**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](#_Toc103018510)

[1.1 实验目的 1](#_Toc103018511)

[1.2 实验内容 1](#_Toc103018512)

[1.3 实验步骤 1](#_Toc103018513)

[1.3.1 Windows系统下进行实验 1](#_Toc103018514)

[1.3.2 Linux系统下进行实验 2](#_Toc103018515)

[1.4 抽象语法树的构建 3](#_Toc103018516)

[1.4.1 产生式 3](#_Toc103018517)

[1.4.2 1+2-3\*2/5的抽象语法树 3](#_Toc103018518)

[1.4.3 (1+2)-(2\*6)的抽象语法树 4](#_Toc103018519)

[1.5 实验总结 4](#_Toc103018520)

[1.5.1 你在编程过程中遇到了哪些难题？ 4](#_Toc103018521)

[1.5.2 你的收获有哪些？ 5](#_Toc103018522)

[二 综合实验2 5](#_Toc103018523)

[2.1 实验目的 5](#_Toc103018524)

[2.2 实验内容 5](#_Toc103018525)

[2.3 实验步骤 5](#_Toc103018526)

[2.3.1 Windows系统下进行实验 6](#_Toc103018527)

[2.3.2 Linux系统下进行试验 6](#_Toc103018528)

[2.4 源码分析 6](#_Toc103018529)

[2.4.1 fb3-2.y源码分析 7](#_Toc103018530)

[2.4.2 fb3-2.l源码分析 9](#_Toc103018531)

[2.4.3 fb3-2.h源码分析 10](#_Toc103018532)

[2.4.4 fb3-2.funcs.c源码分析 12](#_Toc103018533)

[2.5 抽象语法树的建立、遍历与计算 13](#_Toc103018534)

[2.5.1 基本节点ast 13](#_Toc103018535)

[2.5.2 数字节点numval 14](#_Toc103018536)

[2.5.3 比较运算节点ast 14](#_Toc103018537)

[2.5.4 标识符节点symref 15](#_Toc103018538)

[2.5.5 赋值节点symasgn 15](#_Toc103018539)

[2.5.6 内置函数节点fncall 16](#_Toc103018540)

[2.5.7 自定义函数节点ufncall 16](#_Toc103018541)

[2.5.8 创建自定义函数 17](#_Toc103018542)

[2.5.9 表达式求值 18](#_Toc103018543)

[2.6 Debug模式 19](#_Toc103018544)

[2.7 增加函数pow及其重载函数的详细过程分析 19](#_Toc103018545)

[2.7.1 二操作数pow函数代码改动 19](#_Toc103018546)

[2.7.2 二操作数pow的计算逻辑 20](#_Toc103018547)

[2.7.3 二操作数pow函数抽象语法树的建立过程 21](#_Toc103018548)

[2.7.4 二操作数pow的计算结果 22](#_Toc103018549)

[2.7.5 三操作数pow函数 22](#_Toc103018550)

[2.8 实验总结 24](#_Toc103018551)

[三 综合实验3 24](#_Toc103018552)

[3.1 实验目的 24](#_Toc103018553)

[3.2 实验内容 25](#_Toc103018554)

[3.3 实验步骤 25](#_Toc103018555)

[3.3.1 文法修改 25](#_Toc103018556)

[3.3.2 移进/归约冲突的解决 26](#_Toc103018557)

[3.4 实验结果 28](#_Toc103018558)

[3.4.1 Windows系统下的实验结果 28](#_Toc103018559)

[3.4.2 Linux系统下的实验结果 29](#_Toc103018560)

[3.5 实验总结 29](#_Toc103018561)

[四 综合实验4 29](#_Toc103018562)

[4.1 实验目的 29](#_Toc103018563)

[4.2 实验内容 30](#_Toc103018564)

[4.3 项目实现思路及技术栈 30](#_Toc103018565)

[4.4 实验步骤 31](#_Toc103018566)

[4.4.1 修改flex&bison源码，添加对外计算接口 31](#_Toc103018567)

[4.4.2 编译，打包动态链接库 32](#_Toc103018568)

[4.4.3 前端UI界面 32](#_Toc103018569)

[4.4.4 后端逻辑 33](#_Toc103018570)

[4.5 成果展示 34](#_Toc103018571)

[4.6 实验总结 35](#_Toc103018572)

1. 综合实验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. 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. 实验步骤
     1. Windows系统下进行实验

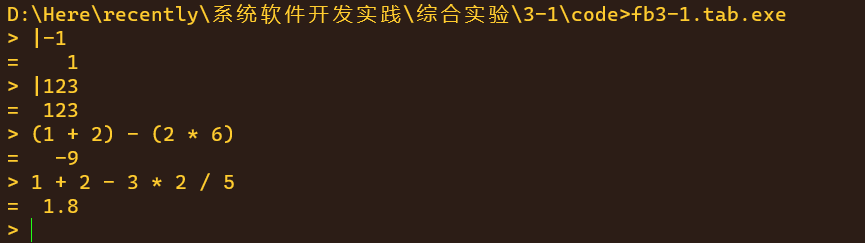
Windows环境下，编译flex与bison文件后，生成了对应的词法/语法分析程序的c语言文件。使用**gcc**编译器编译，得到可执行程序，运行结果顺利，结果如下图所示。

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.

