Formal Language Theory

Alphabet: any set

Any element of an alphabet: **character**, **symbol**

Σ to denote the alphabet.

$\Sigma = \{0, 1\}$ $\Sigma = \{\text{true}, \text{false}\}$ $\Sigma = \{a, b, c\}$

$\Sigma = \{0, 1 \ldots . 7\}$ $\Sigma = \{\#, !, \%, \$\}$

ASCII

$|\Sigma| = 128$ $\Sigma = \{a \ldots z, A \ldots z, 0 \ldots 9, =, :, <, >, \{, \}, \# \ldots\}$

unicode
- UTF-7
- UTF-8
- UTF-16
- UTF-32

String : A sequence of characters from an alphabet, chosen with repetition.

$\Sigma = \{0,1\}$ 01 11110 10101 0 0001

$\Sigma = \{A,C,G,T\}$ AGG GATC CACTTGCTGT

ε : empty string

Language : A set of strings from an alphabet. Σ

$\{0, 00, 000, 0000, \ldots\}$ is a language over $\Sigma = \{0,1\}$

$\{\varepsilon, 0, 1, 01, 00, 10, 11, 000 \ldots\}$ " " " "