

5. Aufgabenblatt vom Donnerstag, den 21. November 2002 zur Vorlesung

Algorithmen und Datenstrukturen (für Bioinformatik)
(Frank Hoffmann)

Abgabe am Donnerstag, den 28. November 2002 bis 16⁰⁰

1. **Nochmal dfa und Regulärer Ausdruck** (4 Punkte) Überlegen Sie sich einen dfa und einen regulären Ausdruck für die Sprache in $\{0, 1\}^*$, die aus den Wörtern mit gleichvielen Teilwörtern der Form 01 und 10 besteht.
2. **Shift-And und agrep** (2 Punkte). Erweitern Sie sowohl die Shift-And-Methode als auch agrep (also die Modifikation von Shift-And auf fest begrenzte Fehlerzahl) dahingehend, dass auch wild cards (die jedes einzelne Zeichen matchen) im Text und/oder Muster behandelt werden können. Wie wird die Effizienz der Algorithmen beeinflusst?
3. **Rechteckige Muster II** (6 Punkte) Erweitern Sie den randomisierten Fingerprint-Algorithmus von Karp und Rabin dahingehend, dass 2-dimensionales Matching von Rechtecken möglich ist. Das zugrundeliegende Alphabet soll zweielementig sein.
4. **Exaktes Matching mit Suffix-Baum** (5 Punkte) Erläutern Sie im Detail, wie man das exakte Matching für ein Muster P der Länge m in $O(m + k)$ mit Hilfe des Suffixbaumes für den Text beantworten kann, wenn k die Anzahl der Vorkommen des Musters im Text ist. Illustrieren Sie dies mit einem geeigneten Beispiel

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.