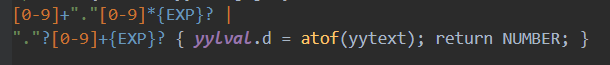
为了弄明白问题所在，我查看flex文件，发现其中匹配数字的模式对应的动作如下图：

图1-2 匹配数字模式的动作

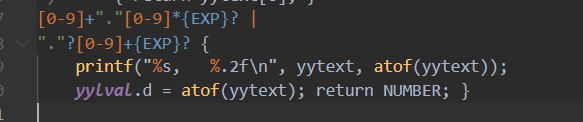
我加入了如下printf语句。目的是为了查看是否成功匹配到数字（输出yytext）、并且是否成功将串转换成浮点数（输出atof的返回值）。

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. Linux系统下进行实验

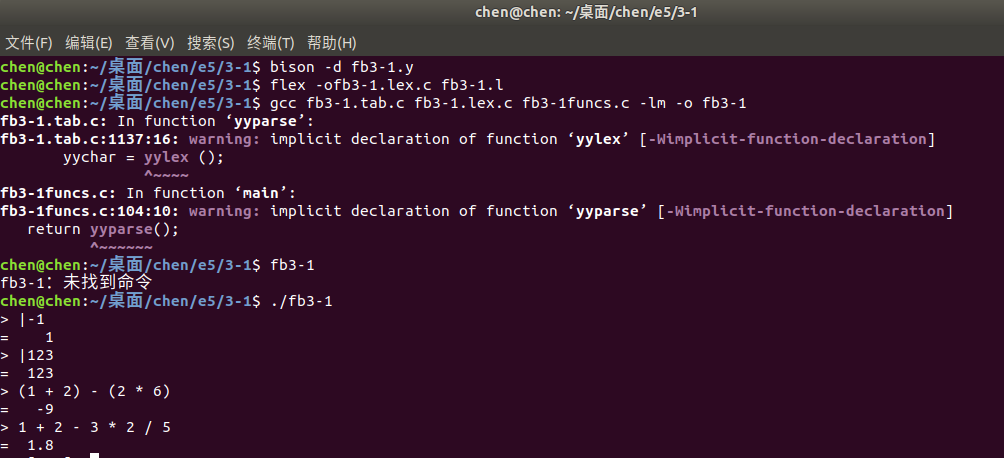
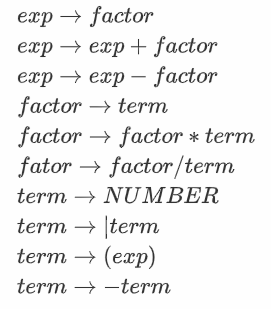
Linux系统下由于不存在cl编译器，默认安装了gcc编译器，因此在编译后的运行阶段没有任何错误。程序结果如下图。

图1-4 Linux系统下的运行结果

* 1. 抽象语法树的构建

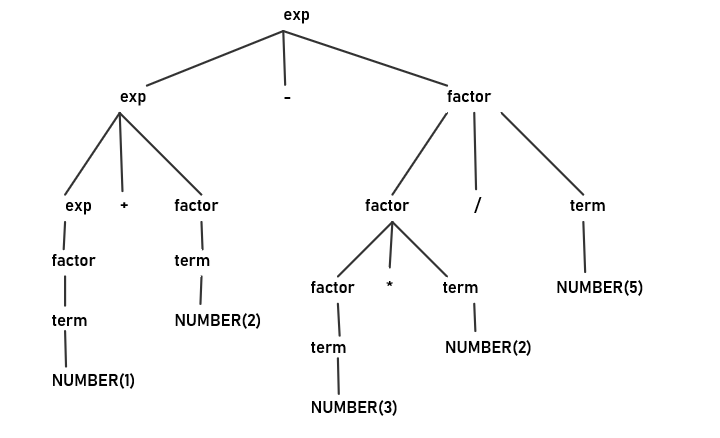
在上次实验中就已经发现，Bison默认采用LALR(1)的语法分析方法。向后多看一个字符，即可以构建整棵抽象语法树。

* + 1. 产生式

查看文件，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：

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. 1+2-3\*2/5的抽象语法树

图1-5 抽象语法树

* + 1. (1+2)-(2\*6)的抽象语法树

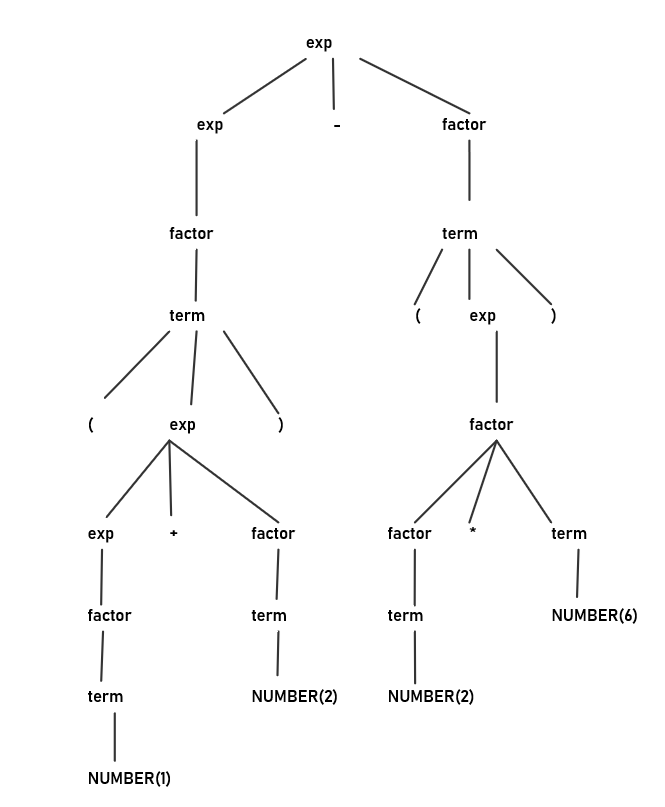


图1-6 抽象语法树

* 1. 实验总结
     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. 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。

