

# Übungen zur Vorlesung Algorithmen auf Sequenzen

TU Dortmund, SS 2008

Prof. Dr. Sven Rahmann

**Blatt 7 vom 11.06.2008**

**Abgabe am Fr 20.06.2008 in der Vorlesung**

**Aufgabe 1** Erweitere die Glushkov-Konstruktion um die Fälle, dass als Metazeichen in regulären Ausdrücken auch  $+$  (mindestens einmal wiederholte Elemente) und  $?$  (optionale Elemente) auftreten können. Im Syntaxbaum handelt es sich dabei (wie bei  $*$ ) um unäre innere Knoten. Gehe auf die First, Follow, Last, Empty und Mindestlängen-Berechnungen ein.

**Aufgabe 2** Zeige, dass die Edit-Distanz eine Metrik ist.

**Aufgabe 3** Die  $q$ -gram Distanz ist keine Metrik: Verschiedene Sequenzen können die gleichen  $q$ -gram Häufigkeiten aufweisen. Gib einen effizienten Algorithmus an, der – gegeben die  $|\Sigma|^q$  Häufigkeitswerte aller  $q$ -grams (das  $q$ -gram Profil) – entscheidet, ob es überhaupt Sequenzen mit diesem  $q$ -gram Profil gibt, genau eine, oder mehrere (Hinweis: Euler).

**Aufgabe 4** Wiederhole die Methoden zur bit-parallelen Simulation von verallgemeinerten Strings ( $x(\min, \max)$ ,  $?$ ,  $+$ ,  $*$ ). Wie sieht der Automat zu  $C-x(2,4)-A-B?-C+-[AB]$  mit  $k = 1$  Fehler aus?