

Theorie der Informatik

7. Formale Sprachen und Grammatiken

Malte Helmert Gabriele Röger

Universität Basel

16. März 2015

Theorie der Informatik

16. März 2015 — 7. Formale Sprachen und Grammatiken

7.1 Einführung

7.2 Alphabete und formale Sprachen

7.3 Grammatiken

7.4 Chomsky-Hierarchie

7.5 Zusammenfassung

7.1 Einführung

Beispiel: Aussagenlogische Formeln

Aus dem Logikteil:

Definition (Syntax der Aussagenlogik)

Sei A eine Menge von atomaren Aussagen. Die Menge der aussagenlogischen Formeln (über A) ist induktiv wie folgt definiert:

- ▶ Jedes **Atom** $a \in A$ ist eine aussagenlogische Formel über A .
- ▶ Ist ϕ eine aussagenlogische Formel über A , dann auch die **Negation** $\neg\phi$.
- ▶ Sind ϕ und ψ aussagenlogische Formeln über A , dann ist es auch die **Konjunktion** $(\phi \wedge \psi)$.
- ▶ Sind ϕ und ψ aussagenlogische Formeln über A , dann ist es auch die **Disjunktion** $(\phi \vee \psi)$.

Beispiel: Aussagenlogische Formeln

Sei S_A die Menge aller aussagenlogischen Formeln über A .

Solche Mengen von Zeichenketten (oder **Wörtern**) nennt man **Sprachen**.

Gesucht: Allgemeine Konzepte, um solche (oftmals unendliche) Sprachen mit endlichen Beschreibungen zu definieren

- ▶ heute: **Grammatiken**
- ▶ nächste Kapitel zusätzlich: Automaten

Beispiel: Aussagenlogische Formeln

Beispiel (Grammatik für $S_{\{a,b,c\}}$)

Grammatikvariablen $\{F, A, N, K, D\}$ mit Startvariable F ,
Terminalsymbole $\{a, b, c, \neg, \wedge, \vee, (,)\}$ und Regeln

$$\begin{array}{lll} F \rightarrow A & A \rightarrow a & N \rightarrow \neg F \\ F \rightarrow N & A \rightarrow b & K \rightarrow (F \wedge F) \\ F \rightarrow K & A \rightarrow c & D \rightarrow (F \vee F) \\ F \rightarrow D & & \end{array}$$

Beginne mit F und ersetze schrittweise eine linke Regelseite durch eine rechte Regelseite bis keine Variablen mehr enthalten sind:

$$\begin{aligned} F &\Rightarrow N \Rightarrow \neg F \Rightarrow \neg D \Rightarrow \neg(F \vee F) \Rightarrow \neg(A \vee F) \Rightarrow \neg(b \vee F) \\ &\Rightarrow \neg(b \vee A) \Rightarrow \neg(b \vee c) \end{aligned}$$

7.2 Alphabete und formale Sprachen

Alphabete und formale Sprachen

Definition (Alphabete, Wörter und formale Sprache)

Ein **Alphabet** Σ ist eine endliche, nicht-leere Menge von **Zeichen** (oder **Symbolen**).

Die Menge Σ^* aller **Wörter über Σ** enthält alle endlichen Folgen von Elementen aus Σ . Das **leere Wort** (die leere Elementfolge) wird dabei mit ε bezeichnet.

Für ein Wort w bezeichnet $|w|$ seine Länge.

Eine **formale Sprache** ist eine Teilmenge von Σ^* .

Beispiel

$$\begin{aligned} \Sigma &= \{a, b\} \\ \Sigma^* &= \{\varepsilon, a, b, aa, ab, ba, bb, \dots\} \\ |aba| &= 3, |b| = 1, |\varepsilon| = 0 \end{aligned}$$

Sprachen: Beispiele

Beispiel (Sprachen über $\Sigma = \{a, b\}$)

- ▶ $S_1 = \{a, aa, aaa, aaaa, \dots\}$
- ▶ $S_2 = \Sigma^*$
- ▶ $S_3 = \{a^n b^n \mid n \geq 0\} = \{\varepsilon, ab, aabb, aaabbb, \dots\}$
- ▶ $S_4 = \{\varepsilon\}$
- ▶ $S_5 = \emptyset$
- ▶ $S_6 = \{w \in \Sigma^* \mid w \text{ enthält doppelt so viele } a \text{ wie } b\}$
 $= \{\varepsilon, aab, aba, baa, \dots\}$
- ▶ $S_7 = \{w \in \Sigma^* \mid |w| = 3\}$
 $= \{aaa, aab, aba, baa, bba, bab, abb, bbb\}$

7.3 Grammatiken

Grammatiken

Definition (Grammatik)

Eine **Grammatik** ist ein 4-Tupel (Σ, V, P, S) mit

- ① Σ endliches Terminalalphabet,
- ② V endliche Menge von Variablen (mit $V \cap \Sigma = \emptyset$),
- ③ $P \subseteq (V \cup \Sigma)^+ \times (V \cup \Sigma)^*$ endliche Menge von Regeln, und
- ④ $S \in V$ Startvariable.

Regelmengen

Was genau bedeutet $P \subseteq (V \cup \Sigma)^+ \times (V \cup \Sigma)^*$?

- ▶ $(V \cup \Sigma)^*$: Alle Wörter über $(V \cup \Sigma)$
- ▶ $(V \cup \Sigma)^+$: Alle nicht-leeren Wörter über $(V \cup \Sigma)$
 Allgemein für Menge X : $X^+ = X^* \setminus \{\varepsilon\}$
- ▶ $\times$: kartesisches Produkt
- ▶ $(V \cup \Sigma)^+ \times (V \cup \Sigma)^*$: Menge aller Paare (x, y) , wobei x nicht-leeres Wort über $(V \cup \Sigma)$ und y Wort über $(V \cup \Sigma)$
- ▶ Statt (x, y) schreiben wir Regeln meist in der Form $x \rightarrow y$.

