



Lab1 实验报告

学生信息

姓名：李子沐；学号：221180029；邮箱：221180029@smail.nju.edu.cn

实验环境和编译方法

实验环境与OJ要求相同，即

- Ubuntu 20.04 64bit
- GCC 7.5.0
- Flex 2.6.4
- Bison 3.6.1

使用makefile编译即可。

使用命令编译方法如下：

在目录 `李子沐_221180029/Code/` 下，输入

```
flex lexical.l
bison -d syntax.y
gcc main.c astnode.c syntax.tab.c -ldl -o parser
```

实现功能

完成实验要求的必做和选做部分，即：

1. 使用flex和bison完成对C--的词法分析和语法分析
2. 输入一个.cmm文件，若无错误，输出要求格式的语法树；若有错误，输出错误类型与行号
3. 输入INT型数据支持八进制，十六进制输入，FLOAT类型数据支持指数形式输入，并处理为十进制数/标准浮点数输出。
4. 可以识别并跳过行注释 `//` 和块注释 `/*...*/`

注：同一行最多处理一个错误。代码中定义全局的errorline，与当前的yylineno对比判断多的错误是否输出。

在识别错误类型（Type A或Type B）外，额外实现了一些的错误类型：

1. 词法错误：通过定义错误的整型和浮点型词法单元，可以检测错误的整型、浮点型输入

3.20更新，删除了十进制整型错误的词法单元，例如0main将识别为错误的ID而非错误的整型。

2. 词法错误：参考实验指导，可以检测不在定义中的字符
3. 语法错误：参考实验指导，实现了三个最基本且粗暴的语法错误检测和恢复
4. 语法错误：面向测试样例识别了一些基本的语法错误，例如 2.cmm 中的 `a[2,5]`
5. 语法错误：根据语法理解，增加测试了19条error产生式，过犹不及！最终文法中保留了8条含error产生式，可以识别基本语法错误。
6. 被注释的error产生式仍然保留在提交的代码中，其主要思想是替换可能出错的词法单元，例如含有SEMI的产生式，都补充一条XXX error来补全分号。然而效果并不好！

注：测试文件全部来自实验指导的样例

实验细节细节

实现步骤主要参考实验指导，关键代码与问题如下。

多叉树

节点定义如下，包含当前词法/语法单元所在行号 `lineno` 等信息。

为了方便操作，这里定义子节点数量最大为 `MAX_CHILDREN_NUM`，未来若有动态分配的需求再修改。

测试遇到的bug：要求产生式生成空串时不输出，由于我在语法分析中为空串分配了语法单元，因此直接打印节点仍然会输出对应的非终结符。通过增加 `nodetype` 成员，通过 `nodetype == SYN_NODE && childnum == 0` 判断产生空串。

3.19更新：重写了多叉树，仅保留了 `create_node` 和 `print_tree` 接口，根据应用场景，删除了parent成员。

```

typedef struct node {
    int lineno;
    /* make sure that name is less than 32 */
    char name[32];                      /* e.g. TYPE */
    /* Attention: the INT and FLOAT is stored as a char array (string) */
    char value[32];                     /* e.g. int */
    struct node* parent;                /* only one parent */
    struct node* childs[MAX_CHILDREN_NUM];
    int childnum;
    int nodetype;
} Node;

```

添加多个子节点

利用 `<stdarg.h>` 的可变参数，实现了 `create_node` 函数，方便产生式使用。

块注释忽略

通过input和while实现块注释忽略，可以识别无效嵌套与EOF，报错仅报错注释的第一行。

yyerror调用(3.20更新)

语法错误在识别到error产生式时会自动调用 `yyerror()`，无需在动作中显式调用 `yyerror()` 函数，作者之前在这里重复报了很多错。

总结

语法报错一直存在问题，卡了好几天发现是error产生式的动作中显式调用了 `yyerror()` 多报了许多错误，修改完立马通过所有测试样例与额外的测试用例，完结撒花~

笨人编程能力较差，上一次独立写代码还是在大一上的二层次程序设计课。构建多叉树的过程遇到了许多弱智问题，例如字符数组要用 `strcpy()` 赋值，总之最后还是磕磕绊绊地完成了实验！

实验过程与想象中不太相同，我以为要手写DFA和LALR分析器，但是在实验指导的帮助下整体体验很棒（比电子的实验课强多了），感谢助教！感谢老师！