编译原理

田玲 教授、博导

lingtian@uestc.edu.cn

第六章 词法分析

主要内容

1.介绍词法分析的过程

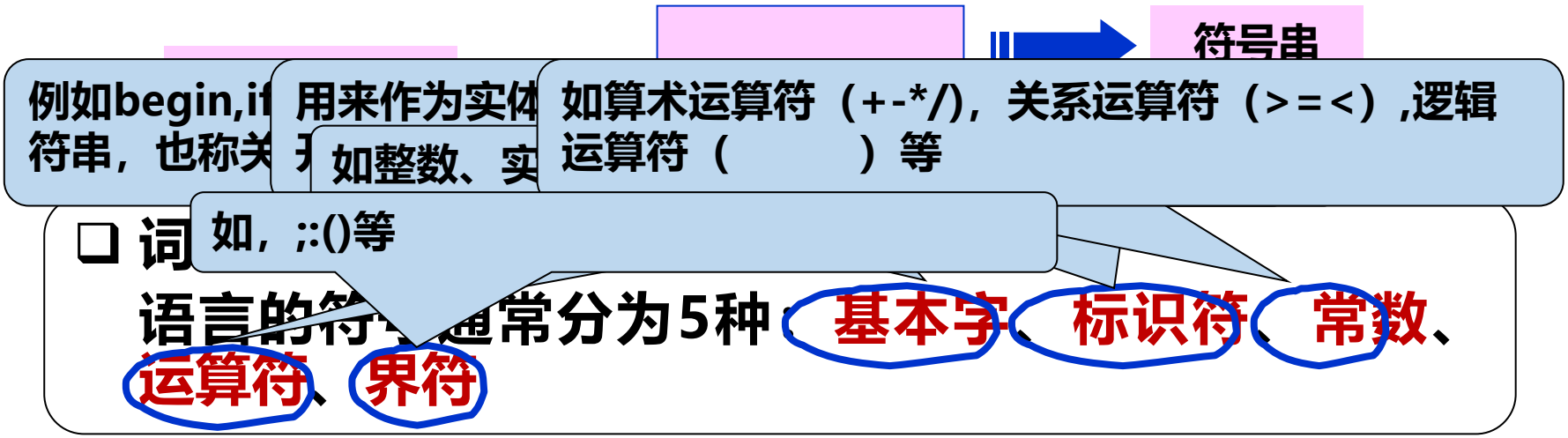
2.讨论词法分析器的设计与实现

3.介绍实现词法分析器的主要工具：状态转换图

第一节 词法分析概述

词法分析器的功能

- 从左到右逐个字符扫描源程序的字符流，分析出一个个单词符号，把由字符串表示的源程序转换成由符号串组成的串，供语法分析器使用；并对识别过程中发现的错误，输出有关信息；



第一节 词法分析概述

❑ 词法分析器符号的输出形式：
二元式 （符号的种别 符号自身的值）

例：扫描语句
A:=B50+10;
的输出为：
(标识符的编码 ‘A’在符号表中的位置)
(‘:=’ 的编码)
(标识符的编码 ‘B50’在符号表中的位置)
(‘+’ 的编码)
(整数的编码 ‘10’在常数表中的位置)

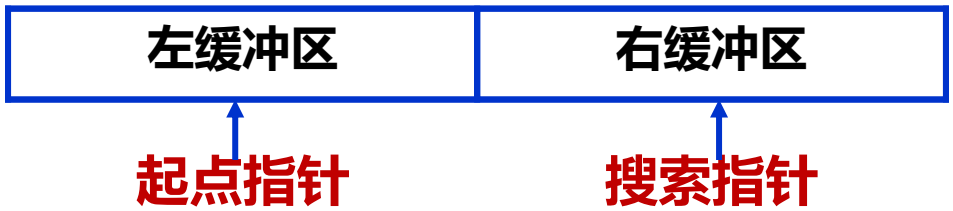
注意：[illegible]本字、运算符、界符一符一种，标识符为一种，常数按类型分种。

