

2.1 Deterministische endliche Automaten

Um die bisherige informale Beschreibung mathematisch zu präzisieren brauchen wir einige Begriffe und Notationen. Das sind neben den schon bekannten Begriffen wie Mengen, Funktionen und Relationen, die neuen Begriffe für formale Sprachen. Diese werden nachfolgend eingeführt. Im Anschluss wird der deterministische endliche Automat eingeführt.

2.1.1 Formale Sprachen

Eine formale Sprache ist eine Menge von Worten. Wir werden also den Begriff eines Wortes brauchen, wofür wir wiederum den Begriff eines Zeichens (als Baustein eines Wortes) brauchen, sowie den Begriff der Konkatenation, so dass einzelne Zeichen hintereinander geschrieben werden können.

Im einzelnen ist ein *Alphabet* eine (total geordnete) endliche Menge von unterschiedlichen *Zeichen* (alternativ: Buchstaben oder Symbole). Wir notieren diese Menge meistens mit Σ . Es ist dann z.B. $\Sigma = \{a, b, c\}$ für das Alphabet mit den Zeichen a , b und c .

Die *Konkatenation* wird mit \cdot notiert und ist die Operation zum Hintereinanderschreiben von Buchstaben. So ergeben sich Worte, auf die die Operation dann erweitert wird. Z.B. ist $a \cdot b = ab$ und dann $ab \cdot c = abc$. Diese Konkatenation wird auf Mengen von Wörtern erweitert, so dass wir z.B.

$$\{ab, aba\} \cdot \{ab, b\} = \{ab \cdot ab, ab \cdot b, aba \cdot ab, aba \cdot b\}$$

erhalten (was das gleiche ist wie $\{abab, abb, abaab\}$, da der \cdot oft weggelassen wird).

Das *leere Wort* (das Wort mit 0 Buchstaben) wird mit λ oder ϵ bezeichnet. Es ist nie in einem Alphabet enthalten, sondern ist das spezielle Wort das aus 0 Symbolen eines Alphabets besteht.

Mit $|w|$ wird die Länge eines Wortes w notiert. Z.B. ist $|001| = 3$, $|00111| = 5$ und $|\lambda| = 0$. Mit $|w|_x$ ist die Anzahl der x in w gemeint, wobei x ein Symbol ist, z.B. ist $|2202|_2 = 3$.

Als weitere Notation schreiben wir $w^1 = w$, $w^2 = w \cdot w$ und $w^k = w^{k-1} \cdot w$ ($k \geq 2$) und als Spezialfall $w^0 = \lambda$. Die hochgestellte Zahl i bei w^i gibt also an, wie oft wir ein Wort hintereinander schreiben wollen. Ähnlich wie oben erweitern wir diese Operation auf Mengen, so dass wir für eine Menge von Worten R dann $R^1 = R$, $R^2 = R \cdot R$ und allgemein $R^n = R^{n-1} \cdot R$ haben (Spezialfall: $R^0 = \{\lambda\}$). In R^i sind also alle Wörter enthalten, die man erhält, wenn man i Wörter aus R auswählt und diese hintereinander schreibt.

Zuletzt zwei besonders wichtige Definitionen, die auf obigem aufbauen. Wir setzen

$$\begin{aligned} R^+ &:= \cup_{i \geq 1} R^i \text{ mit } R^1 := R \text{ und } R^{i+1} = R^i \cdot R \\ R^* &:= R^+ \cup \{\lambda\} \end{aligned}$$

In R^+ sind also jene Wörter, die man erhält, wenn man beliebig oft (aber endlich oft!) Wörter aus R hintereinander schreibt und in R^* ist zusätzlich auf jeden Fall noch λ enthalten (λ kann auch in R^+ enthalten sein, aber nur wenn $\lambda \in R$ gilt).

