6/10/2004
5/10/2005

Review of basic definitions:

- **Alphabet**: finite, nonempty set of symbols ... Σ

  e.g. $\Sigma = \{0, 1\}$

  $\Sigma = \{a, b, \dots, z\}$

  Σ = set of Unicode characters

- **String**: finite sequence of symbols from Σ

  $w = a_1 a_2 \cdots a_n$, with $a_i \in \Sigma$ for $i \in \{1, \dots, n\}$

  e.g. · 01101

  · ciaociao

  · empty string: denoted ε : string with no symbols

  · length of a string = number of (positions for) symbols in the string

  denoted $|w|$ If $w = a_1 \cdots a_n$, then $|w| = n$

  e.g. $|\varepsilon| = 0$ ε is the only string of length 0

  $|\phi| = 1$

  $|\text{ciaociao}| = 8$

  Notice: strictly speaking, the number of symbols in ciaociao is 4

- **Powers of an alphabet**:

  $\Sigma^k = \underbrace{\Sigma \times \Sigma \times \cdots \times \Sigma}_{k \text{ times}}$ ... set of all strings over Σ of length k

  e.g. $\Sigma^0 = \{\varepsilon\}$

  $\{0,1\}^1 = \{0,1\}$ ← what is the difference between lhs and rhs?

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

Closure of an alphabet Σ: Σ^* is the set of all finite strings over Σ

i.e. $\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$

also $\Sigma^+ = \Sigma^1 \cup \Sigma^2 \dots$ — hence $\Sigma^* = \Sigma^0 \cup \Sigma^+$

Note: all strings in Σ^* are finite

Σ^* is an infinite set

e.g. $\Sigma = \{0, 1\}$

$\Sigma^* = \{0, 1\}^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, \dots\}$

Concatenation of two strings:

$x = a_1 a_2 \cdots a_m \in \Sigma^*$

$y = b_1 b_2 \cdots b_n \in \Sigma^*$

$\Rightarrow x \cdot y = a_1 \cdots a_m b_1 \cdots b_n$ (we may omit the $\cdot$)

Note: $\varepsilon \cdot x = x \cdot \varepsilon = x$, i.e. ε is the identity for conc.

$|x \cdot y| = |x| + |y|$

Language L over Σ: is any subset of Σ^* (i.e. $L \subseteq \Sigma^*$)

Note: L contains only finite strings, but it may be infinite

Examples:

$\begin{cases} \Sigma = \{a, b, \dots, z\} \\ L = \text{set of all English words} \end{cases}$ $\begin{cases} \Sigma = \text{Unicode characters} \\ L = \text{compilable Java programs} \end{cases}$

$\begin{cases} \Sigma = \{0, 1\} \\ L = \{\varepsilon, 01, 0011, 000111, \dots\} \end{cases}$ all strings with equal # of 0 and 1, with all 0's preceding the 1's

$\emptyset$ the empty language $(\neq \{\varepsilon\})$