



## 2. Übungszettel Bioinformatik

### Einführung in die Datenbanksysteme Datenbanken für die Bioinformatik

H. Schweppe

## Übungsaufgaben

### Aufgabe 2.1

In dieser Übung (nach A. Robinson) sollen Sie eine fehlerbehaftete XML-Datei so korrigieren, dass sie konform zu einer gegebenen DTD ist. Sie finden die Datei `kraus.xml`, die ein Adressbuch beschreibt auf der Webseite (Übungen). Die DTD ist unter [http://www.inf.fu-berlin.de/lehre/SS05/19517-V/BioFoletc/address\\_book.dtd](http://www.inf.fu-berlin.de/lehre/SS05/19517-V/BioFoletc/address_book.dtd) abgelegt.

Am einfachsten laden Sie die Datei `kraus.xml` in den einen Browser, der XML-Dateien darstellen kann (IE, Mozilla und fast alle anderen). Sie erhalten dann eine Fehlermeldung. Korrigieren Sie sukzessive die Fehler.

### Aufgabe 2.2

Entwerfen Sie ein einfaches XML-Schema, das mindestens die Ausdruckskraft der DTD aus Aufgabe 2.1 hat.

Verändern Sie ihre Lösung `korrekt.xml` so, dass die Datei auf Konformität zum Schema geprüft werden kann.

Ein Beispiel für ein klitzekleines XML-Schema finden Sie unter <http://www.inf.fu-berlin.de/lehre/SS05/19517-V/BioFoletc/Ueb/bank.xsd>

### Aufgabe 2.3

Geben Sie mindestens vier Probleme an, die im Zusammenhang mit der Verwaltung von molekularbiologischen Daten auftreten und die sich in "klassischen" Datenbankanwendungen ("Kontenverwaltung") nicht oder nicht in gleichem Maße stellen.

Erläutern Sie die Probleme ggf. je in maximal zwei Sätzen.

### Aufgabe 2.4

Geben Sie an, welche Werte die folgenden XPath-Ausdrücke haben, die über der korrigierten XML-Datei von A2.1 ausgewertet werden. Adressdatei (möglichst verbal umschrieben, nicht die einzelnen Werte angeben).

```
/address_book/person  
//Person  
/Name  
//Name  
//family/parent::*  
//family/ancestor::*  
//* [count (e-mail)=2]  
//*count(*) = 2  
//[@manager="ma"]
```

Geben Sie XPath-Ausdrücke an für: a) alle Email-Adressen b) Personen, die Untergebene haben. (Achtung: die Modellierung im Adressbuch ist schlecht. Warum?). c) Den Adressbucheintrag der Person mit Nachnamen "Cameron" (ist das überhaupt möglich?)