Formale Sprachen und Automaten

Robin Rausch, Florian Maslowski

11. November 2022

Inhaltsverzeichnis

1	Gru	ndlagen 1
		Alphabet
		Wort
	1.3	Formale Sprachen
	1.4	Kleene Stern
	1.5	Überblick
2	Reg	uläre Sprachen und endliche Ausdrücke
	2.1	Reguläre Ausdrücke
	2.2	Endliche Automaten
		2.2.1 Komplement
		2.2.2 Deterministische endliche Automaten(DEA)
		2.2.3 Nicht-deterministische endliche Automaten(NEA)
		2.2.4 Endliche Automaten und reguläre Ausdrücke
		2.2.5 Transformation DEA zu RA
		2.2.6 RA zu NEA
		2.2.7 NEA zu DEA
		2.2.8 DEA zu RLG
		2.2.9 RLG zu NEA
		2.2.10 Minimierung
	0.0	2.2.11 Produktautomaten
		Nicht-reguläre Sprachen und das Pumping-Lemma
	2.4	Eigenschaften regulärer Sprachen
3		msky Grammatiken und kontextfreie Sprachen 12
	3.1	Typ0 unbeschränkt
	3.2	Typ1 Monoton
	3.3	Typ2 Kontextfreie
	3.4	Typ3 rechtsregulär/-linear
	3.5	Cocke-Younger-Kasami(CYK)-Algorythmus
	3.6	Chomskynormalform CNF
4	Kell	erautomaten 14
5	Turi	ng Maschine
		K-Band zu 1-Band
	5.2	Linear beschränkter Automat



	Duale Hochschule Baden-Württemberg	Formale Sprachen und Automaten					LaTeX Version				
6	6.2 Unentsche6.3 Semi-ents	eit par eidbar cheidbar KorrespondenzProblem (PKP)						. 14 . 14			
7	Berechenbark	eit						15			
8	Komplexität							15			



1 Grundlagen

1.1 Alphabet

Ein Alphabet Σ ist eine nicht-leere Menge von Symbolen(Zeichen, Buchstaben). Beispiel: $\Sigma_{ab}=a,b$

1.2 Wort

Ein Wort w über dem Alphabet Σ (Sigma) ist eine endliche Folge von Symbolen aus Σ . Das Wort w = abaabab wurde beispielsweise aus dem Alphabet Σ_{ab} gebildet.

Die Länge eines Wortes kann durch Betragsstriche angegeben werden. Beispiel: |w|=7 Ebenso kann man die Anzahl bestimmter Symbole in einem Wort bestimmen: $|w|_b=3$ Ein einzelnes Zeichen kann durch eckige Klammern angegeben werden: w[2]=b

Wörter können bliebig konkateniert werden(hintereinanderschreiben ohne abstand): $w_1w_2 = abbabaab$ mit $w_1 = abba$ und $w_2 = baab$.

Wörter dürfen auch potenziert werden: $w^3 = abaabababababababababab = www$ Das leere Wort lautet ε .

1.3 Formale Sprachen

Eine formale Sprache L über einem Alphabet Σ ist eine Menge von Wörtern aus $\Sigma^* : L \subseteq \Sigma^*$. Eine Sprache kann sowohl endlich als auch unendlich sein.

Beispiel: $L_1 = \{w \in \Sigma_{bin}^* | |w| \ge 2 \land w[|w|-1] = 1\}$ ist die Menge aller Binärwörter, an deren vorletzter Stelle 1 steht.

Das Produkt zweier formaler Sprachen: $L_1 \cdot L_2 = \{abac, abcb, bcac, bccb\}$ mit $L_1 = \{ab, bc\}$ und $L_2 = \{ac, cb\}$.

Sprachen können ebenfalls potenziert werden: $L^2 = \{ab, ba\} \cdot \{ab, ba\} = \{abab, abba, baab, baba\}$

1.4 Kleene Stern

Für ein Alphabet Σ und eine formale Sprache $L\subseteq \Sigma^*$ ist der Operator Kleene Stern wie folgt definiert: $L^*=\bigcup_{n \in \mathbb{N}} L^n$.