1. 综合实验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。

* 1. 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；

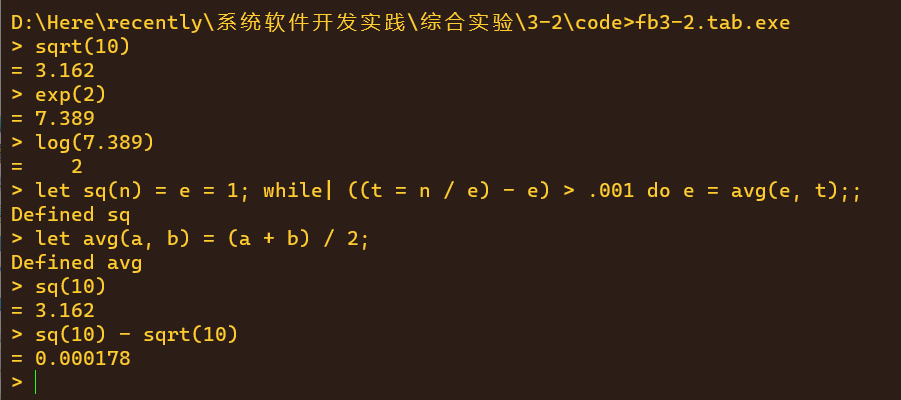
9)撰写实验报告，结合实验结果，给出抽象语法树的构建过程。。

* 1. 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。

* + 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。

图2-1 Windows系统下的运行结果

* + 1. Linux系统下进行试验

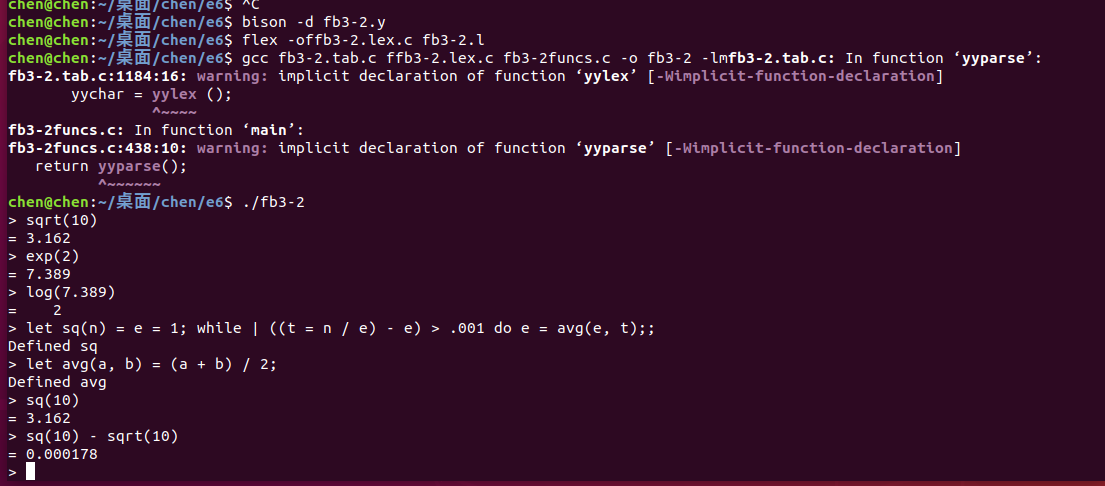
Linux系统下同理，使用gcc编译flex与bison生成的c文件，计算器成功运行。结果如下图。

图2-2 Linux系统下的运行结果

* 1. 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**。

* + 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 { $$ = newflow('W', $2, $4, NULL); }
4. | exp
5. ;
7. list: */\* nothing \*/* { $$ = NULL; }
8. | stmt ';' list { if ($3 == NULL)
9. $$ = $1;
10. else
11. $$ = newast('L', $1, $3);
12. }
13. ;
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 ')' { $$ = newcall($1, $3); }
28. ;
30. explist: exp
31. | exp ',' explist { $$ = newast('L', $1, $3); }
32. ;
33. symlist: NAME { $$ = newsymlist($1, NULL); }
34. | NAME ',' symlist { $$ = newsymlist($1, $3); }
35. ;
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); }
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。

* + 1. 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; }
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。

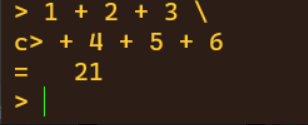
"\\\n"规则则允许我们"换行继续输入"，这里的反斜杠是转义符的含意，当匹配到反斜杠且其后紧跟换行符时，当前表达式的计算不会停止，词法分析器仅输出一个"c>"符号以此表示该行输入紧跟上一行。即，**该计算器具有分多行输入的功能**。

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。

* + 1. 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. };
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 \*/
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**。

* + 1. 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建函数——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。

* 1. 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。

* + 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));
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. }

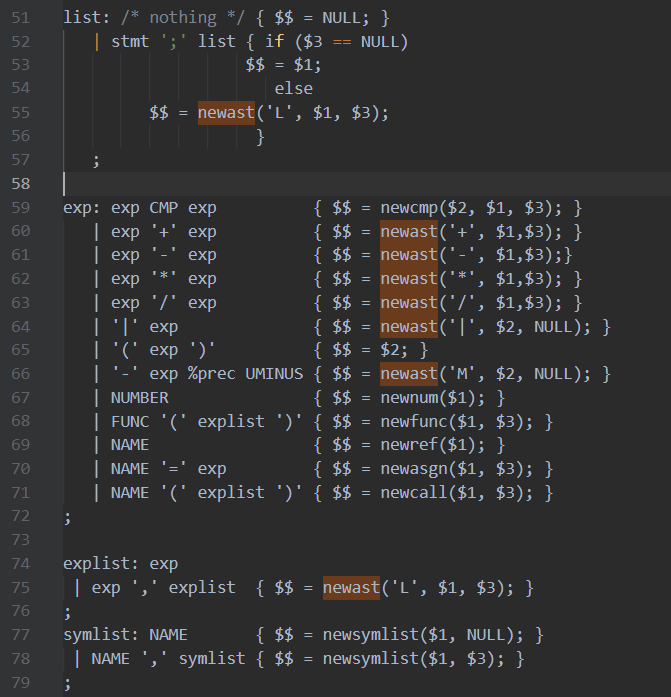
可以看出，ast节点有左右两棵子树，因此推断二元运算符才会创建ast节点。我们在语法分析的.y文件中查找newast函数，得到如下结果。

