Prof. J. Křetínský — Technische Universität München
M. Weininger, T. Meggendorfer

KEINE ABGABE

Einführung in die theoretische Informatik

Sommersemester 2018 – Übungsblatt 6

Selbstständige Vorbereitung

Bereiten Sie sich auf die Tutorgruppen selbstständig vor, indem Sie die Aufgaben 6.1 bis 6.4 ansehen. Insbesondere Aufgabe 6.3 sollten Sie lösen, da Aufgabe 6.5 darauf aufbaut.
Stellen Sie außerdem sicher, dass Sie die Definitionen von Chomsky-Normalform und allen verwandten Begriffen verstanden haben.

AUFGABE 6.1. (*Wichtige Begriffe*) Stufe A

Überprüfen Sie, dass Sie die Folgenden Begriffe korrekt definieren können.

- Chomsky-Normalform
- Greibach-Normalform
- mehrdeutig
- inhärent mehrdeutig

AUFGABE 6.2. (*CFG finden*) Stufe B

Geben Sie eine CFG G an, die genau die Terme über einer Kleene Algebra $\langle K, +, \cdot, ^*, 0, 1\rangle$ erzeugt, die mittels der Konstanten $0, 1, a, b$ gebildet werden können. Die Multiplikation $\cdot$ soll dabei ausgeschrieben werden. Zum Beispiel sollte Ihre Grammatik $a, a+0, a+b+a, a\cdot a+a^*\cdot a+b\cdot a^*+(a^*)$ erzeugen können, jedoch nicht $+a()$, ^{*1} und auch nicht ab.

AUFGABE 6.3. (*Mehrdeutigkeit*) Stufe B

Betrachten Sie die folgende Grammatik G:

$$\begin{array}{rcl}
\mathsf{S} & \to & \text{N-P V-P} \\
\text{N-P} & \to & \text{C-N} \mid \text{C-N P-P} \\
\text{V-P} & \to & \text{C-V} \mid \text{C-V P-P} \\
\text{P-P} & \to & \text{P C-N} \\
\text{C-N} & \to & \text{A N} \\
\text{C-V} & \to & \text{V} \mid \text{V N-P} \\
\text{A} & \to & \texttt{a} \mid \texttt{the} \\
\text{N} & \to & \texttt{girl} \mid \texttt{boy} \mid \texttt{flower} \\
\text{V} & \to & \texttt{touches} \mid \texttt{likes} \mid \texttt{sees} \\
\text{P} & \to & \texttt{with}
\end{array}$$

(a) Zeigen Sie, dass die Grammatik mehrdeutig ist, indem Sie zwei verschiedene Ableitungen für das folgende Wort angeben:

`a girl touches a boy with a flower`

(b) Zeigen Sie, dass die Sprache $\mathsf{L}(\mathsf{G})$ endlich ist.
(c) Zeigen Sie, dass es keine endliche Sprache gibt, die inhärent mehrdeutig ist.
(d) Für Programmiersprachen sind mehrdeutige Grammatiken unerwünscht, da jedes Programm nur eine einzige, eindeutige Bedeutung haben soll. Diskutieren Sie anhand den verschiedenen Bedeutungsmöglichkeiten des in Aufgabenteil (a) angegebenen Satzes, warum man natürliche Sprache nur mit mehrdeutigen Grammatiken sinnvoll repräsentieren kann.

AUFGABE 6.4. (*Kontextfreie Grammatiken finden und beweisen*) Stufe B+D

In Aufgabe 4.6 haben Sie mithilfe des Pumping Lemmas für einige Sprachen gezeigt, dass sie nicht regulär sind. Nun wollen wir, wenn möglich, beweisen, dass die Sprachen kontextfrei sind. Geben Sie dazu für jede der Sprachen, die kontextfrei ist, eine kontextfreie Grammatik an. Beweisen Sie dann, dass Ihre Grammatik genau die angegebene Sprache erzeugt.
Bei Sprachen, die nicht kontextfrei sind, reicht die Feststellung, dass keine kontextfreie Grammatik gefunden werden kann.

