

Einführung in die Computerlinguistik Abschlussprüfung

Laura Kallmeyer

Sommersemester 2019, Heinrich-Heine-Universität Düsseldorf

09.07.2019

Erlaubte Hilfsmittel: Eine Din-A4 Seite mit Notizen. Kein Taschenrechner.

Aufgabe 1 (8 Pkte) Betrachten Sie folgende CFG:

$G = \langle \{S, B, C, A, T\}, \{a, b, c\}, \{S \rightarrow CTB \mid CS \mid A, T \rightarrow CT \mid C, A \rightarrow a, B \rightarrow b \mid CB, C \rightarrow c\}, S \rangle$

1. Geben Sie alle Linksableitungen für das Wort $w = cccb$ an.
2. Welche Sprache wird von G erzeugt?
3. Ist G eine ambige CFG?

Lösung:

1. $S \Rightarrow CTB \Rightarrow cTB \Rightarrow cCTB \Rightarrow ccTB \Rightarrow ccCB \Rightarrow cccB \Rightarrow cccb$
 $S \Rightarrow CTB \Rightarrow cTB \Rightarrow cCB \Rightarrow ccB \Rightarrow ccCB \Rightarrow cccB \Rightarrow cccb$
 $S \Rightarrow CS \Rightarrow cS \Rightarrow cCTB \Rightarrow ccTB \Rightarrow ccCB \Rightarrow cccB \Rightarrow cccb$

3 Pkte

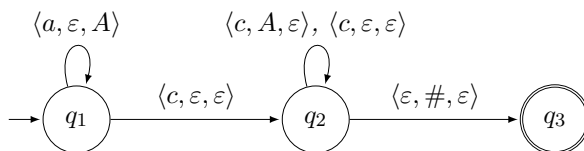
2. $\{c^n b \mid n \geq 2\} \cup \{c^n a \mid n \geq 0\}$

4 Pkte

3. Ja, siehe 1.

1 Pkt

Aufgabe 2 (7 Pkte) Betrachten Sie den Kellerautomaten (PDA) $M = \langle \{q_1, q_2, q_3\}, \{a, c\}, \{\#, A\}, \delta, q_1, \#, \{q_3\} \rangle$ mit der wie folgt festgelegten Überföhrungsfunktion δ :



1. Geben Sie eine mögliche Folge von Konfigurationen ("instantaneous descriptions") an, die der Automat durchlaufen könnte, wenn die Eingabe $w = aacccc$ abgearbeitet wird und der Automat anschließend im Endzustand sein soll. (Es gibt mehrere Möglichkeiten.) Geben Sie dies in einer Tabelle wie folgt an:

Zustand	Resteingabe	Stack (Top ist links)
q_1	$aacccc$	$\#$
q_1	$acccc$	$A\#$
\dots	\dots	\dots

2. Welche Sprache akzeptiert der Kellerautomat bei a) Akzeptanz mit Endzustand ($L(M)$) und bei b) Akzeptanz mit leerem Stack ($N(M)$)?

Lösung:

	Zustand	Resteingabe	Stack (Top ist links)
	q_1	aacccc	#
	q_1	acccc	A#
	q_1	cccc	AA#
1.	q_2	ccc	AA#
	q_2	cc	A#
	q_2	c	#
	q_2	ε	#
	q_3	ε	ε

3 Pkte

2. $L(M) = N(M) = \{a^n c^m \mid m \geq n + 1, n \geq 0\}$

4 Pkte

Aufgabe 3 (5 Pkte) Betrachten Sie nun folgende CFG:

$$G = \langle \{S, B, C\}, \{a, b, c\}, \{S \rightarrow CSB \mid CSa \mid a, B \rightarrow b \mid \varepsilon, C \rightarrow c\}, S \rangle.$$

Transformieren Sie diese in eine äquivalente CFG in Chomsky Normalform, die keine nutzlosen Symbole enthält. (Tipp: Sie müssen zunächst die ε -Produktion entfernen, anschließend Terminal aus gemischten rechten Seiten entfernen, und dann können Sie binarisieren.)