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。

* + 1. 数字节点numval

1. struct ast \*
2. newnum(double d)
3. {
4. struct numval \*a = malloc(sizeof(struct numval));
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。

* + 1. 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));
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。

* + 1. 标识符节点symref

1. struct ast \*
2. newref(struct symbol \*s)
3. {
4. struct symref \*a = malloc(sizeof(struct symref));
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。

* + 1. 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));
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。

* + 1. 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));
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**。

* + 1. 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));
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ufcall节点。创建自定义函数的方法在5.8中介绍。

* + 1. 创建自定义函数

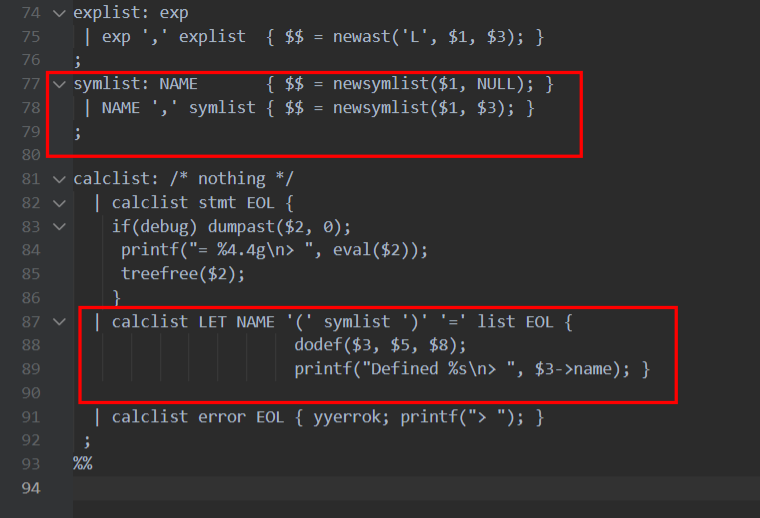
在计算器的语法分析部分中，我们找到了如图所示的两条文法规则。symlist定义了函数参数列表的匹配规则，调用newsymlist方法为symlist中添加参数。calclist定义了标识符Let的使用规则：

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。

* + 1. 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;
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);
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. }
    1. Debug模式

当我们输入debug1的时候，计算器会进入debug的模式1.

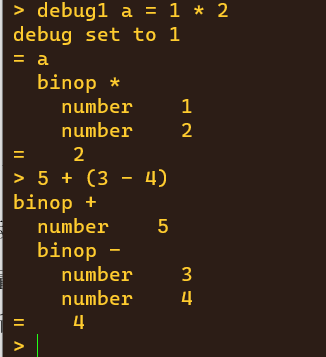
此时，输入任何表达式，会在求值前打印其所建立的抽象语法树的信息。

图2-6 debug模式

* 1. 增加函数pow及其重载函数的详细过程分析
     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.l中，新增保留字匹配模式pow.

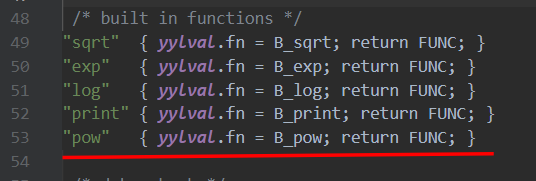
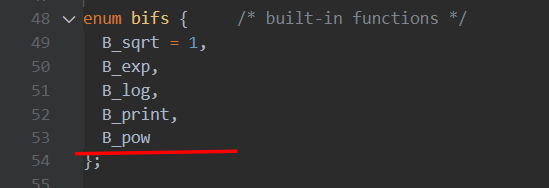
2) fb3-2.h头文件中，为枚举类型bifs增加取值B\_pow.

图2-7 前两处代码改动

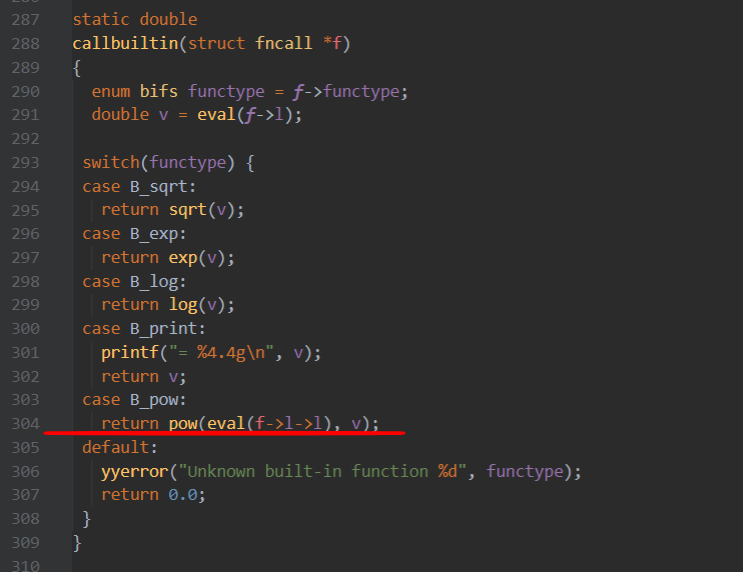
3) fb3-2funcs.c中，为函数callbuiltin中增加case B\_pow，并设置相应的计算方法

