**计算机与电子信息学院/人工智能学院**

《编译原理》课程实验报告



**实验名称： 词法分析**

**专 业： 计算机科学与技术**

**姓 名：**

**学 号：**

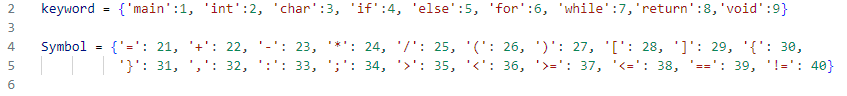
**指导老师： 杨 明**

# 实验内容

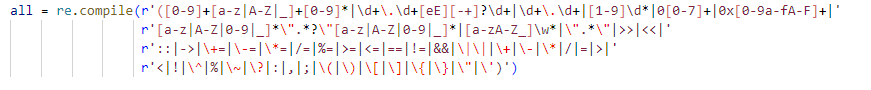
从源代码中分离出合法的单词，包括标识符、整数、保留字等，采用文法和有限自动机等方法实现

# 实验思路

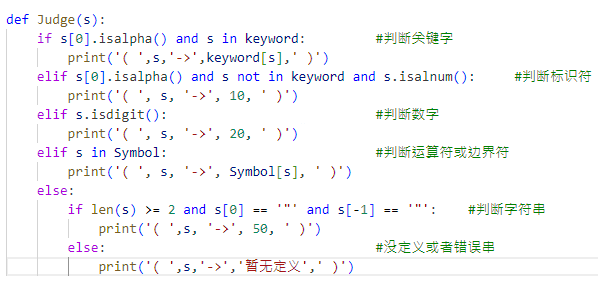
首先，列出 [keyword](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 和 [Symbol](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 两个字典，分别存储了一些预定义的关键字和符号，每个关键字或符号都对应一个唯一的数字，这个数字可以被视为该关键字或符号的类型。



其次，列出 正则表达式 [all](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) ，用于匹配。



最后，Judge 函数接受一个字符串 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 作为参数，然后判断 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 的类型。它首先检查 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 是否是一个关键字，然后检查 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 是否是一个标识符，然后检查 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 是否是一个数字，然后检查 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 是否是一个符号，最后检查 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 是否是一个字符串。如果 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 不满足以上任何一种类型，那么 [s](vscode-file://vscode-app/c:/Users/Juliuskarl/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) 就被视为未定义或错误的字符串。



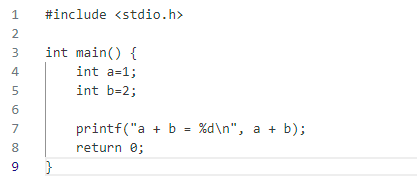
主要流程是通过正则表达式匹配和字典查找，将源代码分解为一系列的标记，并对这些标记进行分类。

# 实验结果

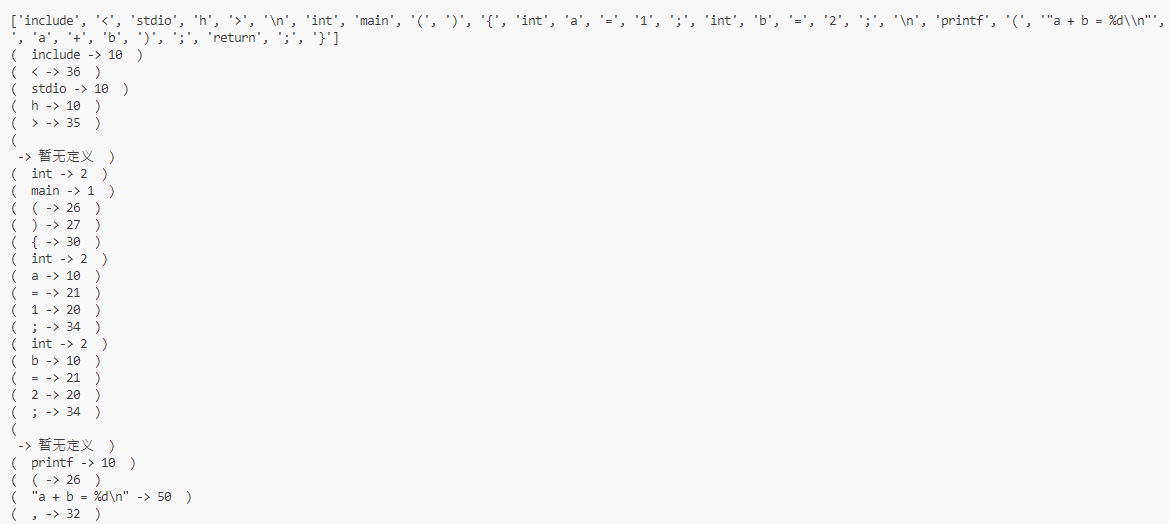
测试函数：



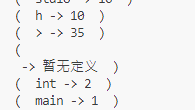
输入text：



输出：



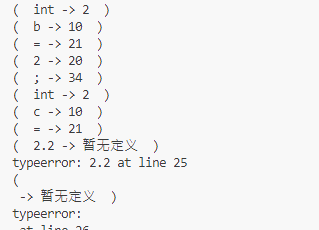
现在在text中加入错误的语句 int c = 2.2；



在验收时老师尤其强调了算法的鲁棒性，但是该代码对于错误的输入，只有“暂无定义”，显然算法的鲁棒性不足，不能够真正达到词法分析器协助改正代码的功能，对于输入错误的种类没有能够细分，因此对代码做出了以下修改：



此时对于形如int c = 2.2 的错误：



如图打印错误类型以及行号。

总结：虽然后来进行了代码的优化，但是仍然不够完美，如果想要打造一个C语言词法分析器还需要进一步学习优化。