课程简介与基础知识

- 课程简介
- 基础知识

基本概念

1. 字母表: 符号 (字符) 的非空有穷集.

$$\begin{aligned}\Sigma_1 &= \{0, 1\},\\ \Sigma_2 &= \{a, b, \ldots, z\},\\ \Sigma_3 &= \{x \mid x \text{ 是一个汉字}\}.\end{aligned}$$

2. 字符串: 由某字母表中符号组成的有穷序列.

若 $\Sigma_1 = \{0,1\}$, 那么 $0, 1, 00, 111001$ 为 Σ_1 上的字符串;

若 $\Sigma_2 = \{a, b, \ldots, z\}$, 那么 $ab, xkcd$ 为 Σ_2 上的字符串.

3. 空串: 记为ε, 有 0 个字符的串.

字母表 Σ 可以是任意的, 但都有 $\varepsilon \notin \Sigma$.

符号使用的一般约定:

- 字母表: $\Sigma, \Gamma, \ldots$
- 字符: $a, b, c, \ldots$
- 字符串: $\ldots, w, x, y, z$
- 集合: $A, B, C, \ldots$

4. 字符串的**长度**: 字符串中符号所占位置的个数, 记为$|\sqcup|$.
若字母表为 Σ, 可**递归定义**为:

$$|w| = \begin{cases} 0 & w = \varepsilon \\ |x| + 1 & w = xa \end{cases},$$

其中 $a \in \Sigma$, w 和 x 是 Σ 中字符组成的字符串.

5. 字符串 x 和 y 的连接: 将首尾相接得到新字符串的运算, 记为 $x \cdot y$ 或 xy . 同样, 可递归定义为

$$x \cdot y = \begin{cases} x & y = \varepsilon \\ (x \cdot z)a & y = za \end{cases},$$

其中 $a \in \Sigma$, 且 x, y, z 都是字符串.

对任何字符串 x, 有 $\varepsilon \cdot x = x \cdot \varepsilon = x$.

连接运算的符号 “$\cdot$” 一般省略.

6. 字符串 x 的 n 次幂$(n \geq 0)$, 递归定义为

$$x^n = \begin{cases} \varepsilon & n = 0 \\ x^{n-1}x & n > 0 \end{cases}.$$

例如, 若 $\Sigma = \{a, b\}$, 那么

$$\begin{aligned} (ba)^2 &= (ba)^1 ba \\ &= (ba)^0 baba \\ &= \varepsilon baba \\ &= baba \end{aligned} \qquad \begin{aligned} ba^2 &= ba^1 a \\ &= ba^0 aa \\ &= b\varepsilon aa \\ &= baa \end{aligned}$$

7. 集合 A 和 B 的连接, 记为 $A \cdot B$ 或 AB, 定义为

$$A \cdot B = \{w \mid w = x \cdot y,\ x \in A \text{ 且 } y \in B\}.$$

8. 集合 A 的 n 次幂$(n \geq 0)$, 递归定义为

$$A^n = \begin{cases} \{\varepsilon\} & n = 0 \\ A^{n-1}A & n \geq 1 \end{cases} .$$

那么, 若 Σ 为字母表, 则 Σ^n 为 Σ 上长度为 n 的字符串集合.
如果 $\Sigma = \{0, 1\}$, 有

$$\Sigma^0 = \{\varepsilon\},\ \Sigma^1 = \{0, 1\},\ \Sigma^2 = \{00, 01, 10, 11\},$$

$$\Sigma^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}, \ldots$$

9. 克林闭包(*Kleene Closure*):

$$\Sigma^* = \bigcup_{i=0}^{\infty} \Sigma^i.$$

10. 正闭包(*Positive Closure*):

$$\Sigma^+ = \bigcup_{i=1}^{\infty} \Sigma^i.$$

显然,

$$\Sigma^* = \Sigma^+ \cup \{\varepsilon\}.$$

语言

定义

若 Σ 为字母表且 $\forall L \subseteq \Sigma^*$, 则 L 称为字母表 Σ 上的语言.

- 自然语言, 程序设计语言等
- $\{0^n 1^n \mid n \geq 0\}$
- The set of strings of 0's and 1's with an equal number of each:

$$\{\varepsilon, 01, 10, 0011, 0101, 1100, \ldots\}$$

- $\emptyset$, $\{\varepsilon\}$ 和 Σ^* 分别都是任意字母表 Σ 上的语言, 但注意 $\emptyset \neq \{\varepsilon\}$

关于语言

唯一重要的约束就是所有字母表都是有穷的.

问题

自动机理论中的典型问题

判断给定的字符串 w 是否属于某个具体的语言 L,

$$w \in L?$$

- 任何所谓问题, 都可以转为语言成员性的问题
- 语言和问题其实是相同的东西

例 1. 若 x 和 y 是 Σ 上的字符串, 请证明 $|xy| = |x| + |y|$.

证明: 通过对 $|y|$ 的归纳来证明

❶ 基础: 当 $|y| = 0$, 即 $y = \varepsilon$

$$\begin{aligned} |x\,\varepsilon| &= |x| && \text{连接的定义} \\ &= |x| + |\varepsilon| && \text{长度的定义} \end{aligned}$$

❷ 递推: 假设 $|y| = n$ $(n \geq 0)$ 时命题成立,
那么当 $|y| = n + 1$, 即 $y = wa$

$$\begin{aligned} |x\,(wa)| &= |(xw)a| && \text{连接的定义} \\ &= |xw| + 1 && \text{长度的定义} \\ &= |x| + |w| + 1 && \text{归纳假设} \\ &= |x| + |wa| && \text{长度的定义} \quad \square \end{aligned}$$

形式化证明: 演绎法, 归纳法和反证法

续例 1. 若 x 和 y 是 Σ 上的字符串, 请证明 $|xy| = |x| + |y|$.

证明: 通过对 y 的归纳来证明

❶ 基础: $y = \varepsilon$ 时

$$\begin{aligned} |x\,\varepsilon| &= |x| && \text{连接的定义} \\ &= |x| + |\varepsilon| && \text{长度的定义} \end{aligned}$$

❷ 递推: 假设 $y = w$ $(w \in \Sigma^*)$ 时命题成立,
那么当 $y = wa$ 时

$$\begin{aligned} |x\,(wa)| &= |(xw)a| && \text{连接的定义} \\ &= |xw| + 1 && \text{长度的定义} \\ &= |x| + |w| + 1 && \text{归纳假设} \\ &= |x| + |wa| && \text{长度的定义} \quad \square \end{aligned}$$