Durch Homom. erkennbare Mengen (5)

Korollar: Zu jeder Sprache $L \in REG(X^*)$ gibt es ein endliches Monoid M und einen Homomorphismus $h : X^* \to M$, so daß für eine geeignete Teilmenge $U \subset M$

$$L = h^{-1}(U)$$

gilt.

Beweis: Wähle $M = synt(L)$ und $U = \{[u] \mid u \in L\}$.

Die Umkehrung dieses Korollars haben wir bereits gezeigt. Somit gilt der

Satz: $L \in REG(X^*) \iff$ es gibt ein endliches Monoid M, eine Teilmenge $U \subset M$ und einen Monoidhomomorphismus $h : X^* \to M$, mit $h^{-1}(U) = L$.

Korollar: Ist $h : Y^* \to X^*$ ein Homomorphismus, dann gilt $h^{-1}(L) \in REG(Y^*)$. für $L \in REG(X^*)$.

Durch Homom. erkennbare Mengen (6)

Diese Charakterisierung der regulären Mengen gibt Anlaß zu der folgenden Definition.

Definition: Ist M ein beliebiges Monoid und $L \subset M$, dann heißt L *recognizable* (durch einen Homomorphismus erkennbar) : $\Longleftrightarrow$ Es gibt einen Homomorphismus $h : M \to M'$ in ein endliches Monoid M' und eine Teilmenge $U \subset M'$ mit $h^{-1}(U) = L$.

Notation: $L \in REC(M)$.

Wir haben gezeigt, daß der

Satz: $REG(X^*) = REC(X^*)$ gilt.

Rationale Mengen über X^* (1)

Definition: Ist M ein Monoid, dann heißt $R \subset 2^M$ **rational abgeschlossen** $: \iff$ R ist abgeschlossen unter der Vereinigung $\bigcup$ von Mengen, dem Komplexprodukt $\cdot$ und dem $*-$Operator.

$RAT(M)$ ist die kleinste, rational abgeschlossene Teilmenge von 2^M, die die Mengen $\{m\}, m \in M$ und $\emptyset$ enthält.

Wir haben bereits gezeigt, daß

$$RAT(X^*) \subset REG(X^*)$$

gilt. Wir wollen nun auch die umgekehrte Relation zeigen.

Satz (Kleene): $REG(X^*) = RAT(X^*)$.

Rationale Mengen über X^* (2)

Beweis: Wir haben

$$L \in REG(X^*) \Rightarrow L \in RAT(X^*)$$

zu zeigen.

$L \in REG(X^*) \Rightarrow$ Es gibt einen endlichen Akzeptor a mit $L = L_a$.

Man zeigt, daß L aus $\emptyset, \{\{x\} \mid x \in X\}$ mittels $\bigcup, \cdot, *$ erzeugt werden kann.

Der Beweis erfolgt durch Induktion nach der Anzahl k der Kanten bei fester Knotenzahl. O.B.d.A. $\#S = 1$, $S \cap T = \emptyset$.

(1) $k = 0 \Rightarrow L_a = \emptyset$, d.h. $L_a \in RAT(X^*)$.

(2) Induktionsschritt: Sei die Behauptung für Akzeptoren mit k Kanten bewiesen, und sei a_k ein solcher Akzeptor. G_k sei der Graph von a_k. Wir fügen zwischen zwei Knoten P_1, P_2 von G_k eine neue Kante s ein, mit $Q(s) = P_1$, $Z(s) = P_2$, $\mu(s) = x$. Der neue Akzeptor sei a_{k+1}.

Rationale Mengen über X^* (3)

Wir haben dann

$$\mathcal{W}(G_{k+1})(S,T) = \mathcal{W}(G_k)(S,T) \cup$$
$$\mathcal{W}(G_k)(S,P_1)\cdot(s\cdot\mathcal{W}(G_k)(P_2,P_1))^*\cdot s\cdot\mathcal{W}(G_k)(P_2,T)$$

$$\begin{tikzcd}
 & P_1 \arrow[r, "s"'] & P_2 \arrow[l, bend right, "\mathcal{W}(G_k)(P_2,P_1)"'] \arrow[dr, bend left, "\mathcal{W}(G_k)(P_2,T)"] & \\
S \arrow[ur, bend left, "\mathcal{W}(G_k)(S,P_1)"] \arrow[rrr, "\mathcal{W}(G_k)(S,T)"'] & & & T
\end{tikzcd}$$

Nach Induktionsannahme gilt

$$\begin{array}{rcl}
\mu(\mathcal{W}(G_k)(S,T)) & \in & RAT(X^*), \\
\mu(\mathcal{W}(G_k)(S,P_1)) & \in & RAT(X^*), \\
\mu(\mathcal{W}(G_k)(P_2,T)) & \in & RAT(X^*), \\
\mu(\mathcal{W}(G_k)(P_2,P_1)) & \in & RAT(X^*), \\
\mu(s) & \in & RAT(X^*).
\end{array}$$

Da $RAT(X^*)$ abgeschlossen ist unter $\cup, \cdot, *$, folgt die Behauptung.

Rechtslineare Sprachen $r - LIN(X^*)$ (1)

N und X sind Alphabete, $A := N \cup X$. Wir setzen $N \cap X = \emptyset$ voraus. Im folgenden heißen endliche Teilmengen $\mathfrak{P} \subset N^* \times A^*$ Produktionssysteme. Weiter ist $S \in N$, und $G = (N, X, \mathfrak{P}, S)$ bezeichnen wir als Chomsky-Grammatik. N heißt Variablen- und X Terminalalphabet.

