp THEN list$ 归约。即移进/归约冲突。

根据 ELSE 与最近的 IF 匹配的原则, 我们选择降低归约操作的优先级, 优先移进操作即可解决冲突。

我们将上述产生式修改为 $stmt \rightarrow IF exp THEN list \%prec LOWER$, 并为符号 ELSE 加入左结合性, 冲突解决。

◆ 状态 51 冲突的解决

日志文件中, 对状态 50 的描述如下:

```
1. state 51
2.
3.     4 stmt: exp . ';'
4.     8 list: exp .
5.     9 exp: exp . CMP exp
6.     10    | exp . '+' exp
7.     11    | exp . '-' exp
```

```

8.      12      | exp . '*' exp
9.      13      | exp . '/' exp
10.
11.     CMP shift, and go to state 26
12.     '+' shift, and go to state 27
13.     '-' shift, and go to state 28
14.     '*' shift, and go to state 29
15.     '/' shift, and go to state 30
16.     ';' shift, and go to state 31
17.
18.     '-'      [reduce using rule 8 (list)]
19.     $default reduce using rule 8 (list)

```

当符号栈中已经存在了归约出的 *exp*，且输入串的下一位是 '-' 时，此时识 DFA 既可以移进 '-' 转为状态 28，又可以使用产生式 $list \rightarrow exp$ 归约。即移进/归约冲突。

解决冲突的办法仍然是降低归约的优先级，为 '-' 增加左结合性即可。

3.4 实验结果

3.4.1 Windows 系统下的实验结果

```

D:\Here\recently\系统软件开发实践\综合实验\3-3\code>fb3-3
> let avg(a, b) {(a + b) / 2;}
Defined avg
> let sq(n){e = 1; while(|((t = n / e) - e) > .001) do {e = avg(e, t);}}
Defined sq
> sq(10);
= 3.162
> sq(10) - sqrt(10);
= 0.000178
> let max(a, b){if(a > b) then a; else b;}
Defined max
> max(4, 5);
= 5
> max(1, 5);
= 5
> max(9, 5);
= 9
> let max3(a, b, c){if(a > b) then { if(a > c) then a;else c;} else {if(b > c) then b; else c;}}
Defined max3
> max3(3, 4, 5);
= 5
> max3(3, 9, 5);
= 9
> max3(10, 9, 5);
= 10
> |

```

图 3-3 windows 系统下实验结果

3.4.2 Linux 系统下的实验结果

```
chen@chen:~/桌面/chen/e7$ ./fb3-3
> let avg(a, b) {(a + b) / 2;}
Defined avg
> let sq(n){e = 1; while(|((t = n / e) - e) > .001) do {e = avg(e, t);}}
Defined sq
> sq(10);
= 3.162
> sq(10) - sqrt(10);
= 0.000178
> let max(a, b){if(a > b) then a; else b;}
Defined max
> max(4, 5);
= 5
> max(1, 5);
= 5
> max(9, 5);
8: error: syntax error
> max(9, 5);
= 9
> let max3(a, b, c){if(a > b) then { if(a > c) then a;else c;} else {if(b > c) then b; else c;}}
Defined max3
> max3(3, 4, 5);
= 5
> max3(3, 9, 5);
= 9
> max3(10, 9, 5);
= 10
> □
```

图 3-4 Linux 系统下实验结果

3.5 实验总结

本次实验总体上较为简单，根据编译原理课程上所学的语法制导翻译相关内容，以及参考书《ANSI C grammar (Yacc).pdf》，可以比较轻松的完成文法的修改；通过 Yacc 的冲突发生时的.output 日志文件，可以快速定位移进/归约冲突的位置，并给出合理的解决方案。通过本次实验，对 Flex 与 Bison 的使用更加熟练，加深了对计算器源码的理解。受益匪浅。

四 综合实验 4

4.1 实验目的

阅读《flex&Bison》第三章。使用 flex 和 Bison 开发一个具有全部功能的桌面计算器，包括如下功能：

- a) 支持变量；
- b) 实现复制功能；
- c) 实现比较表达式（大于小于等）；
- d) 实现 if/then/else 和 do/while 流程控制；

