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

$$
\begin{aligned}
(L+M)+N &= L+(M+N) && \text{(结合律)}\\
L+M &= M+L && \text{(交换律)}\\
L+L &= L && \text{(幂等律)}\\
\emptyset+L = L+\emptyset &= L && \text{(单位元 } \emptyset\text{)}
\end{aligned}
$$

正则表达式的代数定律

- 连接运算

$$
\begin{aligned}
(LM)N &= L(MN) && \text{(结合律)} \\
\varepsilon L &= L\varepsilon = L && \text{(单位元 } \varepsilon\text{)} \\
\emptyset L &= L\emptyset = \emptyset && \text{(零元 } \emptyset\text{)} \\
LM &\neq ML
\end{aligned}
$$

- 分配率

$$
\begin{aligned}
L(M+N) &= LM + LN && \text{(左分配律)} \\
(M+N)L &= ML + NL && \text{(右分配律)}
\end{aligned}
$$

- 闭包运算

$$
\begin{aligned}
(L^*)^* &= L^* \\
\emptyset^* &= \varepsilon \\
\varepsilon^* &= \varepsilon \\
L^* &= L^+ + \varepsilon \\
(\varepsilon + L)^* &= L^*
\end{aligned}
$$

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 $L_1, \ldots, L_n$:

❶ 将变量替换为具体表达式，得正则表达式 $\mathbf{r}$ 和 $\mathbf{s}$，
例如，替换 L_i 为 $\mathbf{a}_i$;

❷ 判断 $\mathbf{L}(\mathbf{r}) \stackrel{?}{=} \mathbf{L}(\mathbf{s})$，如果相等则 $E = F$，否则 $E \neq F$.

例 10. 判断 $(L + M)^* = (L^*M^*)^*$.

❶ 将 L 和 M 替换为 $\mathbf{a}$ 和 $\mathbf{b}$;

❷ $(\mathbf{a} + \mathbf{b})^* \stackrel{?}{=} (\mathbf{a}^*\mathbf{b}^*)^*$;

❸ 因为 $\mathbf{L}((\mathbf{a} + \mathbf{b})^*) = \mathbf{L}((\mathbf{a}^*\mathbf{b}^*)^*)$，所以 $(L + M)^* = (L^*M^*)^*$.

例 11. 判断 $L+ML=(L+M)L$.

❶ 将 L 和 M 替换为 $\mathbf{a}$ 和 $\mathbf{b}$;

❷ 判断 $\mathbf{a}+\mathbf{ba}\stackrel{?}{=}(\mathbf{a}+\mathbf{b})\mathbf{a}$;

❸ 因为 $aa\notin\mathbf{a}+\mathbf{ba}$ 而 $aa\in(\mathbf{a}+\mathbf{b})\mathbf{a}$;

❹ 所以 $\mathbf{a}+\mathbf{ba}\neq(\mathbf{a}+\mathbf{b})\mathbf{a}$, 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

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