4. Übungsblatt - Musterlösungen für Aufgabe 4.4

1. Allgemeines Vorgehen

Ausgabenstellung: Ist eine gegebene Sprache regulär oder kontextfrei? Begründen Sie!

Generelle Lösungsstrategie:

- Fall 1: Die Sprache ist regulär.
  Beweis durch
  - Konstruktion oder Beschreibung eines Automaten, der die Sprache erkennt **oder**
  - Nachweis, dass die Sprache nur endlich viele Wörter enthält. (Das geht natürlich nur, wenn die Sprache tatsächlich nur endlich viele Wörter enthält) **oder**
  - Angabe eines regulären Ausdrucks für die Sprache.
- Fall 2: Die Sprache ist kontextfrei (und nicht regulär).
  Beweis erfolgt in zwei Schritten:
  1. Nachweis, dass die Sprache nicht regulär ist, durch das Pumping Lemma.
  2. Nachweis, dass die Sprache kontexfrei ist, durch Angabe einer kontextfreien Grammatik.

  Hinweis: Die Menge der regulären Sprachen ist eine Teilmenge der kontextfreien Sprachen, d.h. jede reguläre Sprache ist auch kontextfrei. Die Menge der kontextfreien Sprachen ist wiederum eine Teilmenge aller generell möglichen Sprachen.

  Daher genügt es für die Kontextfreiheit einer Sprache nicht, zu zeigen, dass eine Sprache nicht regulär ist (sie könnte ja auch z.B. kontextsensitiv sein), man muss auch zeigen, dass sie tatsächlich „nur“ kontextfrei ist (durch Angabe einer geeigneten Grammatik.)

2. Pumping-Lemma Generelles Schema

Generelle Beweisidee: Man nimmt an, eine Sprache wäre regulär. Dann gibt es auch einen endlichen Automaten, der diese Sprache erkennt. Im Verlauf des Beweises zeigt man, dass es diesen Automaten nicht geben kann. Daher muss die Annahme falsch gewesen sein.

Im Detail: Beweis dass eine Sprache L_1 nicht regulär ist:

- Annahme L_1 ist regulär.
- Dann gibt es einen Automaten, der die Sprache erkennt. Dieser Automat hat eine bestimmte Anzahl von Zuständen, die Anzahl nennen wir k.
- Alle Wörter die mindestens so lang sind wie k, müssen einen Zustand im Automaten mehrfach durchlaufen, enthalten also eine Schleife.

- Man sucht sich jetzt ein Wort aus L_1 aus, das mindestens so lang ist wie k, und zeigt, dass dieses Wort an keiner Stelle eine Schleife enthalten kann. Wichtig: Es reicht tatsächlich **ein einziges Wort**. Dieses Wort muss man dafür unter Umständen geschickt auswählen.
- Dass das augewählte Wort tatsächlich mindestens die Länge k hat, erreicht man z.B., indem einer der Indices k ist, man wählt also z.B. $w = a^k b^k$.
- Dass dieses Wort **nirgendwo** eine Schleife enthalten kann, beweist man, indem man zeigt, dass für jeden möglichen Ort, an dem die Schleife sein könnte, das durch Aufpumpen entstandene Wort nicht mehr zu der Sprache gehört,, d.h. man muss eine Fallunterscheidung für die Lage der Schleife machen (die alle Möglichkeiten abdeckt!) und zeigen, dass **in jedem Fall** durch Pumpen ein Wort entsteht, das nicht mehr in der Sprache liegt.
- Wenn das Wort aber durch einen Automaten erkannt werden würde, müsste es eine Schleife enthalten. → also war die Annahme falsch. Also ist L_1 nicht regulär.

Hinweis für alle, die mehr wissen wollen: Mit der oben beschriebenen Version des Pumping Lemma, lässt sich in vielen, aber nicht in allen Fällen zeigen, dass eine Sprache nicht regulär ist. Eine stärkere Version des Pumping Lemmas, die sich auch für einige der Fälle anwenden lässt, die von der Version oben nicht abgedeckt werden, besagt, dass $|uv| \leq k$ sein muss, das heißt, dass es ausreicht, den aufpumpbaren Teil innerhalb der ersten k Zeichen des Wortes zu suchen. Ein Beispiel für eine Sprache, die sich mit dieser Version des Pumping Lemmas, aber nicht der Obigen, als nichtregulär zeigen lässt, ist $L = \{wc^*w^R | w \in \{a, b\}^*\}$.

Musterlösung für Aufgabe 4.4.

(a) $L_1 = \{wcw^R \; | w \in \{a, b\}^*\}$[1]

L_1 ist kontexfrei.

Beweis, dass L_1 nicht regulär ist, durch das Pumping Lemma:

Wir nehmen an L_1 wäre regulär. Dann gibt es einen endlichen Automaten, der L_1 erkennt. Die Anzahl der Zustände dieses Automaten sei k. Wir wählen jetzt das Wort $x = a^k ca^k$. x liegt in L_1, und ist offensichtlich länger als k. Dieses Wort muss irgendwo eine Schleife, also einen aufpumpbaren Teil enthalten, d.h. man kann es so in uvw zerlegen, dass für jede natürliche Zahl i auch $uv^i w$ zu L_1 gehört. Wo könnte dieser aufpumpbare Teil liegen?

- Fall 1: Der aufpumbare Teil v liegt komplett im Bereich des ersten a^k-Blocks. Dann würde aber $uv^2w = a^{k+|v|}ca^k$ mehr as im ersten Teil als im zweiten Teil enthalten und läge nicht mehr in L_1.