- e) 用户可以自定义函数;
- f) 简单的错误恢复机制。

4.2 实验内容

- 12)修改 flex 与 bison 源代码, 建立用于计算的接口函数
- 13)将 flex 与 bison 源码编译, 生成 Windows 环境下的动态链接库*.dll 与 Linux 环境下的动态链接库*.so
- 14)撰写实验报告, 结合实验结果, 分析整个步骤

4.3 项目实现思路及技术栈

本次实验的要求十分清晰——通过 gcc(或其他编译器)将全部 flex 与 bison 源代码打包生成动态链接库; 为后端的计算逻辑使用合适的语言编写前端的 ui 界面; 实现前端显示逻辑与后端计算逻辑的配合联动。

经过查阅资料与参考, 最后决定主要采用 Java 语言, 完成一个**基于 Tomcat 服务器**的 **Java Web** 项目。前端负责与用户的交互逻辑, 用户需要计算时通过 **get 请求**将表达式发送给后端的 **Java Servlet**, 后端通过调用 dll 动态库文件完成计算, 最后将计算结果反馈至前端。整个 Web 项目由 Maven 管理。在完成项目的构建后, 将生成的 Web 项目的 war 包**部署至阿里云服务器**上。由于服务器绑定了公网 ip, 因此任何用户均可以通过 ip 访问到本项目。

对于编译动态链接库的过程, 仅仅将现有代码编译生成动态链接库是不够的。原因在于, 现有代码的计算逻辑在 main 函数中, 难以被其他的后端程序所调用。因此, 首先需要修改现有代码, 将 main 函数的计算逻辑修改为独立的函数, 作为接口提供给其他的后端程序。之后, 将修改后的全部源文件编译生成动态链接库即可。

项目的前端 UI 界面我采用 **HTML+CSS** 语言, 并基于 **JQuery** 库使用 JavaScript 编写少量的人机交互逻辑; 后端采用 Java 语言, 使用 JNA 加载动态链接库, 完成计算逻辑。

对于前后端的通信, 我采用 Tomcat 服务器, 使用 JavaScript 将当前表达式作为 get 请求的内容发送给后端; 后端采用 Servlet 技术, 捕获 get 请求的参数(表达式), 并调用 dll 动态链接库中的计算方法对表达式计算, 并将结算结果反馈给前端。

最后, 对于这一完整的 Java 前后端项目而言, 所用到的第三方包如 JNA 等, 均使

用 Maven 包管理工具导入包并完成 Web 项目的 war 包的生成。

4.4 实验步骤

本次实验涉及到的语言包括 Java、HTML、CSS、JavaScript。使用的第三方包的版本如下：

表 4.1 开发环境

名称	版本
JDK	openjdk-11
tomcat	9.0.60
Java Servlet API	4.0.1
Java Native Access	5.10.0

4.4.1 修改 flex&bison 源码，添加对外计算接口

我们以综合实验 3-2 的源代码作为基础，进行修改。

观察 fb3-2.y 中的文法，对于表达式 *calclist*，根据词法文法规则进行移进规约后，得到的最终表达式结果直接通过 *printf* 语句输出（图 4-1）这与我们想要的效果不一致。我们在 fb3-2.h 头文件中增加一个字符指针 *char * ans*，用于存放最终的计算结果。并修改文法，将 *printf* 语句替换成 *sprintf* 语句，即将最终的计算结果保存在变量 *ans* 中。

```

calclist: /* nothing */
| calclist stmt EOL {
    if(debug) dumpast($2, 0);
    //printf("= %4.4g\\n> ", eval($2));
    sprintf(ans, "%4.4g", eval($2));
    treefree($2);
}
| calclist LET NAME '(' symlist ')' list EOL {
    dodef($3, $5, $7);
    sprintf(ans, "Defined %s .", $3->name);
    //printf("Defined %s\\n> ", $3->name);
}
| calclist error EOL { yyerrok; printf("> "); }
;
%%

```

图 4-1 语法修改

最后，在 fb3-2.1 中实现对外的计算接口——*do_cal* 函数

```

1. char* do_cal(const char* exp){
2.     size_t len = strlen(exp);

```

```
3.     ans = (char*) malloc (sizeof(char)*(len + 2));
4.     yy_switch_to_buffer(yy_scan_string(exp));
5.     yyparse();
6.     return ans;
7. }
```

为指向计算结果的指针`ans`在内存中开辟空间,通过调用`yy_switch_to_buffer`方法使得 `bison` 的语法分析结果保存在字符串中,最终返回`ans`即可。

