

Übung

Algorithmische Bioinformatik

WS 2007/2008

Allgemeines

Ulf Leser

Wissensmanagement in der
Bioinformatik



Überblick

- Die Übung hat 8 Termine
 - Erster Termin: 1.11.2007
 - Dann ein Termin **alle zwei Wochen**
 - Siehe Webseite für Änderungen
 - Ablauf
 - Vorstellung der Lösungen zur letzten Übung
 - Ausgabe der neuen Aufgabe
- Dazwischen
 - Reserviert für Fragen, Vorträge, Übungen nach Wunsch etc.
- Alles steht auch im Web

Übungsschein

- Der Schein ist Voraussetzung für die Prüfung
- Wir bilden Gruppen a 3-4 Personen
 - Jede Gruppe muss **zu jeder Aufgabe** mindestens 51% der Punkte haben
 - Gruppen bestehen/scheitern nur als Ganzes
- Jede Stunde werden die Lösungen der letzten Aufgabe in der Übung vorgestellt
 - Dabei werden meistens nur 3-4 Gruppen vortragen, da die Lösungen sich ähneln werden (Abweichungen möglich)
 - Wir **losen jede Stunde die vorstellenden Gruppen** aus
 - Aus einer ausgelosten Gruppe muss ein Student vortragen
 - Überlegt vorher, wer das jeweils macht!
 - Immer einen Vortrag parat haben
 - Ein Student, der einmal vorgetragen hat, darf kein zweites Mal (solange noch jemand in seiner Gruppe nicht vorgetragen hat)

Aufgaben

- Die Übung ist implementationsintensiv (aber nicht nur)
 - Expliziter Wunsch vorheriger Teilnehmer
 - Implementiert wird in JAVA
- I.d.R. sind Eingabe, Ausgabe, Aufrufform etc. **exakt vorgegeben** und müssen eingehalten werden
- Aufgaben
 - Exaktes Matching: Naiver und Z-Box Algorithmus; Evaluation der Laufzeiten; Anwendung auf Restriction Enzyme Binding Sites
 - Boyer-Moore mit garantierter linearer Laufzeit; Evaluation
 - Suffixarrays (Aufbau und Suche); Vergleich mit nicht indizierender Suche
 - Approximatives Matching: Globales Alignment, Anwendung auf Mensch/Maus/Schimpanse
 - Globales Alignment in linearem Platz; Vergleich zum vorherigen Alg.
 - Tw. Quasar (Hash-based q-Gramm Index)
 - Phylogenie: UPGMA und Neighbor Joining; Vergleich der resultierenden Stammbäume anhand von 8 Spezies

Wettbewerb

- Wettbewerb
 - 4 der 7 Aufgaben sind als Wettbewerb konzipiert
 - Parallel zu den Punkten für korrekte Lösungen gibt es Punkte für die **schnellste Lösung**
 - Schnellste Lösung: 5 Punkte
 - Zweite: 2 Punkte
 - Dritte: 1 Punkt
 - Die beste Gruppe bekommt am Ende ein kleines Präsent
- Messungen
 - Alle Zeitmessungen werden von uns dreimal wiederholt
 - Ausführung auf AMSEL
 - Alles muss mit dem dort installierten Java laufen
 - Wir messen die SYSTEM Zeit mit UNIX „time“
 - Parallelisierung lohnt sich in dieser Übung also nicht

Gruppen

- Gruppenbildung
 - Gruppe 1: Groth, Mähler, Irshad
 - Gruppe 2: Stigge, Koumbos, Rostin
 - Gruppe 3: Pollex, Winkler, Alberti
 - Gruppe 4: Rheinländer, Starlinger,
 - Gruppe 5: Liske, Dudeck, Eicher
 - Gruppe 6: Bock, Gühler, Wiedstruck
 - Gruppe 7: Kummer, Kruber
 - Gruppe 8:
- Bitte Gruppen **asap** in GOYA eintragen
 - Und zwar mit den Namen „GruppeX“