Lösungen für Aufgabenblatt 2 — 31.10.2019

Einführung in die Computerlinguistik (WS19/20) — Abgabe bis 9.11.2019

Aufgabe 1) Spezifikation von Zeichenketten mit regulären Ausdrücken **Punkte: 6**

Konstruieren Sie für die folgenden regulären Sprachen einen regulären Ausdruck, der genau die Zeichenketten der Sprache erkennt. Falls es Ihrer Meinung nach keine Lösung gibt, erklären Sie bitte Ihren Gedankengang.

1. Über dem Alphabet $\{a, b, c, d, e, f\}$ alle Zeichenketten, die mit c beginnen, *alternativ* mit mindestens einem a oder aber genau einem b fortsetzen, und danach wiederum *alternativ* mit einem e oder mindestens einem f enden. (1 Punkt)

2. Über dem Alphabet $\{a, b, \ldots y, z, A, B, \ldots Y, Z\}$ alle Zeichenketten, die genau einen Vokal enthalten (2 Punkte)

3. Über dem Alphabet $\{a, b\}$, alle Zeichenketten von der Form $a^m b^n$ wobei $m \neq n$ (1 Punkt).

4. Über dem Alphabet $\{a, b\}$ die Menge aller Zeichenketten mit alternierenden as und bs, wobei Sie annehmen dürfen, dass a, b sowie der leere String auch Zeichenketten von diesem Typ sind. (2 Punkte)

Lösung zu 1)

1. $c(a+|b)(e|f+)$

2. [^aeiouAEIOU]*[aeiouAEIOU][^aeiouAEIOU]*

3. Nicht möglich.

4. $a?(ba)^*b?$

Aufgabe 2) Beschreibung regulärer Ausdrücke **Punkte: 4**

Beschreiben Sie natürlichsprachlich und möglichst einfach, die Sprache, die von den folgenden regulären Ausdrücken erzeugt wird. Das Alphabet ist jeweils $\{a, b\}$.

1. $(bb)+$ (1 Punkt)
2. $(ba|a)^*(b|ba)^*$ (3 Punkte)

Lösung zu 2)

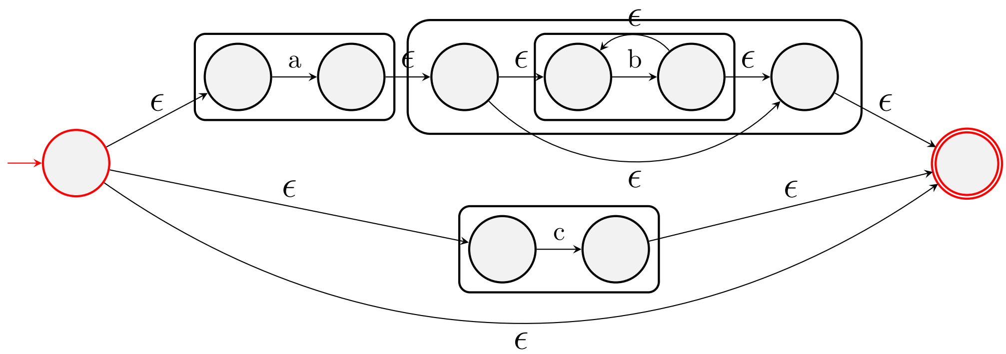
1. Zeichenketten, die nur aus b bestehen, gerade Länge haben und nicht leer sind.
2. Zeichenketten, bei denen jedes aa vor jedem bb vorkommt. (Erster Teil des Ausdrucks sind alle Zeichenketten, die kein bb enthalten. Zweiter Teil sind alle Zeichenketten, die kein aa enthalten).

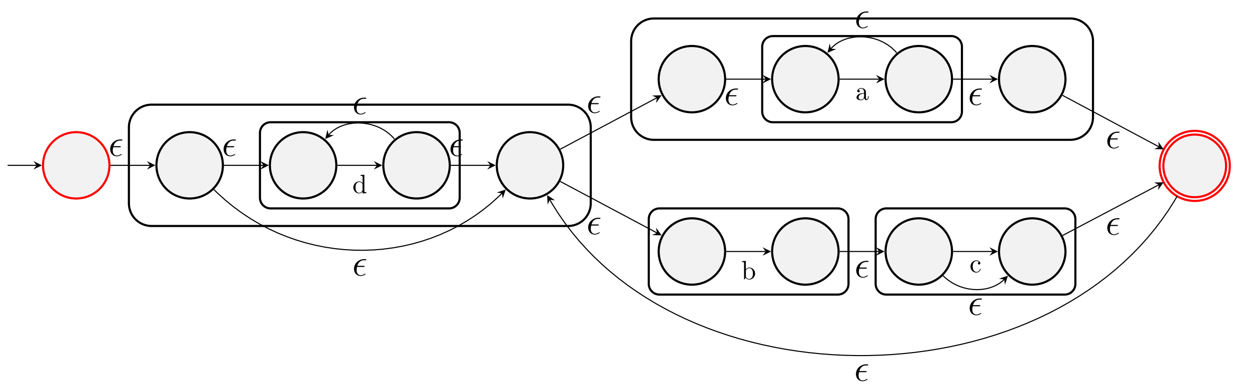
Aufgabe 3) Zeichnen von Automaten mit Thompson-Algorithmus **Punkte: 5**

Übersetzen Sie die folgenden regulären Ausdrücke mithilfe des Thompson-Algorithmus in endliche Automaten. Abzugeben ist die graphische Darstellung des Automaten. Es soll erkenntlich sein, dass Sie genau dem Thompson-Algorithmus gefolgt sind. N.b.: Dadurch werden auch Flüchtigkeitsfehler besser vermieden.