4.4.2 编译,打包动态链接库

动态库在 Windows 系统与 Linux 系统下的命名规则略有不同。如果我们想制作名称为 `cal` 的动态链接库的话,在 Windows 平台下,动态库文件名称应该为 `libcal.dll`,而在 Linux 系统下,名称则是 `libcal.so`。

使用 `gcc`,可以很方便的将 `c` 语言源程序打包成动态链接库文件。分别执行以下命令:

```
gcc -w fb3-2.lex.c fb3-2.tab.c fb3-2funcs.c -c -fpic
```

```
gcc -shared fb3-2.lex.o fb3-2.tab.o fb3-2funcs.o -o libcal.dll
```

进而在当前工作目录下生成 `libcal.dll` 动态链接库文件,如下图所示:

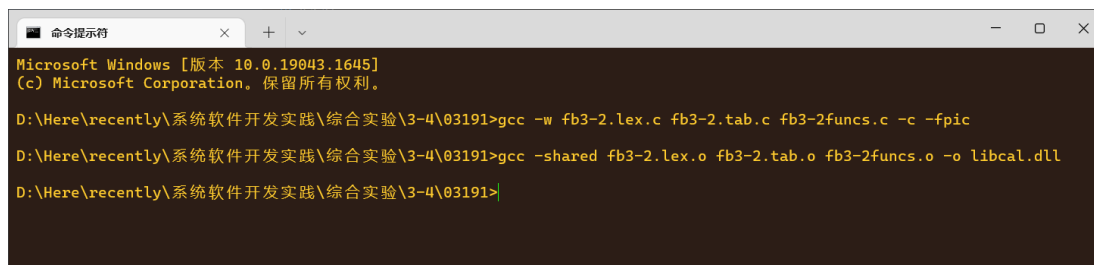


图 4-2 编译,生成动态链接库文件

4.4.3 前端 UI 界面

对于计算器的前端界面,从 `github` 克隆了一个简单的计算器项目作为本项目前端界面的基础,项目地址为: <https://github.com/superzcj/Calculator>。

不过,该项目的 `html`、`css`、`js` 代码全部写在了 `index.html` 中,显得有些许混乱。因此,在阅读源码后,将 `html` 与 `css` 代码分离,保存至不同的文件中。同时,将 `js` 代码全部删除。

最后,基于前端样式,自己编写了计算器的前端交互逻辑。部分核心代码如下:

```

1.      /**
2.      * 等号按下事件
3.      */
4.      equalKeyClicked: function () {
5.
6.          ChenCalculator.exp = $("#answer").val();
7.
8.          console.log("普通模式: " + ChenCalculator.exp);
9.
10.         var URL = "/chencal/cal" + "?exp=" +
            ChenCalculator.checkURL(ChenCalculator.exp) + ";";
11.         $.get(URL, function (data, status) {
12.             console.log("数据: " + data + "\n 状态: " + status);
13.             ChenCalculator.historyFlush(data, 1);
14.             ChenCalculator.afterCal = true;
15.         });
16.     },

```

该部分代码描述了界面中“等号”被按下事件。通过 get 请求，将表达式发送至后端，同时将后端的计算结果更新在计算器结果栏中。

4.4.4 后端逻辑

后端 Java Servlet 的业务逻辑较为简单。接收到前端的 get 请求后，解析请求中的参数，得到待计算表达式(行 4)；通过调用动态链接库(行 8)得到计算结果；最后，将计算结果反馈给前端(行 13)。

```

1.      @Override
2.      protected void doGet(HttpServletRequest req, HttpServletResponse resp)
            throws ServletException, IOException {
3.          System.out.println("get");
4.          String exp = req.getParameter("exp");
5.          if(exp.substring(0, 3).equals("let"))
6.              exp = exp.substring(0, exp.length() - 1);
7.          System.out.println(exp);
8.          String ans = CalculatorLib.lib.do_cal(exp + "\n");
9.          System.out.println(ans);
10.         resp.setHeader("Content-type", "text/html;charset=utf-8");
11.         resp.setCharacterEncoding("utf-8");
12.         PrintWriter writer = resp.getWriter();
13.         writer.print(ans);
14.     }

