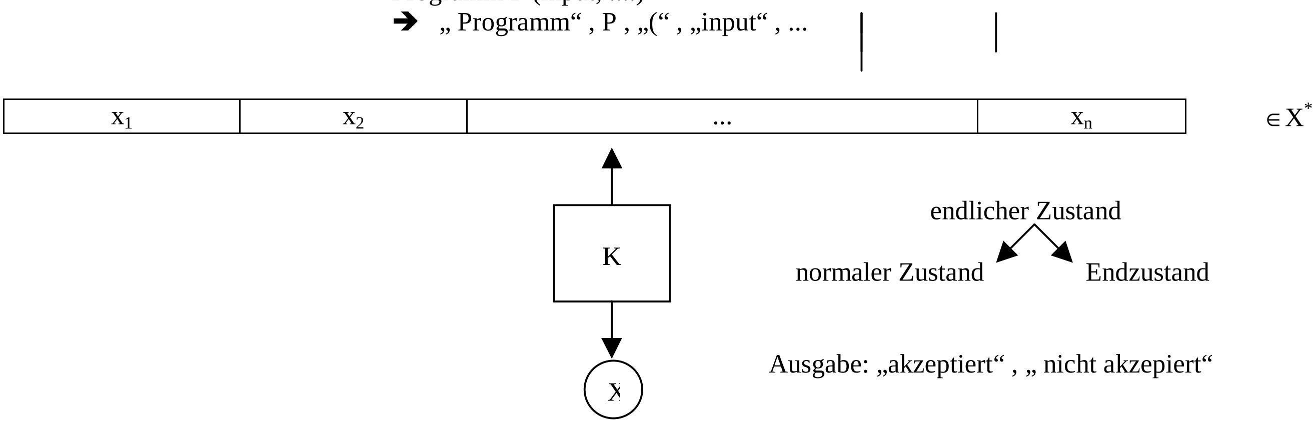
# 4. Endlich Automaten und reguläre Sprachen
## 4.1. Deterministische Akzeptoren

inform. Anwendungen: - Teiltasterkennung
- Übersetzerbau
  - -lexikalische Analyse/Tokenerkennnung
  - Programm P (input, ....)
  - ➔ „ Programm“ , P , „(“ , „input“ , ...




Ausgabe: „akzeptiert“ , „ nicht akzepiert“

**Def. A:**

1) Ein endlich deterministischer Akzeptor A beschreibt man durch ein 5- Tupel $A = (X , S , \delta , s_0 , F)$ mit :
   - X und S unendlich nichtleere Mengen
   - X = Alphabet
   - S = Zustandsmenge
   - $\delta : S \times X \rightarrow S$ ist ein Funktion
   - $\delta$ = Zustandsüberführungsfunktion
   - $s_0 \in S$ ; $s_0$ = Anfangszustandsfuntion
   - $F \subseteq S$; F = Endzustand

2) A lässt sich als abstrakte Maschine beschreiben $M = (I , O , \alpha , \omega , \tau , \pi)$ mit
   - $I = X^*$
   - $O = \emptyset$
   - $K = I \times S \times /N$
   - $\alpha : I \rightarrow K$ mit $\alpha(\omega) = (\omega , s_0 , 1)$
   - $\omega$ = undefiniert
   - $\pi(\omega , s , m) = 1 \Leftrightarrow m = |\omega| +1 \wedge s\in F$
   - $\tau (\omega , s , k) = (\omega , \delta (s , x_k) , k+1)$ für $\omega = x_1 ... x_n$

   Idee: Verbesserung der Schreibweise: Setze $\delta$ fest auf Wörter $\delta^*$

**Def. B:**

Sei $A = (X , S , \delta , s_0 , F)$ ein endlich deterministischer Akzeptor.
Wir erweitern $\delta$ wie folgt zu einer Überführungfunktion $\delta^*$ ( auf Wörter).
$\delta^* : S \times X \text{ -> } S$ mit $\forall a\in X, \omega\in X^*, s\in S$ : i ) $\delta^* (s , \varepsilon) = S$
ii ) $\delta^* (s , a\omega) = \delta^* (\delta(s , a),\omega)$
Bemerkung:
- statt $\delta^*$ schreiben wir oft $\delta$
- $\delta^* (s , a\omega) = \delta (\delta^* (s , a),\omega)$

