

# Inf1B Data and Analysis

## Tutorial 4 (week 6)

Hutchins-Korte, Simpson

5 February 2008

- Please answer all questions on this worksheet in advance of the tutorial, and bring with you all work, including printouts of your XML file and DTD. Tutorials cannot function properly unless you do the work in advance.
- Data & Analysis tutorial exercises are not assessed, but are a compulsory and important part of the course. If you do not do the exercises then you are unlikely to pass the exam.
- Attendance at tutorials is obligatory; please let your tutor know if you cannot attend.
- *Recommended Reading:* Chapter 7.4–7.4.2 (XML) of ‘Database Management Systems’ (Raghu Ramakrishnan and Johannes Gehrke, 2003). Copies of these pages will be handed out during this week’s Thursday lecture (Feb 7th).

## Introduction

In the previous three tutorials you have been working with relational databases. In this tutorial you will be working with a different approach to storing and querying large bodies of data: XML. In Section 1 you will be designing an XML file and the associated DTD. In Section 2 you will look at how the XML approach compares to relational databases.

## 1 Designing an XML file and its DTD

Although designing an XML-file (Question 2) and designing the associated DTD (Question 1) are defined as two separate questions, they should really be considered in parallel, each one informing the other at every step of the design process.

id	name	dob	email
00001	Rose	01/01/1966	rose@inf
00002	Xiang	01/02/1950	xiang@inf
00003	Lev	01/03/1977	lev@phys

Staff

id	code	role
00001	inf1	organiser
00001	inf1	lecturer
00002	math1	organiser
00003	math1	lecturer

Duties

sid	id
s0456782	00001
s0412375	00001
s0378435	00002
s0189034	00003

DirectorOfStudies

sid	name	dob	email
s0456782	John	01/01/1989	john@inf
s0412375	Mary	01/02/1989	mary@inf
s0378435	Helen	01/03/1987	helen@phys
s0189034	Peter	01/04/1985	peter@math

Students

code	name	year
inf1	Informatics 1	1
math1	Mathematics 1	1

Courses

sid	code	mark
s0412375	inf1	80
s0378435	inf1	75
s0456782	inf1	45
s0189034	inf1	95
s0378435	math1	70
s0412375	math1	99
s0456782	math1	15

Takes

Figure 1: Mock Data

## Question 1 - Designing the DTD

Design a DTD for the information in Figure 1. The DTD should capture as much of the information as possible.

## Question 2 - Designing the XML File

Write down the information in Figure 1 in an XML file defined by the DTD you wrote for Question 1.

## Question 3: Checking your XML File

You can check your XML file for *validity* and *well-formedness* using the command `xml-xparse` on a DICE machine. To run the command, you first need to save your XML data, preceded by the DOCTYPE-statement which specifies your DTD, together in a single file. If you name this file `mockdata.xml`, you will be able to run the validity and well-formedness checks by entering the following command into a terminal window on a DICE machine:

```
xml-xparse mockdata.xml
```

## 2 XML vs. Relational Databases (RDBs)

### (Discussion) Question 4 - A Comparison

- (a) What are the strengths and weaknesses of the relational approach versus XML?
- (b) Given these strengths and weaknesses, name a context in which you would prefer to use XML.
- (c) Given these strengths and weaknesses, name a context in which you would prefer to use a RDB.