Beispiel: Sei $L_1 = \{ab, ba\}$, dann $L^* = \{\varepsilon, ab, ba, abab, abba, baba, baba, ababab, ...\}$.

1.5 Überblick

Тур	Sprachklasse	Grammatik	Maschinenmodell
0	rekursiv aufzählbar	unbeschränkt	Turing-Maschine
1	kontextsensitiv	monoton	linear beschränkter Automat
2	kontextfrei	kontextfrei	Kellerautomat
3	regulär	rechtslinear	endlicher Automat



2 Reguläre Sprachen und endliche Ausdrücke

2.1 Reguläre Ausdrücke

Ein regulärer Ausdruck über \varSigma beschreibt eine formale Sprache. Die Menge aller regulären Ausdrücke über \varSigma ist eine formale Sprache.

Beispiel: Sprache aller Wörter über Σ_{abc} , die nur aus genau zwei Symbolen bestehen:

Ausdruck: $r_1 = (a+b+c)(a+b+c)$ Sprache: $\mathcal{L}(r_1) = \{w \in \Sigma^*_{abc} \mid |w| = 2\}$



Operatoren:

$$r_1+r_2\equiv r_2+r_1$$
 Kommutativität von $+$ $(r_1+r_2)+r_3\equiv r_1+(r_2+r_3)$ Assoziativität von $+$ $(r_1r_2)r_3\equiv r_1(r_2r_3)$ Assoziativität von \cdot $\emptyset r\equiv r\emptyset\equiv\emptyset$ Absorbierendes Element für \cdot $\varepsilon r\equiv r\varepsilon\equiv r$ Neutrales Element für \cdot Neutrales Element für $+$ $(r_1+r_2)r_3\equiv r_1r_3+r_2r_3$ Distributivität links $r_1(r_2+r_3)\equiv r_1r_2+r_1r_3$ Distributivität rechts $r+r\equiv r$ Idempotenz von $+$ $(r^*)^*\equiv r^*$ Idempotenz von $*$ $\emptyset^*\equiv \varepsilon$ $\varepsilon^*\equiv \varepsilon$

Nicht alle Operatoren sind für alle Typen zulässig:

	Vereinigung	Konkatenation	Potenz	Kleene-Stern
Wörter	X	$w_1 \cdot w_2$	w^n	X
Sprachen	$L_1 \cup L_2$	$L_1 \cdot L_2$	L^n	L *
Reguläre Ausdrücke	$r_1 + r_2$	$m{r}_1\cdotm{r}_2$	X	r^*

2.2 Endliche Automaten

Endliche Automaten sind eine andere Darstellung einer regulären Sprache. Endliche Ausdrücke lassen sich in Reguläre Ausdrücke umformen. Genauso auch anders herum.

Endliche Automaten erkennen regulären Sprachen. Endliche Ausdrücke lassen sich in Reguläre Ausdrücke umformen. Genauso auch anders herum.

Endliche Automaten lassen sich sowohl deterministisch als auch nicht-deterministisch darstellen.

2.2.1 Komplement

Wird durch vertauschen von End- und Nichtendzuständen erzeugt.

2.2.2 Deterministische endliche Automaten(DEA)

Ein DEA hat endlich viele Zustände. Jeder mögliche Übergang muss hierbei behandelt werden können. D.h. für das Alphabet Σ_{ab} muss von jedem Zustand sowohl ein a, als auch ein



b Übergang gegeben sein. Der Automat beginnt im Startzustand und muss im Endzustand enden. Wenn der Automat sich in einem Nicht-Endzustand befindet, befindet sich das Wort nicht in der Sprache, welche vom Automaten abgebildet wird.