**Def. C:**

Sei $A = (X , S , \delta , s_0 , F)$ ein endlich deterministischer Akzeptor.
Die Menge $L (A) = \{\omega\in X^* | \delta (s_0 , \omega) \in F\}$ heißt die von A akzeptierte (erkannte) Sprache.
Eine Sprache heißt endlich akzeptierbar, wenn ein endlichen deterministischen Akzeptor A gibt mit L (A).

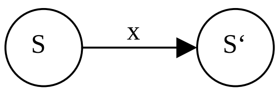
DA ist die Kasse aller Sprachen, die von endlichen deterministischen Akzeptoren akzeptiert werden.

<u>Automatendarstellung:</u> gerichtet markierte Gaphen: $A = (X , S , \delta , s_0 , F)$

$\Rightarrow$ $G = (V , E)$

$V = S$

$E \subseteq S \times X \times S$ mit $e = (s , x , s') \in E$ genau dann wenn $\delta (s , x) = s'$

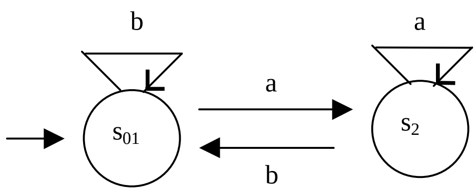




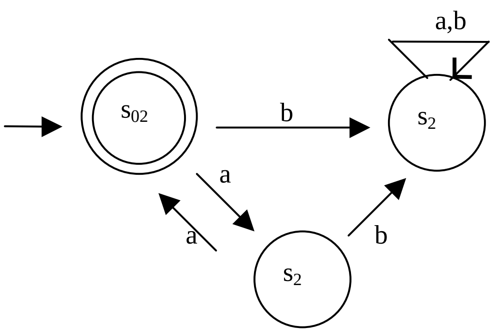
<u>Bsp.:</u>

1) $X = \{a , b\}$, $L = \{\omega a \mid \omega \in X^*\}$ L wird von folgendem Akzeptor $A_1$ erkannt.
$A_1 = (X_1 , S_1 , \delta_1 , s_{01} , F_1)$ ,
$S_1 = \{ s_{01} , s_2 \}$,
$F_1 = \{ s_2\}$

| $\delta$ | a | b |
|---|---|---|
| $s_{01}$ | $s_2$ | $s_{01}$ |
| $s_2$ | $s_2$ | $s_{01}$ |

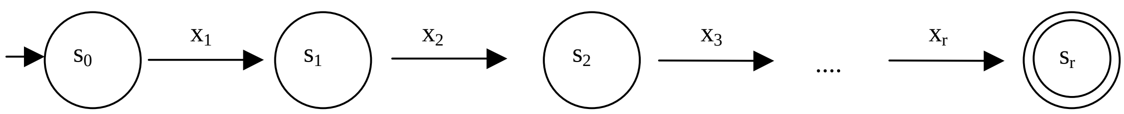



2) $L = \{a^{2n} \mid n \in /N\}$ folgender Akzeptor : $A_1 = (X_2 , S_2 , \delta_2 , s_{02} , F_2)$ ,
mit $X_1 = \{a , b\}$,
$S_2 = \{ s_{02} ,\ s_{12} ,\ s_{22}\}$,
$F_3 = \{ s_{02}\}$



