
编译系统实验指导书

实验一 词法分析与语法分析

实验目的

1. 巩固对词法分析与语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与语法分析。
3. 理解并处理词法分析与语法分析中的异常和错误。

实验内容

编写一个程序对使用C++语言书写的源代码进行词法和语法分析（C++语言的文法参见《编译原理实践与指导教程》附录A），并打印分析结果。实验要求使用词法分析工具GNU Flex和语法分析工具GNU Bison，并使用C语言来完成。

程序要能够查出C++源代码中可能包含的下述几类错误：

- 1) 词法错误（错误类型A）：出现C++词法中未定义的字符以及任何不符合C++词法单元定义的字符；
- 2) 语法错误（错误类型B）。

实验要求

（1）输入格式

程序输入是一个包含C++源代码的文本文件，程序需要能够接收一个输入文件名作为参数。例如，假设程序名为cc、输入文件名为test1、程序和输入文件都位于当前目录下，那么在Linux命令行下运行./cc test1即可获得以test1作为输入文件的输出结果。

（2）输出格式

实验一要求通过标准输出打印程序的运行结果。对于那些包含词法或者语法错误的输入文件，只要输出相关的词法或语法有误的信息即可。在这种情况下，

注意不要输出任何与语法树有关的内容。要求输出的信息包括错误类型、出错的行号以及说明文字，其格式为：

Error type [错误类型] at Line [行号]: [说明文字].

说明文字的内容没有具体要求，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。请严格遵守实验要求中给定的错误分类（即词法错误为错误类型A，语法错误为错误类型B），否则将影响你的实验评分。注意，输入文件中可能包含一个或者多个错误（但输入文件的同一行中保证不出现多个错误），程序需要将这些错误全部报告出来，每一条错误提示信息在输出中单独占一行。

对于那些没有任何词法或语法错误的输入文件，程序需要将构造好的语法树按照先序遍历的方式打印每一个结点的信息，这些信息包括：

1) 如果当前结点是一个语法单元并且该语法单元没有产生 ϵ （即空串），则打印该语法单元的名称以及它在输入文件中的行号（行号被括号所包围，并且与语法单元名之间有一个空格）。所谓某个语法单元在输入文件中的行号是指该语法单元产生的所有词素中的第一个在输入文件中出现的行号。

2) 如果当前结点是一个语法单元并且该语法单元产生了 ϵ ，则无需打印该语法单元的信息。

3) 如果当前结点是一个词法单元，则只要打印该词法单元的名称，而无需打印该词法单元的行号。

a) 如果当前结点是词法单元ID，则要求额外打印该标识符所对应的词素；

b) 如果当前结点是词法单元TYPE，则要求额外打印说明该类型为int还是float；

c) 如果当前结点是词法单元INT或者FLOAT，则要求以十进制的形式额外打印该数字所对应的数值；

d) 词法单元所额外打印的信息与词法单元名之间以一个冒号和空格隔开。

每一条词法或语法单元的信息单独占一行，而每个子结点的信息相对于其父结点的信息来说，在行首都要求缩进2个空格。具体输出格式可参见后续的样例。

(3) 测试环境

程序将在如下环境中被编译并运行：

- 1) GNU Linux Release: Ubuntu 20.04, kernel version 5.1.3.0 - 44-generic
- 2) GCC version 7.5.0
- 3) GNU Flex version 2.6.4
- 4) GNU Bison version 3.5.1

(4) 现场验收

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。如需补查，需在每次实验结束一周内向实验助教提出申请，并在实验助教指定的时间完成补查。

(5) 提交要求

实验一结束后一周内提交如下内容(超过指定时间未提交将影响实验成绩)：

- 1) 以Flex、Bison以及C语言编写的可被正确编译运行的源程序。
- 2) 一份word格式的实验报告，内容包括：
 - a) 程序实现了哪些功能？简要说明如何实现这些功能。
 - b) 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。
 - c) 实验报告的长度不得超过三页。所以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。实验报告中所出现的最小字号不得小于五号字（或英文11号字）。

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）