Beispiellösung - Übungsblatt 12

Diskrete Modellierung

Wintersemester 2023/24

Prof. Dr. Martin Hoefer
Tim Koglin, Conrad Schecker

GOETHE UNIVERSITÄT FRANKFURT AM MAIN
Institut für Informatik
Algorithmen und Komplexität

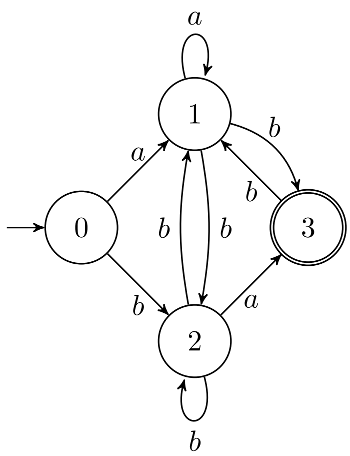
Übungsblatt 12

Ausgabe: 31.01.2024
Abgabe: —

Dieses Übungsblatt wird in der Vorlesung am 06.02. besprochen. Eine Abgabe erfolgt nicht. Die Aufgaben können nicht vorgerechnet werden. Bitte bereiten Sie eine Lösung vor.

Aufgabe 12.1 *NFAs und Potenzmengenkonstruktion*

a) Sei N der rechts abgebildete NFA über dem Alphabet $\Sigma := \{a, b\}$.



i) Welche der folgenden Wörter liegen in $L(N)$, welche nicht? Begründen Sie kurz.

$w_1 := bbb$ $\qquad$ $w_3 := abbbab$

$w_2 := abab$ $\qquad$ $w_4 := baba$

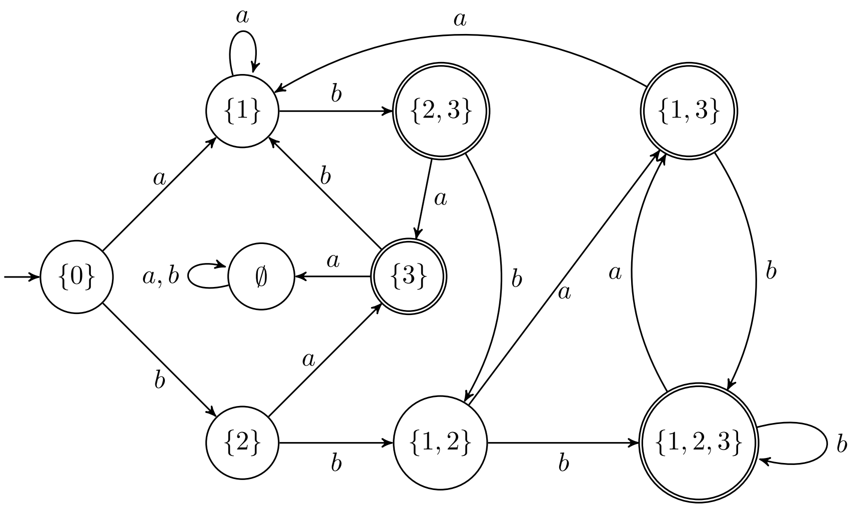
ii) Konstruieren Sie mittels Potenzmengenkonstruktion einen DFA D, der dieselbe Sprache wie N akzeptiert. Berücksichtigen Sie in D nur Zustände, die vom Startzustand von D aus erreichbar sind.

Lösung:

i)
- $w_1 = bbb \in L(N)$ über die Übergänge 0-2-1-3.
- $w_2 = abab \notin L(N)$. Beim Lesen des ersten a-s muss zum Zustand 1 übergegangen werden. Beim Lesen des ersten b-s kann dann entweder der Zustand 3 oder der Zustand 2 erreicht werden. In Zustand 3 stürzt der Automat beim Lesen des nächsten a-s dann direkt ab. In Zustand 2 führt das Lesen des a-s nur zu Zustand 3, von dem aus das Lesen des letzten b-s nur zu Zustand 1 führt. Damit kann nach Lesen des Wortes $abab$ kein Endzustand erreicht werden und somit gilt $abab \notin L(N)$.
- $w_3 = abbbab \in L(N)$ über die Übergänge 0-1-2-2-1-1-3.
- $w_4 = baba \notin L(N)$. Beim Lesen des ersten b-s muss zum Zustand 2 übergegangen werden. Beim Lesen des ersten a-s muss zum Zustand 3 übergegangen werden. Beim Lesen des zweiten b-s muss zum Zustand 1 übergegangen werden. Beim Lesen des zweiten a-s wird im Zustand 1 verblieben, der nicht akzeptierend ist.

