

4. Aufgabenblatt vom Donnerstag, den 14. November 2002 zur Vorlesung

Algorithmen und Datenstrukturen (für Bioinformatik)
(Frank Hoffmann)

Abgabe am Donnerstag, den 21. November 2002 bis 16⁰⁰

1. **Vom dfa zum Regulären Ausdruck** (4 Punkte) Betrachten Sie den dfa mit 2 Zuständen über $\{0, 1\}$, der alle Wörter gerader Länge akzeptiert. Konstruieren Sie den zugehörigen regulären Ausdruck mit der in der Vorlesung vorgestellten Methode des dynamischen Programmierens.
2. **ϵ -Übergänge** (2 Punkte) Beweisen Sie, dass jeder nfa mit ϵ -Übergängen zu einem ohne solche Übergänge äquivalent ist.
3. **Aho-Corasick und Reguläre Ausdrücke** (2 Punkte). Vergleichen Sie den Aho-Corasick-Algorithmus mit dem in der Vorlesung vorgestellten Algorithmus zum Matchen gegen reguläre Ausdrücke.
4. **Rechteckige Muster** (6 Punkte) Gegeben seien $q > 1$ rechteckige paarweise verschiedene Muster gleicher Breite über einem endlichen Alphabet Σ . Die Muster haben Gesamtgröße n , ein rechteckiger Text Größe m . Finden Sie einen effizienten Algorithmus, der alle Muster im Text wiederfindet. Was ist seine Komplexität?
5. **Wild cards** (2 Punkte) Betrachten Sie wild cards, die anstatt einzelner Zeichen Teilstrings beliebiger Länge matchen. Diskutieren Sie kurz die Komplexität des Exakten-Matching-Problems, wenn solche wildcards im Pattern, im Text oder in beiden auftreten.

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.