

UCT 2004 CSC303 Information Management

Practical Assignment 3 : XML

Due: 20 August 2004

Test Data:

5 XML metadata files ({89,254,259,267,268}.XML) plus a listing file (LIST.XML) for Question 3. Contained in testdata.zip.

Recommended Tools:

Xerces for XML DOM parsing and validation, Xalan for XSL Transformation.
Contained in jars.zip.

Sample files:

TEST.XML/XSD/XSL/JAVA/CLASS to demonstrate how to bootstrap the DOM parser and provide a fully working, validifiable and transformable dataset. “validate”, “transform” and “domtest” shell scripts. Contained in script_tools.zip.

Submit:

A single ZIP archive file containing all required files as indicated below.

Question 1

Write a program, using a **DOM 1/2 Parser**, to extract the following from the metadata files:

- Title
- List of authors
- List of URLs of files
- Last subject / sub-subject pair

Submit your program (q1.java) and copies of the output (e.g., 89.out) for each of the 5 metadata files. Store these in the subdirectory “q1”.

Question 2

Design a relational database to store the information contained in these XML metadata files. Then, write a program, using a **DOM 1/2 Parser**, to generate SQL statements to insert those files into the database. Do not create a database – just list the associated SQL statements.

Submit your program (q2.java), the database schema (q2.doc) and the list of SQL statements for each of the 5 metadata files (e.g., 89.sql). Store these in the subdirectory “q2”.

(continued on next page)

Question 3

Write an **XSLT stylesheet** that takes the listing file (LIST.XML) as input and generates a single HTML output document which lists the metadata contained in each of the 5 files in sequence. Your HTML file must also include a list of links at the top to jump immediately to the rendered versions of the metadata files; and a link below each rendering to jump back to the top. These should be implemented using “<a name>” tags.

The structure of your output will therefore be something like:

List of records:

 Abc
 Xyz

Record Abc:

.....

.....

Go back to top

Record Xyz:

.....

.....

Go back to top

Submit your stylesheet (q3.xsl) and the HTML output file (q3.html). Store these in the subdirectory “q3”.

Hint: use the *document* function in XSLT to load an external document.

Question 4

Write an **XML Schema** description that can be used to validate each of the 5 metadata files. Modify each file to include a reference to the schema for validation purposes (using *xmlns:xsi* and *xsi:schemaLocation*) and then validate each file.

Submit your schema (q4.xsd) , the modified XML files (with original names) and copies of the output (e.g., 89.out) to show that you have successfully validated each of the 5 metadata files. Store these in the subdirectory “q4”.

General Notes

In all questions, you are free to use other tools, as long as they are DOM1/2-, XSLT 1.0- and XML Schema 1.0-compliant. The supplied tools will also be loaded on the lab machines. (Talk to the TA first if you opt to use a different DOM parser)

Your assignment will be marked according to the following criteria:

Q1: Correctness of DOM code and output (15%)

Q2: Correctness of DOM code and output (15%)

Q3: Correctness of stylesheet and output (20%)

Q4: Correctness of schema and output (20%)

Documentation (20%) (comments within programs, stylesheet and schema)

Efficiency (10%)