Lesson 2: Deterministic finite state automata

Theme: Deterministic finite state automata.

1 The notion of alphabets and languages

- An alphabet is a finite set of symbols. We usually use the symbol $\Sigma$ to denote an alphabet.
- A (finite) string/word over $\Sigma$ is a finite sequence of symbols from $\Sigma$.
- We will usually write $w = a_1 \ldots a_n$ to denote a word whose label in position $i$ is $a_i$. The length of $w$ is denoted by $|w|$.
- We write $\varepsilon$ to denote the empty string/word, i.e., word of length 0.
- For an integer $n \geq 0$, $\Sigma^n$ denotes all the words over $\Sigma$ of length $n$.
- $\Sigma^*$ denotes the set of all finite words over $\Sigma$, i.e., $\Sigma^* = \bigcup_{n \geq 0} \Sigma^n$.
- A language $L$ over $\Sigma$ is a subset of $\Sigma^*$.

2 Deterministic finite state automata

A deterministic finite state automaton (DFA) is a system $A = (\Sigma, Q, q_0, F, \delta)$, where each component is as follows.

- $\Sigma$ 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 final states.
- $\delta : Q \times \Sigma \to Q$ is the transition function.

Remark 2.1 A DFA $A = (\Sigma, Q, q_0, F, \delta)$ can be visualised as a directed graph as follows.

- The vertices are elements of $Q$.
- There is an edge from $p$ to $p'$ labeled with $a$, if $\delta(p, a) = p'$.

On input word $w = a_1 \ldots a_n$, the run of $A$ on $w$ is the sequence:

$$q_0 \ a_1 \ q_1 \ a_2 \ q_2 \ \cdots \ a_n \ q_n,$$

where $\delta(q_i, a_{i+1}) = q_{i+1}$, for each $i = 0, \ldots, n - 1$. It is called accepting run, if $q_n \in F$. We say that $A$ accepts $w$, if there is an accepting run of $A$ on $w$. The language of all words accepted by $A$ is denoted by $L(A)$.

A language $L$ is called a regular language, if there is a DFA $A$ such that $L(A) = L$.

Remark 2.2 Let $A = (\Sigma, Q, q_0, F, \delta)$ be a DFA.

- The empty string $\varepsilon$ is accepted by $A$ if and only if $q_0 \in F$.
- For every word $w$, there is exactly one run of $A$ on $w$. 

1/2
Theorem 2.3  Regular languages are closed under boolean operations, i.e., intersection, union, and complementation. More formally, it can be stated as follows.

- For every DFA $A$, there is a DFA $A'$ such that $L(A') = \Sigma^* - L(A)$.
- For every two DFA $A_1$ and $A_2$, there is a DFA $A'$ such that $L(A') = L(A_1) \cap L(A_2)$.
- For every two DFA $A_1$ and $A_2$, there is a DFA $A'$ such that $L(A') = L(A_1) \cup L(A_2)$. 