Grundlagen der Theoretischen Informatik

4. Kellerautomaten und kontextfreie Sprachen (I)

9.05.2018

Viorica Sofronie-Stokkermans

e-mail: sofronie@uni-koblenz.de

Bis jetzt

1. Motivation
2. Terminologie
3. Endliche Automaten und reguläre Sprachen
4. Kellerautomaten und kontextfreie Sprachen
5. Turingmaschinen und rekursiv aufzählbare Sprachen
6. Berechenbarkeit, (Un-)Entscheidbarkeit
7. Komplexitätsklassen P und NP

Jetzt

1. Motivation
2. Terminologie
3. Endliche Automaten und reguläre Sprachen
4. Kellerautomaten und kontextfreie Sprachen
5. Turingmaschinen und rekursiv aufzählbare Sprachen
6. Berechenbarkeit, (Un-)Entscheidbarkeit
7. Komplexitätsklassen P und NP

Kellerautomaten und kontextfreie Sprache

Inhalt

- Die von **Kellerautomaten** (Push-Down-Automaten, PDAs) erkannten Sprachen sind genau die vom Typ 2 (**kontextfrei**).
- **Normalformen** für kontextfreie Grammatiken.
- **Pumping-Lemma** für kontextfreie Sprachen.
- Effiziente Algorithmen für Probleme über PDAs

Ableitungsbäume

Zur Erinnerung: kontextfreie Grammatiken

- Kontextfreie Regel:
  Eine Variable wird durch ein Wort ersetzt
  (egal in welchem Kontext die Variable steht)
- Es wird eine **einzelne** Variable ersetzt.
- Das Wort in der Conclusio kann Variablen und Terminale in **beliebiger Mischung** enthalten.

Zur Erinnerung: kontextfreie Grammatiken

Definition (Kontextfreie Grammatik)
Eine Grammatik $G = (V, T, R, S)$ heißt **kontextfreie** gdw

$$\forall (P \to Q) \in R \;\; \big(P \in V \text{ und } Q \in (V \cup T)^*\big)$$

Das heißt, bei jeder Regelanwendung:

- Links eine **einzelne Variable**
- Die Prämisse macht keine Aussage, was der Kontext dieser Variablen ist („kontextfrei“)
- Rechts steht etwas beliebiges

Zur Erinnerung: kontextfreie Sprachen

Beispiel.

- $\{a^n b^n \mid n \in \mathbb{N}\}$
- $\{a^n b a^n \mid n \in \mathbb{N}\}$
- $\{ww^R \mid w \in \{a, b\}^*\}$

Ableitungsbäume

Definition (Ableitungsbaum zu einer Grammatik)

Sei $G = (V, T, R, S)$ eine kontextfreie Grammatik.

Ein **Ableitungsbaum (parse tree)** zu G ist ein angeordneter Baum $B = (W, E, v_0)$

Zudem muss gelten:

- Jeder Knoten $v \in W$ ist mit einem Symbol aus $V \cup T \cup \{\varepsilon\}$ markiert.
- Die Wurzel v_0 ist mit S markiert.
- Jeder innere Knoten ist mit einer Variablen aus V markiert.
- Jedes Blatt ist mit einem Symbol aus $T \cup \{\varepsilon\}$ markiert.
- Ist $v \in W$ ein innerer Knoten mit Söhnen $v_1, \dots, v_k$ in dieser Anordnung und ist A die Markierung von v und A_i die Markierung von v_i, dann ist $A \to A_1 \dots A_k \in R$.
- Ein mit ε markiertes Blatt hat keinen Bruder (denn das entspräche einer Ableitung wie $A \to ab\varepsilon Bc$).

Ableitungsbäume

Ablesen eines Wortes vom Ableitungsbaum

Wenn Wort w von Grammatik G erzeugt wird,
dann gibt es einen Ableitungsbaum mit den Buchstaben von w
als Blätter von links nach rechts.

Merke:

Die Blätter eines Ableitungsbaumes sind angeordnet.

Es gibt eine Ordnung unter den Söhnen eines Knotens.

Ableitungsbäum

Definition

Seien b_1, b_2 Knoten. Dann:

$$b_1 < b_2 \underline{\text{ gdw }} b_1,\ b_2 \text{ sind Brüder, und } b_1 \text{ liegt "links" von } b_2,$$
$$\text{oder } \exists v, v_1, v_2 \in W \quad v \to v_1,\ v \to v_2,\ v_1 < v_2$$
$$\text{und } v_i \text{ ist Vorfahre von } b_i \text{ für } i \in \{1, 2\}.$$

Definition

Sei $\{b_1, \ldots, b_k\}$ die Menge aller Blätter in B mit $b_1 < \ldots < b_k$, und sei A_i die Markierung von b_i.
Dann heißt das Wort $A_1 \ldots A_k$ die **Front** von B.

