universität innsbruck

Einführung in die Theoretische Informatik

Martin Avanzini Christian Dalvit Jamie Hochrainer
Georg Moser Johannes Niederhauser Jonas Schöpf

https://tcs-informatik.uibk.ac.at

universität
innsbruck

Zusammenfassung

Zusammenfassung der letzten LVA

Definition (Alphabet)

Ein Alphabet Σ ist eine endliche, nicht leere Menge von Symbolen

Definition (Wort)

- Eine Zeichenreihe (ein Wort, ein String) ist eine endliche Folge von Symbolen über einem Alphabet Σ
- Die leere Zeichenreihe wird mit ϵ bezeichnet

Definition

Eine Teilmenge L von Σ^* heißt eine formale Sprache über Alphabet Σ

Bemerkung

Die Algebra $\langle \Sigma^*; \cdot, \epsilon \rangle$ ist ein Monoid; das Wortmonoid

Einführung in die Logik

Syntax & Semantik der Aussagenlogik, Formales Beweisen, Konjunktive und Disjunktive Normalformen

Einführung in die Algebra

Algebraische Strukturen, Boolesche Algebra, Universelle Algebra

Einführung in die Theorie der Formalen Sprachen

Grammatiken und Formale Sprachen, Chomsky-Hierarchie, Reguläre Sprachen, Kontextfreie Sprachen, Anwendungen von formalen Sprachen

Einführung in die Berechenbarkeitstheorie und Komplexitätstheorie

Algorithmisch unlösbare Probleme, Turing Maschinen, Registermaschinen, Komplexitätstheorie

Einführung in die Programmverifikation

Prinzipien der Analyse von Programmen, Verifikation nach Hoare universität innsbruck

Grammatiken und Formale Sprachen

Grammatiken und Formale Sprachen

Beispiel

$$
\begin{aligned}
\mathsf{S} &\to \mathsf{Pronomen\ Nomen\ Verb\ Adjektiv}\\
\mathsf{Nomen} &\to \mathsf{Lehrveranstaltungsleiter}\\
\mathsf{Nomen} &\to \mathsf{Vortragender}\\
\mathsf{Pronomen} &\to \mathsf{Unser} \mid \mathsf{Mein}\\
\mathsf{Verb} &\to \mathsf{ist}\\
\mathsf{Adjektiv} &\to \mathsf{lästig} \mid \mathsf{nett} \mid \mathsf{streng} \mid \mathsf{monoton} \mid \mathsf{anspruchsvoll}
\end{aligned}
$$

Es gilt

$$\mathsf{S} \overset{*}{\Rightarrow} \mathsf{Unser\ Lehrveranstaltungsleiter\ ist\ anspruchsvoll}$$

Definition

Eine Grammatik G ist ein Quadrupel $G = (V, \Sigma, R, S)$, wobei

1. V eine endliche Menge von Variablen (oder Nichtterminale)
2. Σ ein Alphabet, die Terminale, $V \cap \Sigma = \varnothing$
3. R eine endliche Menge von Regeln
4. $S \in V$ das Startsymbol von G

Eine Regel ist ein Paar $P \to Q$ von Wörtern, sodass $P, Q \in (V \cup \Sigma)^*$ und in P mindestens eine Variable vorkommt

P nennen wir auch die Prämisse und Q die Konklusion der Regel

Konvention

- Variablen werden groß geschrieben, Terminale klein
- Statt $P \to Q_1$, $P \to Q_2$, $P \to Q_3$ schreiben wir $P \to Q_1 \mid Q_2 \mid Q_3$

Sei $G = (V, \Sigma, R, S)$ eine Grammatik und seien $x, y \in (V \cup \Sigma)^*$

Definition

1. Wir sagen y ist aus x in G direkt ableitbar, wenn gilt:

$$\exists\, u, v \in (V \cup \Sigma)^*,\ \exists\, (P \to Q) \in R \ \text{ sodass } \ (x = uPv \text{ und } y = uQv)$$

2. In diesem Fall schreiben wir kurz $x \underset{G}{\Rightarrow} y$
3. Wenn G aus dem Kontext folgt schreiben wir $x \Rightarrow y$

Definition (Ableitbar)