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。

* + 1. 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文法规则，explist是归约出的参数列表。

观察explist的文法规则，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。

* + 1. 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
* 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。

* 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。
  + 1. 二操作数pow的计算结果

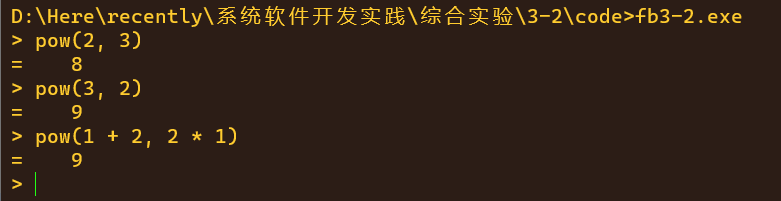
计算结果如下图所示：

图2-9 pow计算结果

* + 1. 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：

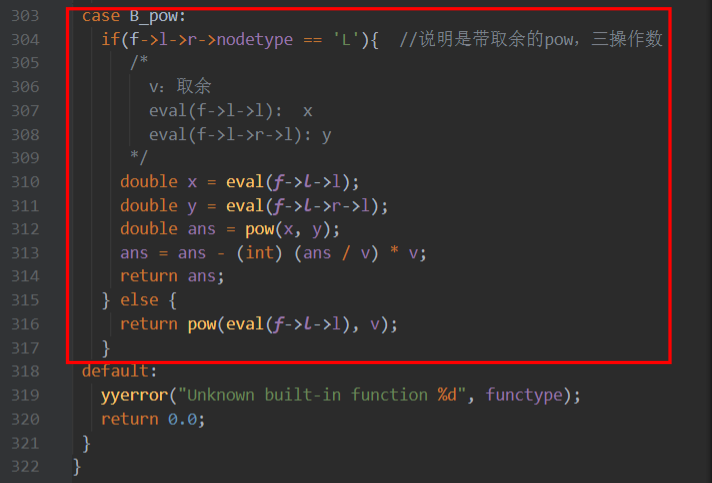
由于是重载函数，对于6.1中的前两处修改不再需要，只需要将第三处即pow的计算逻辑处增加额外的修改即可。

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。

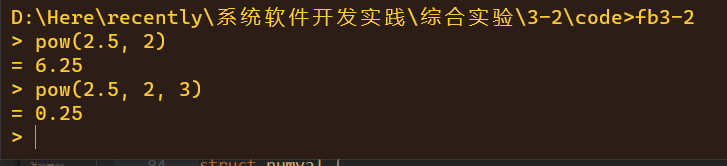
运行结果如图所示。

图2-12 pow重载的运行结果

* 1. 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。

1. 综合实验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。

* 1. 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

* 1. 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。

* + 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); }
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，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]; }

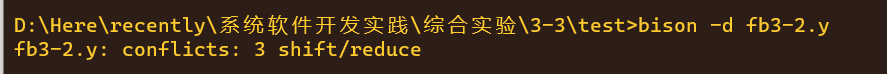
至此，我们的计算器在文法上已经支持大括号结构了。编译结果如下：

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。

* + 1. 移进/归约冲突的解决

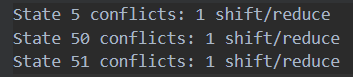
在使用bison -v指令后，得到fb3-2.output日志文件。从中可以看出，识别活前缀的DFA在状态5、50、51发成了移进/归约冲突。

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
3. 19 exp: NAME .
4. 20 | NAME . '=' exp
5. 21 | NAME . '(' explist ')'
7. '=' shift, and go to state 16
8. '(' shift, and go to state 17
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，并将左括号的改为左结合性后，将产生式更改为，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
3. 1 stmt: IF exp THEN list .
4. 2 | IF exp THEN list . ELSE list
6. ELSE shift, and go to state 57
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，此时识DFA既可以移进ELSE转为状态57，又可以使用产生式归约。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，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
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
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
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。

