

Inf1B Data and Analysis

Tutorial 5 (week 7)

18 Feb 2009

- Please answer all questions on this worksheet in advance of the tutorial, and bring with you all work, including printouts of code and other results. 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:* Selected sections from **An Introduction to XML and Web Technologies** by *A. Miller and M. Schwartzbach*. as identified in the lecture slides.

Introduction

In the previous 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. You will be working with the file *restaurants.xml* that you can download from the course webpage.

Question 1

Design a DTD for an XML database of restaurants. The DTD should be designed so that the document *restaurants.xml* is valid with respect to it.

Question 2

Validate the given *restaurants.xml* file with your DTD file. You can check your XML file for validity and well-formedness using the command `xml-xparse` on a DICE machine. Copy the *restaurants.xml* file and include the document type declaration that specifies your DTD, together in a single file. 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 restaurants.xml
```

Question 3

Write path expressions to extract the following information from any XML document matching your restaurants DTD. Note: A good tutorial on XPath is available online at <http://www.w3schools.com/xpath/default.asp>

- (a) Retrieve the addresses of all restaurants.
- (b) Retrieve the names of all restaurants.
- (c) Retrieve the address of the *Kalpna* restaurant.
- (d) Retrieve the phone numbers of all restaurants that serve *Indian* food.
- (e) Retrieve the names of all restaurants which serve *Italian* food.

Question 4

The purpose of this question is to execute your path expressions on the file *restaurants.xml* using the Galax tool available on the DICE machines. Download *restaurants.xq* to your directory and open the file in a text editor. For each of the queries, follow the instructions given below to run Galax. Write down your results.

- (a) Add your query in the allocated space in the file *restaurants.xq*.
- (b) Comment out the other queries (as described in the note below).
- (c) Save the file.
- (d) Open a terminal window.
- (e) Change directory to your current working directory.

- (f) Type the following:
`galax-run restaurants.xq`

Note: **Galax** can only run one query at a time, so you will need to comment out your other queries, before you are able to run a new query. In general, you should always comment out all your queries, except the one you are currently trying to run. You can comment out a line by enclosing it in mirrored (noseless) smiley-faces. E.g. the following line is commented out.

```
(: This line is commented out :)
```

Question 5

How would the same data be naturally represented in a relational database?

Tutorial Discussion

- (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.