

编译技术

# 词法分析

大连理工大学软件学院

## 编译过程



# 词法分析器的功能



# 词法分析



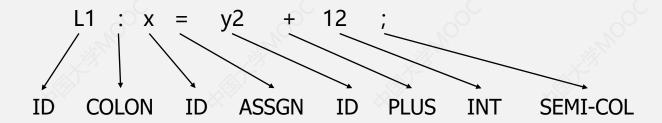
## 词法分析:编译第一步

◆ 看一个中文的句子



通过分词操作, 把句子以单词或者词组为单位进行划分, 得到一个句型。

## C语言的语句例子



编译的词法分析做的工作类似于分词,

把原始的字符串流形式的程序文本转换为词法记号流的形式。

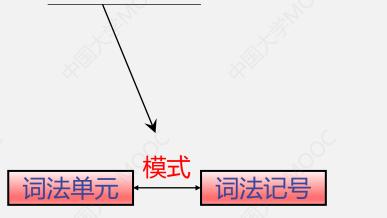
词法单元 又称单词,是编程语言中合法的字符串

例子中 哪些是**词法单元**?

词法记号 满足某种规则的词法单元,采用同一种记法。

## 词法单元与词法记号

满足一个给定规则的词法单元,被记为一个词法记号。



## 词法模式

> C语言的标识符

?

x2, 12, \_12, \_abc 哪些是合法的C标识符?

C语言标识符的规则 (模式):

首字符必须是\_或者字母,由\_、字母或数字组成的字符串

#### 》 常见记号及模式的例子:

词法

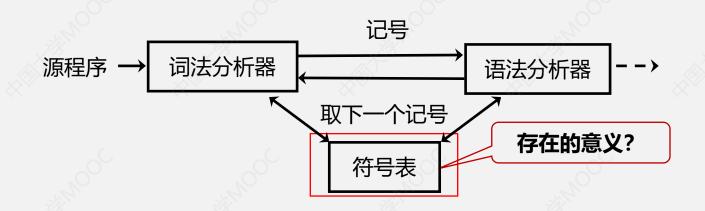
模式



## 词法记号、模式、词法单元

名词	大连 软件 大黑山	表示名称的词
连词	和与或	和与或
词法记号	词法单元例举	模式的非形式化描述
<b>1</b>	<b>1</b>	
词法记号	词法单元例举	模式的非形式描述
relation	< , < = , = ,	< 或 <= 或 = 或
id	sum, count, D5	由字母开头的字母数字串
词法记号	词法单元例举	模式的非形式描述
中国人	胡锦涛 毛泽东	具有中国国籍的人
美国人	奥巴马 克林顿	具有美国国籍的人

## 符号表的作用



### 词法记号的属性

- 如果简单地把词法记号流传给语法分析器,会产生什么后果?
  - 语义被完全摒弃,只剩下一个语法结构。

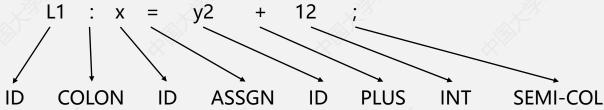
黄蓉是美女 翻译官 Pronoun Verb Noun

说了什么呀????

#### 词法记号的属性



每个词法记号具有一定的含义 (属性)



第一个ID, 名称是L1, 表示的是标号 (Label) 第二个ID, 名称是x, 表示的是一个变量, 类型是int 第三个ID, 名称是y2, 表示的是一个变量, 类型是int

#### 词法记号及属性

```
position := initial + rate * 60的记号和属性值
(id, 指向符号表中position条目的指针)
⟨assign _ op ⟩
 (id, 指向符号表中initial条目的指针)
 <add_op >
 (id, 指向符号表中rate条目的指针)
 \langle mul\_op \rangle
 〈num,整数值60〉
```

## 词法记号的描述与识别

• 词法模式的表示方法,是词法记号描述的核心

词法单元

• 下面是用于描述词法记号的模式语言。

- 字母表:符号的有限集合,例: $\Sigma = \{0,1\}$ 

— 串:符号的有穷序列,例: 0110, ε·

长度为0的空串

- 语言:字母表上的一个串集





## 串的运算

- 连接 xy sε = εs = s
- 积 (指数) s<sup>0</sup>为ε, s<sup>i</sup>为s<sup>i-1</sup>s (i>0)



#### 语言的运算

- 和:  $L \cup M = \{s \mid s \in L$ 或  $s \in M\}$
- 连接: *LM* = {*st* | *s* ∈ *L* 且 *t* ∈ *M*}
- 指数: L<sup>0</sup>是{ε}, L<sup>i</sup>是L<sup>i-1</sup>L
- 闭包: L\* = L<sup>0</sup> ∪ L<sup>1</sup> ∪ L<sup>2</sup> ∪...
- 正闭包:  $L^+ = L^1 \cup L^2 \cup ...$   $L^* = L^0 \cup L^+$
- 例2.2 (p17)
  - L: { A, B, ..., Z, a, b, ..., z }, D: { 0, 1, ..., 9 }
  - LUD, LD, L<sup>6</sup>, L\*, L(LUD)\*, D+



词法分析器对源程序采取非常局部的观点,难以发现下面的错误:

▶ 例外: 如在实数是a.b格式下,可以发现下面的错误 123.



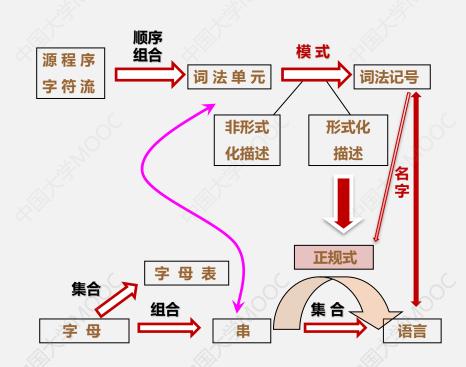
### 词法错误的处理方法

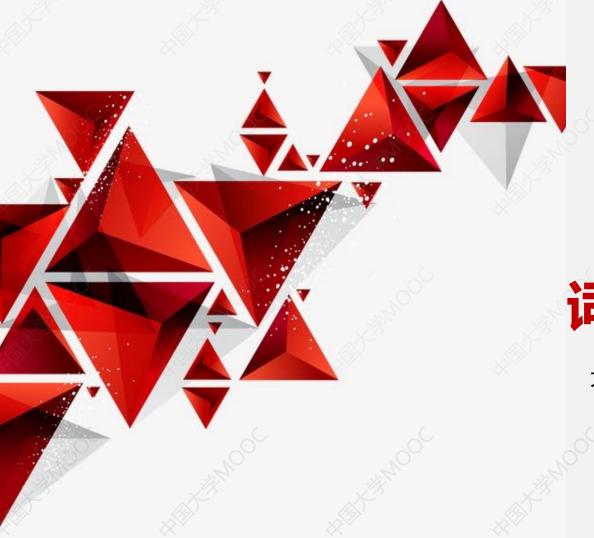
- 恢复策略
  - "紧急方式"
  - 错误修补尝试
    - 删除一个多余的字符
    - » 插入一个遗漏的字符
    - » 用一个正确的字符代替一个不正确的字符
    - > 交换两个相邻的字符

# 词法分析器工作原理:

源代码 词法分析器 一一记号 (token) 流







编译技术

# 词法分析

大连理工大学软件学院