Ein DEA hat endlich viele Zustände. Jeder mögliche Übergang muss hierbei behandelt werden können. D.h. für das Alphabet Σ_{ab} muss von jedem Zustand sowohl ein a, als auch ein b Übergang gegeben sein. Er terminiert wenn das Wort zu ende und Endzustand erreicht ist. Der DEA lässt sich durch folgendes 5-Tupel darstellen:

 $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ mit den Komponenten:

Q ist eine endliche Menge von Zuständen

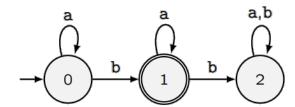
 Σ ist ein endliches Alphabet

 $\delta: Q \times \Sigma \to Q$ ist die Übergangsfunktion

 $q_0 \in Q$ ist der Startzustand

 $F \subseteq Q$ ist die Menge der Endzustände

Beispiel:



 $\mathcal{A}_{b} = (Q, \Sigma, \delta, q_{0}, F)$ mit

- $\mathbb{Q} = \{0, 1, 2\}$
- $\Sigma = \Sigma_{ab}$
- $\delta(0, a) = 0$; $\delta(0, b) = 1$, $\delta(1, a) = 1$; $\delta(1, b) = \delta(2, a) = \delta(2, b) = 2$
- $q_0 = 0$
- $F = \{1\}$

Zustand 2: "Mülleimerzustand" (junk state), d.h. kein Wort wird mehr akzeptiert

2.2.3 Nicht-deterministische endliche Automaten(NEA)

Ein NEA hat endlich viele Zustände. Nicht jeder mögliche Übergang muss hierbei behandelt werden. D.h. für das Alphabet Σ_{ab} reicht es, nur den a-Übergang, bzw. nur den b-Übergang zu besitzen. Der Automat beginnt im Startzustand und muss im Endzustand enden. Wenn der Automat sich in einem Nicht-Endzustand befindet, befindet sich das Wort nicht in der Sprache, welche vom Automaten abgebildet wird. Zudem gibt es ε -Übergänge, diese können jederzeit verwendet werden ohne ein Eingabesymbol.

Der NEA lässt sich durch folgendes 5-Tupel darstellen:

 $\mathcal{A} = (Q, \Sigma, \Delta, q_0, F)$ mit den Komponenten:

Q ist eine endliche Menge von Zuständen

 Σ ist ein endliches Alphabet

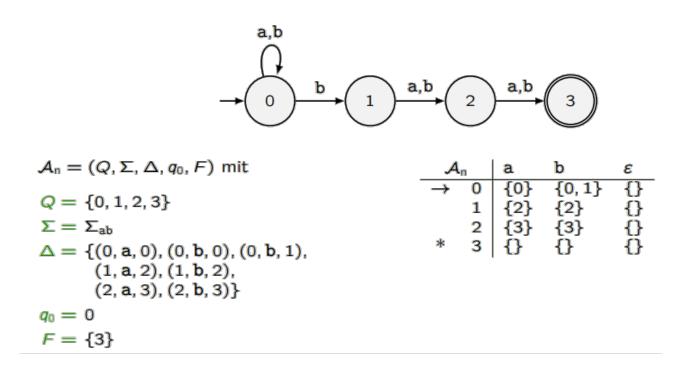
 Δ ist eine Relation über $Q \times (\Sigma \cup \{\varepsilon\}) \times Q$

 $q_0 \in Q$ ist der Startzustand



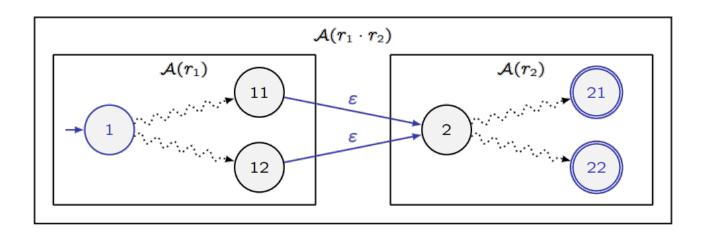
 $F \subseteq Q$ ist die Menge der Endzustände

