
编译原理实验指导书

实验 1 词法分析

实验目的

1. 巩固对词法分析的基本功能和原理的认识。
2. 能够应用自动机的知识进行词法分析。
3. 理解并处理词法分析中的异常和错误。

实验内容

设计实现类高级语言的词法分析器，基本功能如下：

(1) 能识别以下几类单词：

- 标识符（由大小写字母、数字以及下划线组成，但必须以字母或者下划线开头）
- 关键字（①类型关键字：整型、浮点型、布尔型、记录型；②分支结构中的if和else；③循环结构中的do和while；④过程声明和调用中的关键字）
- 运算符（①算术运算符；②关系运算符；③逻辑运算）
- 界符（①用于赋值语句的界符，如“=”；②用于句子结尾的界符，如“;”；③用于数组表示的界符，如“[”和“]”；④用于浮点数表示的界符“.”）
- 常数（无符号整数（含八进制和十六进制数）、浮点数（含科学计数法）、字符串常数等）
- 注释。识别两种风格的注释：一种是使用双斜线“//”进行单行注释，在这种情况下，该行在“//”符号之后的所有字符都将作为注释内容而直接被词法分析程序丢弃掉；另一种是使用“/*”以及“*/”进行多行注释，在这种情况下，在“/*”与之后最先遇到的“*/”之间的所有字符都被视作注释内容而直接被词法分析程序丢弃掉。

(2) 能够进行简单的错误处理，即识别出测试用例中的非法字符。程序在

输出错误提示信息时，需要输出具体的错误类型（即词法错误）、出错的位置（源程序行号）以及相关的说明文字，其格式为：

Lexical error at Line [行号]: [说明文字].

说明文字的内容没有具体要求（例如：非法字符），但是错误类型和出错的行号一定要正确，因为这是判断输出错误提示信息是否正确的唯一标准。

（3）系统的输入形式：要求能够通过文件导入[测试用例](#)。测试用例要涵盖“实验内容”中列出的各类单词。

（4）系统的输出形式：打印输出测试用例对应的token序列，格式如下图所示：

```
■ 输入
  ■ while(num!=100){num++;}
■ 输出
while    <WHILE,    _    >
(        < SLP ,    _    >
num      < IDN ,    num  >
!=       < NE ,    _    >
100      <CONST,    100  >
)        < SRP ,    _    >
{        < LP ,    _    >
num      < IDN ,    num  >
++       < INC ,    _    >
;        < SEMI,    _    >
}        < RP ,    _    >
```

实验要求

（1）要求实验之前完成实验报告中的预习部分（即需求分析、文法设计和系统设计三个板块）。未按时完成预习报告者将扣除相应分数。实验报告中需要重点描述的是程序的亮点，是你认为最个性化、最具独创性的内容，而相对简单的、任何人都可以做的内容则可不提或简单地提一下，尤其要避免大段地在报告里贴代码。

（2）要求当堂完成实验内容，并进行现场验收。未当堂完成验收者将扣除相应分数。

（3）要求实验结束后一周内提交[实验报告及源程序](#)（存档用）。未按时提交者将影响成绩评定。打包的源程序中需要附带一个readme文件，用以说明程序应

该如何被编译？可以使用脚本、**makefile**或逐条输入命令进行编译，请详细说明应该如何编译你的程序。无法顺利编译将导致助教无法对你的程序所实现的功能进行任何测试，从而丢失相应的分数。

实验评分标准

一、课堂表现（10分）

1. 出勤情况（迟到，早退，缺席）
2. 是否遵守课堂纪律

二、操作表现（50分）

1. 功能齐全，结果正确无误（25 分）
2. 程序输出格式规范（5 分）
3. 讲解清楚明了，具有良好的演示效果（10 分）
4. 当堂按时完成（10 分）

三、实验报告（40分）

1. 需求分析（10 分）
2. 设计（20 分）
3. 系统实现及结果分析（10 分）