```

4.5 成果展示

本项目的项目结构如下：

```
root@iZ8vb48stipso6rv9jbbqeZ:/opt/javaProjects/JavaWebBegin/servlet03# tree
.
├── pom.xml
├── src
│   ├── main
│   │   ├── java
│   │   │   ├── com
│   │   │   │   └── chen
│   │   │   │       └── calculator
│   │   │   │           ├── AdvancedModeServlet.java
│   │   │   │           ├── CalculatorLib.java
│   │   │   │           ├── CalculatorServlet.java
│   │   │   │           ├── libcal.dll
│   │   │   │           └── libcal.so
│   │   ├── resources
│   │   └── webapp
│   │       ├── bootstrap.min.css
│   │       ├── css2.css
│   │       ├── css.css
│   │       ├── index2.html
│   │       ├── index.html
│   │       ├── jquery-3.6.0.min.js
│   │       ├── js.js
│   │       └── WEB-INF
│   │           └── web.xml
│   └── test
│       └── java
└── target
    ├── chencal.war
    ├── classes
    │   ├── com
    │   │   └── chen
    │   │       └── calculator
    │   │           ├── AdvancedModeServlet.class
    │   │           ├── CalculatorLib.class
    │   │           ├── CalculatorServlet.class
    │   │           └── Test.class
    │   └── web
    │       ├── ServletDemo1.class
    │       ├── ServletDemo2.class
    │       ├── ServletDemo3.class
    │       └── ServletDemo4.class
```

图 4-3 项目结构

全部源代码保存在./src 中；编译生成的 java 字节码文件与 Web 项目的 war 包保存在./target 中；Maven 的配置文件为./pom.xml。