Beispiel:

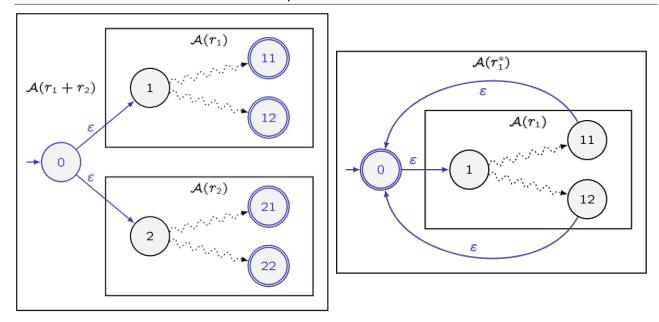


2.2.4 Endliche Automaten und reguläre Ausdrücke

Man kann NEA und RA zusammenführen zu einem Automaten:

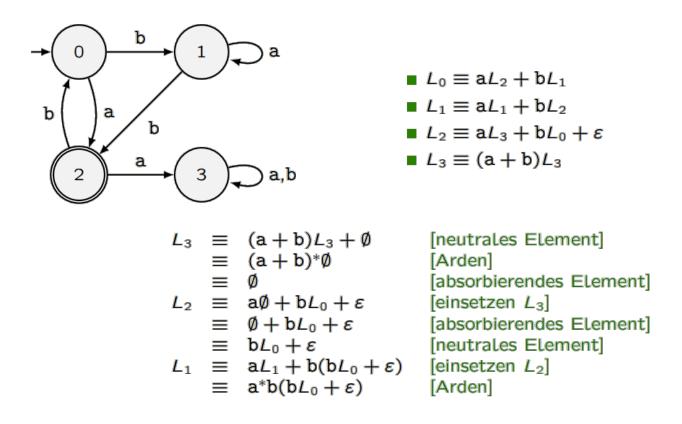






2.2.5 Transformation DEA zu RA

Um einen DEA zu einem RA umzuformen müssen zuerst die Gleichungen der jeweiligen Zustände aufgestellt und vereinfacht werden:



Folglich kann man die gekürzten Gleichungen in einander einsetzen um bis zum Anfangszustand zu kommen:



- $\blacksquare L_0 \equiv aL_2 + bL_1$
- $\blacksquare L_1 \equiv a^*b(bL_0 + \varepsilon)$
- $L_2 \equiv bL_0 + \varepsilon$
- \blacksquare $L_3 \equiv \emptyset$

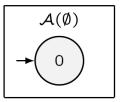
Da L_2 der Endzustand ist, bekommt die Gleichung ε hinzuaddiert! Der vereinfachte Anfangszustand L_0 ist dann der RA zum zugehörigen DEA:

$$L_0 \equiv a(bL_0 + \varepsilon) + b(a^*b(bL_0 + \varepsilon))$$
 [einsetzen L_1, L_2]
 $\equiv abL_0 + a + ba^*bbL_0 + ba^*b$ [Distributivgesetz]
 $\equiv (ab + ba^*bb)L_0 + a + ba^*b$ [Kommutativ-, Distributivgesetz]
 $\equiv (ab + ba^*bb)^*(a + ba^*b)$ [Arden]

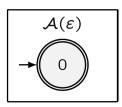
Ergebnis:
$$\mathcal{L}(A) = \mathcal{L}((ab + ba*bb)*(a + ba*b))$$

2.2.6 RA zu NEA

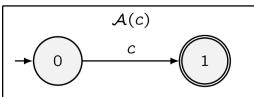
Aus den elementaren RA können einfach NEAs erstellt werden.