ii) Die folgende Tabelle hilft bei der Potenzmengenkonstruktion: Die Übergangsfunktion für die neuen Zustände ergibt sich durch Vereinigung der entsprechenden Mengen in den Zeilen darüber.

q	$\delta(q,a)$	$\delta(q,b)$
$\{0\}$	$\{1\}$	$\{2\}$
$\{1\}$	$\{1\}$	$\{2,3\}$
$\{2\}$	$\{3\}$	$\{1,2\}$
$\{3\}$	$\emptyset$	$\{1\}$
$\emptyset$	$\emptyset$	$\emptyset$
$\{1,2\}$	$\{1,3\}$	$\{1,2,3\}$
$\{2,3\}$	$\{3\}$	$\{1,2\}$
$\{1,3\}$	$\{1\}$	$\{1,2,3\}$
$\{1,2,3\}$	$\{1,3\}$	$\{1,2,3\}$

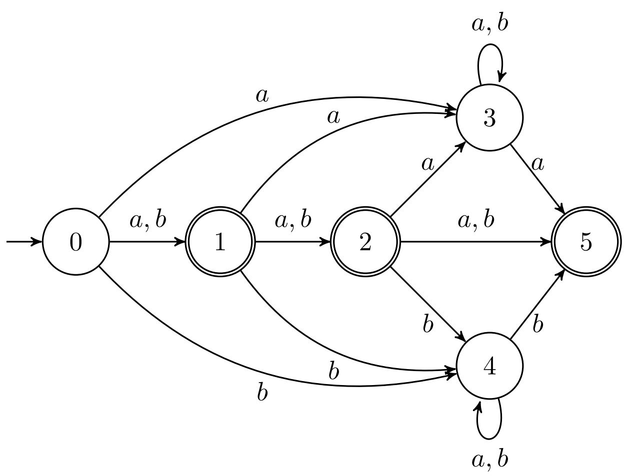


b) Sei L die Sprache der Wörter über dem Alphabet $\{a,b\}$, deren erstes, zweites oder drittes Symbol gleich dem letzten Symbol ist. Konstruieren Sie einen NFA für die Sprache L.

Lösung:

Beachte zunächst, dass bei allen Wörtern der Länge 1, 2 oder 3 immer das erste, zweite oder dritte Symbol gleich dem letzten Symbol ist. Ansonsten muss der NFA aus den ersten drei Symbolen ein Symbol, das identisch mit dem letzten Symbol ist, "erraten".

Der unten abgebildete NFA erkennt die Sprache L.
Zustand 3 kann nur erreicht werden, wenn das erste, zweite oder dritte Symbol ein a war. Genau dann, wenn das letzte Symbol ebenfalls ein a ist, können wir im akzeptierenden Zustand 5 (ohne vorzeitigen Absturz) enden. Analog gilt dies für Zustand 4 und das Symbol b.
Andernfalls können wir nur akzeptierend (ohne vorzeitigen Absturz) terminieren, wenn das Wort die Länge 1, 2 oder 3 hat.



Aufgabe 12.2 *Reguläre Ausdrücke*

a) Gegeben seien die regulären Ausdrücke

$$R_1 := ((aaa)^* \mid (bb)^*)^*,$$
$$R_2 := (aa|ab|ba)(a|b)^*|a|b|\varepsilon \text{ und}$$
$$R_3 := (b|ab^*a)^*|(a|ba^*b)^*$$

i) Welche der folgenden Wörter liegen in $L(R_1)$, $L(R_2)$ bzw. $L(R_3)$, welche nicht?