(a) $\mathsf{L}_1 = \{w \in \{0,1\}^* \mid w = w^R\}$
(b) $\mathsf{L}_2 = \{w \in \{0,1\}^* \mid |w|_0 \geq |w|_1\}$
(c) $\mathsf{L}_3 = \{\varepsilon, \mathsf{a}, \mathsf{a}^{n\cdot m} \mid m \in \mathbb{N}, n \in \mathbb{N}, m > 1, n > 1\}$
(d) $\mathsf{L}_4 = \{\mathsf{a}^{6i}\mathsf{b}^{6i} \mid i \geq 0\}$
(e) $\mathsf{L}_5 = \{\mathsf{a}^{2^i} \mid i \geq 0\}$

AUFGABE 6.5. (*Mehrdeutigkeit*) Stufe B

Zeigen Sie, dass es keine reguläre Sprache gibt, die inhärent mehrdeutig ist.

AUFGABE 6.6. (*Chomsky-Normalform*) Stufe C

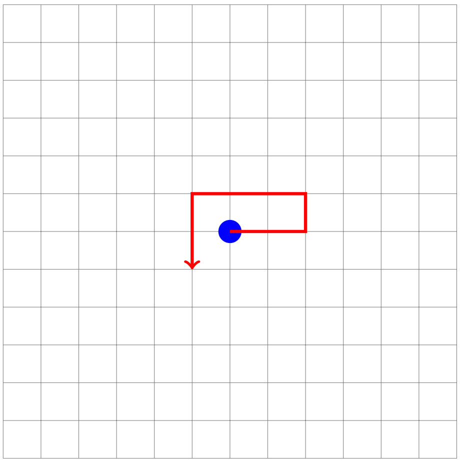
Die CFG G bestehe aus folgenden Produktionen über dem Alphabet $\Sigma = \{\mathsf{a}, \mathsf{b}\}$:

$$
\begin{array}{lcl}
S & \to & ASA \mid \mathsf{a}B \\
A & \to & B \mid S \mid CB \\
B & \to & \mathsf{b} \mid \varepsilon \\
C & \to & \mathsf{a}C \\
D & \to & \mathsf{a}SC\mathsf{b} \mid \mathsf{a}
\end{array}
$$

(a) Beschreiben Sie in eigenen Worten, wann ein Nichtterminal *nützlich* in einer Grammatik ist.
(b) Reduzieren Sie die Grammatik G auf die nützlichen Nichtterminale.
(c) Überführen Sie die reduzierte Grammatik dann in Chomsky-Normalform.
(d) Erklären Sie in eigenen Worten, wie Sie nach Überführen einer Grammatik in CNF überprüfen können, dass Sie keine Fehler gemacht haben.[1]

AUFGABE 6.7. (*Pfeilsprachen*) Stufe B – D

In dieser Aufgabe betrachten wir Sprachen, deren Worte Linienzüge in einem unendlichen zweidimensionalen Gitter von einem fixen Startpunkt aus beschreiben. Die folgende Grafik zeigt einen Ausschnitt aus dem Gitter:

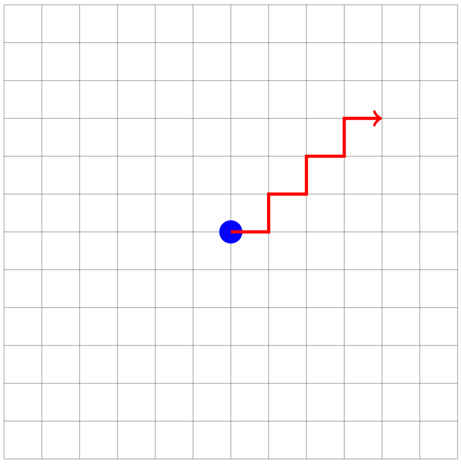
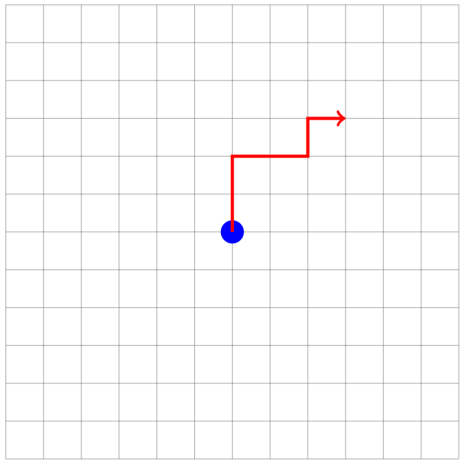