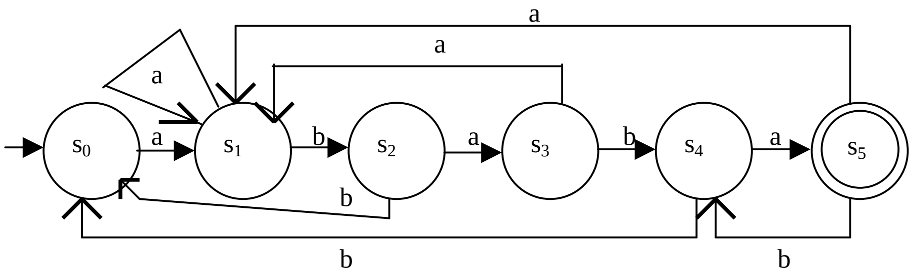

„absorbierender Zustand“

3) X ist ein beliebiges Alphabet, $v = x_1 \ \dots x_r \in X^*$ mit $r \geq 1$, $x_i \in X$ für $i = 1, \dots , r$
Die Sprache $L_v = \{ \omega v \mid \omega \in X^*\}$
Der zugehörige Akzeptor hat $r \neq 1$ Zustände.
Teilstruktur:




Rückwärtssprünge ergänzen, aber nicht alle zurück auf $s_0$ .

Bsp.: $v$ = ababa $\in \{a,b\}^*$




Konstruiere $A_v = (X , S , \delta , s_0 , F)$ für $v = x_1 \; ... \; x_r \in X^*$ mit $r \geq 1$ , $s_0 = \{ s_0 , ... , s_r \}$ , $F = \{s,r\}$ und für alle Fälle $s_i \in S$ $x \in X$ : $\delta( s_i , x) = s_j$ , wobei j der größte Index aus $\{0 , 1 , ... , r\}$ ist , so dass ein $u \in X^*$ gibt mit $x_1 ... x_j x = u \, x_1 ... x_j$ Bsp: $\underline{x_1 \; x_2} \; x_3 ... x_i x$

$\Downarrow$

$u \quad x_1 ... x_j$

(Gehe zu den größten zuständen zurück , wo das längste Restwort , zu v , ergänzt werden kann.)

Zeige : $L_v = L\,(A_v)$

Beweise unter Einbeziehung einer Interpretation der Zustände $s_i$ .

<u>Beh. 1:</u> Es sei $\omega \in X^*$ . Dann gilt für alle $i = 0 , 1 , ... , r_i$ $\delta(s_0 , \omega) = s_i \Rightarrow \omega = \omega' \, x_1 ... x_i$ für ein $\omega' \in X^*$ , und alle $m > i$ und $\omega^n \in X^*$ gilt $\omega \neq \omega' \, x_1 ... x_m$

{Erreicht man $s_i$ , so sind die letzten i Symbole $x_1 ... x_j$ Anfangsstücke von $\omega$ aber nicht mehr.}

<u>Bew. 1:</u> Induktion über die Länge von $\omega$

<u>Induktionsvoraussetzung:</u>

$|\omega| = 0$ , $\omega = \varepsilon$

Aussage für i = 0 trivialerweise richtig.

<u>Induktionsannahme:</u>

Beh. 1 sei für alle Wörter bis zu einer festen Länge k bewiesen.

<u>Induktionsbehauptung:</u>

Beh. 1 gilt auch für Wörter der Länge k + 1.

Betrachte dazu $\omega \in X^*$ mit $|\omega| = k$ und $x \in X$ (k +1-tes Zeichen)

Dann gilt zu Bemerkung zu Def. B

$\delta( s_0 , \omega x) = \delta(\underset{s_i}{\underline{\delta( s_0 , \omega)}} , x) = s_j \Rightarrow \delta( s_i , x) = s_j \Rightarrow \omega = \omega' \, x_1 ... x_i$ für ein $\omega' \in X^*$ und $\omega \neq \omega' \, x_1 \ldots x_m$ für alle $\omega^n \in X^*$ und $n > i$

$\Rightarrow$ $\omega = \omega' \, x_1 ... x_i$ für ein $\omega' \in X^*$ und $\omega \neq \omega' \, x_1 ... x_i$

Def. von $\delta$ für alle $\omega^n \in X^*$ und $m > i$

J ist der größte Index, zu dem s ein $u \in X^*$ mit $x_1 ... x_j x = u \, x_1 ... x_j$ gibt.

