

# Theoretische Informatik

Zusammenfassung

SoSe2024

## Inhaltsverzeichnis

<b>1</b>	<b>Allgemein</b>	<b>3</b>
1.1	Alphabete und Wörter . . . . .	3
1.2	Grammatiken . . . . .	3
<b>2</b>	<b>Chomsky-Hierarchie</b>	<b>4</b>
2.1	Typ 0 ( $\mathcal{L}_0$ ) - Phrasenstrukturgrammatiken . . . . .	4
2.2	Typ 1 ( $\mathcal{L}_1$ ) - Kontextsensitive Grammatiken . . . . .	4
2.3	Typ 2 ( $\mathcal{L}_2$ ) - Kontextfreie Grammatiken . . . . .	4
2.4	Typ 3 ( $\mathcal{L}_3$ ) - Reguläre Grammatik . . . . .	4
<b>3</b>	<b>Deterministischer Endlicher Automat (DEA)</b>	<b>5</b>
<b>4</b>	<b>Nicht-deterministischer Endlicher Automat (NEA)</b>	<b>6</b>
<b>5</b>	<b>Äquivalenz von DEA und NEA</b>	<b>7</b>
5.1	Satz von Rabin und Scott . . . . .	7
5.2	Potenzmengenkonstruktion ( $\text{NEA} \rightarrow \text{DEA}$ ) . . . . .	7
<b>6</b>	<b>Regex</b>	<b>8</b>
6.1	Satz von Kleene . . . . .	8
<b>7</b>	<b>Pumping Lemma</b>	<b>9</b>
<b>8</b>	<b>Satz von Myhill und Nerode</b>	<b>10</b>
<b>9</b>	<b>Minimalautomaten</b>	<b>11</b>
9.1	Table-Filling-Algorithmus . . . . .	11
<b>10</b>	<b>Kontextfreie Sprachen (<math>\mathcal{L}_2</math>)</b>	<b>12</b>
10.1	Chomsky Normalform (CNF) . . . . .	12
10.2	Greibach Normalform . . . . .	12
10.3	Konvertierung . . . . .	12

<b>11 Kellerautomaten</b>	<b>13</b>
<b>12 CYK-Algorithmus</b>	<b>14</b>
<b>13 Turing-Maschine</b>	<b>15</b>
13.1 Linear beschränkte Turing-Maschine . . . . .	16
<b>14 Satz von Kuroda</b>	<b>16</b>

# 1 Allgemein

## 1.1 Alphabete und Wörter

- Ein Alphabet  $\Sigma$  ist eine endliche Menge unterscheidbarer Symbole
- Element  $\sigma \in \Sigma$  ist ein Zeichen des Alphabets  $\Sigma$
- Jedes Element  $\omega \in \Sigma^*$  ist ein Wort über  $\Sigma$
- $\varepsilon$  = Leeres Wort
- $\Sigma^*$ : Menge aller Wörter über  $\Sigma$
- $\Sigma^+$ : Menge aller Wörter über  $\Sigma$  mit mind. 1 Element
- $|\omega|$ : Länge eines Wortes ( $|\varepsilon| = 0$ )

## 1.2 Grammatiken

Eine Grammatik  $G$  ist ein 4-Tupel  $(V, \Sigma, P, S)$ :

- $V$ : endliche Menge an Nicht-Terminal-Symbolen
- $\Sigma$ : endliche Menge an Terminal-Symbolen ( $V \cap \Sigma = \emptyset$ )
- $P$ : endliche Menge an Produktionsregeln
- $S$ : Startsymbol ( $S \in V$ )

## 2 Chomsky-Hierarchie

### 2.1 Typ 0 ( $\mathcal{L}_0$ ) - Phrasenstrukturgrammatiken

- Beliebige Kombination aus T- und NT-Symbolen

### 2.2 Typ 1 ( $\mathcal{L}_1$ ) - Kontextsensitive Grammatiken

- $|l| \leq |r|$
- Länge des Wortes steigt
- $S \rightarrow \varepsilon$  erlaubt, wenn S auf **keiner** rechten Seite einer Regel steht!

**Beispiel:**

$$S \rightarrow S' \mid \varepsilon$$

$$S' \rightarrow aS'Bc \mid abc$$

$$cB \rightarrow Bc$$

$$bB \rightarrow bb$$

Das Nichtterminal S' braucht man nur, damit die Bedingung der Sonderregel erfüllt ist. Das Nichtterminal B wird mal zur Satzform Bc und mal zu bb, je nachdem ob B im **Kontext** c oder b steht.

