

Zusammenfassung zur Vorlesung Basismodul Computerlinguistik

Universelle Levenshtein-Automaten & Themenschwerpunkte in der Klausur

20.01.2022

- 1 Universelle Levenshtein-Automaten: Eine Tafelübung
- 2 Themenschwerpunkte der Klausur

Universelle Levenshtein-Automaten: Eine Tafelübung

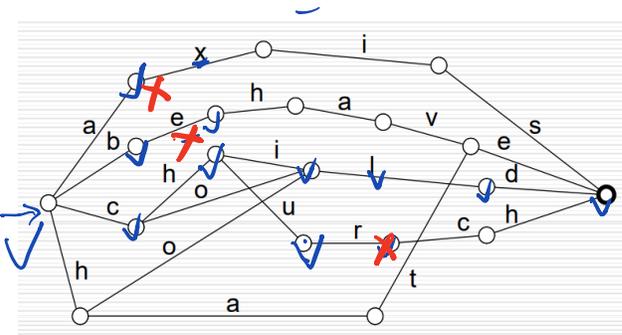
Aufgabe: Gegeben seien ein Lexikonautomat \mathcal{L} und Patternwort *child*. Verwende den Universellen Levenshtein-Automaten mit der Schranke 1 $A^{\vee}(1)$ und finde alle Wörter im \mathcal{L} , deren Levenshtein-Abstand mit dem Patternwort kleiner gleich 1 ist.

$\$child$

$$l = 2 \times 1 + 2 = 4$$

$$\chi(c, \$cho) = 0/00$$

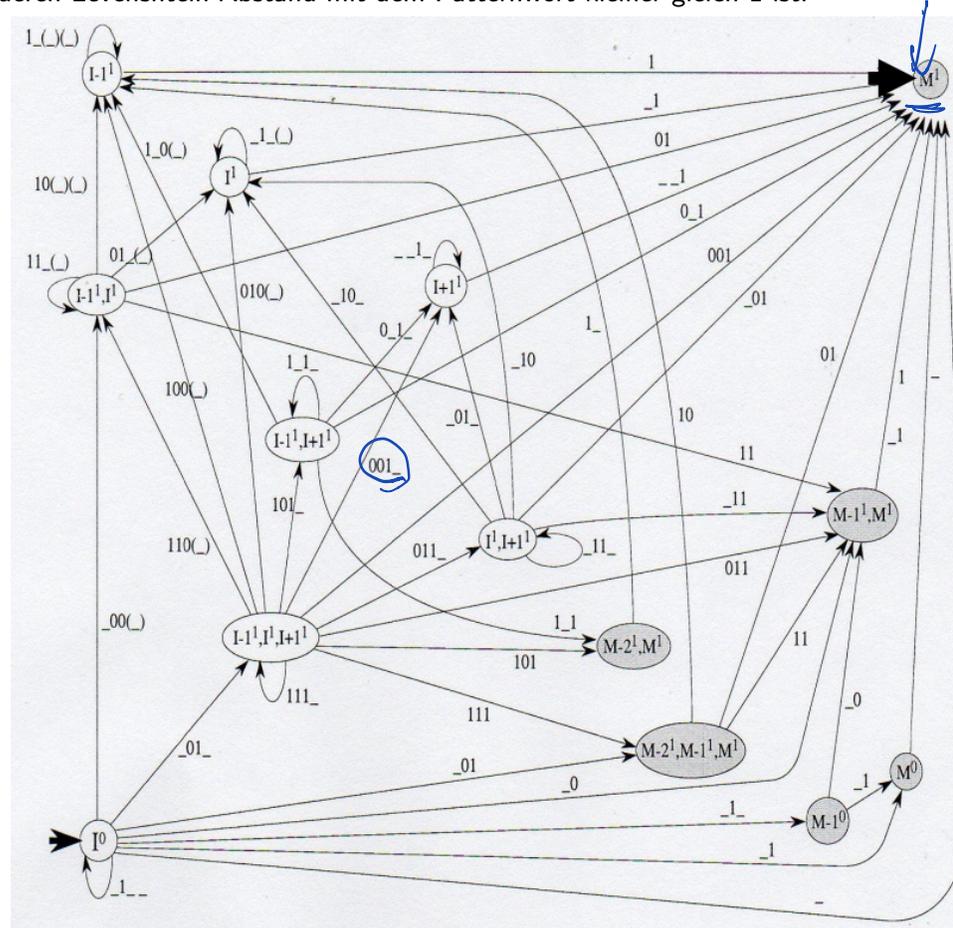
$$\chi(c, chob) = 00/0$$



$$\chi(c, hold) = 000/0$$

$$\chi(d, old) = 00/$$

Output: *child, cold.*



Themenschwerpunkte der Klausur

Alphabet Σ Σ^n Σ^* Sprache
Wort $w \in \Sigma^*$

Basiswissen

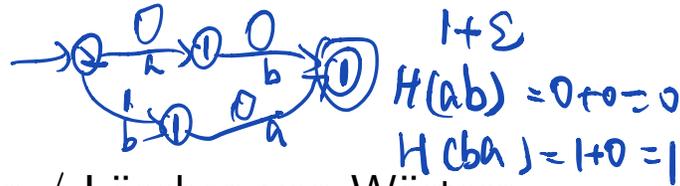
- Grundbegriffe: Sprache, Endlicher Automat $A = (\Sigma, Q, S, F, \Delta)$.
- Determinisierung eines NDEA (Potenzmengenkonstruktion)
- DEA-Minimierung $DEA \rightarrow mDEA$
- Nerode-Konstruktion $Sprache \rightarrow mDEA$
Rechtssprache

Themenschwerpunkte der Klausur

Lexikonautomat

Eingabe: Sortierte Wortliste.

- **Daciuk-Algorithmus:** Minimierung eines Lexikonautomaten
- Perfektes Hashing
- Tarjan-Tabellen
- Nachträgliches Hinzufügen / Löschen von Wörtern



Zeilennr. Label Zustand

Zeilennr.	Label	Zustand
1	a	1
2	b	2
3		
4		
5		
6		

→ Zustandszelle
→ Übergangszelle

Themenschwerpunkte der Klausur

Universelle Levenshtein-Automaten

- Generelles Verständnis
- Nachfolgestand zu berechnen
- Für Schranke k aus gegebenem Wort Bitvektoren berechnen

1. $C = 2k + 2$ (max. Länge)

2. $\underbrace{\text{PATTERN}}_k$

$I-3^3$
 $I-1^2$
 $I+1^1$



Textkorrektur

- Verständnis

- Wenn nicht explizit gefordert: partiellen Automaten entwerfen
- Können auch Textaufgaben drankommen
- Lösung lieber doppelt checken, wenn alles fertig ist