

Introduction

Reading: Sections 1.5 – 1.7

Definitions

- Symbol – An atomic unit, such as a digit, character, lower-case letter, etc. Sometimes a word.
- Alphabet – A finite set of symbols, usually denoted by Σ .

$$\Sigma = \{0, 1\}$$

$$\Sigma = \{0, a, 4\}$$

$$\Sigma = \{a, b, c, d\}$$

- String – A finite length sequence of symbols, presumably from some alphabet.

$$w = 0110$$

$$y = 0aa$$

$$x = aabcaa$$

$$z = 111$$

special string: ε

- Operations:

concatenation: $wz = 0110111$

length: $|w| = 4$ $|\varepsilon| = 0$ $|x| = 6$

reversal: $y^R = aa0$

- We are interested in *sets* of strings.
- Some special sets of strings:

Σ^*	All strings of (zero or more) symbols from Σ	(Kleene closure)
Σ^+	$\Sigma^* - \{\epsilon\}$	(positive closure)

- Example:

$$\Sigma = \{0, 1\}$$

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

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

Formal Definition of a Language

- A (formal) language is a set of strings from some alphabet.
- Alternatively, a (formal) language is any subset L of Σ^*
- Examples:

$$\Sigma = \{0, 1\}$$

$$\begin{aligned} L1 &= \{x \mid x \in \Sigma^* \text{ and } x \text{ contains an even number of } 0\text{'s}\} \\ &= \{\epsilon, 1, 111, 00, 100, 010, 001, 0000, 00100, \dots\} \end{aligned}$$

$$\Sigma = \{0, 1, 2, \dots, 9, .\}$$

$$\begin{aligned} L2 &= \{x \mid x \in \Sigma^* \text{ and } x \text{ forms a finite length real number}\} \\ &= \{0, 1.5, 9.326, \dots\} \end{aligned}$$

- Examples:

$$\Sigma = \{a, b, c, \dots, z, A, B, \dots, Z\}$$

$$\begin{aligned} L3 &= \{x \mid x \in \Sigma^* \text{ and } x \text{ is a Java reserved word}\} \\ &= \{\text{while, for, if, } \dots\} \end{aligned}$$

$$\Sigma = \{\text{ASCII characters}\}$$

$$L4 = \{x \mid x \in \Sigma^* \text{ and } x \text{ is a syntactically correct Java program}\}$$

$$\Sigma = \{\text{English words}\}$$

$$L5 = \{x \mid x \in \Sigma^* \text{ and } x \text{ is a syntactically correct English sentence}\}$$

- Even arbitrary sets of strings are languages:

$\{01, 10010, 1111, 0010\}$

$\{\varepsilon, 0, 11, 101010\}$

- Some special languages:

$\{\}$ The empty set/language, containing no strings, also denoted \emptyset .

$\{\varepsilon\}$ A language containing one string, the empty string.

Σ^* All strings of symbols from Σ (Kleene closure)

Σ^+ $\Sigma^* - \{\varepsilon\}$ (positive closure)