### 2.3 Typ 2 ( $\mathcal{L}_2$ ) - Kontextfreie Grammatiken

Beim Ableiten in Typ-1-Grammatiken muss man immer aufpassen, dass das Nichtterminal auch im richtigen Kontext steht. Das Erzeugen von Sätzen ist viel leichter, wenn die Grammatik kontextfrei ist.

Eine Grammatik G ist vom Typ 2, wenn sie vom Typ 1 ist und zusätzlich auf der linken Seite jeder Regel genau **ein** Nichtterminal steht!

- $l \in V$
- $X \rightarrow \varepsilon$  immer erlaubt

### 2.4 Typ 3 ( $\mathcal{L}_3$ ) - Reguläre Grammatik

Eine Grammatik G ist vom Typ 3, wenn sie vom Typ 2 ist und zusätzlich folgende Regeln hat:

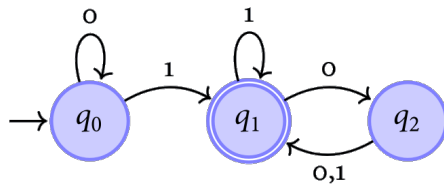
- $A \rightarrow b$
- $A \rightarrow bC$
- $A \rightarrow \varepsilon$

### 3 Deterministischer Endlicher Automat (DEA)

Eine DEA  $M$  ist ein 5-Tupel  $(Q, \Sigma, \delta, q_0, F)$ :

- $Q$ : endliche Zustandsmenge
- $\Sigma$ : endliches Alphabet
- $\delta: Q \times \Sigma \rightarrow Q$  Übergangsfunktionen
- $q_0$ : Startzustand
- $F$ : Menge der akzeptierten Endzustände

Beispiel:



- $Q = \{q_0, q_1, q_2\}$
- $\Sigma = \{0, 1\}$
- $q_0 = q_0$
- $F = q_2$
- $\delta$ :

$$\delta(q_0, 0) = q_0$$

$$\delta(q_0, 1) = q_1$$

$$\delta(q_1, 0) = q_2$$

$$\delta(q_1, 1) = q_1$$

$$\delta(q_2, 0) = q_1$$

$$\delta(q_2, 1) = q_1$$

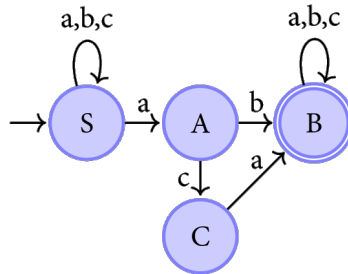
## 4 Nicht-deterministischer Endlicher Automat (NEA)

Eine NEA  $M$  ist ein 5-Tupel  $(Q, \Sigma, \delta, q_0, F)$ :

- $Q$ : endliche Zustandsmenge
- $\Sigma$ : endliches Alphabet
- $\delta: Q \times \Sigma \rightarrow Q$  Übergangsfunktionen
- $q_0$ : Menge der Startzustände
- $F$ : Menge der akzeptierten Endzustände

Beispiel:

$S \rightarrow aS \mid bS \mid cS \mid aA$   
 $A \rightarrow bB \mid cC$   
 $B \rightarrow aB \mid bB \mid cB \mid \varepsilon$   
 $c \rightarrow aB$



## 5 Äquivalenz von DEA und NEA

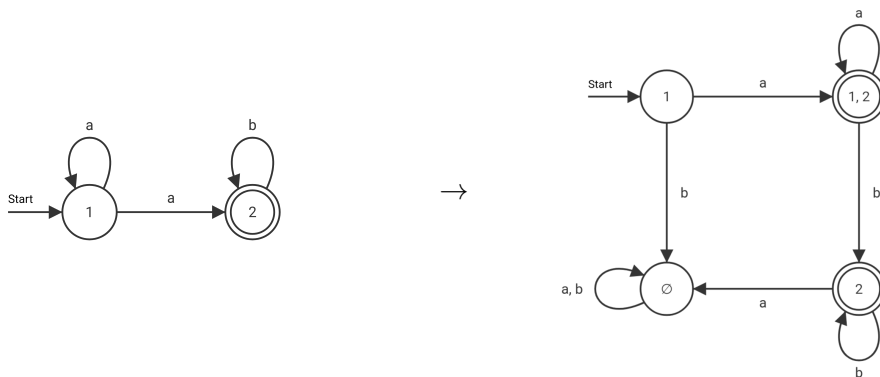
### 5.1 Satz von Rabin und Scott

Jede von einem NEA akzeptierte Sprache L ist auch von einem DEA akzeptierbar.

