1 Wiederholung

Eigenschaft	CH3	CH2
Berechnungsmodell (alg. Umwandlung)	DFA, NFA, ε-NFA, (RLG), regulärer Ausdruck	PDA, CFG
Entscheidbarkeit	alles	teilweise
Abschluss	alles	teilweise
Ist L CHX?	PL: L CH3 $\Rightarrow \dots$ MN: L CH3 $\Leftrightarrow \equiv_L$-Index endl.	PL: L CH2 $\Rightarrow \dots$

Warum Sprachen? formales (= computerlesbares) Modell
In idealer Welt können Computer alles automatisch erledigen und terminieren immer.
Praxis: klappt nur für einfachste Probleme
CH3: endlicher Speicher, Suchen mit regulären Ausdrücken, Compilerbau
CH2: mächtiger, viele positive Eigenschaften, aber nichtdeterministisch, unbeschränkter (aber endlicher) Speicher, Modellierung von Eingabesprachen (Parser, Programmiersprache)

2 Kontextfreie Sprachen und Kellerautomaten (S. 37-70)

§1 Kontextfreie Grammatiken (S. 38-43)

Def. 1.1 Eine *kontextfreie Grammatik* ist ein 4-Tupel $G = (N, T, P, S)$ mit

- N Alphabet von *Nichtterminalsymbolen* (Großbuchstaben),
- T Alphabet von *Terminalsymbolen* (Kleinbuchstaben) mit $N \cap T = \emptyset$,
- $S \in N$ *Startsymbol*, und
- $P \subseteq N \times (N \cup T)^*$ endliche Menge von *Produktionen* oder *Regeln*.

Wir definieren die *Ableitungsrelation* $\vdash_G$ auf $(N \cup T)^*$:

- $uAx \vdash_G uvx$ wenn $A \to v$.
- $\vdash_G^n$, $\vdash_G^*$
- Die *von G erzeugte Sprache* ist $L(G) = \{w \in T^* \mid S \vdash^* w\}$.

Def. 1.2 Eine *Sprache* $L \subseteq T^*$ heißt *kontextfrei*, falls es eine kontextfreie Grammatik G mit $L = L(G)$ gibt.
Es gibt Nichtdeterminismus an zwei Stellen:

(a) die Wahl der Regel

(b) die Wahl des nächsten Symbols

Der zweite Nichtdeterminismus kann folgendermaßen eliminiert werden:

(a) immer das linkeste Nichtterminal ersetzen (*Linksableitung*, Def. 1.3)

(b) betrachte einen *Ableitungsbaum*

Def. 1.4 Ein *Ableitungsbaum* von A nach w in G ist ein Baum mit folgenden Eigenschaften:

(i) Jeder Knoten ist mit einem Symbol aus $N \cup T \cup \{\varepsilon\}$ beschriftet. Die Wurzel ist mit A beschriftet und jeder innere Knoten ist mit einem Symbol aus N beschriftet.

(ii) Wenn ein mit B beschrifteter innerer Knoten k Nachfolgeknoten besitzt, die in der Reihenfolge von links nach rechts mit den Symbolen $\beta_1, \ldots, \beta_k$ beschriftet sind, dann gilt

  a) $k = 1$ und $\beta_1 = \varepsilon$ und $B \to \varepsilon \in P$
  oder

  b) $k \geq 1$ und $\beta_1, \ldots, \beta_k \in N \cup T$ und $B \to \beta_1 \ldots \beta_k \in P$.

(iii) Das Wort w entsteht, indem man die Symbole an den Blättern von links nach rechts konkateniert.

Def. 1.5

(i) Eine kontextfreie *Grammatik* G heißt *eindeutig*, wenn es zu jedem Wort $w \in T^*$ höchstens einen Ableitungsbaum bzw. höchstens eine Linksableitung von S nach w in G gibt. Andernfalls heißt G *mehrdeutig*.

(ii) Eine kontextfreie *Sprache* $L \subseteq T^*$ heißt *eindeutig*, wenn es eine eindeutige kontextfreie Grammatik G mit $L = L(G)$ gibt. Andernfalls heißt L *(inhärent) mehrdeutig*.

Es ist *unentscheidbar*, ob eine gegebene CFG mehrdeutig ist. Es gibt eindeutige, nichtdeterministische Grammatiken (z.B. gerade Palindrome).

rechtslineare Grammatik (S. 93f): $P \subseteq N \times (T \cdot N \cup \{\varepsilon\})$

§4 Abschlusseigenschaften (S. 59-60)

Satz 4.1 Die Klasse der kontextfreien Sprachen ist abgeschlossen unter den Operationen

(i) Vereinigung, CFG

(ii) Konkatenation, CFG

(iii) Iteration, CFG

(iv) Durchschnitt mit regulären Sprachen. später

Dagegen ist die Klasse der kontextfreien Sprachen *nicht* abgeschlossen unter den Operationen

(v) Durchschnitt, $a^m b^n c^n \cap a^n b^n c^m$

(vi) Komplement. (i), (v), de Morgan

§5 Transformation in Normalformen (S. 61-63)

Def. 5.1 Eine *ε-Produktion* ist eine Regel der Form $A \to \varepsilon$. Eine kontextfreie Grammatik G heißt *ε-frei*, falls es in G

(i) entweder überhaupt keine ε-Produktion gibt

(ii) oder nur die ε-Produktion $S \to \varepsilon$ und S dann nicht auf der rechten Seite irgendeiner Produktion in G auftritt.

Satz 5.2 Jede kontextfreie Grammatik lässt sich in eine äquivalente ε-freie Grammatik transformieren.
Für alle Regeln $A \to \varepsilon$ füge für A auf einer rechten Seite $B \to w_1 A w_2$ die Regel $B \to w_1 w_2$ hinzu und entferne zum Schluss alle Regeln $A \to \varepsilon$. Für das Startsymbol muss eine Extraregel eingefügt werden.
Bsp.: $S \to A, A \to \varepsilon \mid aAbAc$
Es gibt $A \to \varepsilon$, also erhalte:
$S \to \varepsilon \mid aAbAc \mid abAc \mid aAbc \mid abc, A \to \varepsilon \mid aAbAc \mid abAc \mid aAbc \mid abc$
Entferne $S \to \varepsilon$ und $A \to \varepsilon$. Die erste Regel wird durch $S' \to \varepsilon \mid S$ ersetzt für neues Startsymbol S'.

Def. 5.3 Eine kontextfreie Grammatik G ist in *Chomsky-Normalform*, wenn folgendes gilt:

- G ist ε-frei (so dass höchstens $S \to \varepsilon \in P$ erlaubt ist),

- jede Produktion in P anders als $S \to \varepsilon$ ist von der Form

$$A \to a \quad \text{oder} \quad A \to BC,$$

wobei $A, B, C \in N$ und $a \in T$ sind.

Satz. 5.4 Jede kontextfreie Grammatik lässt sich in eine äquivalente Grammatik in Chomsky-Normalform transformieren.