Wir sagen y ist aus x in G ableitbar, wenn $k \in \mathbb{N}$ und $w_0, w_1, \ldots, w_k \in (V \cup \Sigma)^*$ gibt, sodass

$$x = w_0 \Rightarrow w_1 \Rightarrow \ldots \Rightarrow w_k = y$$

Wir schreiben $x \overset{*}{\underset{G}{\Rightarrow}} y$, beziehungsweise $x \overset{*}{\Rightarrow} y$

Sprache einer Grammatik

Definition

- Die vom Startsymbol S ableitbaren Wörter heißen Satzformen
- Elemente von Σ^* heißen Terminalwörter
- Satzformen, die Terminalwörter sind, heißen Sätze

Definition (Sprache einer Grammatik)

Die Menge aller Sätze

$$\mathrm{L}(G) = \{x \in \Sigma^* \mid \mathrm{S} \overset{*}{\underset{G}{\Rightarrow}} x\}$$

heißt die von der Grammatik G erzeugte Sprache

Zwei Grammatiken G_1 und G_2 heißen äquivalent, wenn $\mathrm{L}(G_1) = \mathrm{L}(G_2)$

Klassen von Grammatiken

Definition (rechtslinear)

Grammatik $G = (V, \Sigma, R, S)$ heißt rechtslinear, wenn für alle Regeln $P \to Q$ gilt:

1. $P \in V$
2. $Q \in \Sigma^* \cup \Sigma^+ V$

Beispiel

- Die Grammatik $G_1 = (\{B\}, \{0,1\}, R, B)$ ist rechtslinear, wobei R wie folgt definiert:

$$B \to 0 \mid 1 \mid 0B \mid 1B$$

- Es gilt:

$$\mathrm{L}(G_1) = \{0,1\}^+$$

Definition (kontextfrei)

Grammatik $G = (V, \Sigma, R, S)$ heißt kontextfrei, wenn für alle Regeln $P \to Q$ gilt:

1. $P \in V$
2. $Q \in (V \cup \Sigma)^*$

Beispiel

- Die Grammatik $G_2 = (\{K\}, \{(,)\}, R, K)$ ist kontextfrei, wobei R wie folgt definiert:

$$K \to \epsilon \mid (K) \mid KK$$

- Es gilt:

$$K \Rightarrow KK \Rightarrow (K)K \Rightarrow (\epsilon)K = ()K \Rightarrow ()(K) \Rightarrow ()(KK) \overset{*}{\Rightarrow} ()(()(()))$$

Definition (kontextsensitiv)

Grammatik $G = (V, \Sigma, R, \mathsf{S})$ heißt kontextsensitiv, wenn für alle Regeln $P \to Q$ gilt:

1. entweder es existieren $u, v, w \in (V \cup \Sigma)^*$ und $A \in V$, sodass

$$P = uAv \text{ und } Q = uwv \text{ wobei } |w| \geqslant 1$$

2. oder $P = \mathsf{S}$ und $Q = \epsilon$

Wenn $\mathsf{S} \to \epsilon \in G$, dann kommt S nicht in einer Konklusion vor

Beispiel

$G_3 = (\{\mathsf{S}, B, C, H\}, \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}, R, \mathsf{S})$ ist kontextsensitiv, wobei R:

$$\begin{aligned}
\mathsf{S} &\to \mathsf{a}\mathsf{S}BC \mid \mathsf{a}BC & HC &\to BC & \mathsf{b}C &\to \mathsf{bc} \\
CB &\to HB & \mathsf{a}B &\to \mathsf{ab} & \mathsf{c}C &\to \mathsf{cc} \\
HB &\to HC & \mathsf{b}B &\to \mathsf{bb} & &
\end{aligned}$$

$\mathrm{L}(G_3) = \{\mathsf{a}^n\mathsf{b}^n\mathsf{c}^n \mid n \geqslant 1\}$

Definition (beschränkt)

Grammatik $G = (V, \Sigma, R, \mathrm{S})$ heißt **beschränkt**, wenn für alle Regeln $P \to Q$ gilt:

1. entweder $|P| \leqslant |Q|$ oder
2. $P = \mathrm{S}$ und $Q = \epsilon$

Wenn $\mathrm{S} \to \epsilon \in G$, dann kommt S nicht in einer Konklusion vor

Beispiel

Die Grammatik $G_4 = (\{\mathrm{S}, X, Y, T\}, \{\mathrm{a}\}, R, \mathrm{S})$ sei wie folgt definiert:

$$
\begin{array}{rcl@{\qquad}rcl}
\mathrm{S} & \to & YT \mid \mathrm{a} \mid \mathrm{aa} & X\mathrm{a} & \to & \mathrm{aa}X \\
Y & \to & XY \mid \mathrm{aa} & X\mathrm{a}T & \to & \mathrm{aa}T \\
 & & & X\mathrm{aa}T & \to & \mathrm{aaaa}
\end{array}
$$

