



编译原理

实验 1 报告

学 号：20188068

班 号：1802

姓 名：孔天欣

提交日期：2020 年 10 月 6 日

成 绩：

东北大学秦皇岛分校

**【实验目的和要求】**

设计、编写、调试一个具体的词法分析程序，加深对词法分析原理的理解。

【实验内容】

通过对 PL/0 词法分析程序 (GETSYM) 的分析，并在此基础上按照课本 1.4 节中给出的 PL/0 语言的语法描述，编写一个 PL/0 语言的词法分析程序。然后使用该词法分析程序实现对课本 12 页的图 1.19 中的源程序的词法分析。

要求：输入为字符串（即待进行词法分析的源程序），输出为二元式，表示为：（单词种别、单词自身的值）。详见 P38。有一定检查词法错误的能力，例如发现 2A 这类不能作为单词的字符串。

【实验环境】

计算机，任意程序设计语言。

【实验结果】

完整实验报告（包括算法描述、程序结构、调试情况、设计技巧、心得体会），提交纸质报告。源代码（含注释）和可执行程序，提交电子文件。

【算法描述】

本词法分析器程序通过读取源代码文件，逐个读取字符并组成字符串，然后分析字符串的性质。在分析过程中采用了有穷自动机原理，能够较好地进行单词的匹配以及语法正确性的初步检查。

根据自动机原理，词法分析器在保持跳过空格、回车、换行的多余字符的情况下读入有效字符，读取完毕后，依次分析其是否是标识符、保留字，或者是无符号整数，又或者是赋值或比较符，如果以上都不是，最后分析其是否是单字符的语法单词，如果不是说明语法错误。

【程序结构】

本程序采用 Java 语言设计，分为三个类：

1. CompilerBoot 类，这个类是程序入口，用于启动词法分析程序。
2. WordParser 类，词法分析器的主要内容都在本类中，主要工作是提取、分析单词，并提供返回单词的属性、编号和单词的值。
3. Symbol 枚举类，包含了类 PL/0 语言的所有语法单词。

【调试情况】

对于教科书 p12 图 1.18 中代码的词法分析内容如下图所示（片段）：



```
D:\Java\Java-jdk\bin\java.exe "-javaagent:D:\IntelliJ IDEA 2019.3\lib\idea_rt.jar=51514:D:\IntelliJ IDEA 2019.3\bin" -
孔天欣 20188068 180235
27,const
1,a
18,:=
2,20
16,;
28,var
1,b
15,,
1,c
29,procedure
1,p
16,;
20,begin
1,c
18,:=
1,b
3,+
1,a
16,;
21,end
16,;
```

如果将代码第一行 `const a := 20;` 改为 `const a := 2X;` ,那么程序输出如下:

```
D:\Java\Java-jdk\bin\java.exe "-javaagent:D:\IntelliJ IDEA 2019.3\lib\idea_rt.jar=51812:D:\IntelliJ IDEA 2019.3\bin" -
孔天欣 20188068 180235
27,const
1,a
18,:=
Syntax Error:Word "2X" isn't Valid.
16,;
28,var
1,b
15,,
1,c
29,procedure
1,p
16,;
20,begin
```

【设计技巧】

本程序首先采用枚举类将所有单词都列入其中,清晰明了。同时将每一个单词分析的过程都封装为一个方法,分工明确,提高了词法分析器的可拓展性。程序还采用自动机原理进行单词分析,提高了效率,同时能够匹配全面。对于单个字符的单词,使用哈希表来提高访问时间效率。对于保留字列表,预先按照字典序排序,可以方便地凭借二分搜索快速查找对应单词。

【心得体会】

通过本次实验,本人初步学习了词法分析器的设计方法和技巧,并成功实现了一个简单的具备词法分析功能的类 PL/0 词法分析程序,进一步提高了词法分析理论方面的水平。本人意识到编译原理的学习离不开实际的实践,同时实践的难度也并非想象中的大。