Besonders wichtig ist für uns der Spezialfall Σ^* für ein Alphabet Σ . Dabei ist Σ^* dann die *Menge aller endlichen Wörter* (über dem Alphabet Σ). Jede (Teil-)Menge $L \subseteq \Sigma^*$ heißt *formale Sprache*.

Wir fassen die vielen Definitionen noch einmal zusammen.

Definition 2.1.1 (Zeichen, Worte, Sprachen).

1. Ein **Alphabet** ist eine (total geordnete) endliche Menge von unterschiedlichen **Zeichen** (alternativ: Buchstaben oder Symbole).
2. Die **Konkatenation** \cdot ist die Operation zum Hintereinanderschreiben von Buchstaben. So ergeben sich Worte auf die die Operation dann erweitert wird. (Z.B. ist $a \cdot b = ab$ und dann $ab \cdot c = abc$.) Die Konkatenation wird auf Mengen von Wörtern erweitert, wobei dann $R \cdot S = \{u \cdot v \mid u \in R, v \in S\}$ gilt.
3. Das **leere Wort** (Wort mit 0 Buchstaben) wird mit λ oder ϵ bezeichnet (das leere Wort ist nie in einem Alphabet enthalten!).
4. Sei Σ ein Alphabet, $x \in \Sigma$ ein Symbol und w ein Wort bestehend aus Symbolen aus Σ . Mit $|w|$ wird die **Länge** von w notiert, also z.B. $|001| = 3$ und $|00111| = 5$. Außerdem ist $|\lambda| = 0$. Ferner wird mit $|w|_x$ die **Anzahl** der x in w notiert. Z.B. ist mit $\Sigma = \{0, 1, 2\}$ dann $|2202|_2 = 3$, $|2202|_1 = 0$ und $|2202|_0 = 1$.
5. Wir schreiben $w^1 = w$, $w^2 = w \cdot w$ und $w^k = w^{k-1} \cdot w$ ($k \geq 2$) und als Spezialfall $w^0 = \lambda$. Für eine Menge von Worten R schreiben wir ebenso $R^1 = R$, $R^2 = R \cdot R$ und $R^k = R^{k-1} \cdot R$ und als Spezialfall $R^0 = \{\lambda\}$. Ferner setzen wir

$$\begin{aligned} R^+ &:= \cup_{i \geq 1} R^i \text{ mit } R^1 := R \text{ und } R^{i+1} = R^i \cdot R \\ R^* &:= R^+ \cup \{\lambda\} \end{aligned}$$

6. Sei Σ ein Alphabet, dann bezeichnen wir mit Σ^* die **Menge aller endlichen Wörter** (über dem Alphabet Σ).
7. Jede (Teil-)Menge $L \subseteq \Sigma^*$ heißt **formale Sprache**.

Fragen

1. Was ergibt sich bei $\{a, bc\} \cdot \{de, fg\}$?
 - a) $\{abc, defg\}$
 - b) $\{abcde, abcfg\}$
 - c) $\{abcde, abcfg, bcade, bcafg\}$
 - d) $\{ade, afg, bcde, bcfg\}$

2. Was ergibt sich bei $\{abc, a\} \cdot \{bc, \lambda\}$?
 - a) $\{abca, bc\}$
 - b) $\{abcbc, abc\}$
 - c) $\{abcabc\}$
 - d) $\{abcbc, abc, a\}$

3. Sei $R = \{a, b\}$ was ist R^2 ?
 - a) $\{\lambda, a, b, aa, ab, ba, bb\}$
 - b) $\{a, b, aa, ab, ba, bb\}$
 - c) $\{aa, ab, ba, bb\}$
 - d) $\{\{aa\}, \{ab\}, \{ba\}, \{bb\}\}$

4. Sei $R = \{ab, ba, a, b\}$ was ist R^* ?
 - a) Menge aller endlichen Worte aus a und b
 - b) Menge aller *ungeraden* Worte aus a und b
 - c) Menge aller *geraden* Worte aus a und b
 - d) Menge aller unendlichen Worte aus a und b