Ableitungsbäume

Theorem.
Sei $G = (V, T, R, S)$ eine kontextfreie Grammatik.
Dann gilt für $w \in T^*$:

$(S \Longrightarrow_G^* w)$ <u>gdw</u> Es existiert ein Ableitungsbaum zu G mit Front w.

Beweis. Einfach aus den Definitionen.

Ableitungsbäume: Beispiel

Grammatik für die Menge aller aussagenlogischen Formeln über den Variablen $\{x, x_0, x_1, x_2, \ldots\}$:

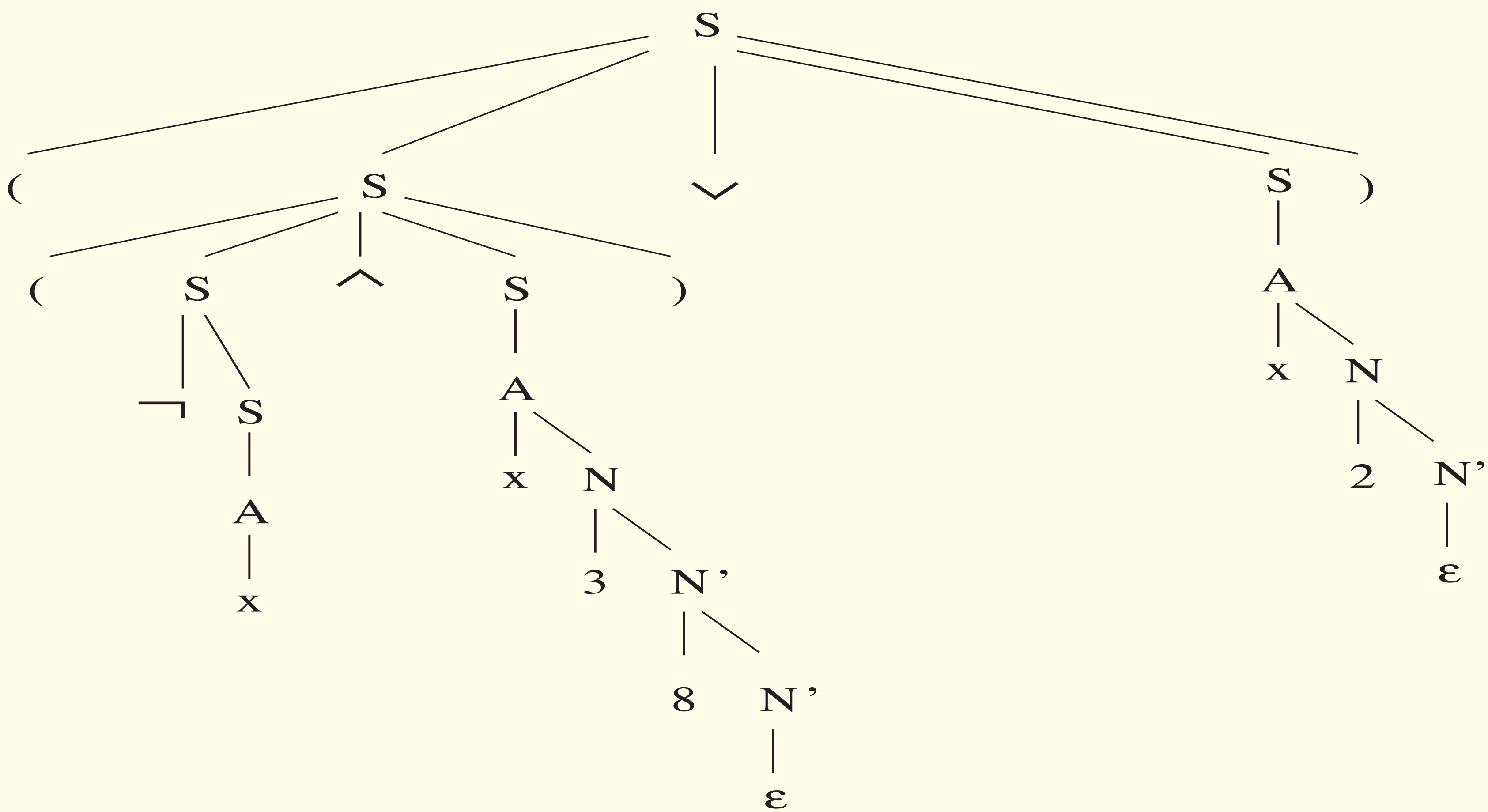
$$G = (\{S, A, N, N'\}, \{x, 0, \ldots, 9, (,), \wedge, \vee, \neg\}, R, S)$$

