# Alphabets and Languages

Lecture notes based on "Elements of the theory of computation" by H.R. Lewis and C. H. Papadimitriou.

# Alphabets and Languages

"*In computational practice, data are encoded in the computer's memory as strings of bits or other symbols appropriate for manipulation by a computer. The mathematical study of the theory of computation must therefore begin by understanding the mathematics of strings of symbols.*".

**Alphabet:** a finite set of symbols, e.g., $\{a, b, \ldots, z\}$, $\{0, 1\}$.

**String:** A string over an alphabet is a finite sequence of symbols from the alphabet, e.g., "apple", "0011", "a", "e" (empty string). The set of all strings (including the empty string) over $\Sigma$ is denoted by $\Sigma^\star$. The length of a string is the length of the sequence, e.g. $|abc| = 3$. A string $w \in \Sigma^\star$ can be considered as a function $w : \{1, \ldots, |w|\} \to \sigma$, where $w(j)$ is the symbol at the $j-$th position.

**Language:** Any subset $L$ of $\Sigma^\star$ for an alphabet $\Sigma$ is called a language over $\Sigma$.

## String operations

**Concatenation:** Two strings $x, y$ over the same alphabet, e.g. $x, y \in \Sigma^\star$, can be combined. $w = x \circ y$, or simply $w = xy$. $|w| = |x| + |y|$, $w(i) = x(i)$ for $i \le |x|$, and $w(i) = y(i - |x|)$ for $i > |x|$.

$w \circ e = e \circ w = w$. Concatenation is associative, $x(yz) = (xy)z$

**Substring:** A string $v$ is a substring of $w$ if $w$ can be written as $w = xvy$. If $w = vx$ then $v$ is a prefix of $w$, and if $w = xv$, then $v$ is a suffix of $w$.

$w^i$**:** For each string $w \in \Sigma^\star$ and natural number $i \in \mathbb{N}$, $w^i$ is defined as (definition by induction):

$$w^0 = e$$
$$w^{i+1} = w^i \circ w \text{ for each } i \ge 0$$

**Reversal** The reverse of a string $w$, denoted by $w^R$, is the string spelled backwards, $w^R(i) = w(|w| - i + 1)$.
Inductive definition:

- If $|w| = 0$, then $w^R = w = e$
- If $|w| = n + 1$ for some $n \in \mathbb{N}$, then $w = ua$ for some $a \in \Sigma$, and $w^R = au^R$.

**Theorem 1** *For any two strings $x, w$, $(wx)^R = x^R w^R$.*

## Languages

Given an alphabet $\Sigma$, any subset of $\Sigma^\star$ is called a language. E.g. $\Sigma = \{0, 1\}$, $L_1 = \{0, 1\}$, $L_2 = \{e, 0, 1, 00, 01, 10, 11\}$, $L_3 = \{w \in \Sigma^\star \mid |w| \le 2\}$, $L_4 = \{w \in \Sigma^\star \mid w \text{ has an equal number of 0's and 1's}\}$.
Representation:

$$L = \{w \in \Sigma^\star \mid w \text{ has the property } P\}$$

### Language operations

Languages are sets, so set operations (union, intersection, difference) can be used on languages.

**Complement:** $\overline{L} = \Sigma^\star \setminus L$

**Concatenation:** $L_1, L_2$ are languages over $\Sigma$. $L = L_1 \circ L_2$ (or simply $L = L_1 L_2$) is defined as

$$L = \{w_1 w_2 \mid w_1 \in L_1 \text{ and } w_2 \in L_2\}$$

**Kleene star:** The Kleene star of a language $L$, denoted by $L^\star$, is the set of strings obtained by concatenating 0 or more strings from $L$.

Note that $\Sigma^\star$ and $L^\star$ definitions are consistent.

$L^+ = LL^\star$

See problems for section 1.7.