

Einführung in die Theoretische Informatik

Sommersemester 2020 – Übungsblatt 3

AUFGABE 3.1. (HTML 1:0 RegEx)

1,5 Punkte

Glaubt man diesem berühmten Stack Overflow-Post, so ist HTML keine reguläre Sprache. So berühmt der Post auch sein mag, gibt er jedoch keinen expliziten Beweis für diese Behauptung. Wir bleiben skeptisch, bis ein Beweis vorhanden ist: Zeigen Sie, dass HTML in der Tat nicht mit einem regulären Ausdruck – wie in der Vorlesung definiert – geparsed werden kann. Sie dürfen die Aussage für eine vereinfachte Variante $HTML^{}$ über dem Alphabet $\Sigma := \{<, >, /, \mathbf{b}\}$ zeigen, die ausschließlich aus balancierten $<\mathbf{b}>...</\mathbf{b}>$ Elementen besteht. Balanciert bedeutet dabei, dass

- (a) jeder geöffnete $<\mathbf{b}>$ Tag auch wieder mit $</\mathbf{b}>$ geschlossen wird und
- (b) ein \mathbf{b} -Element immer erst geöffnet werden muss, bevor es geschlossen werden kann.

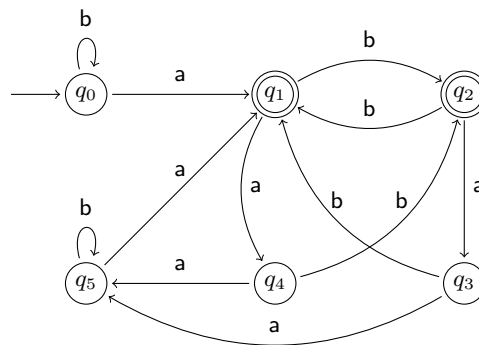
Hier ein paar Beispiele:

- $<\mathbf{b}></\mathbf{b}> \in HTML^{}$
- $<\mathbf{b}></\mathbf{b}><\mathbf{b}></\mathbf{b}> \in HTML^{}$
- $<\mathbf{b}><\mathbf{b}></\mathbf{b}><\mathbf{b}></\mathbf{b}></\mathbf{b}> \in HTML^{}$
- $<\mathbf{b}></\mathbf{b}></\mathbf{b}><\mathbf{b}> \notin HTML^{}$
- $<\mathbf{b}></\mathbf{b}><\mathbf{b}> \notin HTML^{}$
- $<\mathbf{b}/> \notin HTML^{}$

AUFGABE 3.2.

1 Punkt

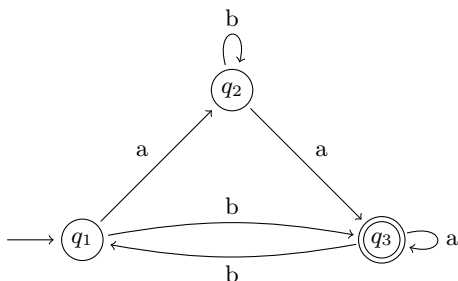
Minimieren Sie den folgenden DFA.



AUFGABE 3.3.

1 Punkt

Gegeben sei folgender Automat $M = (\{q_1, q_2, q_3\}, \{a, b\}, \delta, q_1, \{q_3\})$:



Berechnen Sie mit dem Gauß-Verfahren und Ardens Lemma einen regulären Ausdruck α mit $L(\alpha) = L(M)$.

AUFGABE 3.4. (Rekursive Prozedur)

1,5 Punkte

Geben Sie eine rekursive Prozedur $insert(x, r)$ an, die für einen gegebenen regulären Ausdruck r einen neuen regulären Ausdruck berechnet, sodass gilt $L(insert(x, r)) = \{uxv \mid \exists w \in L(r).w = uv\}$. Die Sprache $L(insert(x, r))$ soll also alle Wörter enthalten, die man durch das Einfügen eines Zeichens x in ein Wort aus der Sprache $L(r)$ erhalten kann. Für Ihre Definition sollten Sie das folgende Gerüst verwenden:

-
- $insert(x, \emptyset) =$
 - $insert(x, \mathbf{a}) =$
 - $insert(x, \varepsilon) =$

- $insert(x, \alpha\beta) =$
- $insert(x, \alpha \mid \beta) =$
- $insert(x, \alpha^*) =$

Beweisen Sie mittels struktureller Induktion, dass Ihre Definition korrekt ist, wobei Sie den Fall $r = \alpha\beta$ weglassen dürfen.

Parsing HTML with regex summons tainted souls into the realm of the living.

— bobince