

# 正则表达式

- 正则表达式
- 自动机和正则表达式
- 正则表达式的代数定律

# 正则表达式的代数定律

## 定义

含有变量的两个正则表达式, 如果以任意语言替换其变量, 二者所表示的语言仍然相同, 则称这两个正则表达式**等价**. 在这样的意义下, 正则表达式满足一些**代数定律**.

- 并运算

$$(L + M) + N = L + (M + N) \quad (\text{结合律})$$

$$L + M = M + L \quad (\text{交换律})$$

$$L + L = L \quad (\text{幂等律})$$

$$\emptyset + L = L + \emptyset = L \quad (\text{单位元 } \emptyset)$$

# 正则表达式的代数定律

- 连接运算

$$(LM)N = L(MN) \quad (\text{结合律})$$

$$\epsilon L = L\epsilon = L \quad (\text{单位元 } \epsilon)$$

$$\emptyset L = L\emptyset = \emptyset \quad (\text{零元 } \emptyset)$$

$$LM \neq ML$$

- 分配率

$$L(M + N) = LM + LN \quad (\text{左分配律})$$

$$(M + N)L = ML + NL \quad (\text{右分配律})$$

- 闭包运算

$$(L^*)^* = L^*$$

$$\emptyset^* = \epsilon$$

$$\epsilon^* = \epsilon$$

$$L^* = L^+ + \epsilon$$

$$(\epsilon + L)^* = L^*$$

$$Q = P^+$$

$$\begin{aligned} R &= Q + R^+ \\ &= QP^+ \end{aligned}$$

Ardan theorem

# 发现与验证正则表达式的定律

## 检验方法

要判断表达式  $E$  和  $F$  是否等价, 其中变量为  $L_1, \dots, L_n$ :

- ① 将变量替换为具体表达式, 得正则表达式  $r$  和  $s$ ,  
例如, 替换  $L_i$  为  $a_i$ ;
- ② 判断  $L(r) \stackrel{?}{=} L(s)$ , 如果相等则  $E = F$ , 否则  $E \neq F$ .

例 10. 判断  $(L + M)^* = (L^*M^*)^*$ .

- ① 将  $L$  和  $M$  替换为  $a$  和  $b$ ;
- ②  $(a + b)^* \stackrel{?}{=} (a^*b^*)^*$ ;
- ③ 因为  $L((a + b)^*) = L((a^*b^*)^*)$ , 所以  $(L + M)^* = (L^*M^*)^*$ .

例 11. 判断  $L + ML = (L + M)L$ .

- ① 将  $L$  和  $M$  替换为  $a$  和  $b$ ;
  - ② 判断  $a + ba \stackrel{?}{=} (a + b)a$ ;
  - ③ 因为  $aa \notin a + ba$  而  $aa \in (a + b)a$ ;
  - ④ 所以  $a + ba \neq (a + b)a$ , 即  $L + ML \neq (L + M)L$ .
-

## 注意

这种方法仅限于判断正则表达式, 否则可能会发生错误.

例 12. 若用此方法判断  $L \cap M \cap N \stackrel{?}{=} L \cap M$ , 以  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  替换  $L, M, N$ , 有

$$\{a\} \cap \{b\} \cap \{c\} = \emptyset = \{a\} \cap \{b\},$$

而显然

$$L \cap M \cap N \neq L \cap M.$$