Wir haben Startpunkt blau markiert. Linienzüge beschreiben wir im Folgenden als eine Sequenz von Pfeilen, d.h. als Worte über dem Alphabet $\Sigma = \{\rightarrow, \leftarrow, \uparrow, \downarrow\}$. Die Pfeile beschreiben dabei (vom Startpunkt aus gesehen) einen ein Kästchen langen Schritt entlang des Gitters. Wir stellen daher den im Bild rot eingezeichnete Linienzug durch das Wort $w = \rightarrow\rightarrow\uparrow\leftarrow\leftarrow\leftarrow\downarrow\downarrow$ dar.

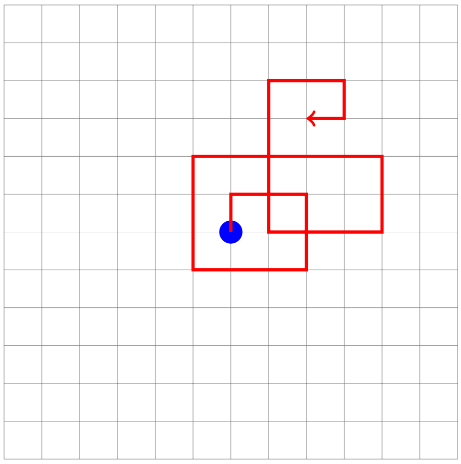
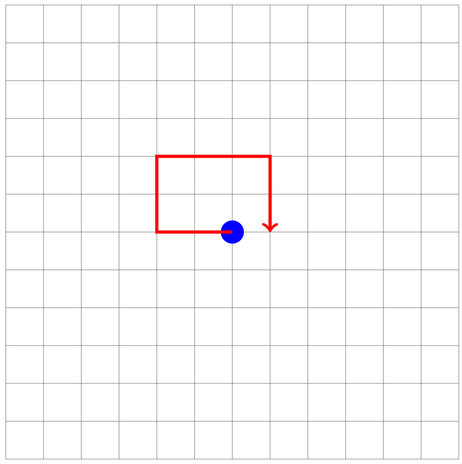
[1] Hier geht es nicht um einen formalen Beweis, dass die beiden Sprachen gleich sind, sondern um eine Strategie, wie Sie bei Aufgaben wie im vorherigen Aufgabenteil Fehler vermeiden.

(a) Betrachten Sie die folgenden natürlich sprachlichen Beschreibungen zusammen mit jeweils einem Beispiel, welches in der Sprache liegt (auf der linken Seite), und einem Beispiel, das kein Element der Sprache ist (auf der rechten Seite). *Geben Sie für jede der Sprachen eine formale Definition der Form* $\{w \in \Sigma^* \mid \dots\}$ *an.*[2]

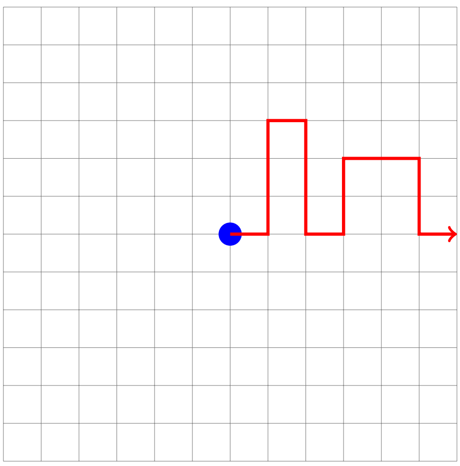
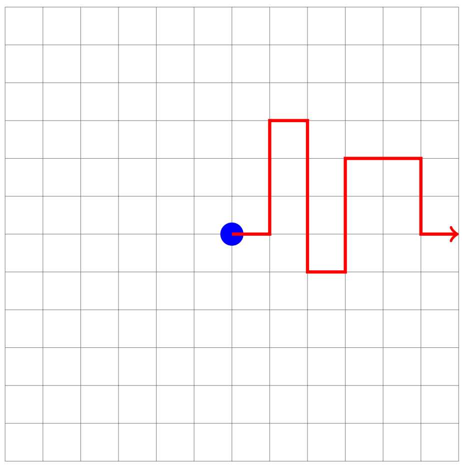
(i) die Sprache aller Treppen über dem Alphabet $\Sigma' = \{\rightarrow, \uparrow\}$