第二节 词法分析器的结构

- **扫描缓冲区**：完成预处理后，词法分析器从输入缓冲区输入一个固定长度的[illegible]到另一个缓冲区，这个缓冲区称为扫描缓[illegible]

去掉对于程序无意义的字符，比如程序员所加的注释，无用的空白、制表符、回车换行符，以及按格式输入语言的续行及行结束符等；

- **扫描缓冲区的结构**：
  - ✓**起点指针**：用来指示正在扫描的单词的起点；
  - ✓**搜索指针**：用于向前搜索，寻找单词的结束；
  - ✓**双缓冲区结构**：设置左右两个缓冲区，当左缓冲区读完后，新读入的字符存入右缓冲区；反之，存放在左缓冲区；



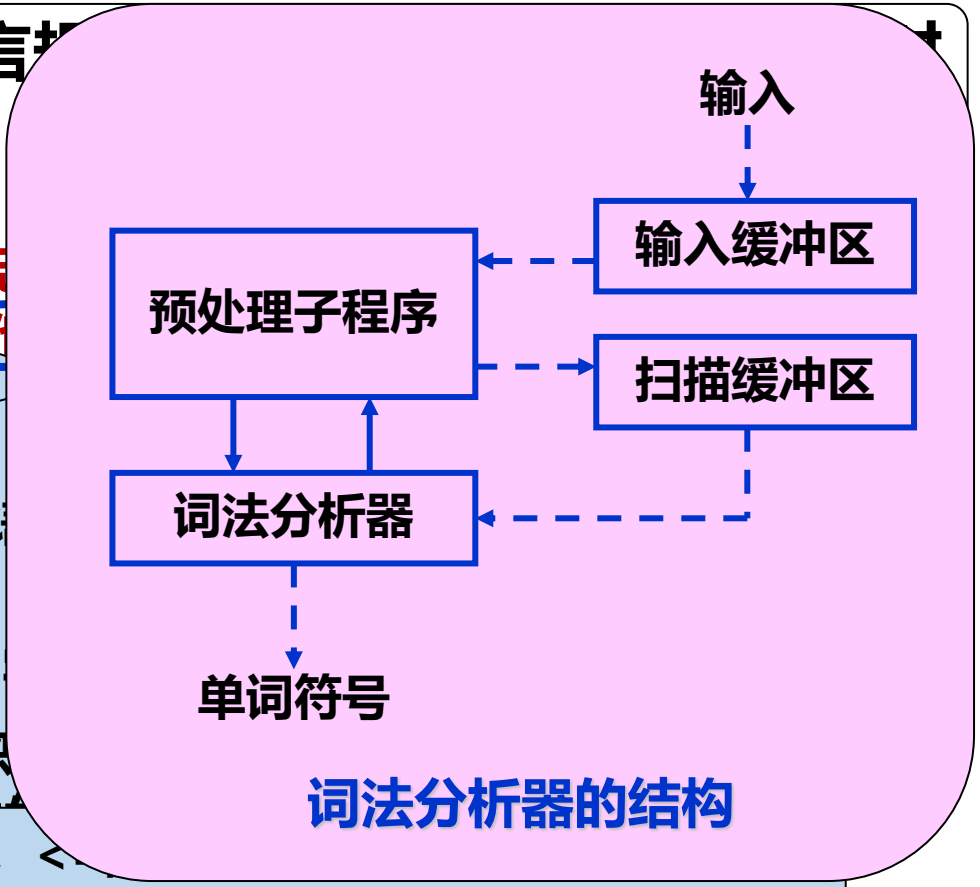
第二节 词法分析器的结构

- **符号的识别**：根据语言
  不同类型的单词符号，
  - ① 基本字的识别
  - ✓ 如果基本字有特定标
  - ✓ 如果允许基本字作
  - ✓ 本书的词法分析器
    识时
    符

