

第三章 正则表达式

2025年6月17日 20:22

正则表达式

$$\emptyset^0 = \emptyset^* = \{\varepsilon\}, \quad \emptyset^n = \emptyset, (n \geq 1)$$

优先级: $() > * > \cdot > +$

有穷自动机和正则表达式

定理: 若 $L = L(A)$ 是某 DFA A 的语言, 那么存在正则表达式 R 满足 $L = L(R)$.

(对任意 DFA A , 总能找到正则表达式 R , 使得 A 接受字符串和 R 匹配字符串一致)

证明: 状态消除法

化简规则: r 和 s 为两个正则表达式, \emptyset 为零元, r 为单位元

$$(\varepsilon + r)r^* = r^*$$

$$r + rs^* = rs^*$$

分治: 递归求解

状态消除法: 删除状态, 用新路径替换被删路径, 设计新正则表达式, 产生等价自动机.

定理: 正则表达式定义的语言, 都可被有穷自动机识别

定理: 任何正则表达式 r , 都存在等价的 ε -NFA A , 即 $L(A) = L(r)$, 并且 A 满足:

- ① 仅有一个接收状态
- ② 没有进入开始状态的边
- ③ 没有离开接受状态的边

正则表达式的代数定律

基本代数定律:

$$(L^*)^* = L^*; \quad \emptyset^* = \varepsilon; \quad \varepsilon^* = \varepsilon; \quad (\varepsilon + L)^* = L^*$$

校验方法: 代入法