

词法分析

状态转换图

- 状态转换图是一张有限方向图
 - 节点表示状态
 - 箭弧为输入的字符和字符类
 - 包含有限个状态，一个初态，至少一个终态

词法分析器

任务

对源程序进行扫描，产生单词符号

- 特点
- 输入源程序，输出单词符号
 - 单词符号种类
 - 基本字
 - 标识符
 - 常数
 - 运算符
 - 界符

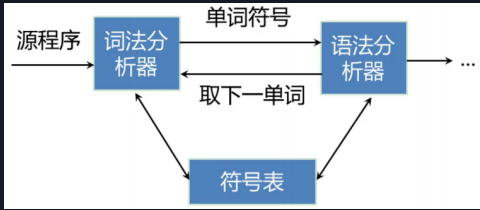
别称

扫描器

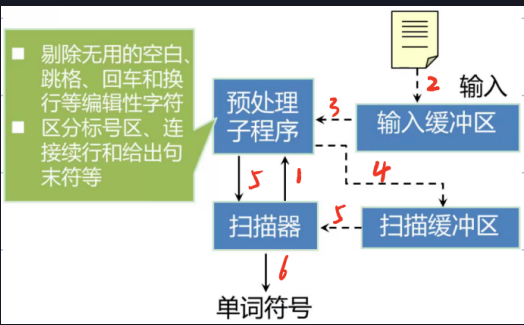
输出

表示形式

- 单词种别
- 单词自身的值



词法分析器可作为语法分析器的子程序



结构

正规表达式与有限自动机

正规式和正规集

正规式是一个表达式，正规集是正规式的集合

1、 定义

- (1) ϕ 是正规式，表示空集 $\{\}$;
- (2) ϵ 是正规式，表示 $\{\epsilon\}$;
- (3) $a \in \Sigma$ 都是正规式, 表示 $\{a\}$; (设 Σ 为有穷字母表)
- (4) 若 e_1 和 e_2 是正规式，表示的正规集分别是 L_1 和 L_2 ，
则 $e_1|e_2$ 是正规式，正规集为 $L_1 \cup L_2$ (或)
 $e_1 \cdot e_2$ 是正规式，正规集为 $L_1 \cdot L_2$ (链接)
 e_1^* 是正规式，正规集为 L_1^* (闭包)。

优先顺序为 $*$, \cdot , $|$

定义

正规式的等价

两个正规式的正规集相同，则两者等价

作用

用于识别正规集

确定的有限自动机(DFA)

形式

- 状态转换图
- 状态转换矩阵
- 形式化描述

形式定义(五元组)

DFA M 是一个五元组
 $M = (S, \Sigma, f, S_0, Z)$

其中:

- (1) S 是有穷状态集
- (2) Σ 是有穷字母表
- (3) f 是从 $S \times \Sigma$ 至 S 的状态转换函数(单值)
- (4) S_0 是唯一的初态
- (5) Z 是终态集 (终态集可空)

非确定的有限自动机(NFA)

有多值映射必为NFA

NFA表达式不唯一

实现过程

- 1.正规式 \rightarrow NFA
- 2.NFA确定化(子集构造法)
 - 不带 ϵ 弧的NFA确定化
 - 带 ϵ 弧的NFA确定化