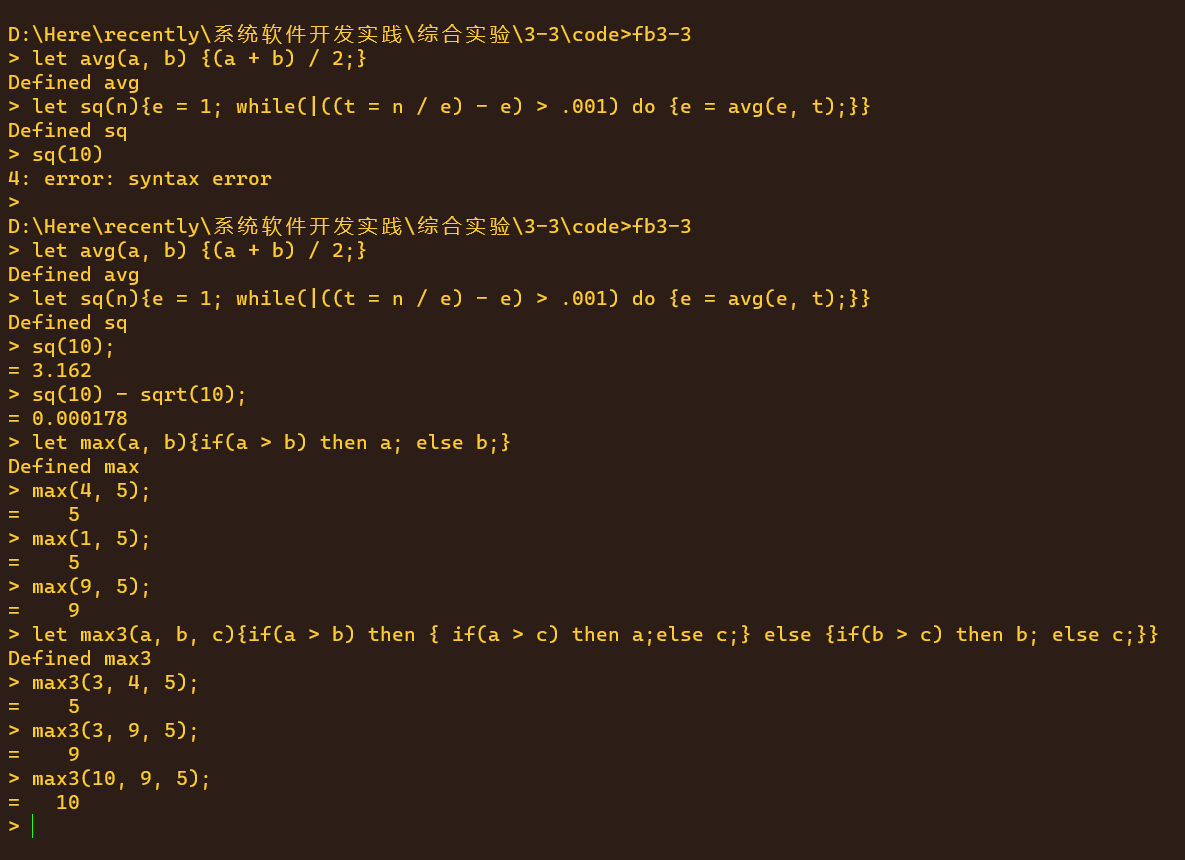
* 1. 实验结果
     1. Windows系统下的实验结果

图3-3 windows系统下实验结果

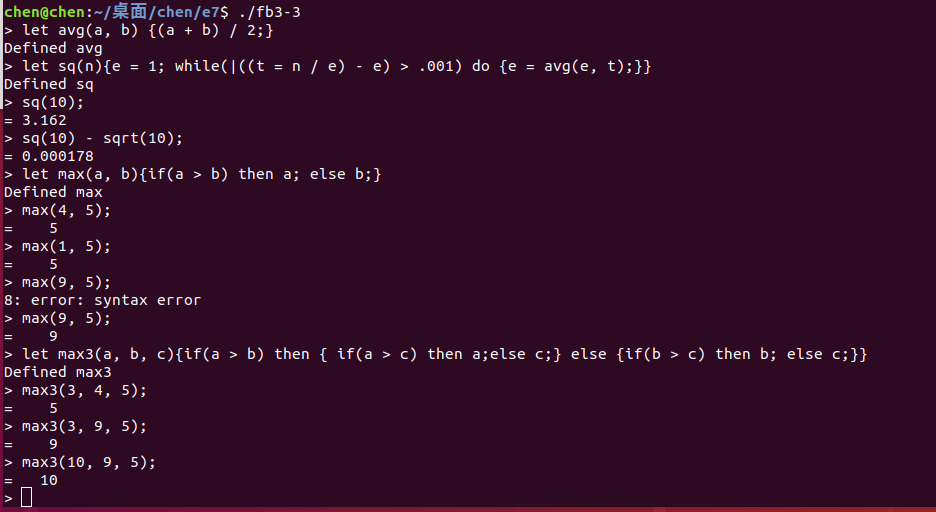
* + 1. Linux系统下的实验结果

图3-4 Linux系统下实验结果

* 1. 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。

1. 综合实验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。

* 1. 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

* 1. 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。

* 1. 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 |

* + 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。

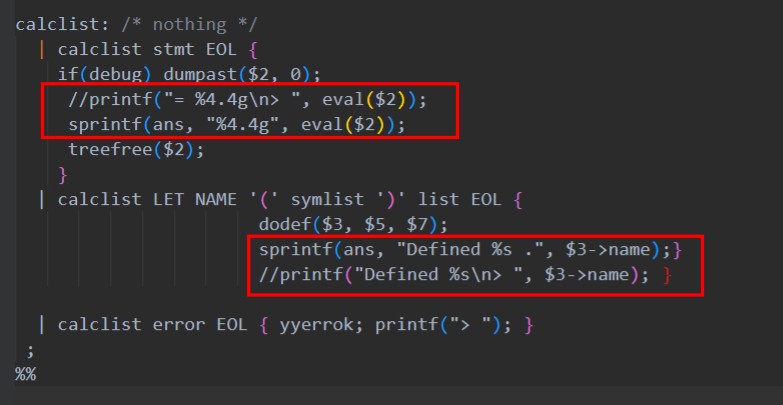
观察fb3-2.y中的文法，对于表达式，根据词法文法规则进行移进规约后，得到的最终表达式结果直接通过语句输出（图4-1）这与我们想要的效果不一致。我们在fb3-2.h头文件中增加一个字符指针，用于存放最终的计算结果。并修改文法，将语句替换成语句，即将最终的计算结果保存在变量中。

图4-1 语法修改

最后，在fb3-2.l中实现对外的计算接口——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在内存中开辟空间，通过调用方法使得bison的语法分析结果保存在字符串中，最终返回即可。