Lösung:

Ursprungsgrammatik:

$$G = \langle \{S, B, C\}, \{a, b, c\}, \{S \rightarrow CSB \mid CSa \mid a, B \rightarrow b \mid \varepsilon, C \rightarrow c\}, S \rangle.$$

Entfernen von $B \rightarrow \varepsilon$ führt zu

$$G' = \langle \{S, B, C\}, \{a, b, c\}, \{S \rightarrow CSB \mid CS \mid CSa \mid a, B \rightarrow b, C \rightarrow c\}, S \rangle. \quad (2 \text{ Pkte})$$

Anschließendes Entfernen von Terminalen in rechten Seiten länger als 1:

$$G'' = \langle \{S, B, C, A\}, \{a, b, c\}, \{S \rightarrow CSB \mid CS \mid CSA \mid a, A \rightarrow a, B \rightarrow b, C \rightarrow c\}, S \rangle. \quad (1 \text{ Pkt})$$

Binärisierung:

$$G''' = \langle \{S, B, C, A, X, Y\}, \{a, b, c\}, \{S \rightarrow CX \mid CS \mid CY \mid a, X \rightarrow SB, Y \rightarrow SA, A \rightarrow a, B \rightarrow b, C \rightarrow c\}, S \rangle. \quad (2 \text{ Pkte})$$

Aufgabe 4 (9 Pkte)

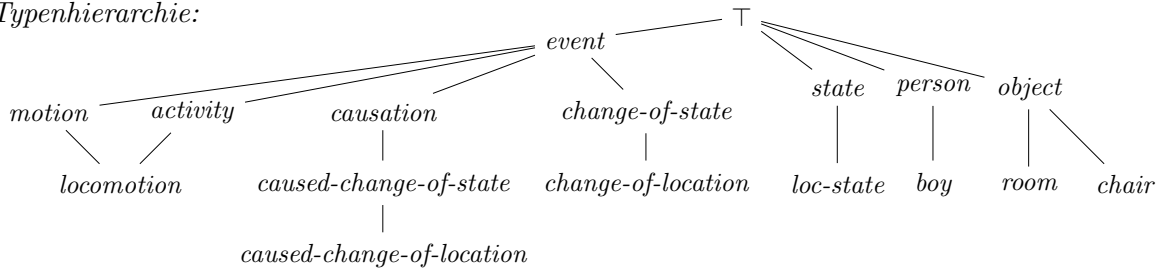
- Geben Sie für folgende als Attribut-Wert Matrix (AVM) notierte ungetypte Merkmalsstruktur den entsprechenden Graphen an.

$$\left[\begin{array}{l} \text{EXPERIENCER } \underline{1} \left[\text{NAME } \textit{Otto} \right] \\ \text{CAUSE} \left[\begin{array}{l} \text{AGENT } \left[\text{NAME } \textit{Hans} \right] \\ \text{THEME } \underline{1} \end{array} \right] \\ \text{EFFECT} \left[\text{EXPERIENCER } \underline{1} \left[\text{ANIMATE } \textit{yes} \right] \right] \end{array} \right]$$

- Betrachten Sie folgende getypte Merkmalsstrukturen:

$$S_1 = \left[\begin{array}{l} \textit{causation} \\ \text{CAUSE} \left[\begin{array}{l} \textit{activity} \\ \text{ACTOR} \quad \boxed{1} \textit{boy} \\ \text{THEME} \quad \boxed{2} \textit{chair} \\ \text{DIRECTION} \quad \boxed{3} \textit{room} \end{array} \right] \\ \text{ACTOR} \quad \boxed{1} \\ \text{EFFECT} \left[\begin{array}{l} \textit{change-of-location} \\ \text{THEME} \quad \boxed{2} \\ \text{RESULT-STATE} \left[\begin{array}{l} \textit{state} \\ \text{LOCATION} \quad \boxed{3} \end{array} \right] \end{array} \right] \end{array} \right] S_2 = \left[\begin{array}{l} \textit{caused-change-of-state} \\ \text{CAUSE} \quad \textit{motion} \\ \text{EFFECT} \left[\begin{array}{l} \textit{change-of-state} \\ \text{THEME} \quad \boxed{1} \textit{object} \\ \text{RESULT-STATE} \left[\begin{array}{l} \textit{loc-state} \\ \text{THEME} \quad \boxed{1} \\ \text{LOCATION} \quad \textit{room} \end{array} \right] \end{array} \right] \\ \text{THEME} \quad \boxed{1} \end{array} \right]$$

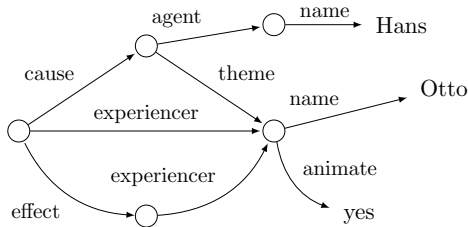
Typenhierarchie:



Berechnen Sie $S_1 \sqcup S_2$. Berücksichtigen Sie dabei auch neue Typen, die sich eventuell ergeben. Falls die Unifikation nicht möglich ist, erklären Sie, warum dies so ist.