$$\mathcal{A}(\varepsilon) = (\{0\}, \Sigma, \{\}, 0, \{0\})$$



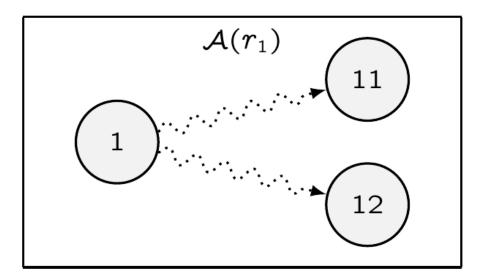
3
$$\mathcal{A}(c) = (\{0, 1\}, \Sigma, \{(0, c, 1)\}, 0, \{1\})$$
 für alle $c \in \Sigma$



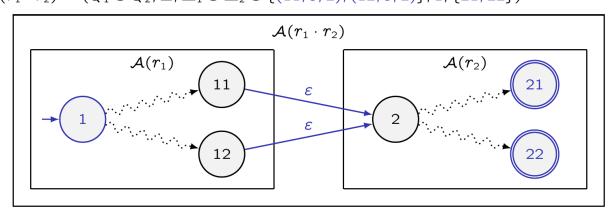


Bei komplexen RAs werden die RAs als "BlackBox"dargestellt, dabei werden die Übergänge gepunktet dargestellt.

$$\mathcal{A}(r_1) = (Q_1, \Sigma, \Delta_1, 1, \{11, 12, \ldots\})$$



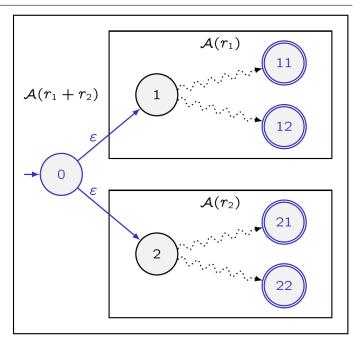
- $A(r_1) = (Q_1, \Sigma, \Delta_1, 1, \{11, 12, \ldots\})$
- $A(r_2) = (Q_2, \Sigma, \Delta_2, 2, \{21, 22, \ldots\})$
- $A(r_1 \cdot r_2) = (Q_1 \cup Q_2, \Sigma, \Delta_1 \cup \Delta_2 \cup \{(11, \varepsilon, 2), (12, \varepsilon, 2)\}, 1, \{21, 22\})$



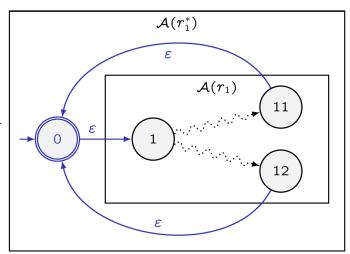
- Startzustand 1
- \blacksquare ε -Übergänge von 11 und 12 zu 2
- Endzustände 21 und 22



- $A(r_1) = (Q_1, \Sigma, \Delta_1, 1, \{11, 12\})$
- $A(r_2) = (Q_2, \Sigma, \Delta_2, 2, \{21, 22\})$
- **5** $A(r_1 + r_2) = (Q, \Sigma, \Delta, 0, F)$
 - $Q = Q_1 \cup Q_2 \cup \{0\}$
 - $\Delta = \Delta_1 \cup \Delta_2 \cup \{(0, \varepsilon, 1), (0, \varepsilon, 2)\}$ $F = \{11, 12, 21, 22\}$
- Neuer Startzustand 0
- ε-Übergänge von 0 zu 1 und 2
- Endzustände 11, 12, 21 und 22



- $A(r_1) = (Q_1, \Sigma, \Delta_1, 1, \{11, 12\})$
- **6** $A(r_1^*) = (Q, \Sigma, \Delta, 0, \{0\})$
 - $Q = Q_1 \cup \{0\}$
- Neuer Startzustand 0
- \blacksquare ε -Übergänge
 - von 0 zu 1
 - von 11 und 12 zu 0
- Endzustand 0



