Atitit.词法分析的原理 理论

[1. 分词 .词法分析](#_Toc4675)**[lexical analysis](#_Toc4675)** [1](#_Toc4675)

[1.1. 分词主要流程 1](#_Toc1744)

[1.2. 分词的属性如下表token 2](#_Toc20686)

[1.3. 词法分析器主要包括：构造转换图与转换表、设计词法分析器算法。 2](#_Toc27154)

[2. 状态转换表 2](#_Toc11770)

[3. 词性划分 2](#_Toc4822)

[3.1. Token的类型，根据程序设计语言的特点，单词可以分为五类：关键字、标识符、常量、运算符、界符。以 3](#_Toc29559)

[3.2. 词法分析的第一阶段即](#_Toc293)**[扫描器](#_Toc293)** [3](#_Toc293)

[3.3. 词法分析的第二阶段](#_Toc29149)**[评估器](#_Toc29149)**[（Evaluator） 3](#_Toc29149)

[4. Qa 4](#_Toc6430)

[4.1. 词法分析的逻辑结果包括空格吗 4](#_Toc8973)

[5. 参考资料 4](#_Toc604)

# 分词 .词法分析**lexical analysis**

## 分词主要流程

字符切分，对比字典词典得到分词，并查阅词典，标注词性。完善其他属性，输出

当词法分析器发现了一个标识符的词素时，要将这个词素添加到符号表中。

其他任务：

* 过滤掉源程序中的注释和空白。
* 将编译器生成的错误信息与源程序的位置联系起来。记录行号等。

词法分析器的两个级联的处理阶段：

1. 扫描阶段：简单处理，删除注释，压缩空白字符。
2. 词法分析阶段：处理扫描阶段的输出并生成词法单元。

## 分词的属性如下表token

参考antlr..

|  |  |  |
| --- | --- | --- |
| Col | 列数 |  |
| Filename | 所属文件名 |  |
| Line | 行数 |  |
| Text | 分词 |  |
| Type | 词性 |  |

作者::  ★(attilax)>>>  绰号:老哇的爪子 （ 全名：：Attilax Akbar Al Rapanui 阿提拉克斯 阿克巴 阿尔 拉帕努伊 ） 汉字名：艾龙，  EMAIL:1466519819@qq.com

转载请注明来源： http://blog.csdn.net/attilax

## 词法分析器主要包括：构造转换图与转换表、设计词法分析器算法。

词法分析器的核心就是依据转换图识别单词。不过，事实并非完全如此。由于某些程序设计语言的词法定义缘故，仅仅依据程序2-1的算法是不足以完成词法分析的

# 状态转换表

Java dsl的state change table class.meth(p1,p2)

|  |  |  |  |
| --- | --- | --- | --- |
| curChar | curStat | afterStat | demo |
| ( | Not strStart | brkStart |  |
| ) | Not strStart | brkEnd |  |
| . | brkEnd |  | Meth().meth2() |
| “ | brkStart | strStart |  |
| “ | strStart | strEnd |  |
| “ | commaStat | strStart |  |
| , | brkStart,strend,commastat | commaStat | Meth(p1,p2) |
|  |  |  |  |
|  |  |  |  |

# 词性划分

## Token的类型，根据程序设计语言的特点，单词可以分为五类：关键字、标识符、常量、运算符、界符。以

读者可能对"单词"感到有点疑惑，不明白到底什么才是词法分析中所说的"单词"。试图回答这个问题就必须了解几个基本概念。这里，引入几个程序设计语言相关的名词。

（1）标识符：用户自定义的变量名、函数名等字符串。

（2）关键字：具有特殊含义的标识符。

（3）运算符：例如＋、－、\*、/ 等。

（4）常量：例如3.24、92等。

（5）界符：具有特殊含义的符号，如分号、括号等。

## [词法分析](http://baike.baidu.com/view/487040.htm" \t "http://baike.baidu.com/_blank)的第一阶段即**扫描器**

[词法分析](http://baike.baidu.com/view/487040.htm" \t "http://baike.baidu.com/_blank)的第一阶段即**扫描器**，通常基于[有限状态自动机](http://baike.baidu.com/view/157853.htm" \t "http://baike.baidu.com/_blank)

## [词法分析](http://baike.baidu.com/view/487040.htm" \t "http://baike.baidu.com/_blank)的第二阶段**评估器**（Evaluator）

，[语法分析器](http://baike.baidu.com/view/5538632.htm" \t "http://baike.baidu.com/_blank)需要第二阶段的**评估器**（Evaluator）。评估器根据语素中的字符序列生成一个“值”，这个“值”和语素的类型便构成了可以送入[语法分析器](http://baike.baidu.com/view/5538632.htm" \t "http://baike.baidu.com/_blank)的单词。一些诸如括号的语素并没有“值”，评估器函数便可以什么都不返回。整数、[标识符](http://baike.baidu.com/view/390932.htm" \t "http://baike.baidu.com/_blank)、字符串的评估器则要复杂的多。评估器有时会抑制语素，被抑制的语素（例如空白语素和注释语素）随后不会被送入[语法分析器](http://baike.baidu.com/view/5538632.htm" \t "http://baike.baidu.com/_blank)。

# Qa

## 词法分析的逻辑结果包括空格吗

好像不包括，参考 atitit.c词法分析器的详细原理.docx。。参考请问编译原理词法分析器是干什么用的，怎么用？？？-CSDN论坛-CSDN.NET-中国最大的IT技术社区.html， 也是不包括

分号逗号是包括的。

词法分析的逻辑结果包括分界符号码

看起来好像包括

包括字符串双引号吗，好像包括，参考 请问编译原理词法分析器是干什么用的，怎么用？？？-CSDN论坛-CSDN.NET-中国最大的IT技术社区.html

for(int i=0;++i;i<10)printf("Hello world");

printf  
(  
"Hello聽world"  
)  
;

# 参考资料

词法分析器的作用 - cyendra - 博客园.html