例如，FORTRAN
DO 88 K=1， 10
DO88K=1,10
前两个是DO循环
DO88K是一个标识

例如，<
于.EQ.等

- ② **标识符的识别**：标识符以
  中通常后跟界符，识别时
- ③ **常数的识别**：常数
  别出常数后，要将其
- ④ **运算符的识别**：一个字
  个字符组成的运算符，要将其合成后再确定；
- ⑤ **界符的识别**：单字界符；

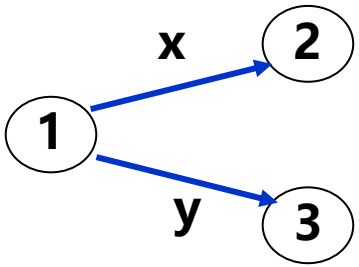


词法分析器的结构

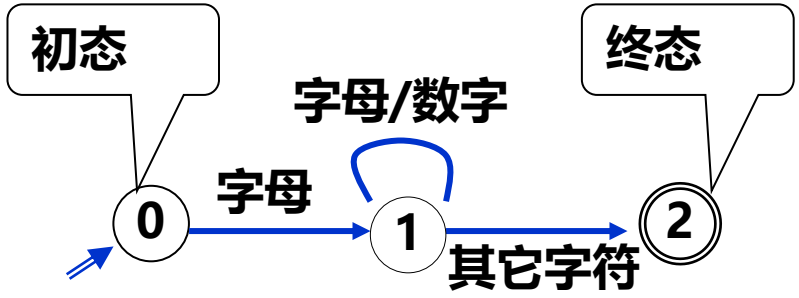
第三节 状态转换图

- **状态转换图**（state transition diagram），简称转换图,是一张有限方向图，是设计词法分析器的有效工具；它由如下成分构成：
  - ✓**结点**(node)：圆圈表示结点，代表状态（state）
  - ✓**有向边**（弧）:连接结点，边上的标记字符表示该状态下可能接收或识别的字符；

例子



状态转换图



识别标识符的转换图

第四节 词法分析器的设计

以第四章的小语言为例，设计一个词法分析器。下面是它所有的单词符号

单词符号	种别编码	助忆符	内码值
begin	01	$BEGIN	-
end	02	$END	-
integer	03	$INTEGER	-
if	04	$IF	-
then	05	$THEN	-
else	06	$ELSE	-
function	07	$FUNCTION	-
read	08	$READ	-
write	09	$WRITE	-

单词符号	种别编码	助忆符	内码值
标识符	10	$ID	字符串
常数	11	$INT	二进制值
=	12	$EQ	-
<>	13	$NE	-
<=	14	$LE	-
<	15	$LT	-
>=	16	$GE	
>	17	$GT	-

第四节 词法分析器的设计

单词符号	种别
-	
*	
:=	
(	
)	
;	