$w_1 := aabb$ $\quad w_2 := bbaaabb$ $\quad w_3 := ababab$

ii) Beschreiben Sie die Sprachen $L(R_1)$, $L(R_2)$ und $L(R_3)$ umgangssprachlich.

Lösung:

i) $\boldsymbol{L(R_1)}$ **:**

- w_1 ist nicht in $L(R_1)$, denn der vordere Block von a's hat Länge 2, die Länge ist also kein Vielfaches von 3.
- w_2 ist in $L(R_1)$, denn jede nicht erweiterbare Folge von a's bzw. b's hat eine Länge, die ein Vielfaches von 3 bzw. 2 ist.
- w_3 ist nicht in $L(R_1)$, da beispielweise der vordere Block von a's Länge 1 hat, also die Länge kein Vielfaches von 3 ist.

$\boldsymbol{L(R_2)}$ **:**

- w_1 ist in $L(R_2)$, da es nicht auf bb beginnt.
- w_2 ist nicht in $L(R_2)$, da es auf bb beginnt.
- w_3 ist in $L(R_2)$, da es nicht auf bb beginnt.

$\boldsymbol{L(R_3)}$ **:**

- w_1 ist in $L(R_3)$, denn es kommen gerade viele a's (und auch b's) vor.
- w_2 ist ebenfalls in $L(R_3)$, denn es kommen gerade viele b's vor.
- w_3 ist nicht in $L(R_3)$, denn es kommen ungerade viele a's und ungerade viele b's vor.

ii) $L(R_1)$ enthält genau die Worte über dem Alphabet $\{a, b\}$, in denen alle nicht erweiterbaren Folgen von a's bzw. b's eine Länge haben, die ein Vielfaches von 3 bzw. 2 ist.

$L(R_2)$ enthält genau die Wörter über dem Alphabet $\{a, b\}$, die nicht mit bb beginnen.

$L(R_3)$ enthält genau die Worte über dem Alphabet $\{a, b\}$, welche gerade viele a's oder gerade viele b's enthalten.

b) Geben Sie für die folgenden Sprachen je einen (möglichst kurzen) regulären Ausdruck an, der die Sprache beschreibt.

i) $L_1 := \{w \in \{a, b\}^* : w \text{ enthält genau ein } a\}$

ii) $L_2 := \{w \in \{0, 1\}^* : w \text{ enthält weder das Teilwort 01 noch das Teilwort 10}\}$

iii) $L_3 := \{w \in \{a, b\}^* : w \text{ enthält nicht das Teilwort } bab\}$

iv) $L_4 := \{w \in \{a, b\}^* : w \text{ enthält mindestens ein } a \text{ und ein } b.\}$

Sie müssen Ihre Antworten nicht begründen.

Lösung:

i) $R_1 := b^*ab^*$
Die Worte bestehen aus beliebig vielen b's (oder keinem) gefolgt von einem einzigen a und wiederum beliebig vielen b's (oder keinem).

ii) $R_2 := 0^*|1^*$
Es darf keinen Wechsel von 0 auf 1 oder von 1 auf 0 im Wort geben. Die Sprache enthält somit genau die Worte, die nur aus 0en oder nur aus 1en bestehen (auch das leere Wort gehört dazu).

iii) $R_3 := a^*(b|aaa^*)^*a^*$
Am Anfang oder Ende des Wortes können beliebig viele a's vorkommen. Mitten im Wort dürfen beliebig viele b's vorkommen, a's zwischen b's allerdings nur in Blöcken, die mindestens zwei lang sind.

