

# Zusammenfassung zur Vorlesung Basismodul Computerlinguistik

## Wiederholung

27.01.2022

# Minimierung eines totalen DEA

**Aufgabe:** Gegeben sei ein DEA mit totaler Übergangsfunktion:  $\Sigma = \{a, b\}$ ,  
 $Zustände = \{1, 2, 3, 4, 5, 6\}$ ,  $F = \{1, 4, 5\}$ ,  $s = 1$ ,  
 $\Delta = \{(1, a, 3), (1, b, 2), (2, a, 4), (2, b, 1), (3, a, 1), (3, b, 2), (4, a, 4), (4, b, 3),$   
 $(5, a, 3), (5, b, 6), (6, a, 4), (6, b, 5)\}$ .  
Berechne für den DEA einen äquivalenten minimalen DEA.

# Nerode-Konstruktion

**Aufgabe:** Es sei  $\Sigma$  das Alphabet  $\{a, b\}$ . Die Menge  $L$  enthalte alle Wörter mit genau **2** a's und **2** b's. Berechne den minimalen DEA für  $L$  mittels Nerode-Äquivalenzklassen.

**Aufgabe:** Gegeben sei ein Lexikon  $\{bac, abc, bacb, abcb\}$ .

- 1 Berechne den minimalen Lexikonautomaten für das Lexikon nach dem Daciuk-Algorithmus.
- 2 Notiere auf dem Bild für den Lexikonautomaten die Zahlen, die für die Berechnung von Perfekten Hashingwerten nötig sind.
- 3 Speichere den Lexikonautomaten als Tarjan-Tabelle.
- 4 Wie würde der Automat aussiehen, wenn man das Wort *abcbe* hinzufügt, und dann das Wort *bac* löscht.



**Aufgabe:** Betrachte den universellen Levenshtein-Automaten mit der Fehlerschranke  $b = 3$ . Zum aktuellen Zeitpunkt  $t$  werden die Zustände  $l - 2^2, l + 1^2, l^3$  aktiviert. Es sei 01110001 der Bitvektor für  $t$ .

- 1 Welche Zustände werden zum nächsten Zeitpunkt  $t + 1$  aktiviert?
- 2 Welche davon sind echte aktive Zustände mit Berücksichtigung der Normalisierung (d.h. Löschung aller redundanten Zustände)?

**Aufgabe:** Es sei  $v$  das Pattern-Wort *colour* und  $w$  das Wort *color*. Wir möchten einen Universellen Levenshtein-Automaten verwenden, um zu testen, ob der Levenshteinabstand zwischen  $v$  und  $w$  kleinergleich 2 ist.

- 1 Welche Schranke soll der Universelle Levenshtein-Automat haben?
- 2 Wie sieht die Folge der Automaten-Eingaben aus?