# Languages

Suradet Jitprapaikulsarn

## Alphabet

An ***Alphabet*** A is a finite nonempty set of symbols. E.g.

- $A_1 = \{ 0, 1 \}$
- $A_2 = \{ a, b, \ldots, z \}$
- $A_3 = \{ 0, \ldots, 9 \}$
- $A_4$ = set of all ASCII characters
- $A_5$ = set of all Unicode characters





## String

- A "***string***" or "***word***" w is a finite sequence of symbols from an alphabet. E.g.
- $w_1$ = 110101 is a string from $A_1$
- $w_2$ = hello is a string from $A_2$
- $w_3$ = 191 is a string from $A_3$
- $w_4$ = abc123 is a string from $A_4$
- $w_5$ = กขค阿123 is a string from $A_5$

## The empty string

- The ***empty string*** ε is the string with zero occurrence of symbols



## Length of strings

- Length of string = number of positions in a string
- number of positions ≠ number of symbols
- E.g.
- $|w_1| = 6$
- $|w_2| = 5$
- $|w_3| = 3$
- $|w_4| = 6$
- $|w_5| = 7$
- $|\varepsilon| = 0$



## Powers of Alphabet

- $A^n$ = the set of strings of length n
- $A^0 = \{ \varepsilon \}$
- $A^* = A^0 \cup A^1 \cup A^2 \cup \ldots$ Kleene Closure
- $A^+ = A^1 \cup A^2 \cup \ldots$ Positive Closure

- Note that $A^1 \neq A$

## Concatenation

- Let $x = a_1a_2 \ldots a_n$
- Let $y = b_1b_2 \ldots b_m$
- Then $xy = a_1a_2 \ldots a_nb_1b_2 \ldots b_m$
- and $|xy| = n + m$



## Language

- A language *L* over an alphabet A is a subset of A*