Chapter 1

Regular Languages

1.1 Deterministic Finite State Automata

Definition 1.1.1. An **alphabet** Σ is a finite set of symbols.

- A **string** over Σ is a finite sequence of symbols from Σ.
- The **length** of a string w, denoted by $|w|$, is the number of symbols it contains.
- The string of length 0 is called the **empty string**, denoted by ϵ.

Definition 1.1.2. Let Σ be an alphabet.

- For any nonnegative integer n, Σ^n denotes the set of words of length n.
- Σ^* denotes the set of all strings over Σ.
- A **language** over Σ is a subset of Σ^*.

Definition 1.1.3. A **deterministic finite state automaton** is a system $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$, where each component is as follows.

- Σ is the alphabet.
- Q is a finite set of **states**.
- $q_0 \in Q$ is the **initial** state.
- $F \subseteq Q$ is the set of **accepting** states.
- δ is the **transition function** from $Q \times \Sigma$ to Q.

Definition 1.1.4. The **run** of DFA $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$ on an input string $w = a_1 \cdots a_n$ over Σ is the sequence of states

$$r = (r_0, r_1, \ldots, r_n)$$

where $r_0 = q_0$ and $\delta(r_{i-1}, a_i) = r_i$ for each $i \in \{1, \ldots, n\}$.

- r is an **accepting** run if $r_n \in F$.
- We say that $\mathcal{A}$ **accepts** w if the run of $\mathcal{A}$ on w is an accepting run.

- The language of all strings accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$.
- A language L is **regular** if there is a DFA $\mathcal{A}$ with $L = L(\mathcal{A})$.

Remark. For DFA $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$, the empty string ϵ is accepted by $\mathcal{A}$ if and only if $q_0 \in F$.

1.2 Nondeterministic Finite State Automata

Definition 1.2.1. A **nondeterministic finite state automaton** is a system $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$, where each component is as follows.

- Σ is the alphabet.
- Q is a finite set of **states**.
- $q_0 \in Q$ is the **initial** state.
- $F \subseteq Q$ is the set of **accepting** states.
- $\delta \subseteq Q \times \Sigma \times Q$ is the **transition relation**.

Definition 1.2.2. A **run** of NFA $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$ on an input string $w = a_1 \cdots a_n$ over Σ is the sequence of states

$$r = (r_0, r_1, \ldots, r_n)$$

where $r_0 = q_0$ and $(r_{i-1}, a_i, r_i) \in \delta$ for each $i \in \{1, \ldots, n\}$.

- r is an **accepting** run if $r_n \in F$.
- We say that $\mathcal{A}$ **accepts** w if there is an accepting run of $\mathcal{A}$ on w.
- The language of all strings accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$.

Theorem 1.2.3. For every NFA $\mathcal{A}$, there is a DFA $\mathcal{A}'$ with $L(\mathcal{A}) = L(\mathcal{A}')$.

Proof. To be completed. $\square$