

# Lab1实验报告

---

## 1.程序功能

实现基本的SysY的词法分析器Lexer，能够获取到ID、空白符、关系符、注释、括号、整数等基本词素。  
能够获取并输出所有的token，若发生错误，能够检测并输出错误所在的位置。

## 2.实现过程

1.利用java + antlr4根据对应词法要求实现SysYLexer.g4文件，把要识别的词法转化为正则表达式的形式。编译g4文件得到lexer.java

2.为了获取所有的token，需要调用lexer类的getAllTokens函数。而为了满足输出错误词法所在的位置等信息，则需要巧妙地重写ErrorListener类。

3.有趣的地方在于调用getAllTokens的过程中，会不断地调用nextToken，这个过程中就会触发检查，如果发生错误就会通知Errorlistener。一旦错误发生，ErrorListener中的syntaxError函数就会被调用。由于传入参数中包含了错误相关的信息，那么只需要输出即可。同时在该类中添加一个标记错误产生的字段，这样在main里面输出时只要该字段表明有错误产生，那么就不再输出tokens。

4.其他实现：

0x开头的16进制数以及0开头的8进制数转化为10进制：利用Integer.ValueOf传入进制的基即可转换

## 3.遇到过的bug

1.一开始无法把08识别成两个数字0和8，原因是INTEGER的正则表达式不正确，必须对十六进制、八进制、十进制分别写再用或连接，因为8进制后面只能跟0-7这几个数字，是不会出现08的情况的。