ToC Class-1: Alphabets and Languages

Computer Science Department
RKMVERI

2025-07-23

Alphabet and String

- **Alphabet (Σ):** A **Finite** set of **symbols**.

- **Alphabet (Σ):** A **Finite** set of **symbols**.
  - {1,2,3}
  - {a,b,c,d}
  - {1,2,a,b,c}
  - **Binary alphabet:** $\{0, 1\}$

- **Alphabet (Σ):** A **Finite** set of **symbols**.
  - {1,2,3}
  - {a,b,c,d}
  - {1,2,a,b,c}
  - **Binary alphabet:** $\{0, 1\}$
- **String over an alphabet:** Finite sequence of symbols from the alphabet.

- **Alphabet (Σ):** A **Finite** set of **symbols**.
  - {1,2,3}
  - {a,b,c,d}
  - {1,2,a,b,c}
  - **Binary alphabet:** $\{0, 1\}$
- **String over an alphabet:** Finite sequence of symbols from the alphabet.
  - AUTOMATA is a string over alphabet $\{A, \ldots, Z\}$.
  - 110000110 is a string over alphabet $\{0, 1\}$.
  - A is a string over alphabet $\{A, \ldots, Z\}$
  - **Empty string:** String without any symbols (we will use e to denote empty string).

Kleene closure (*)

- Let Σ be an alphabet.
- **Kleene closure** (Σ^*): Set of all strings, including empty string, over Σ.

- Let Σ be an alphabet.
- **Kleene closure** (Σ^*): Set of all strings, including empty string, over Σ.
  - $\{0,1\}^* = \{e, 0, 1, 00, 01, 10, 11, 000, \ldots\}$.

- Let Σ be an alphabet.
- **Kleene closure** (Σ^*): Set of all strings, including empty string, over Σ.
  - $\{0, 1\}^* = \{e, 0, 1, 00, 01, 10, 11, 000, \ldots\}$.
  - Finite or Infinite ?
  - Countable or Uncountable ?

- Let Σ be an alphabet.
- **Kleene closure** (Σ^*): Set of all strings, including empty string, over Σ.
  - $\{0,1\}^* = \{e, 0, 1, 00, 01, 10, 11, 000, \ldots\}$.
  - Finite or Infinite ?
  - Countable or Uncountable ?
  - Σ^* is Infinite but Countable.
- **Enumeration** of Σ^*.
  - Let $\Sigma = \{a_1, a_2, \ldots, a_n\}$.
  - All strings of length k is enumerated before all strings of length $k+1$.
  - Length k strings are enumerated **lexicographically,** that is $a_{i_1} \ldots a_{i_k}$ proceeds $a_{j_1} \ldots a_{j_k}$ if $0 \leq m \leq k-1$, $a_{i_m} = a_{j_m}$ and $a_{i_k} < a_{j_k}$.
  - If $\Sigma = \{0, 1\}$, then

$$e, 0, 1, 00, 01, 10, 11, 000, \ldots$$

Operations on Strings

- **Length:** Number of symbols in the string.
  - $|101| = 3$.
  - $w = aabbab$, $|w| = 6$.
  - $|e| = 0$.

- **Length:** Number of symbols in the string.
  - $|101| = 3$.
  - $w = aabbab$, $|w| = 6$.
  - $|e| = 0$.
- **Occurrence:** Let $w \in \Sigma^*$, can be considered as a function

$$w : \{1, \ldots, |w|\} \to \Sigma$$

  such that, $w(i)$ is the symbol in the i^{th} position of w, where $i \in \{1, \ldots, |w|\}$.
  - If $w = occurrence$, then $w(1) = o, w(2) = c, w(3) = c, w(4) = u, w(5) = r, w(6) = r, w(7) = e, w(8) = n, w(9) = c$.

- **Concatenation:** Let $x, y \in \Sigma^*$, then $x \cdot y$ or xy is the string w where:
$$|w| = |x| + |y|,$$
$$w(i) = x(i) \text{ where } i \in \{1, \ldots, |x|\},$$
$$w(|x| + i) = y(i) \text{ where } i \in \{1, \ldots, |y|\}.$$
    - $001 \cdot 110 = 001110$.
    - $w \cdot e = e \cdot w = w$.
    - **Associative:** $w(xy) = (wx)y$.