2.2.7 **NEA zu DEA**

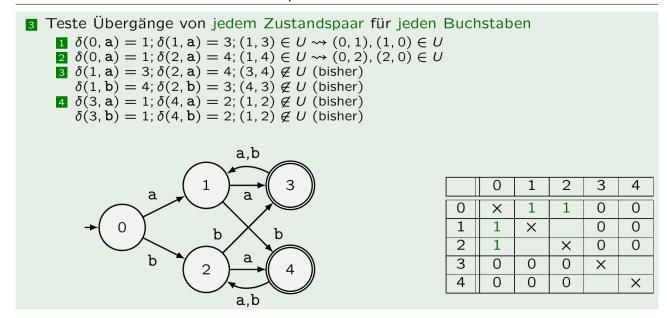
2.2.8 DEA zu RLG

2.2.9 **RLG zu NEA**

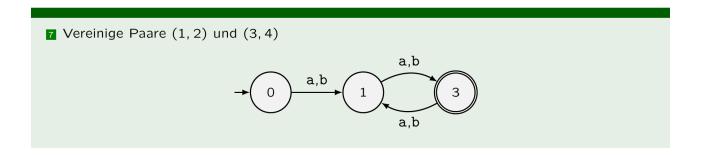
2.2.10 Minimierung

Es wird überprüft ob Zustände erreichbar sind, dafür wird vom Startzustand aus jeder Übergang verfolgt und jeder erreichbare Zustand markiert. Alle unmarkierten Zustände können dann einfach aus dem Automaten entfernt werden, da diese Mülleimer Zustände sind. Anschließend muss überprüft werden, ob die Zustände unterscheidbar sind. Dafür wird eine Tabelle erstellt, welche alle Zustände beinhaltet, die Diagonale kann dabei entfernt werden, im Anschluss werden bei den Endzuständen alle nichtendzustände entfernt. Danach müssen die Übergänge getestet werden. Dazu wird für jeder freie Feld in der Tabelle die Übergänge geprüft.





Hier wurden die Übergänge für (1,0),(2,0),(2,1),(4,3) geprüft. Bei (1,0) führt der Übergang a beim Zustand 0 zu 1 und beim Zustand 1 zu 3, (1,3) ist aber schon markiert, deshalb wird in die Zelle eine eins geschrieben. Wenn der Übergang nicht schon markiert wurde kann eine 0 in die Zelle geschrieben werden.



Die Zustände bei denen eine 1 Steht können zusammengefasst werden, damit ist die Minimierung auch fertig.

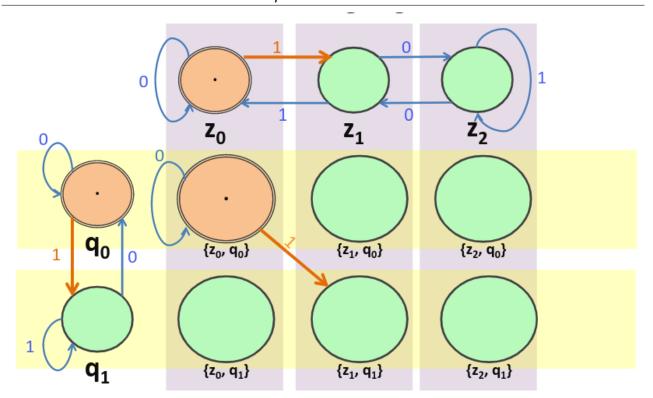
2.2.11 Produktautomaten

Kreuzprodukt der Zustände von A_1 und A_2 erstellen.

Für jeden Zustand die dazugehörigen Übergänge betrachten.

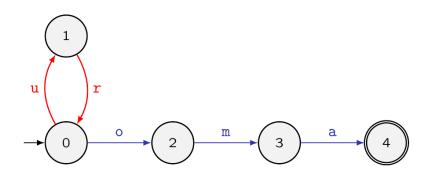
Dabei wird zum Beispiel für den Zustand (z_0, q_0) überprüft zu welchem Zustand der Übergang 1 führt. Der Zustand z_0 führt mit dem Übergang 1 zum Zustand z_1 und z_0 tund z_0 heißt das der Zustand z_0 mit dem Übergang 1 zum Zustand z_0 führt.