Lösung:

1.



4 Pkte

$$2. S_1 \sqcup S_2 = \left[\begin{array}{l} \textit{caused-change-of-state} \\ \text{CAUSE} \left[\begin{array}{l} \textit{locomotion} \\ \text{ACTOR} \quad \boxed{1} \textit{boy} \\ \text{THEME} \quad \boxed{2} \textit{chair} \\ \text{DIRECTION} \quad \boxed{3} \textit{room} \end{array} \right] \\ \text{ACTOR} \quad \boxed{1} \\ \text{EFFECT} \left[\begin{array}{l} \textit{change-of-location} \\ \text{THEME} \quad \boxed{2} \\ \text{RESULT-STATE} \left[\begin{array}{l} \textit{loc-state} \\ \text{LOCATION} \quad \boxed{3} \\ \text{THEME} \quad \boxed{2} \end{array} \right] \end{array} \right] \\ \text{THEME} \quad \boxed{2} \end{array} \right]$$

5 Pkte

Aufgabe 5 (5 Pkte) Betrachten Sie nun folgende CFG:

$$G = \langle \{S, C\}, \{a, c, d\}, \{S \rightarrow aSC \mid Cc \mid d, C \rightarrow c\}, S \rangle$$

Geben Sie die erfolgreiche Folge von Stack, Resteingabe und Analysestack an, die sich bei einem Top-Down Parsing der Eingabe $w = aacccc$ ergibt. Die Produktionen seien dabei in der Reihenfolge ihres Auftretens oben nummeriert, $S \rightarrow aSC$ bekommt also z.B. den Index $\langle S, 1 \rangle$.

Lösung:

Resteingabe	predict stack	Analysestack
$aacccc$	S	–
$aacccc$	aSC	$\langle S, 1 \rangle$
$acccc$	SC	$a \langle S, 1 \rangle$
$acccc$	$aSCC$	$\langle S, 1 \rangle a \langle S, 1 \rangle$
$cccc$	SCC	$a \langle S, 1 \rangle a \langle S, 1 \rangle$
$cccc$	$CcCC$	$\langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
$cccc$	$ccCC$	$\langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
ccc	cCC	$c \langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
cc	CC	$cc \langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
cc	cC	$\langle C, 1 \rangle cc \langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
c	C	$c \langle C, 1 \rangle cc \langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
c	c	$\langle C, 1 \rangle c \langle C, 1 \rangle cc \langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$
ε	ε	$c \langle C, 1 \rangle c \langle C, 1 \rangle cc \langle C, 1 \rangle \langle S, 2 \rangle a \langle S, 1 \rangle a \langle S, 1 \rangle$

5 Pkte

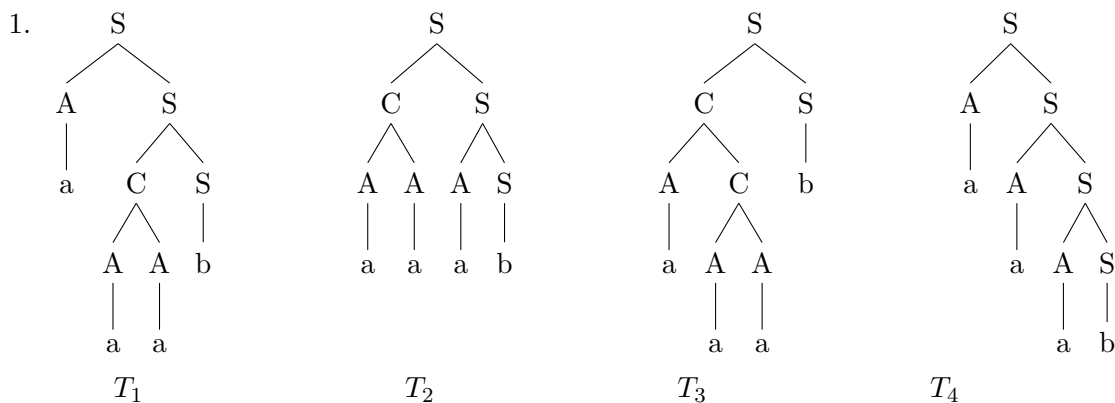
Aufgabe 6 (12 Pkte) Betrachten Sie folgende PCFG:

$G = \langle \{S, A, T, B\}, \{a, b\}, P, S \rangle$ wobei P folgende Produktionen enthält (die Zahlen davor sind die jeweiligen Produktionswahrscheinlichkeiten):

$$\begin{aligned} 0.1 : S &\rightarrow AS & 0.4 : S &\rightarrow CS & 0.5 : S &\rightarrow b \\ 0.2 : C &\rightarrow AC & 0.4 : C &\rightarrow AA & 0.6 : C &\rightarrow c & 1 : A &\rightarrow a \end{aligned}$$

- Geben Sie sämtliche Parsbäume für $aaab$ an und berechnen Sie ihre Wahrscheinlichkeiten.
- Welchen Wert haben die Inside Wahrscheinlichkeiten $\alpha_{S,2,4}$ und $\alpha_{C,2,4}$ bezogen auf diese PCFG und die Eingabe $aaaab$?
- Welchen Wert hat die Outside Wahrscheinlichkeit $\beta_{C,2,4}$ ebenfalls bezogen auf diese PCFG und die Eingabe $aaaab$?

Lösung:



4 Pkte

Wahrscheinlichkeiten:

$$T_1, T_2: 0.1 \cdot 0.4 \cdot 0.4 \cdot 0.5 \cdot 1 \cdot 1 \cdot 1 = 80 \cdot 10^{-4}$$

$$T_3: 0.4 \cdot 0.2 \cdot 0.4 \cdot 0.5 \cdot 1 \cdot 1 \cdot 1 = 160 \cdot 10^{-4}$$

$$T_4: 0.1 \cdot 0.1 \cdot 0.1 \cdot 0.5 \cdot 1 \cdot 1 \cdot 1 = 5 \cdot 10^{-4}$$

4 Pkte

2. $\alpha_{S,2,4} = 0$, da alle aus S abgeleiteten Ketten notwendig mit b enden, es gibt also keinen Baum mit Wurzel S und Span aaa .

$\alpha_{C,2,4} = 0.2 \cdot 0.4 \cdot 1 \cdot 1 \cdot 1 = 0.08$ (es gibt nur einen möglichen Baum, siehe den linken Teil in T_3 oben).

2 Pkte

3. Mögliche Ableitungen $S \xrightarrow{*} aCb$ kann man in T_1 und T_3 oben ablesen:

$$\beta_{C,2,4} = 0.4 \cdot 0.5 \cdot 0.1 + 0.2 \cdot 0.4 \cdot 0.5 = (20 + 40) \cdot 10^{-3} = 6 \cdot 10^{-2} \quad 2 \text{ Pkte}$$

Aufgabe 7 (4 Pkte) Betrachten Sie nochmals die PCFG aus Aufgabe 6. Geben Sie die CYK Chart (= Viterbi Chart) für ein Parsing von aab an, einschließlich Angabe von Produktionen und Zwischenlängen.

Eläutern Sie den Rechenwert für das Feld mit Anfangsposition 1 und Länge 3.

Lösung:

l				
3	$\langle 0.08 : S \rightarrow CS, 2 \rangle$			
2	$\langle 0.4 : C \rightarrow AA, 1 \rangle$	$\langle 0.05 : S \rightarrow AS, 1 \rangle$		
1	$\langle A, 1 \rangle$	$\langle A, 1 \rangle$	$\langle S, 0.5 \rangle$	
	1	2	3	i

Eintrag für S bei $i = 1$ und $l = 3$:

$$\begin{aligned} & \max\{0.4 \cdot 0.5 \cdot 0.4 \text{ (Produktion } S \rightarrow CS), 1 \cdot 0.05 \cdot 0.1 \text{ (Produktion } S \rightarrow AS)\} \\ & = \max\{8 \cdot 10^{-2}, 5 \cdot 10^{-3}\} = 0.08 \end{aligned}$$

4 Pkte