

Übungen zur Vorlesung „Molekularbiologische Datenbanken“

Aufgabenblatt 4: SWISSPROT, Versionierung

Abgabetermin: **08.06.2004** bis **17 Uhr** über e-mail oder in RUD25 IV.104

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, 20031129.dat und 20040404.dat, die in der Datenbank gespeichert werden sollen. Die Datei 20040404.dat ist dabei die aktuellste Version. Die Daten sind auf der Web-page als zip oder auf paprika unter /daten/mdb/... als flat-files gespeichert.

Von den Daten in SwissProt sollen die **ID**, die **Accession-Nummern**, das **Erstellungsdatum** (creation date), die **Description**, die **Sequenz**, die darin erwähnten **Features** (mit der Struktur der Einträge) sowie die **Database References** gespeichert werden. Auch sollen die beiden Versionen und in Zukunft noch weitere verfügbar sein.

Zu den Features:

Features wird in Sequenzdatenbanken alles bezeichnet, dass als zusätzliche Information gegeben wird. In Swiss-Prot haben die Features folgende Struktur:

<i>Columns</i>	<i>Data</i>
1-2	FT
6-13	'From' endpoint
22-27	'To' endpoint
35-75	Description

Aufgaben

1. Erweitern Sie die bestehende Datenbank um diese Informationen und zeigen Sie dies anhand eines konzeptuellen Modells mit UML und eines relationalen Modells.
6 P
2. Erstellen Sie Routinen, die das Laden einer neuen Version ermöglichen. Laden Sie anschließend die Daten von Version 1 und Version 2 in die Datenbank. (Muss in Zukunft für n Versionen funktionieren)
4 P
3. Wie viele Einträge wurden verändert, ergänzt, und gelöscht von der ersten zur zweiten Version?
2 P
4. Wie viele der Proben auf dem Array haben einen entsprechenden Eintrag in der aktuellen Version von SwissProt?
2 P
5. Nach wie vielen der Proteine, die in der aktuellen Version von SwissProt vertreten sind, können wir auf unseren gespeicherten Arrays Tests durchführen?
2 P
6. Da ein Mitarbeiter eines Labors die Proteine in SwissProt nicht mit der Accession number gespeichert hat, sondern mit der ID (die sich für eine Accession auch ändern kann), möchte er wissen, für welche Accession-Nummern sich die Identifier von der alten zur neuen Version verändert haben und die entsprechenden Identifier dazu.
2 P
7. 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

Für die Präsentation zeigen Sie bitte, wie Sie die verschiedenen Versionen speichern, wie das Laden einer neuen Version funktioniert und die Ergebnisse der Queries.