iv) Jedes Wort in L_4 muss ein a und ein b enthalten. Es gibt zwei Möglichkeiten: Das Wort enthält erst ein a und irgendwann folgt ein b, oder es enthält erst ein b und irgendwann folgt ein a. Dieser naive Ansatz führt auf den regulären Ausdruck

$$R_4 = (a|b)^*a(a|b)^*b(a|b)^* \mid (a|b)^*b(a|b)^*a(a|b)^*.$$

Das lässt sich nach dem Distributivgesetz vereinfachen zu

$$R_4' = (a|b)^*\Big(a(a|b)^*b \Bigm| b(a|b)^*a\Big)(a|b)^*.$$

Weiter können wir uns überlegen, dass der Ausdruck $(a|b)^*$ am Anfang (bzw. am Ende, einer von beiden) überflüssig ist. Entweder das Wort beginnt mit einem a und wir müssen auf das b warten oder umgekehrt. Außerdem: Um nach einem a auf das b zu warten, brauchen wir nicht den Ausdruck $(a|b)^*$ zwischen a und b, sondern nur ein a^* – analog für den anderen Fall (erst b, dann a). Es ergibt sich der folgende vereinfachte Ausdruck:

$$R_4'' = (aa^*b \mid bb^*a)(a|b)^*.$$

Äquivalent dazu ist die folgende Betrachtung: Falls ein Wort $w \in \{a,b\}^*$ mindestens ein a und ein b enthält, dann muss es entweder das Teilwort ab oder das Teilwort ba enthalten. An irgendeiner Stelle muss es nämlich einen "Wechsel" geben, sonst würde w nur aus demselben Symbol bestehen (oder es würde $w = \varepsilon$ gelten). Entsprechend ergibt sich der folgende reguläre Ausdruck:

$$R_4''' = (a|b)^*(ab|ba)(a|b)^*$$

Aufgabe 12.3 *Kontextfreie Grammatiken*

a) Sei $G := (\Sigma, V, S, P)$ die kontextfreie Grammatik mit $\Sigma := \{a, b, c\}$, $V := \{S, T\}$ und

$$P := \{\, S \to aScc \mid T \\ T \to bT \mid \varepsilon \,\}.$$

Beschreiben Sie die Sprache $L(G)$ umgangssprachlich oder mathematisch.

Lösung:

$$L(G) = \{a^m b^n c^{2m} \,:\, m, n \in \mathbb{N}\}$$

b) Gegeben sei die Sprache

$$L_1 := \{a^m b^n \,:\, m, n \in \mathbb{N},\ n \le 2m\}.$$

Konstruieren Sie eine kontextfreie Grammatik $G_1 := (\Sigma, V, S, P)$ mit $L(G_1) = L_1$. Sie müssen Ihre Antwort nicht begründen.

Lösung:

Idee: mit jedem a kann kein, ein oder zwei b erzeugt werden.

$$\Sigma := \{a, b\}$$
$$V := \{S\}$$
$$P := \{\, S \to aS \mid aSb \mid aSbb \mid \varepsilon \,\}$$

c) Gegeben sei die Sprache

$$L_2 := \{w \in \{0,1\}^* : \text{für jedes Präfix } v \text{ von } w \text{ gilt: es gibt nicht mehr Nullen als Einsen in } v\}.$$

Konstruieren Sie eine kontextfreie Grammatik $G_2 := (\Sigma, V, S, P)$ mit $L(G_2) = L_2$. Begründen Sie kurz die Korrektheit Ihrer Grammatik.

Hinweis: Die Präfixe eines Wortes $w_1 w_2 \dots w_n \in \{0,1\}^n$ sind alle Worte $w_1 \dots w_i$ mit $0 \le i \le n$.

Lösung:

$$\Sigma := \{0, 1\}$$
$$V := \{S\}$$
$$P := \{\, S \to 1S0 \mid SS \mid 1 \mid \varepsilon \,\}$$

Die von der Grammatik erzeugten Worte erfüllen die angegebene Eigenschaft, denn mit jeder 0 wird auch eine 1 im Wortteil davor erzeugt.

Andererseits ist auch jedes Wort der Sprache ableitbar. Intuition: Wörter in L_2 mit genauso vielen 0en wie 1en können als Klammerausdrücke aufgefasst werden: 1en sind öffnende Klammern und 0en sind schließende Klammern. So kann man sich intuitiv überlegen, dass alle solchen Wörter mit den Regeln $S \to 1S0$, $S \to SS$ und $S \to \varepsilon$ gebildet werden können. Die übrigen Wörter in L_2 entstehen durch „einstreuen" weiterer 1en, was durch die Regel $S \to 1$ ermöglicht wird.