<u>Beh. 2:</u> Vorausgesetzt sei $\omega = \omega' \, x_1 ... x_i$ für ein $\omega' \in X^*$ und $\omega \neq \omega^n \, x_1 \ldots x_i$ für alle $\omega^n \in X^*$ und $n > i$

J ist größter Index, zudem es ein $u \in X^*$ gibt mit $x_1 ... x_j x = u \, x_1 ... x_j$

Dann gilt $\omega x = \omega' \, x_1 ... x_i x = \omega' u \, x_1 ... x_j = s' x_1 ... x_j = v' x_1 ... x_j$ mit $v' = \omega' u$

Zu Zeigen ,

<u>Induktionsbehauptung</u>: $\omega x = v' x_1 ... x_j$ (ist ein $\omega' \in X^*$ und $\omega x \neq v^n x_1 ... x_n$ für alle $v^n \in X^*$ und alle $m > j$)

Gegenteil von Beh. 2 (Widerspruchsbeweis):

Zusätzlich gelte $\omega x = v^n x_1 ... x_m$ für ein $m \geq j$ und ein $v^n \in X^*$ .

Dann muss $i \geq m - 1$ sein , da sonst die Voraussetzung verletzt wäre. Folglich gibt es $\omega'$ , $u' \in X$ mit $\omega = v^n x_1 ... x_{m-1} = \omega' x_1 ... x_n = \omega' u' x_1 ... x_{m-1}$

Wir erhalten $x_1 ... x_i = u' x_1 ... x_{m-1} \, x_m \Rightarrow$ Da y maximal , mit dieser Eigenschaft, ist , folgt $j \geq m$

$\Rightarrow$ Beh. 2 bewiesen $\Rightarrow$ Beh. 1 bewiesen

<u>Beh. 3:</u> $L\,(A_v) = L_v$

<u>Beweis:</u> „$\subseteq$" : Sei $\omega \in L\,(A_v)$ , d.h. $\delta(s_0 , \omega) = s_r$.

Nach Beh. 1 kann man $\omega$ als $\omega = \omega' x_1 ... x_r = \omega' v \Rightarrow L_v$

„$\geq$" : Sei $\omega \in L_v$ , d.h. $\omega = \omega' x_1 ... x_r$ . Sei $\delta(s_0 , \omega) = s_i$

Wäre $i < r$ , dann würde wegen Beh. 1 gelten :

$\omega \neq \omega^n x_1 ... x_r$ für alle $\omega^n \in X^*$

zu $\omega = \omega' x_1 ... x_r \Rightarrow i = r$ folglich $\omega \in L(A_v)$

Anwendung : Teilwortbestimmung/substring Test
Problem : Ermittle ob v Teilwort eines Wortes $\omega$ ist , ob Wort $\omega'$ , $\omega^n$ existiert , so dass $\omega = \omega' \vee \omega^n$
Lösung naiv: Teste subzessiv , ob $\omega = v$ , $\omega = xv$ , $\omega = xxv$ , ...
Laufzeit: $0(|v| * |\omega|)$

Automatenlösung: Zu v konstruiere $\delta$- Tabelle , die aus |v| +1 Zeilen , |x| Spalten: Laufzeit: $0(|v| * |x|)$ tatsächlich länger. Schnell zu bestimmen ist $\delta(s , x)$ nur im Fall x = x +1 ($s_{i+1}$) sonst ist max. langes Anfangswort von v zu finden , das Endwort von $x_1 \ldots x_i x$ ist. (Probiere: $0(i^2)$)

Summe: $|X| \sum_{i=0}^{|v|} 0(i^2) = 0(|X| * |v|^3)$nur

Konstruktion der $\delta$- Tabelle

$A_v$ arbeitet in $0(|\omega|)$ Summe: $0 (|\omega| + |X| * |v|^3)$
(kann man $\delta$ schneller bestimmen) Besser für häufiges Enden eines festen Wortes v.