**编译原理课程大作业报告一**

**词法分析器说明文档**



**成员： 涂远鹏-1652262**

**成员： 黎盛烜-1652130**

**指导老师： 丁志军**

**日期： 2018年11月2日**

1. **软件使用说明**

软件界面如图1所示：



图1：词法分析器初始界面

单击“打开文件”可添加代码文件至词法分析器，并将类C代码显示在左上方的文本框一中；之后单击“词法分析”，可在右方的表格中显示词法分析的结果；单击“代码高亮”可在左下方的文本框二中高亮显示中间代码文件。具体实现效果如图2：

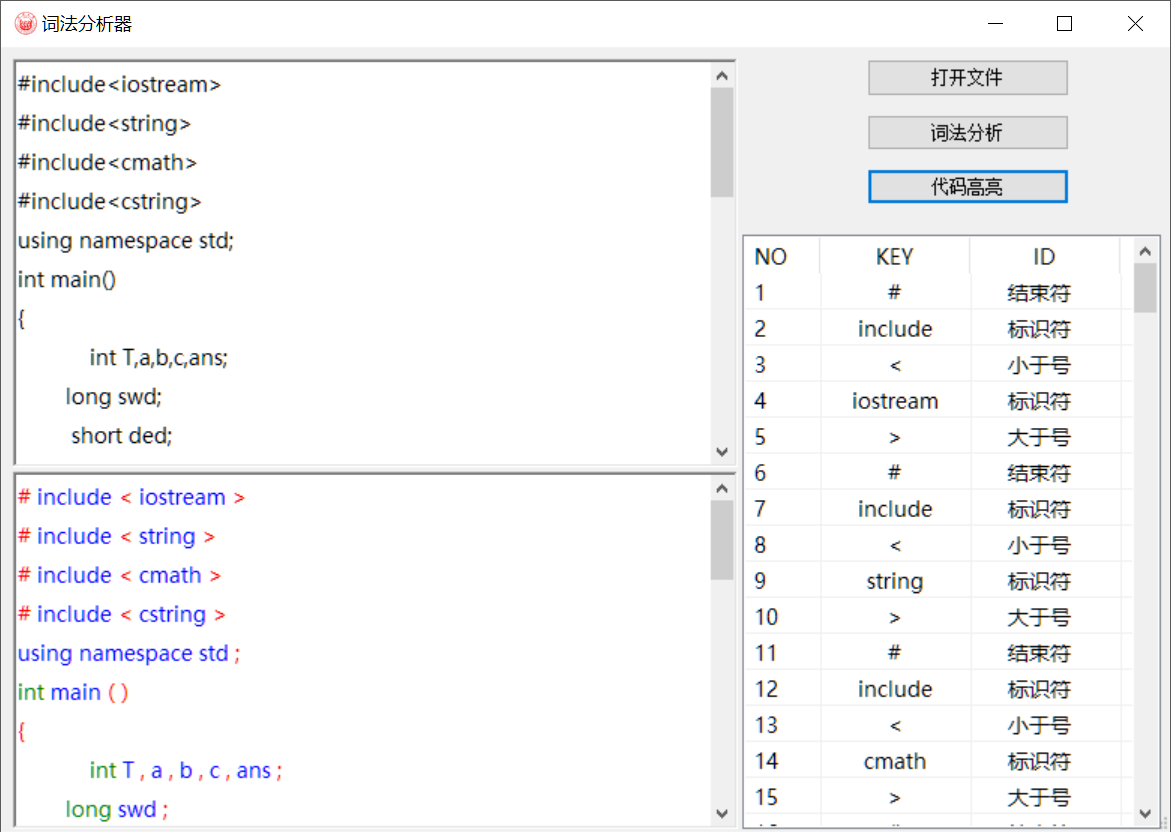


图2：词法分析器运行界面

1. **设计思路**
   1. **总体设计思路**

本词法分析器的总体思想是使用具有25个状态的自动机，对于需要分析的代码进行逐行逐字符扫描，通过状态机的不同对应状态获取各个词；对于获取的词，为其添加相关信息后得到一个数据结构Date，加入到队列token中，在本行扫描结束后，将队列返回结果输出到对应位置，准备进行下一次扫描。

* 1. **运行流程**
     1. 打开软件后，单击“打开文件”按钮，调用相关click事件，读取选定文件并将其添加到对应文本框中；
     2. 在已经读取代码文件的情况下，单击“词法分析”按钮，调用相关click事件，对代码文件进行扫描，每次获取一行代码字符串line，调用方法analysis对其进行逐字符扫描；
        1. 使用状态转移，根据终止状态来判断各个词的类型，为每一个扫描出的词添加Type类型的说明标识和中文说明，存储为Date结构，插入队列token中；
        2. 在一行的字符串扫描结束后，将token队列返回，调用listView相关方法将token中的内容添加到表格中；
        3. 获取新的一行代码字符串line，进行下一次词法分析，直到所有文本分析结束。
     3. 在已经读取代码文件的情况下，单击“代码高亮”按钮，调用相关click事件，对代码文件进行扫描，每次获取一行代码字符串line，调用方法analysis对其进行逐字符扫描；
        1. 先判断所读取的line的内容，保证之后输出结果的tab或空格能够对应；
        2. 使用状态转移，根据终止状态来判断各个词的类型，为每一个扫描出的词添加Type类型的说明标识和中文说明，存储为Date结构，插入队列token中；
        3. 在一行的字符串扫描结束后，将token队列返回，对token中的每一个元素进行分析，根据Type类型修改字体颜色，将对应内容通过尾部追加的方式加入到富文本框中；
        4. 所有元素追加完成后，输出\n换行；
        5. 获取新的一行代码字符串line，进行下一次词法分析，直到所有文本分析结束。
  2. **状态转移表**