mit der Regelmenge

$$\begin{aligned} R = \{S & \rightarrow (S \ \wedge \ S) \mid (S \ \vee \ S) \mid \neg S \mid A \\ A & \rightarrow x \mid xN \\ N & \rightarrow 1N' \mid 2N' \mid \ \ldots \ \mid 9N' \mid 0 \\ N' & \rightarrow 0N' \mid 1N' \mid \ \ldots \ \mid 9N' \mid \varepsilon\} \end{aligned}$$

Ableitungsbäume: Beispiel

Ableitungsbaum für $((\neg x \wedge x38) \vee x2)$



Ableitungsbäume: Beispiel

Der Ableitungsbaum steht für viele **äquivalente** Ableitungen, darunter diese:

$$
\begin{array}{rcrc}
S \Rightarrow & & (S \vee S) & \Rightarrow \\
((S \ \wedge \ S) \vee S) & \Rightarrow & ((\neg S \ \wedge \ S) \vee S) & \Rightarrow \\
((\neg A \ \wedge \ S) \vee S) & \Rightarrow & ((\neg x \ \wedge \ S) \vee S) & \Rightarrow \\
((\neg x \ \wedge \ A) \vee S) & \Rightarrow & ((\neg x \ \wedge \ xN) \vee S) & \Rightarrow \\
((\neg x \ \wedge \ x3N') \vee S) & \Rightarrow & ((\neg x \ \wedge \ x38N') \vee S) & \Rightarrow \\
((\neg x \ \wedge \ x38) \vee S) & \Rightarrow & ((\neg x \ \wedge \ x38) \vee A) & \Rightarrow \\
((\neg x \ \wedge \ x38) \vee xN) & \Rightarrow & ((\neg x \ \wedge \ x38) \vee x2N') & \Rightarrow \\
((\neg x \ \wedge \ x38) \vee x2) & & &
\end{array}
$$

Links- und Rechtsableitung

Definition (Linksableitung)

Eine Ableitung

$$w_1 \Longrightarrow_G w_2 \Longrightarrow_G \ldots \Longrightarrow_G w_n$$

heißt **Linksableitung** falls w_{i+1} durch Ersetzen der linkesten Variable in w_i entsteht für alle $i < n$.

Die **Rechtsableitung** ist analog definiert.

Mehrdeutigkeit

Definition (Mehrdeutigkeit)

Eine cf-Grammatik G heißt **mehrdeutig**

<u>gdw</u>

es gibt ein Wort $w \in L(G)$,
zu dem es in G **zwei verschiedene Linksableitungen** gibt.

Eine Sprache $L \in \mathbf{L}_2$ heißt **inhärent mehrdeutig**

<u>gdw</u>

alle kontextfreien Grammatiken für L sind mehrdeutig.

Bemerkung

Eine Grammatik G ist mehrdeutig <u>gdw</u> :

es gibt zwei verschiedene Ableitungsbäume in G mit gleicher Front.

Mehrdeutigkeit: Beispiele

Eindeutige Grammatik für aussagenlogische Formeln:

$$\begin{array}{lcl} S & \to & (S \wedge S) \mid (S \vee S) \mid \neg S \mid A \\ A & \to & x \mid xN \\ N & \to & 1N' \mid 2N' \mid \ldots \mid 9N' \mid 0 \\ N' & \to & 0N' \mid 1N' \mid \ldots \mid 9N' \mid \varepsilon \end{array}$$

Mehrdeutigkeit: Beispiele

Eindeutige Grammatik für aussagenlogische Formeln:

