**利用C++编写一个Python语言核心的编译器**

**目标**

实现面向对象的编程

**学习目标**

学习，归纳，总结Python语言的核心

熟练使用C++进行编程

**文件接口**

读取文件

**预处理**

将源代码按照行进行分割

将一行中的每个部分分为一个词素Token

**词法元素的分类**

运算符：+, -, \*, /, %, =

变量名：

类型名：//Python中不必说明类型，但是仍然存在数据类型int, float, char, str, bool, list, dict, tuple等

保留字段：return, for, while, if, else , elif, raise, and, or, True, False, import … as …

字符串：'watafak'

特殊符号

转义字符：'\?'

分隔符：冒号，逗号，

缩进：4个空格或者'\t'

换行符

特殊中的特殊符号

下标括号

…

如何Return好几次DENT？

**模拟对一行进行分词**

|  |
| --- |
| if a > 0:  a = 3 + b |

对于Line1来说，应该得到<if>, <a>, <>>, <0>, <:>这几个词素

对于Line2来说，应该得到<\t>, <a>, <=>, <35>, <+>, <b>这些词素

如果不采用有穷状态自动机而只使用简单地按空格分类，将会出现大问题

采用有穷自动机可以减少扫描的遍数

使用FLEX进行词法分析器的构造，关键是缩进的INDENT和DEDENT

NEWLINE的下一行没有制表符咋办？

如何从yyFlexLexer中得到下一个字符

https://github.com/python-git/python/blob/master/Doc/reference/lexical\_analysis.rst

使用行号和作用域用于符号表管理

变量中的声明和作用域，可以用递进的层数进行判别

Python的语法<https://github.com/python-git/python/blob/master/Grammar/Grammar>

语法分析

Line55 pass\_stmt: 'pass'

Line57 break\_stmt: 'break'

Line58 continue\_stmt: 'continue'

Line102 comp\_op: '<'|'>'|'=='|'>='|'<='|'<>'|'!='|'in'|'not' 'in'|'is'|'is' 'not'

Line41 vfpdef: NAME

Line