Formal kann man diesen zweiten Teil mit Hilfe der folgenden Eigenschaft zeigen. Für jedes Wort $w \in L_2$ der Länge mindestens 2 trifft einer der folgenden zwei Fälle zu.

- Fall 1: Es gilt $w = uv$ für zwei Wörter $u, v \in L_2$ die beide mindestens Länge 1 haben.
- Fall 2: Es gilt $w = 1u0$ für ein Wort $u \in L_2$.

Aufgrund dieser Eigenschaft lässt sich induktiv zeigen, dass jedes Wort $w \in L_2$ auch von der obigen Grammatik erzeugt wird, da die Regel $S \to SS$ induktiv Wörter konstruieren kann, für die Fall 1 zutrifft, während die Regel $S \to 1S0$ induktiv Wörter konstruieren kann, für die Fall 2 zutrifft.

Der Vollständigkeit halber hier noch der komplette Beweis für $L_2 \subseteq L(G_2)$. Wir beweisen induktiv folgende Aussage für alle $n \in \mathbb{N}$:

$$\text{Für alle } w \in \Sigma^* \text{ mit } |w| \geq n \text{ gilt: } w \in L_2 \implies w \in L(G_2).$$

Für $n = 0$ und $n = 1$ ist die Aussage korrekt, da die einzigen Wörter dieser Längen in L_2 die Wörter ε und 1 sind und diese von G_2 erzeugt werden können.

Nehme nun an, dass die Aussage für $n \in \mathbb{N}$ gilt und betrachte ein beliebiges Wort $w \in L_2$ mit $|w| = n+1$. Sei $w =: w_0 w_1 \ldots w_n$ mit $w_i \in \Sigma$ für alle $i \leq n$.

Fall 1: Es gibt eine Zerlegung $w = uv$ für zwei Wörter $u, v \in L_2$, die mindestens Länge 1 haben. Nach Induktionsvoraussetzung gilt dann $u, v \in L(G_2)$ und somit kann die Grammatik G_2 mit der Regel $S \to SS$ auch das Wort $w = uv$ erzeugen. Es gilt also $w \in L(G_2)$.

Fall 2: Für jede Zerlegung $w = uv$ mit $|u|, |v| \geq 1$ gilt: $u \notin L_2$ oder $v \notin L_2$. Nach Definition von L_2 gilt $w_0 = 1$ (sonst hätte das Präfix der Länge 1 mehr 0en als 1en). Außerdem muss $w_n = 0$ gelten, denn sonst ließe sich das Wort in $u = w_0 w_1 \ldots w_{n-1} \in L_2$ und $v = w_n \in L_2$ zerlegen. $\lightning$

Weiterhin können wir zeigen, dass der Mittelteil des Wortes in L_2 liegt, also $w_1 w_2 \ldots w_{n-1} \in L_2$: Angenommen dies wäre nicht der Fall. Dann gibt es ein Präfix $w_1 \ldots w_i$ dieses Wortes, das eine 0 mehr als 1en enthält. Somit hat $w_0 w_1 \ldots w_i = 1 w_1 w_2 \ldots w_i$ genauso viele 0en wie 1en. Das heißt aber: Hätte das restliche Teilwort $w_{i+1} w_{i+2} \ldots w_n$ ein Präfix, das mehr 0en als 1en enthält, dann hätte das Teilwort $w_0 w_1 \ldots w_j$ ebenfalls mehr 0en als 1en und somit wäre $w \notin L_2$. $\lightning$ Es ist also $w = uv$ mit $u = w_0 w_1 \ldots w_i$ und $v = w_{i+1} w_{i+2} \ldots w_n$ eine Zerlegung von w, die im aktuell betrachteten Fall ausgeschlossen ist. $\lightning$

Wir haben also gezeigt, dass $w = 1u0$ für ein Wort $u \in L_2$. Nach Induktionsvoraussetzung liegt u auch in $L(G_2)$ und damit erlaubt die Regel $S \to 1S0$, auch das Wort w abzuleiten. Somit gilt $w \in L(G_2)$.