- **Substring:** A string $v \in \Sigma^*$ is substring of string $w \in \Sigma^*$ if and only if there exists $x, y \in \Sigma^*$ such that $w = xvy$.
  - ab**abba**bb,aab**bba**,**abbab**bba,**abbabba**.
- **Prefix:** If $w = xv$ then x is prefix of w.
  - **abba**bb,**bba**,**abbab**bba.
- **Suffix:** If $w = xv$ then v is prefix of w.
  - ab**abba**,aab**bba**.
- **Power:** For $w \in \Sigma^*$ and a natural number i, w^i :

$$w^0 = e$$

$$w^{i+1} = w^i \cdot w \text{ for } i \geq 0$$

  - $w = ab$, $w^3 = ababab$.

- **Reversal:** Let $w \in \Sigma^*$,
  If $|w| = 0$, then $w^R = w = e$.
  If $|w| = n + 1$, then $w = va$ for some $a \in \Sigma$, and $w^r = av^r$.
  - $automata^R = atamotua$.

An inductive proof

- **Claim:** Let $w, x \in \Sigma^*$, then $(w \cdot x)^R = x^R \cdot w^R$.
  - Example: $(catdog)^R = (dog^R) \cdot (cat)^R = godtac$.
- **Proof:** We use **induction on length of** x.
  - **Induction Base:** $|x| = 0$, then $x = e$ and,

$$(w \cdot x)^R = (w \cdot e)^R = w^R = e \cdot w^R = e^R \cdot w^R = (e \cdot w)^R = (x \cdot w)^R$$

  - **Induction Hypothesis:** If $|x| \leq n$, then $(w \cdot x)^R = x^R \cdot w^R$.
  - **Induction step:** Let $|x| = n + 1$, then $x = v \cdot a$ for some $v \in \Sigma^*$ and $a \in \Sigma$. Here $|v| = n$.

$$\begin{aligned}
(w \cdot x)^R &= (w \cdot (v \cdot a))^R, && \text{Since } x = v \cdot a \\
&= ((w \cdot v) \cdot a))^R, && \text{From associativity of concatenation} \\
&= a(wv)^R, && \text{From definition of reversal} \\
&= a \cdot v^R \cdot w^R, && \text{From induction hypothesis} \\
&= (v \cdot a)^R \cdot w^R, && \text{From definition of reversal} \\
&= x^R \cdot w^R, && \text{Since } x = v \cdot a
\end{aligned}$$

Language

- **Language** over Σ is a subset of Σ^*.
  - $\{a, abba, ababa, aaaaa, bbbbb\}$ is a language over $\{a, \ldots, z\}$
  - $\{w \in \{0,1\}^* : w$ has equal number of 0 and 1$\}$.
  - $\{w \in \Sigma^* : w = w^R\}$.

Operations on Language

- **Complement:** $L \subseteq \Sigma^*$, then $\overline{L} = \Sigma^* \setminus L$.
- **Concatenation:** $L = L_1 \cdot L_2$ or $L = L_1 L_2$ if,

$$L = \{w \in \Sigma^* : w = x \cdot y \text{ where } x \in L_1, y \in L_2\}$$

- **kleene star:** Consider alphabet Σ and a language $L \subseteq \Sigma^*$. Then,

$$L^* = \{w \in \Sigma^* : w = w_1 \cdot \ldots \cdot w_k \text{ where } k \geq 0 \text{ and } w_1, \ldots, w_k \in L\}$$

  - Let $L = \{01, 1, 100\}$, then, $011100011 \in L^*$, which is $01 \cdot 1 \cdot 100 \cdot 01 \cdot 1$.

- **Claim:** $L = \emptyset$, then $L^* = ?$

- **Claim:** $L = \emptyset$, then $L^* = e$.
- **Proof:** From definition $w_1 \cdot \ldots \cdot w_k$, with $k \geq 0$, and $w_1, \ldots, w_k \in \emptyset^*$.
- Here, $k = 0$, and concatenation of zero strings is e.

- **Claim:** $L_1 \subseteq \Sigma^*$, and $L_2 \subseteq \Sigma^*$. If $L_1 \subseteq L_2$ then $L_1^* \subseteq L_2^*$.
- **Proof:** Let $w \in L_1^*$ such that $w = w_1 \cdot \ldots \cdot w_k$ and $w = w_1, \ldots, w_k \in L_1$.
    - Since $L_1 \subseteq L_2$ it follows that $w_1, \ldots, w_k \in L_2$. Hence $w \in L_2^*$.
    - For all $w \in L_1^* \Rightarrow w \in L_2^*$.
    - Hence $L_1^* \subseteq L_2^*$.

- L^+**:** Consider alphabet Σ and a language $L \subseteq \Sigma^*$. Then,

  $L^* = \{w \in \Sigma^* : w = w_1 \cdot \ldots \cdot w_k \text{ where } k \geq 1 \text{ and } w_1, \ldots, w_k \in L\}$