$\mathrm{L}(G_4) = \{\mathrm{a}^{2^n} \mid n \geqslant 0\}$

Definition

Eine formale Sprache L heißt

- regulär (vom Typ 3)
  wenn $\exists$ rechtslineare Grammatik G, $L = \mathrm{L}(G)$
- kontextfrei (vom Typ 2)
  wenn $\exists$ kontextfreie Grammatik G, $L = \mathrm{L}(G)$
- kontextsensitiv (vom Typ 1)
  wenn $\exists$ kontextsensitive Grammatik G, $L = \mathrm{L}(G)$

Bemerkung

- formale Sprache $L \subseteq \Sigma^*$
- Grammatik ist endliche Beschreibung von L
- Art der Beschreibung bestimmt Typ der Sprache

Definition

Eine formale Sprache L heißt

- beschränkt wenn $\exists$ beschränkte Grammatik G, $L = \mathrm{L}(G)$
- rekursiv aufzählbar (vom Typ 0)
  wenn $\exists$ Grammatik G, $L = \mathrm{L}(G)$

Satz (Chomsky-Hierarchie)

Es gelten die folgenden Inklusionen

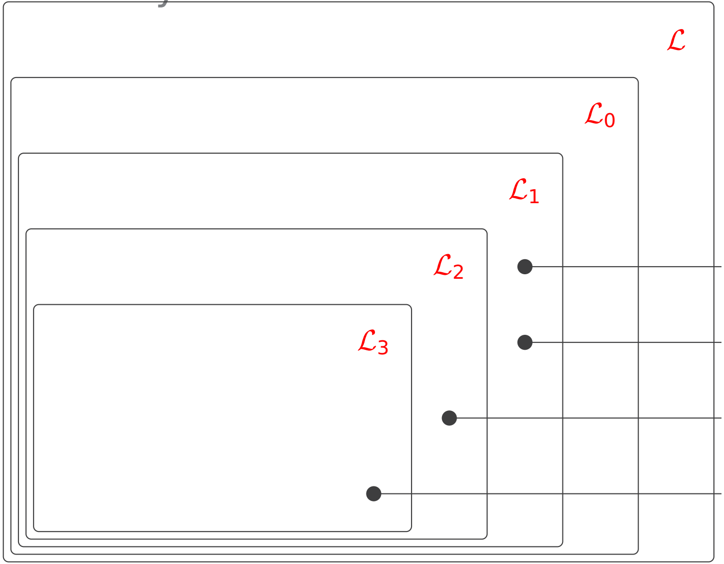
$$\mathcal{L}_3 \subsetneq \mathcal{L}_2 \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_0 \subsetneq \mathcal{L}$$

- $\mathcal{L}_i$ *die Klasse der Sprachen von Typ* i
- $\mathcal{L}$ *Klasse der formalen Sprachen*

Satz

Eine Sprache L ist kontextsensitiv gdw. L beschränkt ist

Chomsky-Hierarchie



$\mathrm{L}(G_4) = \{\mathrm{a}^{2^n} \mid n \geqslant 0\}$

$\mathrm{L}(G_3) = \{\mathrm{a}^n\mathrm{b}^n\mathrm{c}^n \mid n \geqslant 1\}$

$\mathrm{L}(G_2) =$ „Klammerausdrücke“

$\mathrm{L}(G_1) = \{0, 1\}^+$

Feedback zum Typ einer Formalen Sprache

Wie sollte man denn die Sprachen beschreiben? In natürlicher Sprache ist es unexakt. Und wenn man auf eine Menge der Sprache kommt: Wie kann man dann sicher sein, dass es sich bei der erzeugten Sprache tatsächlich um diese Menge handelt? Müsste man das beweisen? Wenn ja: Wie?

Zudem ist es mir schwer gefallen, „leichtere“ Grammatiken für gegebene Sprachen zu finden. Auch hier: Wie kann man sicher sein, dass eine gegebene Grammatik tatsächlich die gesamte Sprache erzeugt? Und wie kann man sicher sein, dass die gefundene Grammatik tatsächlich die „minimalste“ ist?