### 5.2 Potenzmengenkonstruktion (NEA $\rightarrow$ DEA)

Potenzmenge ist die Menge aller Teilmengen

Beispiel:



1. Potenzmenge bilden,  $\mathcal{P}\{1, 2\} = \{\{\emptyset\}, \{1\}, \{2\}, \{1, 2\}\}$
2. Jede Teilmenge ist Zustand des DEA
3. Wohin kommt man von von jedem NEA Zustand?
  - $\delta(\{1\}, a) = 1 \wedge 2 \rightarrow \{1, 2\}$
  - $\delta(\{1\}, b) = \{\emptyset\}$
  - $\delta(\{2\}, a) = \{\emptyset\}$
  - $\delta(\{2\}, b) = \{2\}$
  - $\delta(\{2\}, b) = \{2\}$
  - $\delta(\{1, 2\}, a) = \delta(\{1\}, a) \cup \delta(\{2\}, a) = 1 \wedge 2 \cup \emptyset = \{1, 2\}$
  - $\delta(\{\emptyset\}, a) \wedge \delta(\{\emptyset\}, b) = \emptyset \rightarrow \{\emptyset\}$
4. Startzustand bleibt gleich
5. Jede Teilmenge, in der der ursprüngliche Endzustand vorkommt, wird wieder zum Endzustand,  $F = 2 \rightarrow F' = \{\{2\}, \{1, 2\}\}$

## 6 Regex

- `.` : Beliebiges einzelnes Zeichen außer Zeilenumbruch
- `*` : Null oder mehr Wiederholungen
- `+` : Eine oder mehr Wiederholungen
- `{n}` : Genau n Wiederholungen
- `{n,}` : Mindestens n Wiederholungen
- `{n, m}` : Zwischen n und m Wiederholungen
- `[ ]` : Zeichenklasse (z.B. `[a-z]` für alle Kleinbuchstaben)
- `[ ^ ]` : Negierte Zeichenklasse

Email-Adressen: `[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}`

### 6.1 Satz von Kleene

Die Menge der durch reguläre Ausdrücke (Regex) beschreibbaren Sprachen ist genau die Menge der regulären Sprachen.

→ Alle endlichen Sprachen sind durch reguläre Ausdrücke beschreibbar



## 7 Pumping Lemma

Das Pumping-Lemma wird verwendet, um zu beweisen, dass eine Sprache sicher nicht regulär ist.

!!!TODO!!!

## 8 Satz von Myhill und Nerode

Eine Sprache  $L$  ist genau dann regulär, wenn der Index  $R_L$  endlich ist!

## 9 Minimalautomaten

### 9.1 Table-Filling-Algorithmus

Nur für DEAs!

1. Eventuell nicht erreichbare Zustände entfernen
2. Tabelle aus Zuständen erstellen  $\{q, q'\}$ ,  $q \neq q'$
3. Zustandspaar markieren, bei dem immer **ein** Zustand ein Endzustand ist
4. Wiederhole solange, bis keine Markierungen mehr dazu kommen
  - Für jedes unmarkierte Paar  $\{q, q'\}$  und jeden Buchstaben  $a \in \Sigma$ 
    - Wenn  $\{\delta(q, a), \delta(q', a)\}$  markiert ist, dann das Paar  $\{q, q'\}$  selbst markieren
5. Verschmelze unmarkierte Zustandspaare

## 10 Kontextfreie Sprachen ( $\mathcal{L}2$ )

### 10.1 Chomsky Normalform (CNF)

Regeln müssen folgende Formen haben:

- $A \rightarrow BC$
- $A \rightarrow a$
- $S \rightarrow \varepsilon$

### 10.2 Greibach Normalform

Eine  $\varepsilon$ -freie, kontextfrei Grammatik mit folgenden Regeln:

- $A \rightarrow aB_1B_2B_3\dots B_k$
- $k \geq 0$

### 10.3 Konvertierung

!!!TODO!!!

## 11 Kellerautomaten

Ein Kellerautomat (PDA)  $M$  ist ein 6-Tupel  $(Q, \Sigma, \Gamma, \delta, q_0, \#)$ :

- $Q$ : endliche Zustandsmenge
- $\Sigma$ : endliches Bandalphabet
- $\Gamma$ : endliches Kelleralphabet
- $\delta$ : Übergangsfunktionen
- $q_0$ : Startzustand ( $q_0 \in Q$ )
- $\#$ : Ursprüngliches Kellersymbol ( $q_0 \in \Gamma$ )

Akzeptanz:

- Kein akzeptierender Endzustand!
- Akzeptanzkriterien für Wörter  $x \in |\Sigma|^*$ :
  1. Wort komplett gelesen
  2. Keller (Stack leer)