2.3 Nicht-reguläre Sprachen und das Pumping-Lemma

- 1. Das gesuchte Wort s besteht aus Prolog u, Zyklus v und Epilog w
- 2. Der Zyklus hat mindestest die Länge 1 $v \neq \epsilon$
- 3. Prolog und Zyklus zusammen haben höchstens die Länge k
- 4. Eine beliebige Anzahl von Zyklus-Durchläufen erzeugt ein Wort der Sprache L



C hat 5 Zustände

k = 5

■ uroma hat 5 Buchstaben

 $s = \mathtt{uroma}$

■ Es gibt eine Zerlegung $s = u \cdot v \cdot w$

 $v = ur \quad w = oma$

v = ur

lacksquare so dass $|v|
eq \varepsilon$

$$|\varepsilon \cdot \mathrm{ur}| = 2 \le 5$$

 $\begin{tabular}{l} \blacksquare \begin{tabular}{l} und & |u\cdot v| \leq k \\ \hline & und & \forall h \in \mathbb{N} (u\cdot v^h\cdot w \in \mathcal{L}(\mathcal{C})) \\ \end{tabular}$

 $(ur)^*oma \subseteq \mathcal{L}(C)$



Hier ein Beispiel:

$$L = \{a^n b^m \mid n < m\}$$

Das Wort ist damit a^nb^m , dieses Word muss in u, v und w aufgeteilt werden um das Pumping Lemma anzuwenden.

$$x = a^n b^{n+1} \in L$$

$$u = a^{n-k}$$

$$v = a^k$$

$$w = b^m$$

Wir teilen u, v und w so auf, dass wir nun das Pumping Lemma anwenden können, wobei k>0 ist.

$$x = a^{n-k} (a^k)^i b^{n+1}$$

Laut dem Pumping Lemma können wir jetzt k beliebig wählen um das erzeugte Wort sollte immer noch ein Teil der Sprache L sein. i steht dabei für die Iterationen, der des Zyklus. Also wählen wir i=3 und Formen die gleichen um.

$$x = a^{n+k(-1+3)}b^{n+1}$$

$$x = a^{n+2k}b^{n+1}$$

Weil k > 0 wissen wir, das das $x \notin L$ und somit auch das L keine reguläre Sprache ist.

2.4 Eigenschaften regulärer Sprachen

3 Chomsky Grammatiken und kontextfreie Sprachen

Gramatiken erzeugen formale Sprachen dar.

$$G=(N, \sum, P, S)$$
 mit:

- **N** Nichtterminalsymbole. Diese können für Regeln verwendet werden, aber dürfen nicht selbst im abgeleiteten Wort stehen.
- **P** Ableitungsregeln. Bsp.: $P=\{S \rightarrow Aa | \varepsilon, A \rightarrow a\}$
- **S** Startsymbol (ist nichtterminel)

3.1 Typ0 unbeschränkt

3.2 Typ1 Monoton

 $\alpha \to \beta$ mit $|\alpha| \le |\beta|$ und Ausnahme $S \to \varepsilon$, wenn S auf keiner rechten Seite ist.

3.3 Typ2 Kontextfreie

 $A \rightarrow \beta$ mit $A \in N$ und $\beta \in V^*$



3.4 Typ3 rechtsregulär/-linear

 $A \rightarrow cB \text{ mit } A \in N; B \in N \cup \{\varepsilon\}; c \in \sum \cup \{\varepsilon\}$

3.5 Cocke-Younger-Kasami(CYK)-Algorythmus

Entscheidet Wortproblem für kontextfreie Grammatiken in CNF. Bsp.:

$$w = abacba$$

Regeln:

5	\rightarrow	a
В	\rightarrow	b
В	\rightarrow	С
S	\rightarrow	SA
Α	\rightarrow	BS
В	\rightarrow	BB
В	\rightarrow	BS