(ii) die Sprache aller im Uhrzeigersinn laufenden Spiralen über dem Alphabet Σ, die vom Startpunkt aus zuerst nach oben laufen

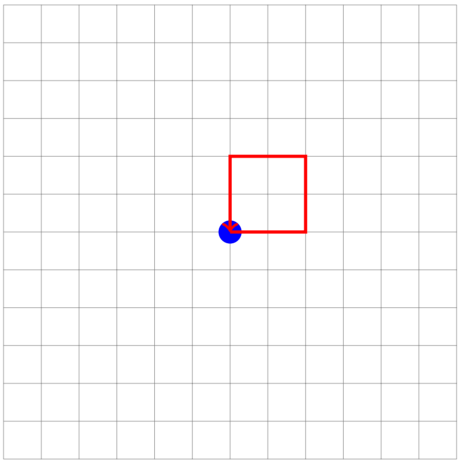



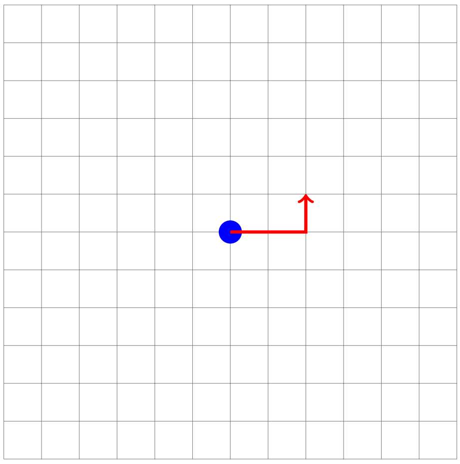
(iii) die Sprache aller “Skylines” über dem Alphabet $\Sigma'' = \{\rightarrow, \uparrow, \downarrow\}$

[2] Das heißt insbesondere, dass Sie in diesem Aufgabenteil keinen Automaten, keinen regulären Ausdruck, keine Grammatik oder ähnliches angeben sollen, die die Sprache beschreiben.

(iv) die Sprache aller Quadrate über dem Alphabet Σ





Hinweis: Die Sprachen sind mithilfe der Beispiele nicht eindeutig bestimmt! Ziel der Aufgabe ist es, die intuitive Beschreibung (z.B. "Sprache aller Quadrate") zusammen mit den Beispielen in eine möglichst allgemeine Sprachdefinition zu bringen.

(b) *Stellen Sie Vermutungen auf, ob die obigen Sprachen regulär, kontextfrei bzw. kontextsensitiv sind.* Begründen Sie Ihre Antwort möglichst anschaulich anhand des Beispiels. Überlegen Sie dazu, was es für Linienzüge heißt, regulär, kontextfrei bzw. kontextsensitiv zu sein.

(c) *Geben Sie zu jeder der Sprachen* L *aus Aufgabenteil (a) eine Grammatik* G *an.*

(d) Verallgemeinern Sie Aufgabenteil (a), iv) zu der Sprache aller Rechtecke, indem Sie zunächst Beispiele von Wörtern angeben, die in der Sprache liegen bzw. nicht in der Sprache liegen, und dann die Aufgabenteile (a) bis (c) auch für diese Sprache lösen.

AUFGABE 6.8.

Stufe D/E

(a) Erklären Sie, was man unter einer Familie von Grammatiken versteht und geben Sie ein eigenes Beispiel dazu an.

(b) Zeigen Sie, dass jede Grammatik G der Größe $\mathcal{O}(n)$, wenn man sie in CNF **entsprechend den Folien** übersetzt maximal nur quadratisch größer ($\mathcal{O}(n^2)$) werden kann.

(c) Zeigen Sie, dass es eine Familie von Grammatiken G_n der Größe $\mathcal{O}(n)$ gibt, die die folgende Behauptung erfüllt: *Überführt man* G_n *in CNF, indem man die Schritte der Folien umordnet nach (3), (4), (1), (2), dann ist die erzeugte Grammatik* G'_n *ebenfalls in CNF, jedoch exponentiell größer als* G.