详情见附件1：状态转移表。

* 1. **状态转移图**

本词法分析器使用的自动机对应的状态转移图如下图3：

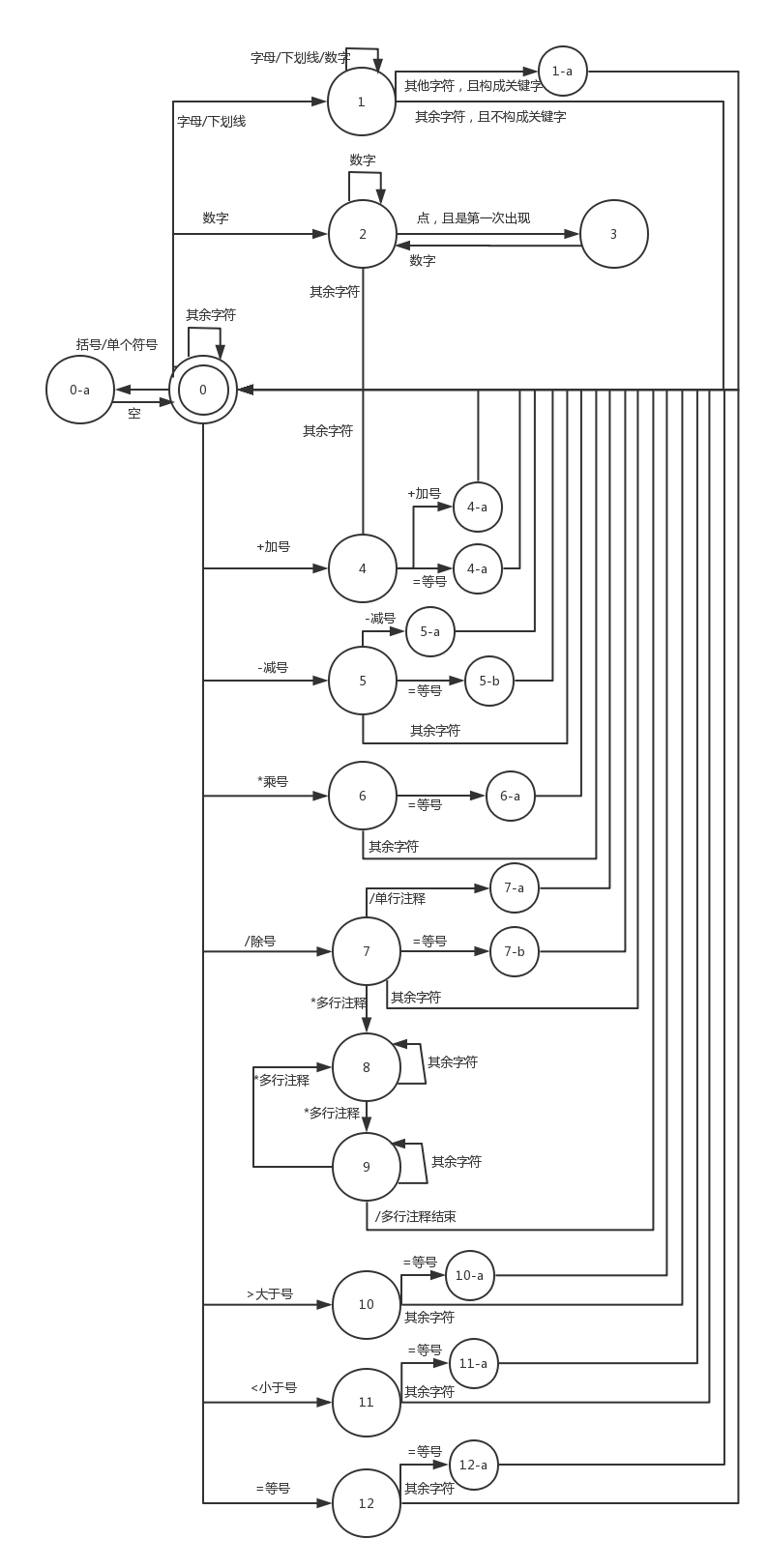


图3：词法分析器状态转移图

* 1. **状态说明**

|  |  |
| --- | --- |
| 状态名 | 说明 |
| 0 | 初始状态，终结状态，表示未读入字符或是刚结束一个单词的识别。 |
| 0-a | 读入一个字符，并且这个字符是括号或#等单个字符即为单词的字符 |
| 1 | 读入字母或是下划线 |
| 1-a | 此时读入的字母已经组成了一个关键字 |
| 2 | 读入数字 |
| 3 | 在已经读入数字的情况下第一次读入点 |
| 4 | 读入加号+ |
| 4-a | 连续读入加号，表示++ |
| 4-b | 连续读入加号等号，表示+= |
| 5 | 读入减号- |
| 5-a | 连续读入减号，表示-- |
| 5-b | 连续读入减号等号，表示-= |
| 6 | 读入乘号\* |
| 6-a | 连续读入乘号等号，表示\*= |
| 7 | 读入除号/ |
| 7-a | 连续读入除号，表示单行注释// |
| 7-b | 连续读入除号等号，表示除等于/= |
| 8 | 连续读入除号乘号，表示多行注释/\* |
| 9 | 在多行注释状态下读入乘号\* |
| 10 | 读入大于号> |
| 10-a | 连续读入大于号等号，表示大于等于>= |
| 11 | 读入小于号< |
| 11-a | 连续读入小于号等号，表示小于等于<= |
| 12 | 读入等号= |
| 12-a | 连续读入两个等号，表示等于比较 |

1. **附录**

附录1：状态转移表.xlsx

附录2：状态转移图.png