Technische Universität München
Fakultät für Informatik
Lehrstuhl für Effiziente Algorithmen
Dr. Volker Heun

Wintersemester 2001/2002
Übungsblatt 6
28. November 2001

Algorithmische Bioinformatik

Hinweis: *Die als freiwillig gekennzeichneten Aufgaben werden nicht bewertet und sind für die Klausurzulassung ohne Bedeutung.*

Freiwillige Aufgabe 1

Implementiere die Algorithmen zum Suchen in Texten (Naive Methode, Knuth-Morris-Pratt, Boyer-Moore mit Good-Suffix-Rule bzw. Bad-Character-Rule). Erweitere die Implementierungen so, dass die Anzahl der ausgeführten Vergleiche ebenfalls ermittelt wird.

Vergleiche nun die verschiedenen Algorithmen auf biologischen Sequenznen und zufällig konstruierten Sequenzen über den Alphabeten $\{A, C, G, T\}$ und $\{A, \ldots, Z\}$.

DNS- und Protein-Sequenzen kann man u.a. von folgenden Quellen erhalten:

- The Genome Database: `http://www.gdb.org/`
- National Center for Biotechnical Information: `ftp://ftp.ncbi.nih.gov/` bzw. `ftp://ftp.ncbi.nih.gov/genbank/genomes/`
- Human Genome Working Draft at UC Santa Cruz: `http://genome.ucsc.edu/`
- SWISS-PROT: `http://www.expasy.org/` bzw. `ftp://ftp.expasy.org/databases/swiss-prot/release/`

Als Suchmuster kann man sowohl zufällig gewählte Sequenzen als auch Teilstücke der bekannten Sequenzen verwenden. Variiere die Längen der Suchmuster (z.B. zwischen 5 und 100). Wähle insbesondere auch lange Texte ($\geq 10.000.000$).

Hinweise:

Die Klausur findet am **4. Februar 2002** um **18:00 Uhr** im **Großen Physikhörsaal (1200)** statt!

Wegen des Dies Academicus entfällt die Übung am 6.12.!