

Übungen zur Vorlesung „Molekularbiologische Datenbanken“

Aufgabenblatt 2: Microarray-Modellierung

Abgabetermin: **11.05.2004 bis 17 Uhr** per e-mail oder in RUD25 IV.104

Bitte geben das konzeptionelle Modell mit UML,
das relationale Model und DTD zum Datenaustausch
im **PDF/Word-format** ab

Maximal: 20 Punkte, Erforderlich: 15 Punkte
Namen nicht vergessen!

Bei einem EU Projekt zur Genexpression sollen die Daten zentral verwaltet werden. Die Datenbank muss Informationen über die Experimente enthalten und auch die Ergebnisse speichern. Mit den Laboren haben wir vereinbart, dass die Informationen über die Experimente in XML übermittelt werden.

Informationen über Teilnehmer

Es beteiligen sich eine Anzahl von Laboren, von denen jeweils der Name des Labors (Universität, Max-Planck-Institut, usw.), die Adresse und die URL gespeichert werden sollen. Eine Labor-Id wird bei uns intern vergeben und an die Labors weitergegeben.

Um Daten eines Experiments übermitteln zu können, muss ein Submitter seine Submitter-ID angeben. Von einem Submitter ist sein Name, bei welchem Labor er beschäftigt ist, die Abteilung, die Telefon- und Faxnummer und seine e-mail Adresse bekannt.

Informationen über den Microarray

Für jedes Experiment wird ein Microarray benötigt. Wir haben momentan zwar nur einen Lieferanten, doch möchten wir auch die Daten über ihn in der Datenbank in einer eigenen Tabelle speichern. Über den Array-Hersteller ist sein Name, seine email-Adresse, das Labor in dem der Hersteller beschäftigt ist, der Firmensitz und die Telefonnummer bekannt.

Wir haben in unserer Datenbank einige Arrays gespeichert. Jeder Array hat eine eigene ID, außerdem wird der Titel, der Typ, der Organismus, von dem die Sequenzen auf dem Array stammen, eine Beschreibung, ein web-link zu weiteren Daten, ein Submission Date und die Hersteller-ID zu dem Array abgelegt.

Für jede Probe auf dem Array wird die ID, der Titel, die GenBank AccessionNumber, auf welchem Chromosom das Gen lokalisiert ist (map) und die Gene Ontology Annotationen, die sich in biological process, cellular component und molecular function aufteilen, gespeichert. Identische Proben mit der selben ID können auf unterschiedlichen Arrays verwendet worden sein.

Information zu Protokollen

Um Protokolle zentral zu verwalten, werden Angaben über die Protokolle auch mit in der Datenbank gespeichert. Dazu wird eine Protokoll-ID vergeben und jeweils eine Beschreibung des Experiments sowie der Protokoll-Typ in der Datenbank abgelegt.

Informationen zu den verwendeten Samples und deren Meßergebnisse

Ein Patientensample hat eine eindeutige ID, außerdem sind der Titel, der Organismus, von dem das Sample stammt sowie der Submitter bekannt.

Ein Experiment wird durch das verwendete Sample, die dafür verwendeten Protokolle (multivalued-Attribute!), dem Submission-Date und der Chip, auf dem der Array erstellt worden ist, beschrieben. Ein Experiment kann zu einer bestimmten Serie gehören (gleiches Untersuchungsziel (Krebsart), Zellzyklus, ...)

Für jedes Experiment wird der Wert der Average Difference Intensity jeder Probe des Arrays (Wert des roten, grünen oder gelben Spots) in der Datenbank gespeichert.

Aufgabe:

1 Erstellen Sie anhand der Vorgaben ein konzeptionelles Modell mit UML.

7 P

2 Überführen Sie dieses in ein relationales Modell und

7 P

3 Erstellen Sie für den Datenaustausch eine DTD (Document Type Definition) und überprüfen Sie diese in einem XML-Validator (unter www.xmlspy.com gibt es eine Windows-version von xmlspy5.0 zur Validierung).

6 P