启动 tomcat 服务器，项目的界面如下：

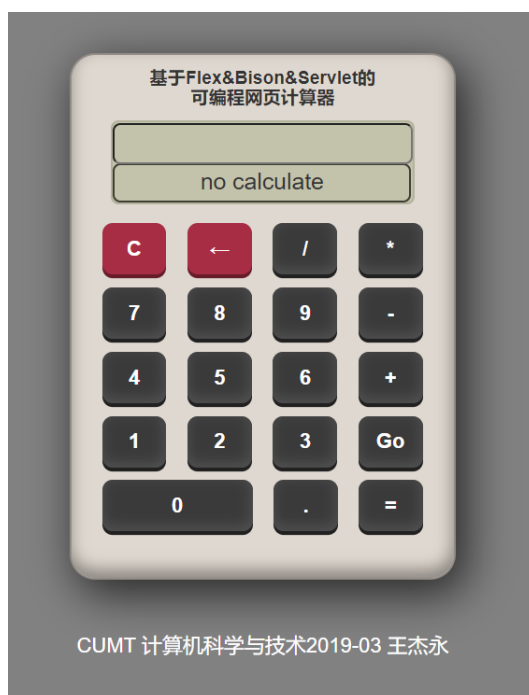


图 4-4 计算器界面

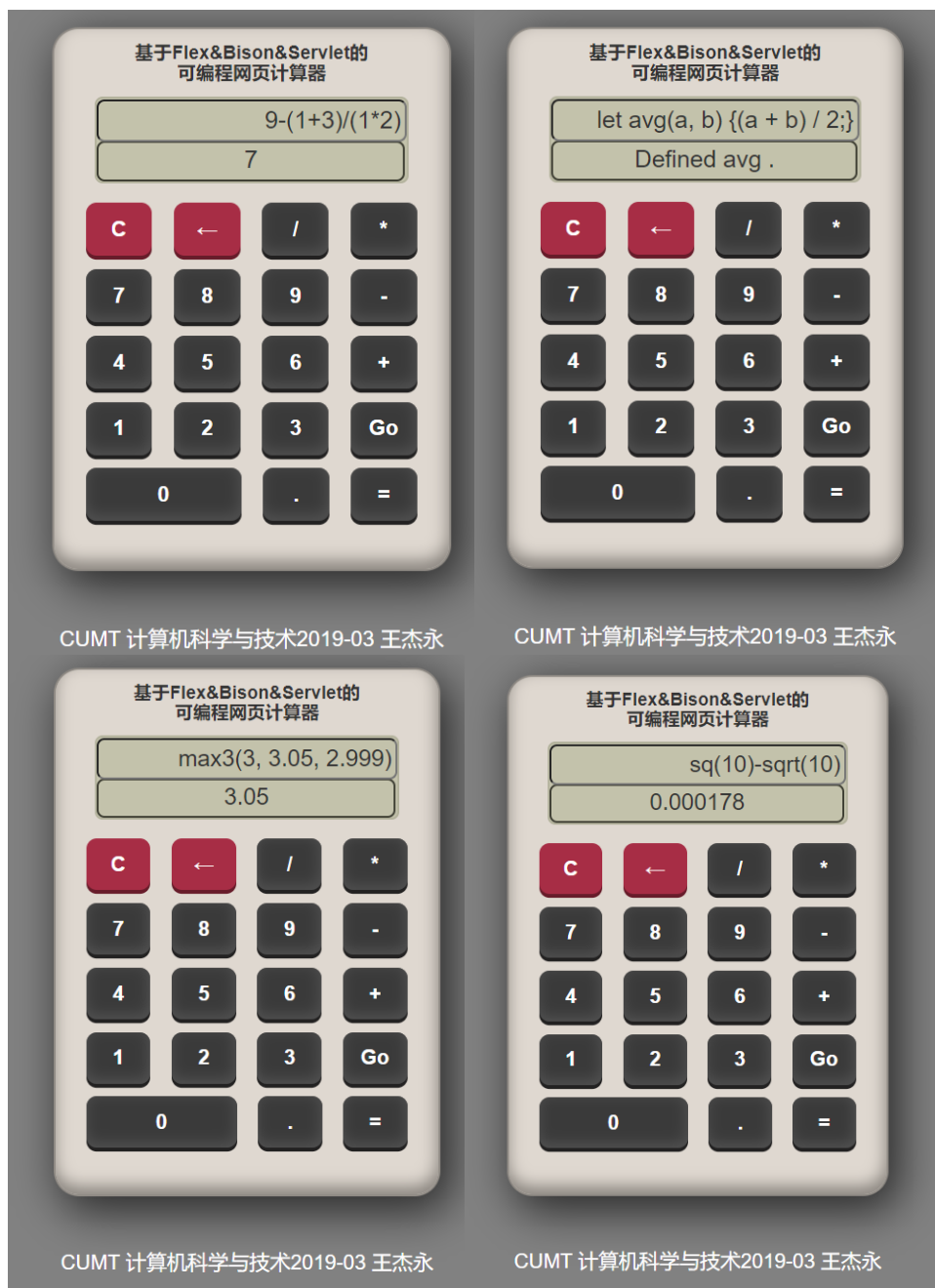


图 4-5 部分计算结果展示

4.6 实验总结

自己对本次项目的投入时间还是非常多的。

本次项目像是打开了一扇新的大门——曾经认为计算器只能通过高层次的算法去实现（栈，逆波兰式的计算），但现在自己才意识到，原来可以通过编译原理这一更底层一点的形式去完成一个计算器。

通过对提供的完整的计算器词法、语法分析以及语法制导翻译模式相关的源码的阅

读，自己的阅读代码能力得到了很大的提升。在较为充分的理解透原作者计算逻辑的语法后，尝试修改源码，增加了内置函数 `pow(x, y)` 及其重载的三操作数形式 `pow(x, y, c)`（图 2-12），得到了深深的成就感。

为了将底层的计算逻辑打包动态库，又一次深入源码，进行部分修改，独立编写对外的接口函数；受到老师提供的 PHP 计算器的思路影响，最后决定开发前后端分离的 JavaWeb 项目。通过之前学习的 URL 地址解析的原理以及 HTML、CSS、JS 的基础知识，我对前端部分没有太大的困难；相反，由于一开始确定了将计算逻辑放在后端以及项目采用 JavaWeb 的架构，这对于仅仅是听过看过相关知识的我有些许困难。通过若干天的学习，我掌握了 Tomcat 服务器的搭建、Maven 包管理工具的使用、Servlet 的基本原理等相关知识，在每天枯燥的计算机视觉代码的阅读中增添了些许乐趣。

最后，在完成了整个项目的搭建后，又花费了一些时间学习了在 Linux 系统的服务器上部署项目。当从自己的手机上打开项目网址看到自己的计算器时，感到这些天的努力并没有白费。

通过本次课程设计，我加深了对编译原理的理解，激发了自己对计算机体系结构的兴趣；同时掌握了一些高层的软件开发技术。

感谢本次课程设计。