

Übungen zur Vorlesung „Molekularbiologische Datenbanken“

Aufgabenblatt 4: SWISSPROT, Versionierung

Abgabetermin: **24.06.2003 bis 17 Uhr** über e-mail oder in RUD25 IV.102

Bitte geben sie konzeptuelle Modell, das relationale Modell,
die Trigger-Funktionen, die Queries und deren Ergebnisse schriftlich ab.

Maximal: 20 Punkte, Erforderlich: 15 Punkte.

Namen nicht vergessen!

Um die Expressionsdaten in der Datenbank einem Protein zuzuordnen, integrieren wir auch noch SWISSPROT. Einige unserer Partner wollen, dass die Datenbank auch Daten von alten Releases speichert.

Sie versorgen uns mit 2 SWISSPROT-flat-files, 20021129.dat und 20030404.dat, die in der Datenbank gespeichert werden sollen. Die Datei 20030404.dat ist dabei die aktuellste Version. Die Daten sind auf der Web-page als zip oder auf jazz unter /vol/home-vol3/wbi/trissl/uebung/aufgabe4 als flat-files gespeichert.

Von den Daten in SWISSPROT sollen die ID, die Accession-Number, last_update und die Description sowohl als auch die DatabaseReferences gespeichert werden. Auch sollen die beiden Versionen und in Zukunft noch weitere verfügbar sein.

1. Erweitern Sie die bestehende Datenbank um diese Informationen und zeigen Sie dies anhand eines konzeptuellen Modells mit UML und eines relationalen Modells. Laden Sie anschließend die Daten von Version 1 und Version 2 in die Datenbank. (Muß in Zukunft für n Versionen funktionieren)
10 P
2. Wie viele der Proben auf dem Array haben einen entsprechenden Eintrag in der aktuellen Version von SWISSPROT?
2 P
3. Nach wie vielen der Proteine, die in der aktuellen Version von SWISSPROT vertreten sind, können wir auf unseren gespeicherten Arrays Tests durchführen?
2 P
4. Da ein Mitarbeiter eines Labors die Proteine in SWISSPROT nicht mit der Accession number gespeichert hat, sondern mit der ID (die sich auch ändern kann), möchte er wissen, welche ID's sich von der Alten zur Neuen Version verändert haben.
2 P
5. An uns wird eine Anfrage gestellt, für welche Proteine, die das Wort 'apoptosis' in ihrer Description enthalten, es auch Proben auf unseren gespeicherten Arrays gibt. Geben Sie dazu die Protein und die GenBank Accession number, die ID der Probe und die dazugehörige Array ID an.
2 P
6. Bitte erstellen Sie ein dump-file ihrer Datenbank und geben sie dessen Größe an.
2 P