

Übungen zur Vorlesung „Molekularbiologische Datenbanken“

Aufgabenblatt 5: Prositate, automatische Update

Abgabetermin: **22.06.2004** bis **17 Uhr** über e-mail oder in RUD25 IV.104
Abgabe: Relationales Modell, SQL Queries und deren Antworten sowie das Programm zum
automatischen Update
Maximal: 20 Punkte, Erforderlich: 15 Punkte.
Namen nicht vergessen!

Wie schon erwähnt, wollen wir an den Prositate Daten das automatische Update unserer In-house Datenbank erproben.

Erstellen Sie dazu zuerst ein relationales Schema, um die Daten in unsere bestehende Datenbank zu integrieren. Von den Prositate Daten sollen nur die Einträge gespeichert werden, die Pattern enthalten. Laden Sie die Daten und beantworten Sie einige Fragen. Um nun immer aktuelle Daten in unserer Datenbank vorrätig zu haben, wollen wir automatisch prüfen, ob eine neue Version vorhanden ist. Dabei ist es notwendig, das Datum und die Uhrzeit des neuen Releases mit den Werten des aktuell in der Datenbank vorhandenen Releases zu vergleichen und gegebenenfalls die Daten zu ersetzen. In diesem Fall muss nur das jeweils aktuelle Release vorhanden sein.

Aufgaben

1. Erstellen Sie das relationale Modell, um die Prositate Daten in die bestehende Datenbank aufzunehmen. Von den Prositate Einträgen sollen der Identifier, die Accession, die Description, das eigentliche Pattern, und die Referenzen zu Swiss-Prot. **5 P**
2. Laden Sie die Daten und beantworten Sie die folgenden Fragen
 - a) Wie viele Pattern-Einträge hat Prositate von dem File prositate-new.dat? **2 P**
 - b) Wie viele dieser Pattern haben einen Verweis zu Swiss-Prot Daten in unserer Datenbank? Nehmen Sie dazu die Swiss-Prot Accession (P02023, ..) als eindeutiges Zuordnungskriterium. **2 P**
 - c) Ein Eintrag in Prositate kann sowohl über die angegebenen Links als auch über eine Suche des Patterns in den Sequenzen mit Swiss-Prot verknüpft werden. Um festzustellen, wie vollständig bzw. richtig die angegebenen Links sind, betrachten wir das Beispiel BZIP_BASIC. Wir wollen die gefundenen Links aus Prositate mit den Verknüpfungen über Pattern vergleichen. Geben Sie dazu als Ergebnis die Art der Verknüpfung und den Swiss-Prot Identifier wie sie in Prositate angegeben ist an, sowie die gefundenen Sequenzen mit dem dazugehörigen Swiss-Prot Identifier. **3 P**
3. Erstellen Sie ein Programm, das auf der FTP-Seite prüft, ob die für uns interessanten Daten verändert wurden (das Datum und die Uhrzeit müssen sich, im Vergleich zu der letzten Version geändert haben) und laden Sie diese dann in die Datenbank. Um zu sehen, ob das funktioniert, soll es in der Übung am 23.06. vorgeführt werden. **8 P**