4

Words and Finite Automata

In this chapter, we recall the notion of finite automata, which we consider to be a sequential model without any communication. Actions are processed in a linear manner and therefore arranged as words.

4.1 Words

Words can be represented in many different ways. For example, a word can be seen as a string, i.e., a sequence of symbols. Such a sequence gives rise to a totally ordered set, a special case of a partially ordered one, which, as we have seen, has a graph-theoretical counterpart. In the following, let Σ be an alphabet.

Definition 4.1 (Word). *A* word *over Σ is a structure $\underline{w} = (\{1,\ldots,n\}, \lhd, \lambda)$, $n \in \mathbb{N}$, where $1,\ldots,n$ are the* letter positions *of $\underline{w}$, $\lhd$ is the* successor relation *on $\{1,\ldots,n\}$, which contains the pairs $(i, i+1)$ with $i \in \{1,\ldots,n-1\}$, and λ is a mapping $\{1,\ldots,n\} \to \Sigma$.*

The set of words over Σ is denoted by $\mathbb{W}(\Sigma)$ or simply by $\mathbb{W}$ if Σ is clear from the context. Note that a word over Σ is just a graph over Σ and a singleton, which is, to some extent, extraneous. Actually, $\mathbb{W}(\Sigma)$ is simply $\mathbb{DAG}_H(\Sigma, -)$ restricted to graphs $(V, \lhd, \lambda)$ such that $\lhd^*$ forms a total order. Recall that we do not distinguish isomorphic structures. We can therefore identify a word $(\{1,\ldots,n\}, \lhd, \lambda) \in \mathbb{W}$ with the sequence $a_1 \ldots a_n \in \Sigma^*$ where $a_i = \lambda(i)$ for each $i \in \{1,\ldots,n\}$. For example, $(\{1,2,3\}, \{(1,2),(2,3)\}, \{1 \mapsto a, 2 \mapsto b, 3 \mapsto a\}) \in \mathbb{W}(\{a,b\})$ is equated with the string $aba \in \{a,b\}^*$. Recall that the empty word, which corresponds to the structure $(\emptyset, \emptyset, \emptyset)$, is denoted by ε. It appears as the unit word in the free monoid $\mathbb{W}$. Recognizability and rationality coincide in finitely generated free monoids, which is, as commonly known, Kleene's theorem.

Given a word $\underline{w} = (\{1,\ldots,n\}, \lhd, \lambda) \in \mathbb{W}$ with $n \geq 1$, we denote by *first*$(\underline{w})$ and *last*$(\underline{w})$ the events 1 and n, respectively.

Theorem 4.2 (Kleene [53]).

$$\mathcal{REC}_{\mathbb{W}} = \mathcal{RAT}_{\mathbb{W}}$$

Exercise 4.3. Let $\Sigma = \{a, b\}$. Provide rational expressions $\alpha_1, \alpha_2, \alpha_3$ of $\mathbb{W}(\Sigma)$ such that the following hold:

$L(\alpha_1) = \{\underline{w} \in \mathbb{W}(\Sigma) \mid \underline{w}$ does not end with $ab\}$,
$L(\alpha_2) = \{\underline{w} \in \mathbb{W}(\Sigma) \mid ab$ occurs in $\underline{w}$ exactly once$\}$,
$L(\alpha_3) = L(\alpha_1) \cup L(\alpha_2)$.

Note that a recognizable word language is also called *regular*. As $\mathbb{W}(\Sigma) \subseteq \mathbb{DG}(\Sigma, -)$, the monadic second-order formulas that can be applied to words over Σ are those from $\mathrm{MSO}(\Sigma, -)$. Recall that the corresponding atomic entities are

- $\lambda(x) = a$,
- $x \lhd y$,
- $x \in X$, and
- $x = y$.

The definition of their semantics arises from the general case of graphs (cf. Chap. 3).

Exercise 4.4. Describe each of the following word languages over $\Sigma = \{a, b\}$ by an $\mathrm{MSO}(\Sigma, -)$-sentence relative to $\mathbb{W}(\Sigma)$:

(a) $\{a\underline{w}a \mid \underline{w} \in \mathbb{W}(\Sigma)\}$,
(b) $\{\underline{w} \in \mathbb{W}(\Sigma) \mid |\underline{w}|_b > 3\}$,
(c) $\{\underline{w} \in \mathbb{W}(\Sigma) \mid |\underline{w}|$ is even$\}$.

4.2 Finite Automata

We now recall a well-known automata model, which is tailored to words.

Definition 4.5 (Finite Automaton). *A* finite automaton *over Σ is a structure (S, Δ, s^{in}, F) where*

- *S is its nonempty finite set of* states*,*
- *$\Delta \subseteq S \times \Sigma \times S$ is the set of* transitions*,*
- *$s^{in} \in S$ is the* initial state*, and*
- *$F \subseteq S$ is the set of final states.*

Let $\mathcal{A} = (S, \Delta, s^{in}, F)$ be a finite automaton over Σ. A *run* of $\mathcal{A}$ on a word $\underline{w} = (\{1, \ldots, n\}, \lhd, \lambda) \in \mathbb{W}(\Sigma)$ is a mapping $\rho : \{0, 1, \ldots, n\} \to S$ such that $\rho(0) = s^{in}$ and, for any $i \in \{1, \ldots, n\}$, $(\rho(i-1), \lambda(i), \rho(i)) \in \Delta$. We call ρ *accepting* if $\rho(n) \in F$. The *language* of $\mathcal{A}$, denoted by $L(\mathcal{A})$, is the set $\{\underline{w} \in \mathbb{W} \mid$ there is an accepting run of $\mathcal{A}$ on $\underline{w}\}$. Note that ε is contained in