Regeln: Beispiele

Beispiel

Sei $\Sigma = \{a, b, c\}$ und $V = \{X, Y, Z\}$.

Folgende Regeln sind in $(V \cup \Sigma)^+ \times (V \cup \Sigma)^*$:

$$\begin{aligned} X &\rightarrow XaY \\ Yb &\rightarrow a \\ XY &\rightarrow \varepsilon \\ XYZ &\rightarrow abc \\ abc &\rightarrow XYZ \end{aligned}$$

Ableitung

Definition (Ableitung)

Sei (Σ, V, P, S) eine Grammatik. Ein Wort $v \in (V \cup \Sigma)^*$ kann von Wort $u \in (V \cup \Sigma)^+$ **abgeleitet** werden ($u \Rightarrow v$), falls

- ① $u = xyz$, $v = xy'z$ mit $x, z \in (V \cup \Sigma)^*$ und
- ② es existiert eine Regel $y \rightarrow y' \in P$.

Wir schreiben: $u \Rightarrow^* v$ falls v in endlich vielen Schritten (d.h. durch die Anwendung von n Regeln für $n \in \mathbb{N}_0$) von u abgeleitet werden kann.

Ableitung: Mehrdeutigkeit

Beispiel

$G = (\{a, b, c\}, \{S, A, B\}, P, S)$, wobei P folgende Regeln enthält:

$$\begin{array}{ll} S \rightarrow aB & A \rightarrow ab \\ S \rightarrow Ac & B \rightarrow bc \end{array}$$

Wort abc kann von S abgeleitet werden: $S \Rightarrow aB \Rightarrow abc$

Alternative Ableitung: $S \Rightarrow Ac \Rightarrow abc$

- ▶ Es kann verschiedene Ableitungen für ein Wort geben.
- ▶ Dies gilt selbst, wenn man immer nur die erste vorkommende Variable ersetzt (sogenannte **Linksableitung**).

Erzeugte Sprache einer Grammatik

Definition (Sprache)

Die von einer Grammatik $G = (\Sigma, V, P, S)$ erzeugte **Sprache**

$$\mathcal{L}(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

ist die Menge aller Wörter aus Σ^* , die von S mit endlich vielen Regelanwendungen abgeleitet werden können.

Grammatiken

Beispiele: Tafel

7.4 Chomsky-Hierarchie

Chomsky-Hierarchie

Grammatiken werden in die **Chomsky-Hierarchie** eingegliedert.

Definition (Chomsky-Hierarchie)

- ▶ Jede Grammatik ist vom Typ 0 (beliebige Regeln erlaubt).
- ▶ Grammatik ist vom Typ 1 (kontextsensitiv), falls für alle Regeln $w_1 \rightarrow w_2$ gilt: $|w_1| \leq |w_2|$.
- ▶ Grammatik ist vom Typ 2 (kontextfrei), falls für alle Regeln $w_1 \rightarrow w_2$ gilt: $w_1 \in V$ (einzelne Variable).
- ▶ Grammatik ist vom Typ 3 (regulär), falls zusätzlich $w_2 \in \Sigma \cup \Sigma V$.

Spezialfall: Regel $S \rightarrow \varepsilon$ ist immer erlaubt, wenn S Startsymbol ist und auf keiner rechten Seite irgendeiner Regel auftritt.

Chomsky-Hierarchie

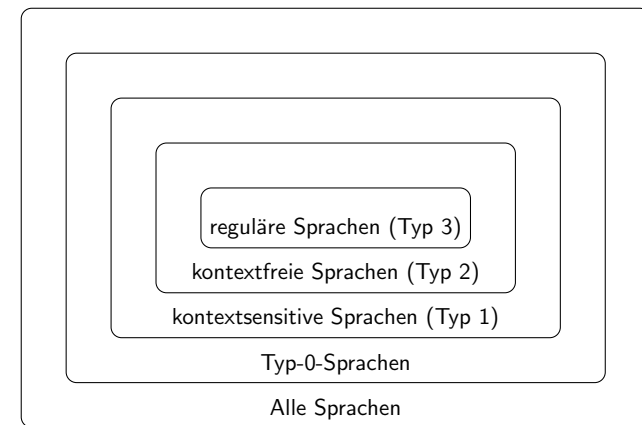
Beispiele: Tafel

Chomsky-Hierarchie

Definition (Typ-0–3 Sprachen)

Eine Sprache $L \subseteq \Sigma^*$ heisst vom Typ 0 (Typ 1, Typ 2, Typ 3), falls es eine Typ-0- (Typ-1-, Typ-2-, Typ-3-) Grammatik G gibt mit $\mathcal{L}(G) = L$.

Chomsky-Hierarchie



Nicht alle Sprachen sind durch Grammatiken beschreibbar.

7.5 Zusammenfassung

Zusammenfassung

- ▶ **Sprachen** sind Mengen von Zeichenketten.
- ▶ **Grammatiken** sind eine Möglichkeit Sprachen zu beschreiben.
- ▶ Erzeugte **Sprache einer Grammatik** ist die Menge der **ableitbaren** variablenfreien Wörter.
- ▶ **Chomsky-Hierarchie** teilt Grammatiken in Stufen unterschiedlicher Erzeugungsmächtigkeit ein.

Nächstes Kapitel

- ▶ Mehr zu regulären Sprachen
- ▶ Automaten als alternative Darstellungsform von Sprachen