1. $(ab^*|c)?$ (2 Punkte)
2. $d^*(a+|bc?)+$ (3 Punkte)

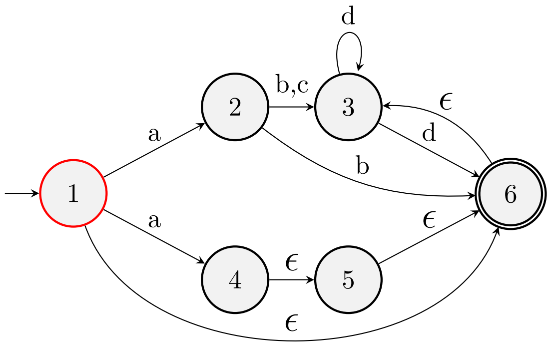
Lösung zu 3)

1. 

2. 

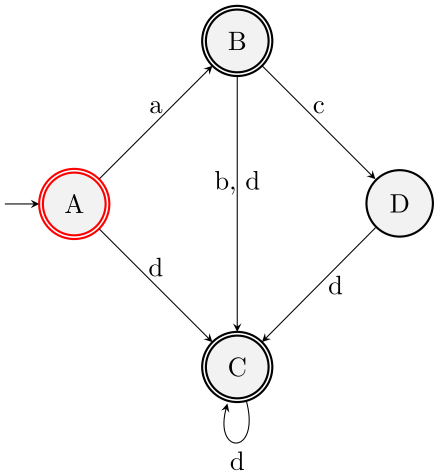
Aufgabe 4) Automatentransformation **Punkte: 5**

Wandeln Sie den untenstehenden nichtdeterministischen endlichen Automaten durch die Potenzmengenkonstruktion (Powerset Construction) in einen deterministischen endlichen Automaten um. Reichen Sie sowohl die Tabelle zur Potenzmengenkonstruktion als auch eine graphische Darstellung des entstehenden deterministischen Automaten ein. Im Automaten stehen durch Komma verbundene Zeichen für ein logisches *oder*.



Lösung zu 4)

NFA	DFA	a	b	c	d
{1,3,6}	**A**	B	-	-	C
{2,3,4,5,6}	**B**	-	C	D	C
{3,6}	**C**	-	-	-	C
{3}	D	-	-	-	C



Aufgabe 5) Bonusaufgabe **Punkte: 4**

1. Sei das Alphabet $\Sigma = \{a, b\}$ gegeben. Sei L die Sprache aller Zeichenfolgen über Σ, die mit ab beginnen. Schreiben Sie einen regulären Ausdruck für das Komplement von L, d.h. die Sprache $\bar{L} := \Sigma^* - L$, die genau die Zeichenfolgen über Σ enthält, die nicht in L sind. (2 Punkte).

2. Jeder FSA ist äquivalent zu einem (eventuell nicht-deterministischem) FSA mit nur einem Endzustand. Zeigen Sie das, indem Sie zu einem beliebigem FSA A mit mehr als einem Endzustand einen FSA A' mit nur einem Endzustand generieren, der die gleiche Sprache akzeptiert. (2 Punkte).

Lösung zu 5)

1. $a?|b(a|b)^*|aa(a|b)^*$, d.h. das leere Wort und a, alles was mit b beginnt und alles was mit aa beginnt.

2. Man wähle willkürlich einen der Endzustände von A aus, nenne diesen s_f und markiere ihn als neuen alleinigen Endzustand und schaffe epsilon-Übergänge der anderen alten Endzustände zu s_f.

Aufgabe 6) Unbepunktet und unkorrigiert: Zum weiteren Üben **Punkte:** –

- Alphabet $\{a, b\}$. Ist a in der Sprache, die vom regulären Ausdruck $(b^*a) + b^*$ erzeugt wird, enthalten? Begründen Sie kurz Ihre Antwort.
- Alphabet $\{a, b\}$. Sind die beiden Sprachen, die von $(ab)^*a$ bzw $a(ba)^*$ erzeugt werden gleich? Begründen Sie kurz Ihre Antwort.
- Alphabet $\{a, b\}$. In welcher Relation stehen die beiden Sprachen, die von $(aba)|(ba)|(aaa^*)$ bzw $a^*(ba|b|aa)$ erzeugt werden? Begründen Sie kurz Ihre Antwort.
- Alphabet $\{a, b\}$. Simplifizieren Sie den folgenden regulären Ausdruck weitmöglichst, wobei natürlich dieselbe Sprache generiert werden soll: $a|(b(b^*))a$
- Verwandeln Sie die mit dem Thompson-Algorithmus erzeugten NFAs der Aufgabe 3 in DFSAs.
- Kreiieren Sie FSAs für die regulären Ausdrücke, die Sie in Aufgabe 1 geschrieben haben.

Lösung zu 6)

Lösungen unvollständig. Nur Lösungen ohne Begründung. Bei Bedarf bitte in Tutorium diskutieren.

1. Ja.
2. Ja.
3. Die erste ist eine echte Teilmenge der zweiten.
4. b^*a