$i \setminus j$	1	2	3	4	5	6
1	5	{}	5	{}	{}	5
2		В	A, B	В	В	A, B
3			5	{}	{}	5
4				В	В	A, B
5					В	A, B
6						5
w = 1	a	b	a	С	b	a

Umgekehrte Regeln:

$$SA \leftarrow S$$
 $BS \leftarrow A, B$
 $BB \leftarrow B$

- 1. Wort unter Tabelle schreiben
- 2. Umgekehrte Regeln bilden (optional)
- 3. Herleitungen in Tabellen-Diagonale eintragen
- 4. Herleitungen der Teilwörter als Menge eintragen (alle Möglichkeiten)

3.6 Chomskynormalform CNF

Zur Entscheidung des Wortproblems. Nur Regeln mit Syntax:

1. $A \to BC$ oder 2. $A \to a$ mit $A, B, C \in Nichtterminal symbole <math>\land a \in Terminal symbole$ und $S \to \epsilon$, wenn S auf keiner Rechten Seite ist.

Vorgehen:

- 1. Epsilon-Regeln entfernen
 - (a) Erstelle Liste L mit $A \to \epsilon$ -Regeln und allen Regeln, die auf Nichtterminalsymbole in L zeigen.
 - (b) Ergänze rechte Seite der Regeln mit sich selbst mit eingesetztem ϵ für Nichtterminalsymbole \in L
 - (c) Wenn $S \in L$ füge Regel $S_0 \to S | \epsilon$ ein
- 2. Kettenregel entfernen
 - (a) Erstelle Listen für Kettenregeln ausgehend von jeweiligen Nichtterminalen
 - (b) Ergänze Regeln mit allen Kettenregeln
 - (c) Regeln kürzen
- 3. überflüssige Symbole entfernen
- 4. Einzelne Nichtterminalsymbole auf rechter Seite ersetzen
- 5. Rechte Seite mit Hilfssymbolen kürzen



4 Kellerautomaten

Endlicher Automat mit unendlichem Stack, auf welchem nur der oberste/neueste Buchstabe gelesen, »gepusht« oder »gepopt« werden kann.

Syntax: $A=(Q, \Sigma, \Gamma, \Delta, q_0, Z)$

- Γ Stack-Alphabet
- q_0 Startzustand
- **Z** Startstacksymbol

5 Turing Maschine

Automaten mit endlosem Einleseband.

Terminiert wenn Endzustand erreicht und Einleseband nicht verschiebbar ist.

$$M=(Q, \sum, \Gamma, \Delta, q_0, F)$$
 mit:

- Γ Vereinigung aus mindestens Blank-Symbol und Terminalsymbole: $\Gamma \supseteq \sum \cup \{\Box\}$
- Δ Übergangsrelationen Syntax: IST-Zustand IST-Inhalt Neuer-Inhalt Verschiebung Neuer-Zustand Bsp.: 0 $\binom{a}{\Box}$ $\binom{a}{a}$ $\binom{r}{l}$ 1
- F Endzustände

5.1 K-Band zu 1-Band

5.2 Linear beschränkter Automat

Touring Automat mit beschränktem Band durch Start- und Endsymbol (Band: "> abaab < ")

6 Entscheidbarkeit

- 6.1 Entscheidbar
- 6.2 Unentscheidbar

Halteproblem:

6.3 Semi-entscheidbar

6.4 Postsches KorrespondenzProblem (PKP)

Endliche Folge an Wortpaaren, mit nichtleeren Wörtern, über endlichem Alphabet.

Syntax:
$$P=((l_1,r_1),(l_2,r_2),...(l_n,r_n))$$
 Lösung: $l_{i_1} \cdot l_{i_2} \cdot ... \cdot l_{i_n} = r_{i_1} \cdot r_{i_2} \cdot ... \cdot r_{i_n}$



7 Berechenbarkeit

8 Komplexität