Rheinische Friedrich-Wilhelms-Universität Bonn
Institut für Informatik IV
Prof. Dr. N. Blum, C. Rick

# Diplomarbeit zu vergeben

## Empirische Untersuchungen über Algorithmen zur Berechnung einer längsten gemeinsamen Teilsequenz in linearem Platz.

Ein grobes Maß für die Ähnlichkeit zweier Zeichenketten ist die Länge einer längsten gemeinsamen Teilsequenz, d.h. einer Sequenz, die aus beiden Zeichenketten durch Löschen einiger (nicht notwendig aufeinanderfolgender) Zeichen gewonnen werden kann.

Man kennt mittlerweile eine ganze Reihe von Algorithmen unterschiedlicher Effizienz zur Berechnung dieses Wertes. Der benötigte Speicherplatz ist linear in der Länge der Eingabesequenzen beschränkt. Möchte man jedoch neben der Länge auch eine gemeinsame Teilsequenz rekonstruieren, wird der Speicherplatzbedarf in der Regel quadratisch (d.h. proportional zum Produkt der Längen der Eingabesequenzen).

Mittels eines Divide & Conquer Schemas ist es gelungen, für einige Algorithmen sowohl ihre (asymptotische) Effizienz als auch den linearen Platzbedarf bei der Rekonstruktion einer längsten gemeinsamen Teilsequenz aufrechtzuerhalten.

Allerdings scheint sich der Aufwand bei der Anwendung des Divide & Conquer Schemas je nach Algorithmus erheblich zu unterscheiden. Es ist daher denkbar, daß in der Praxis ein sehr guter Algorithmus zur Berechnung der Länge bei der Rekonstruktion einer längsten gemeinsamen Teilsequenz in linearem Platz schlechter abschneidet als ein einfacherer Algorithmus, für den das Divide & Conquer Schema leichter anwendbar ist. In der Diplomarbeit soll der Frage nachgegangen werden, ob diese Vermutung stimmt. Auch Verfahren, die hierbei zunächst schlecht abschneiden, lassen sich ggf. durch geeignete Kombination verschiedener Methoden in der Praxis verbessern. Die Diplomarbeit sollte als Ergebnis aufzeigen, welches Verfahren unter welchen Umständen und in welcher Realisierung am besten zur Rekonstruktion einer längsten gemeinsamen Teilsequenz geeignet ist.

Bei guter Bearbeitung der Aufgabenstellung ist eine Präsentation der Ergebnisse auf einem Workshop bzw. die Veröffentlichung in einer Fachzeitschrift möglich.

**Zeitplan:**

- Einarbeitung/Literaturstudium: 2 – 3 Monate
- Programmierung und Durchführung der Experimente: 2 – 3 Monate
- Auswertung und Verfassen der Diplomarbeit: 2 – 3 Monate

Interessenten wenden sich bitte an C. Rick (Zimmer N106).

Postanschrift:
Universität Bonn
Institut für Informatik IV
Römerstr. 164, D-53117 Bonn

Telefon: (02 28) 73 –41 20
Fax: (02 28) 73 –45 71
Email: rick@cs.uni-bonn.de