* + 1. 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：

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：

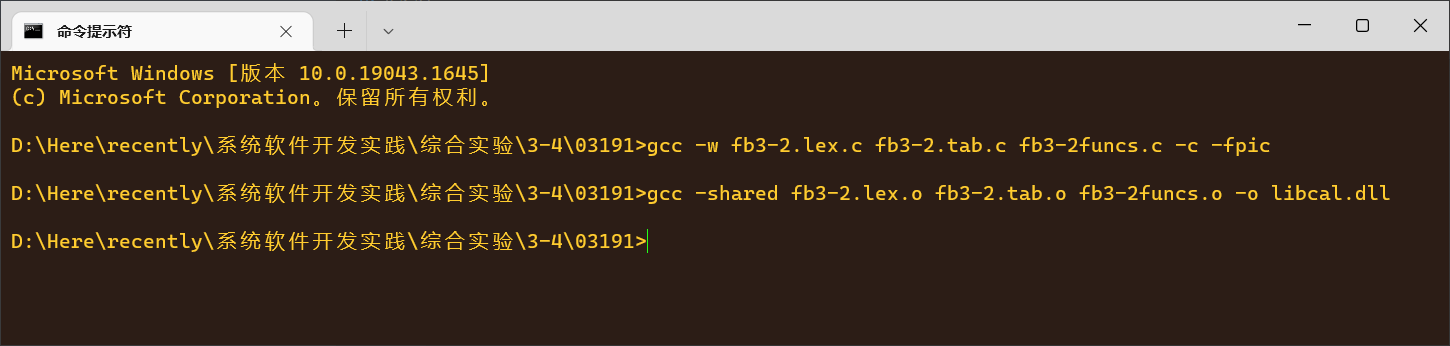


图4-2 编译，生成动态链接库文件

* + 1. 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. *\* 等号按下的事件*
3. *\*/*
4. equalKeyClicked: **function** (){
6. ChenCalculator.exp = $("#answer").val();
8. console.log("普通模式：" + ChenCalculator.exp);
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”被按下的事件。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。