[1]Erläuterung: w^R ist die Spiegelung von w, d.h. es enthält die Zeichen von w in umgekehrter Reihenfolge. Worte von L_1 sind also z.B. *c*, *abcba*, *bbbaabacabaabbb*

- Fall 2: v enthält das c. Dann würde aber uv^2w zwei cs enthalten und läge damit nicht mehr in L_1.
- Fall 3: Der aufpumpbare Teil liegt komplett im Bereich des zweiten a^k-Blocks. Dann liegt analog zu Fall 1 uv^2w nicht mehr in L_1.

Unser Wort lässt sich also nicht so zerlegen, dass man den Mittelteil aufpumpen kann, also ist die Annahme, dass L_1 regulär ist, falsch.

Beweis, dass L_1 kontextfrei ist, durch Angabe einer kontexfreien Grammatik:

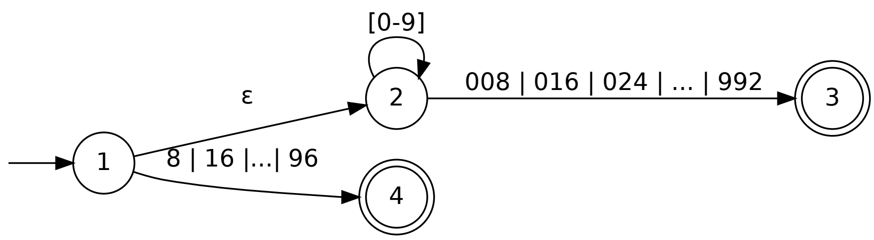
$S \rightarrow aSa$
$S \rightarrow bSb$
$S \rightarrow c$

(b) $L_2 = \{w|w \in \{0,1,2,3,4,5,6,7,8,9\}^* \text{ und } w \text{ ist durch 8 teilbar}\}$

L_2 ist regulär.

Beweis durch Angabe eines Automaten:

Eine Zahl ist genau dann durch 8 teilbar, wenn die letzten drei Ziffern der Zahl eine durch acht teilbare Zahl bilden. Ein endlicher Automat, der alle (endlich vielen) durch 8 teilbaren Suffixe $(000, 008, 016, 024, ..., 984, 992)$ und davor beliebige Ziffern erkennt, erkennt also alle durch 8 teilbaren Zahlen. Dieser Automat sieht z.B. folgendermaßen aus:



(c) $L_3 = \{a^*b^mcb^ma|m \in \mathbb{N}, m \geq 42\}$

L_3 ist kontexfrei.

Beweis, dass L_3 nicht regulär ist, durch das Pumping Lemma:

Wir nehmen an L_3 wäre regulär. Dann gibt es einen endlichen Automaten, der L_3 erkennt. Die Anzahl der Zustände dieses Automaten sei k. Wir wählen jetzt das Wort $x = b^lcb^la$ mit $l \geq 42$ und $l \geq k$. x liegt in L_3, und ist offensichtlich länger als k. Dieses Wort muss irgendwo eine Schleife, also einen aufpumpbaren Teil enthalten, d.h. man kann es so in uvw zerlegen, dass für jede natürliche Zahl i auch uv^iw zu L_3 gehört. Wo könnte dieser aufpumpbare Teil liegen?

- Fall 1: Der aufpumbare Teil v liegt komplett im Bereich des ersten b^k-Blocks. Dann würde aber $uv^2w = b^{k+|v|}cb^ka$ mehr bs im ersten Teil als im zweiten Teil enthalten und läge nicht mehr in L_3.

- Fall 2: v enthält das c. Dann würde aber uv^2w zwei cs enthalten und läge damit nicht mehr in L_3.
- Fall 3: Der aufpumpbare Teil liegt komplett im Bereich des zweiten b^k-Blocks. Dann liegt analog zu Fall 1 uv^2w nicht mehr in L_3.
- Fall 4: v enthält das a (und kein c). Dann würde aber uv^2w zwei as nach dem c enthalten und läge damit nicht mehr in L_3.

Unser Wort lässt sich also nicht so zerlegen, dass man den Mittelteil aufpumpen kann, also ist die Annahme, dass L_1 regulär ist, falsch.

Hinweis: Der „Trick“ bei dieser Aufgabe liegt darin, das Wort x so zu wählen, dass vor dem ersten b-Block keine as mehr liegen, die man sonst aufpumpen könnte.

Beweis, dass L_3 kontextfrei ist, durch Angabe einer kontexfreien Grammatik:

$$\begin{array}{ll} S \rightarrow & Ab^{42}Bb^{42}a \\ A \rightarrow & aA \\ A \rightarrow & \epsilon \\ B \rightarrow & bBb \\ B \rightarrow & c \end{array}$$

(d) $L_4 = \{a^m bbbc^{2m}c^* | m \in \mathbb{N}, m \leq 5\}$

L_4 ist regulär.

Beweis durch Beschreibung eines endlichen Automaten:

Man kann einen endlichen Automaten zu L_4 angeben, der zunächst die 5 Möglichkeiten $a^1b^3c^2$, $a^2b^3c^4$, $a^3b^3c^6$, $a^4b^3c^8$ oder $a^5b^3c^{10}$ erkennt und dann mit einer reflexiven Kante eine beliebige Anzahl von cs.

Beweis durch Angabe eine regulären Ausdrucks:

$L_4 = (abbbcc|aabbbcccc|aaabbbcccccc|aaaabbbcccccccc|aaaaabbbcccccccccc)c^*$