Nicht-Determinismus:

- Mehrere simultane Übergänge möglich
- Spontane Übergänge ( $a = \varepsilon$ ) möglich

Konfiguration eines PDA gegeben durch 3-Tupel  $(Q, \Sigma^*, \Gamma^*)$ :

- $q \in Q$ : Momentaner Zustand
- $w' \in \Sigma^*$ : Noch zu lesender Anteil der Eingabe
- $\gamma \in \Gamma^*$ : Aktueller Kellerinhalt

Übergangsfunktion:

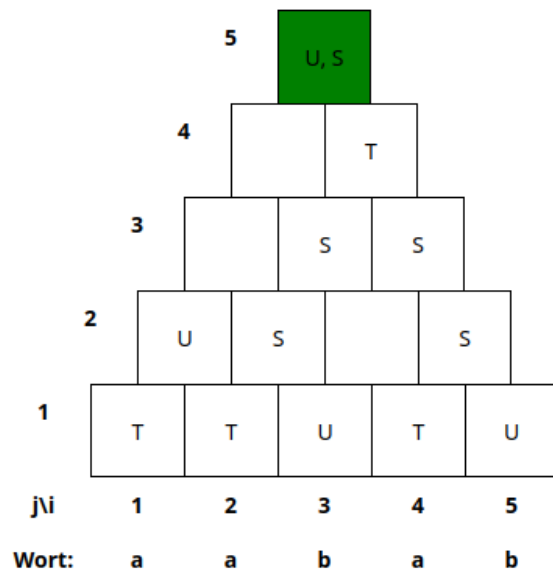
- $\delta(q, a, A) \ni (q', B_1 B_2 \dots B_k)$
- Wenn Automat in Zustand  $q$  ist, das Symbol  $a$  liest und  $A$  oben auf Stack liegt, wechselt er in Zustand  $q'$  und ersetzt das  $A$  auf dem Stack durch  $B_1 B_2 \dots B_k$

## 12 CYK-Algorithmus

Beispiel:

- $S \rightarrow ST \mid TU \mid US$
- $T \rightarrow SS \mid a$
- $U \rightarrow TT \mid b$

Wort: **aabab**



Nur wenn S (Startsymbol) ganz oben in der Pyramide steht, wird das Wort akzeptiert!

## 13 Turing-Maschine

Eine Turing-Maschine  $M$  ist ein 7-Tupel  $(Q, \Sigma, \Gamma, \delta, q_0, \square, F)$ :

- $Q$ : endliche Zustandsmenge
- $\Sigma$ : endliches Eingabealphabet
- $\Gamma$ : endliches Arbeitsalphabet ( $\Sigma \subset \Gamma$ )
- $\delta$ : Übergangsfunktionen
- $q_0$ : Startzustand ( $q_0 \in Q$ )
- $\square$ : Blank-Symbol ( $\square \in \Gamma - \Sigma$ )
- $F$ : Menge der Endzustände

Übergangsfunktionen:

- Deterministisch:  $\delta(q, a) = \delta(q', b, x)$ 
  1.  $M$  befindet sich in Zustand  $q$  und liest  $a$  vom Band
  2.  $M$  geht in Zustand  $q'$  über und ersetzt das  $a$  mit einem  $b$
  3.  $M$  führt Kopfbewegung  $x \in \{l, n, r\}$
- Nicht-deterministisch:  $\delta(q, a) \ni \delta(q', b, x)$

Konfiguration einer TM ist ein Wort  $k \in \Gamma^* Q \Gamma^*$ :

$$k = \alpha q \beta$$

$$k = \alpha_1 \dots \alpha_m q \beta_1 \dots \beta_n$$

- $q$ : Aktueller Zustand
- $\alpha$ : Wort links des Schreib/Lese-Kopfes
- $\beta$ : Wort rechts des Kopfes

Startkonfiguration  $q_0 \vec{w}$ :

- $w \in \Sigma^*$  steht auf Band
- $M$  in Zustand  $q_0$
- S/L-Kopf steht auf erstem Buchstaben von  $w$
- Restliches Band mit Blanks  $\square$  befüllt

### 13.1 Linear beschränkte Turing-Maschine

Eine Turing Maschine  $M$  ist linear beschränkt, wenn  $|\vec{w}| < \infty$ , also endlich ist.

## 14 Satz von Kuroda

Die von linear beschränkten, nicht-deterministischen Turing-Maschinen akzeptierbaren Sprachen sind genau die kontextsensitiven Sprachen  $\mathcal{L}_1$