* + 1. 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. }
    1. 成果展示

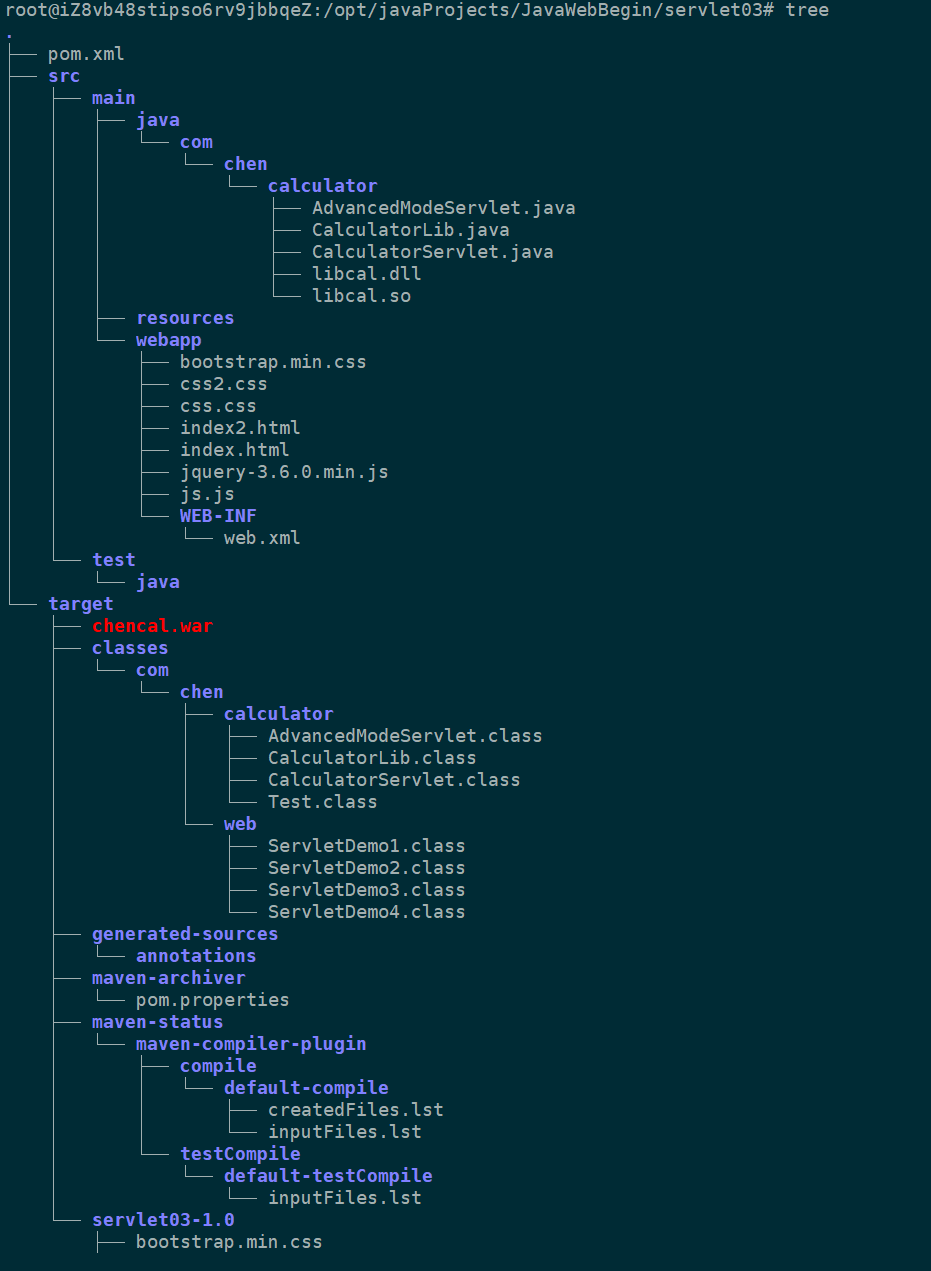
本项目的项目结构如下：

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。

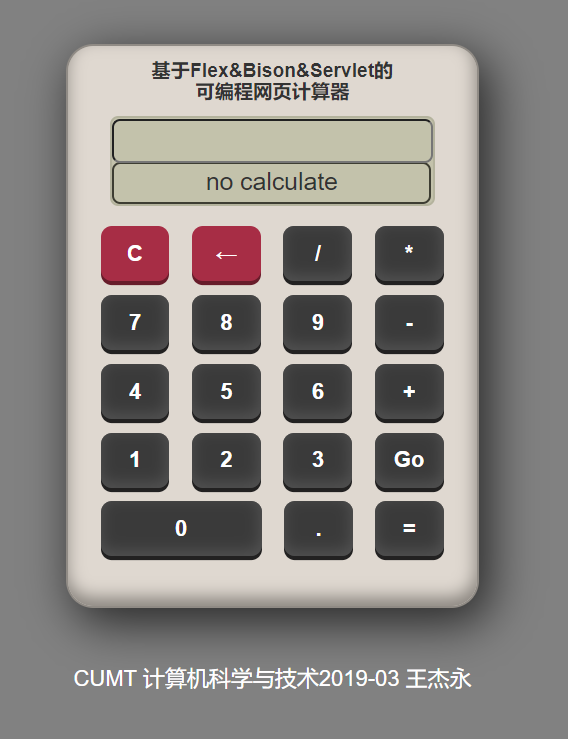
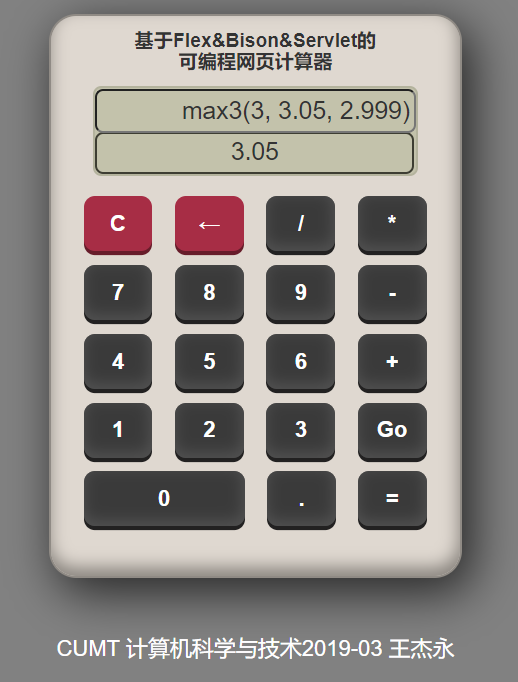
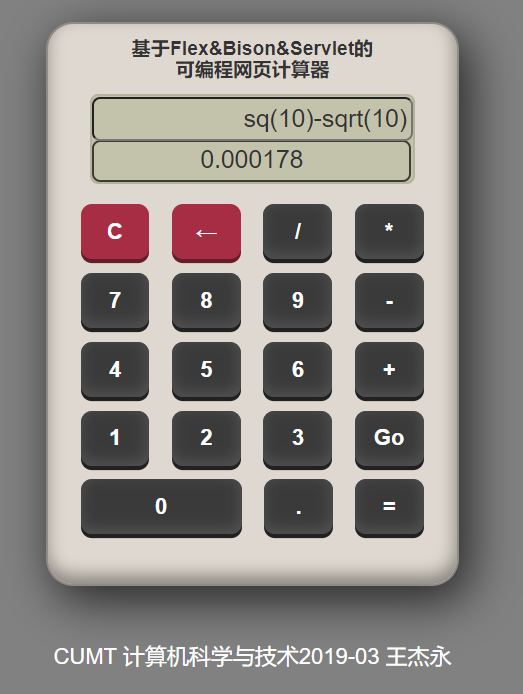
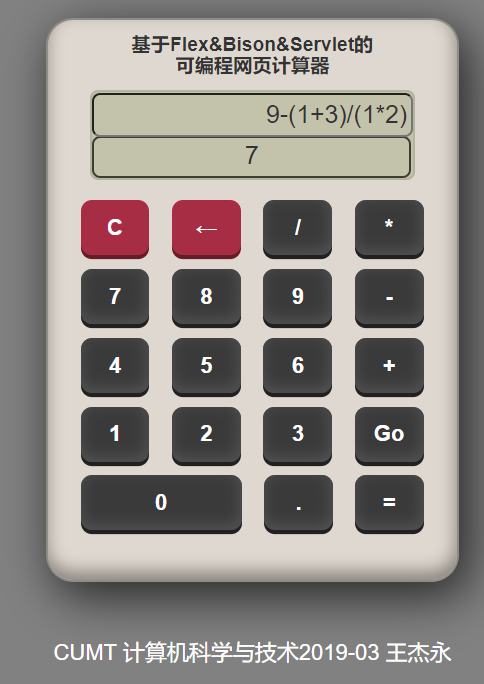
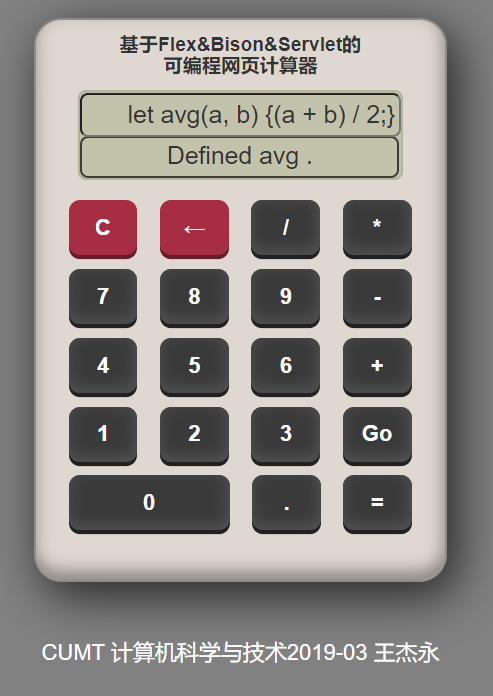
启动tomcat服务器，项目的界面如下：

图4-4 计算器界面

图4-5 部分计算结果展示

* 1. 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文法后，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。