Ist $(p, q) \in \mathfrak{P}$, dann schreiben wir auch

$$p \xrightarrow{\mathfrak{P}} q \quad \text{oder kurz} \quad p \longrightarrow q.$$

Wir setzen die Relation $\mathfrak{P}$ wie folgt zu einer Relation $\mathfrak{P}^*$ fort. $\mathfrak{P}^*$ ist minimal mit

(1) $w \xrightarrow{\mathfrak{P}^*} w$, für $w \in A^*$,

(2) $u \xrightarrow{\mathfrak{P}^*} v \Rightarrow wuw' \xrightarrow{\mathfrak{P}^*} wvw'$,

(3) $w_1 \xrightarrow{\mathfrak{P}^*} w_2,\ w_2 \xrightarrow{\mathfrak{P}^*} w_3 \Rightarrow w_1 \xrightarrow{\mathfrak{P}^*} w_3$.

Anstelle von $w \overset{\mathfrak{P}^*}{\to} w'$ schreiben wir oft $w \overset{*}{\to} w'$.

Rechtslineare Sprachen $r - LIN(X^*)$ (2)

Im Falle

$$w = upv \xrightarrow{*} uqv = w'$$

mit $(p, q) \in \mathfrak{P}$ schreiben wir auch $\longrightarrow$, um auszudrücken, daß w durch eine **einmalige** Anwendung von $\mathfrak{P}$ in w' übergeführt werden kann.

Eine Folge

$$w = w_1 \longrightarrow w_2 \longrightarrow \ldots \longrightarrow w_k = w'$$

heißt **Ableitung** von w' aus w. $k-1$ heißt die **Länge** der Ableitung.

$$L(G) := \{w \in X^* \mid S \xrightarrow{*} w\}$$

heißt „die von G erzeugte Sprache“.

Notiz: Chomsky versuchte, Grammatiken natürlicher Sprachen mittels „Chomsky“-Grammatiken zu modellieren.

Hier betrachten wir zunächst nur den Sonderfall

$$\mathfrak{P} \subset N \times N \cdot X \cup N \times (X \cup \{1\}).$$

G heißt in diesem Fall **rechtslinear**.

Rechtslineare Sprachen $r - LIN(X^*)$ (3)

Definition: $L \subset X^*$ heißt rechtslineare Sprache $:\Longleftrightarrow$ Es gibt eine rechtslineare Chomsky-Grammatik G mit $L = L(G.)$ $r - LIN(X^*) := \{L \subset X^* | L$ ist rechtslineare Sprache$\}$.

Satz: $r - LIN(X^*) = REG(X^*)$.

Beweis: (1) $r - LIN(X^*) \subset REG(X^*)$.

Setze:

$$\begin{array}{rcl} K & := & N \cup \{t\}, t \notin N, \\ E & := & \{(n, n', x) \mid (n, x \cdot n') \in \mathfrak{P}\} \\ & \cup & \{(n, t, x) \mid (n, x) \in \mathfrak{P}\}. \end{array}$$

$$Q(n, t, x) := n, Z(n, t, x) := t, \mu(n, t, x) := x$$

$$Q(n, n', x) := n, Z(n, n', x) := n', \mu(n, n', x) := x.$$

$$G := (E, K, Q, Z),$$

$$a := (\mathcal{W}(G), S, t, \mu, X).$$

Rechtslineare Sprachen $r - LIN(X^*)$ (4)

Man zeigt nun $L_a = L(G)$.

(2) $REG(X^*) \subset r - LIN(X^*)$.

Sei $L \in REG(X^*)$,

Wir konstruieren zu a eine rechtslineare Grammatik: O.B.d.A. nehmen wir an, daß a vollständig ist.

Wir setzen

$$\begin{array}{lll} N & := & K, \\ \mathfrak{P} & := & \{(z, x \cdot z') \in N \times N \cdot X \mid \text{ Es gibt } s \in E \\ & & \text{mit } Q(s) = z, Z(s) = z', \mu(s) = x\} \\ & \cup & \{(z, x) \in N \times X \mid \text{ Es gibt } s \in E \\ & & \text{mit } Q(s) = z, Z(s) \in T, \mu(s) = x\} \\ & \cup & \{(s, 1) \mid \text{ für } 1 \in L_a\}. \end{array}$$

Für die Grammatik

$$G := (N, X, \mathfrak{P}, S)$$

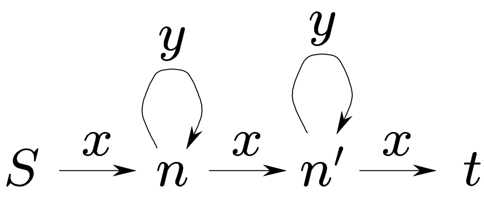
gilt

$$L(G) = L_a.$$

Rechtslineare Sprachen $r - LIN(X^*)$ (5)

Beispiel 1: $S \longrightarrow x \cdot n$, $n \longrightarrow y \cdot n$, $n \longrightarrow x \cdot n'$, $n' \longrightarrow y \cdot n'$, $n' \longrightarrow x$

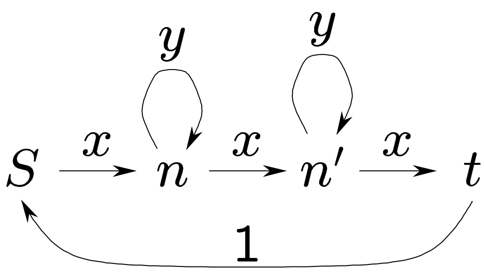
Akzeptor:



Die Sprache: $L = xy^*xy^*x$.

Beispiel 2: Ergänze die obige Grammatik durch die Produktion $t \longrightarrow S$.

Zugehöriger Akzeptor:



Die akzeptierte Sprache: $L = (xy^*xy^*x)$