

Algorithmen und Datenstrukturen (für Bioinformatik)
(Frank Hoffmann)

Abgabe am Donnerstag, den 14. November 2002 bis 16⁰⁰

1. **Nichtdeterministische Automaten** (3+3 Punkte)

(a) Konstruieren Sie zu dem in Fig.1 dargestellten nfa einen äquivalenten dfa.

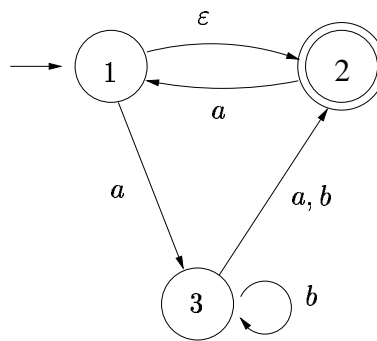


Abbildung 1:

(b) Konstruieren Sie für nachfolgende reguläre Ausdrücke über dem Alphabet $\Sigma = \{0, 1\}$ nfa's, die dieselben Sprachen beschreiben.

$$(((00)^*(11)) \cup 01)^*, \quad (0 \cup 1)^*000(0 \cup 1)^*$$

2. **Berechnung des Match-Count** (6 Punkte).

Gegeben seien zwei Strings $A[1 \dots n]$ und $B[1 \dots n]$ der Länge n und einen Parameter $1 \leq k \leq n$. In A und B gibt es jeweils $n - k + 1$ Teilstrings der Länge k . Die Anzahl der Paare (α, β) von solchen Teilstrings, wobei α ein Teilstring von A und β ein Teilstring von B ist (jeweils der Länge k), ist daher $\Theta(n^2)$. Der *match-count* des Paares (α, β) ist die Anzahl der Positionen in denen diese beiden Teilstrings übereinstimmen. Der match-count aller Paare kann offensichtlich in $O(kn^2)$ berechnet werden.

Geben Sie einen Algorithmus an, der alle match-counts in $O(n^2)$ berechnet.

3. **Medianbestimmung** (4 Punkte)

Geben Sie einen $O(\log n)$ Algorithmus an, der den Median von $2n$ Elementen $X[1..n]$, $Y[1..n]$ findet, falls die n -elementigen Listen X und Y bereits sortiert sind. Begründen Sie die Laufzeit.

Hinweis: Die Abgabe der Implementierung des Aho–Corasick–Algorithmus hat bis zum 19.12., 16 Uhr zu erfolgen. Die Aufgabe wird mit 16 Punkten plus 4 mögliche Zusatzpunkte bewertet.