1.全局字符变量

2.字符数组token

3.读一字符的函数getchar

4.读入非空白字符的函数getnbc

5.连接字符串函数concat

6.判定字母函数letter

7.判定数字函数digit

8.回退一字符子程序retract

9.查保留字子程序reserve

10.处理标识符函数symbol

11.处理常数表的函数constant

12.出错处理函数error

把char中的字符连

布尔函数；若char

布尔函数；若char

将刚读入char中的字

用token中的字符串转换成二进制，存入常数表，

其中num是种别编码，val或者是token

位置，或者

的位置

处理出现的词法错误

字符是否为

若是，调用

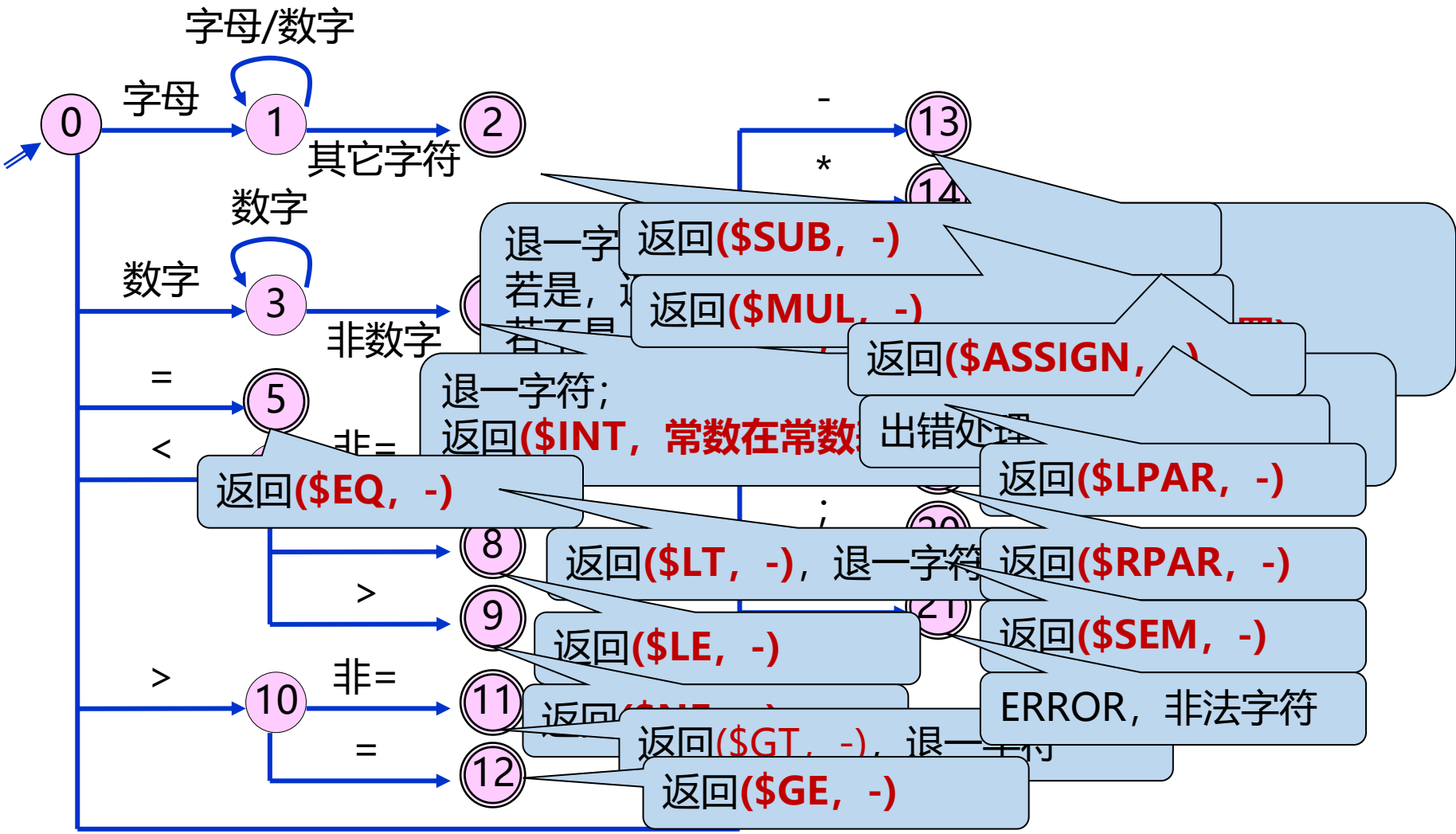
r读入下一字符，

中的字符不是空白符

,val)

error

词法分析器的状态转换图



❑ 人

以上状

```
  '<'  :   begin
             getchar;
             if(char =  '='  ) then return($LE, -);
             else if (char =  '>'  ) then return($NE, -);
            else begin retract; return($LT, -) end;
       end;
  '>'  :   begin
             getchar;
             if(char =  '='  ) then return($GE, -);
             else begin retract; return($GT, -) end;
           end;
  ':'  :   begin
             getchar;
             if(char =  '='  ) then return($ASSIGN, -);
       else error;
  '('  :   return($LPAR, -);
  ')'  :   return($RPAR, -);
  ';'  :   return($SEm, -);
other: error;
end of case;
goto start;
```