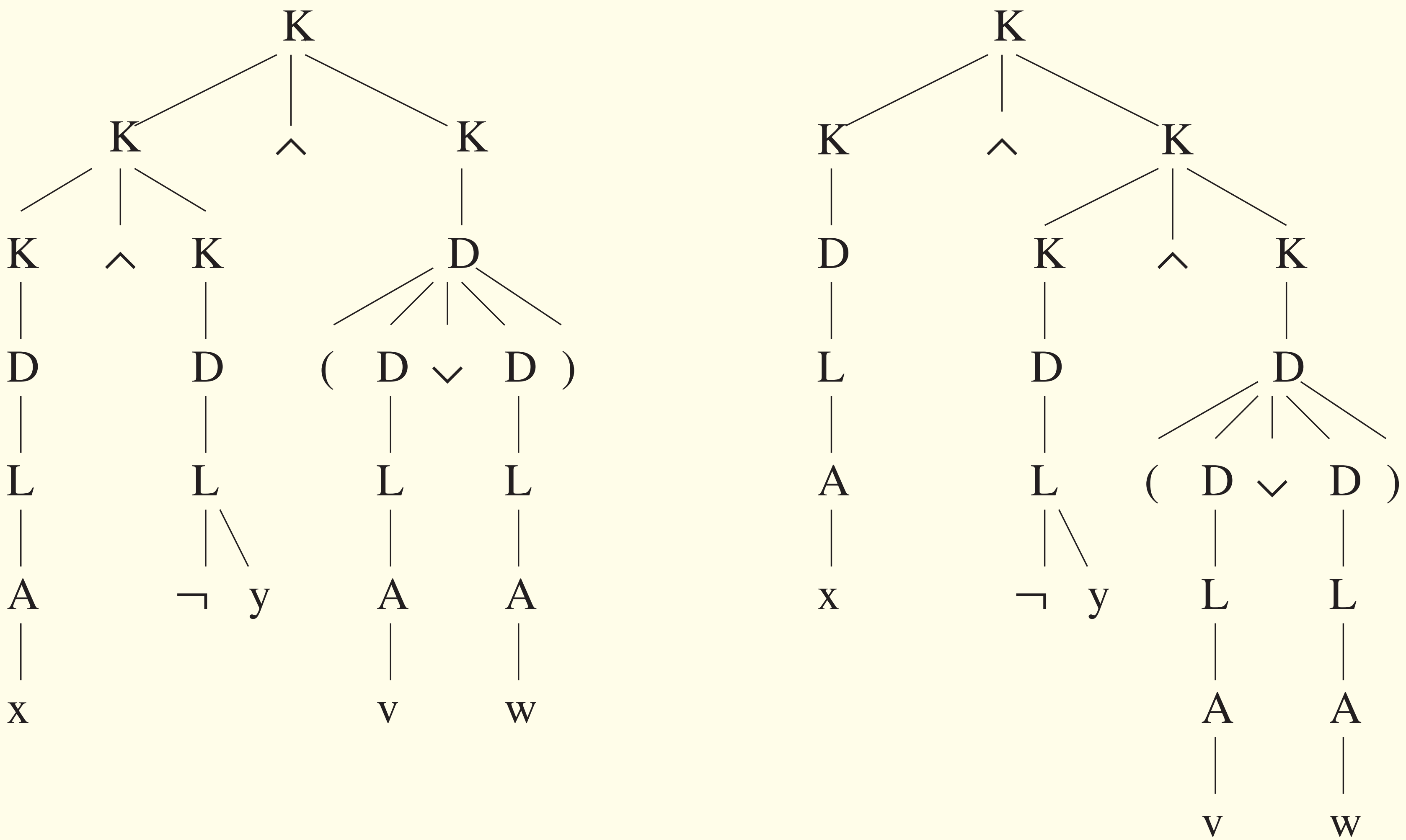
$$\begin{array}{lcl} S & \to & (S \ \wedge \ S) \mid (S \ \vee \ S) \mid \neg S \mid A \\ A & \to & x \mid xN \\ N & \to & 1N' \mid 2N' \mid \ldots \mid 9N' \mid 0 \\ N' & \to & 0N' \mid 1N' \mid \ldots \mid 9N' \mid \varepsilon \end{array}$$

Mehrdeutige Grammatik für aussagenlogische Formeln in KNF:

$$\begin{array}{lcl} K & \to & K \wedge K \mid D \\ D & \to & (D \vee D) \mid L \\ L & \to & \neg A \mid A \\ A & \to & v \mid w \mid x \mid y \mid z \end{array}$$

Regel mit Klammer-Ersparnis!

Mehrdeutigkeit: Beispiele



Mehrdeutigkeit: Beispiele

Inhärente Mehrdeutigkeit

Die Sprache

$$L := \{a^i b^j c^k \mid i = j \text{ oder } j = k\}$$

ist **inhärent mehrdeutig**.

Für einen Beweis siehe [Wegener “Theoretische Informatik” 1993 S.168-170], cf. Beispiel 6.1.7 Erk, Priese “Theoretische Informatik”.