Atitit 词法分析的方法attilax总结

[1.1. 词法分析器读入程序员写的程序，然后对字符流做切分成记号流。 1](#_Toc11139)

[1.2. 2.6 词法分析 1](#_Toc19464)

[1.3. 再看第二个问题，词法分析器怎么实现呢？目前来说一般有两种方案： 2](#_Toc19882)

[1.4. 如何区别关键字和标识符呢？ 2](#_Toc20383)

[1.5. 使用递归的方法来进行词法分析,这里使用了 while 循环... 3](#_Toc25565)

[1.6. 。由词法分析识别出的常数和由用户定义的名字，分别在常数表和符号表中予以登记，在编译的各个阶段都要频繁的使用符号表。  3](#_Toc3264)

[1.7. 。常数可归为一类，也可按整型，实型，字符型，布尔型等分类，标识符类似处理 3](#_Toc1395)

[1.8. 状态转换图，简称转换图，是设计词法分析器的有效工具。 3](#_Toc5296)

[1.9. 每一个符号要用一个类别标识。如下图 4](#_Toc16554)

[1.10. 13、error  出错处理的函数，处理出现的词法错误。有一类词法错误可以在词法分析时发现，如出现字母表以外的非法字符、不合规则的常数、标识符的前缀为保留字等。但还有一类词法错误，例如，把if写成fi，词法分析会将fi当作标识符处理，le ngth中间多了一个空格，词法分析会将le和ngth当作两个标识符处理。这类词法错误往往要推迟到语法分析时才能发现，不属于函数处理的范畴。 5](#_Toc7838)

## 词法分析器读入程序员写的程序，然后对字符流做切分成记号流。

## 2.6 词法分析

　　2.6.1 剔除空白和注释  
　　2.6.2 预读  
　　2.6.3 常量  
　　2.6.4 识别关键字和标识符  
　　2.6.5 词法分析器  
　　2.6.6 2.6节的练习

## 再看第二个问题，词法分析器怎么实现呢？目前来说一般有两种方案：

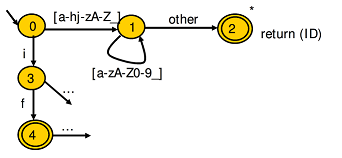
　　　　1.手工编码实现法。这种方法相对复杂，容易出错，但是非常流行的方法，如GCC,LLVM等。

　　　　2.词法分析器的生成器。这种方法可快速原型，代码量少，但较难控制细节。

## 如何区别****关键字****和****标识符****呢？

因为在很多语言中关键字和标识符是有交集的，但从词法分析上看，关键字是标识符的一部分

**第一种**，我们需要将上面的转移图稍加修改：（以识别if为例）

此处还应注意一个问题，假如输入ifxy呢？并不是关键字，因此这里还需判断。

再看**第二种**，使用关键字表，即Hash表。对语言中所有的关键字构建哈希表，对所有的关键字和标识符，先统一按标识符的转移表进行识别，识别完成后，再查哈希表看是否为关键字。通过构造合理的哈希表，可以再O(1)时间内完成。

## 使用递归的方法来进行词法分析,这里使用了 while 循环...

原则上应该递归 nextToken比较好些吧。貌似很多源码这样写。

## 。由词法分析识别出的常数和由用户定义的名字，分别在常数表和符号表中予以登记，在编译的各个阶段都要频繁的使用符号表。

## 输入缓冲（预读问题）。可以使用带有“哨兵标记”的双缓冲区方案来解决，

对于这个问题，可以使用带有“哨兵标记”的双缓冲区方案来解决，同时使用Lexeme和Forward两个指针进行词素的预读，具体不进行赘述。

## 。常数可归为一类，也可按整型，实型，字符型，布尔型等分类，标识符类似处理

。在这种情况下，每一类别中的常数或标识符将由第二元单词的属性值来区别。通常将常数在常数表中的位置编号作为常数的属性值，从而将标识符在符号表中的位置编号作为标识符的属性值。

## 状态转换图，简称转换图，是设计词法分析器的有效工具。

状态转换图是有限有向图，图中的结点代表状态，结点间的有向边代表状态之间的转换关系，有向边上标记的字符表示状态转换的条件。   
状态的数量是有限的，其中必有一个初始状态，若干个终止状态。大部分终止状态可对应一类单词符号的成功识别，所以也被成为识别状态。在识别状态下，可以给出相应单词的类别编码和属性值。某些终止状态是在多识别了一个字符后才成为识别状态的，对于这种情况，多识别的字符应予以退回，在终态上标以“\*”作为区别。   
有了状态转换图后，就可以方便地设计和实现词法分析器了。

## 每一个符号要用一个类别标识。如下图

及其编码，其中助记符是用于方便书写和记忆的。



## 13、error  出错处理的函数，处理出现的词法错误。有一类词法错误可以在词法分析时发现，如出现字母表以外的非法字符、不合规则的常数、标识符的前缀为保留字等。但还有一类词法错误，例如，把if写成fi，词法分析会将fi当作标识符处理，le ngth中间多了一个空格，词法分析会将le和ngth当作两个标识符处理。这类词法错误往往要推迟到语法分析时才能发现，不属于函数处理的范畴。

《编译原理（第2版）——计算机科学丛书》(（美）阿霍　等著)【简介\_书评\_在线阅读】 - 当当图书.html

编译原理学习--词法分析（1） - Anthony007 - 博客园.html

Java 实现词法分析器（编译原理） - u012577528的专栏 - 博客频道 - CSDN.NET.html