

## Übungen zur Vorlesung „Molekularbiologische Datenbanken“

### Aufgabenblatt 2: Microarray-Modellierung

---

Abgabetermin: **20.05.2003** bis **17 Uhr** über Goya oder in RUD25 IV.102

Bitte geben das konzeptionelle Modell mit UML,  
das relationale Model und DTD zum Datenaustausch ab.

Bitte geben Sie im PDF-Format oder auf Papier 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.

#### 1 Informationen über Teilnehmer

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

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

#### 2 Informationen über den Microarray

2.1 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-Submitter ist sein Name, seine email-Adresse, die Organisation, der Firmensitz und die Telefonnummer bekannt.

2.2 Wir haben in unserer Datenbank einige Arrays gespeichert. Jeder Array hat eine eigene ID, ausser dem 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 Submitter-ID zu dem Array abgelegt.

2.3 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.

### 3 Information zu Protokollen

3.1 Um Protokolle, die die einzelnen Labors verwenden 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.

### 4 Informationen zu den verwendeten Samples und deren Meßergebnisse

4.1 Ein Patientensample hat eine eindeutige ID, außerdem sind das Submission-Datum, der Titel, der Organismus, von dem das Sample stammt, eine Beschreibung der Experimente, die verwendeten Protokolle (multivalued-Attribute!), sowie der Submitter und die Plattform, auf dem der Array erstellt worden ist, bekannt.

4.2 Für jeden Meßpunkt des Arrays wird der Wert der Average Difference Intensity 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](http://www.xmlspy.com) gibt es eine Windows-version von xmlspy5.0 zur Validierung).

**6 P**