

## **UCT 2005 CSC303 Information Management**

### **Practical Assignment 3 : XML**

#### *Test Data:*

5 XML metadata files ({89,254,259,267,268}.XML), 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:

- Full name of contributor (first names, space, last names)
- List of multimedia\_content fields
- List of URLs and sizes for each of the files

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**

Write a program, using a **DOM 1/2 Parser**, to convert an XML file into a stream of whitespace-separated words, removing all XML markup.

Submit your program (q2.java) and copies of the output (e.g., 89.out) for each of the 5 metadata files. Store these in the subdirectory “q2”.

### Question 3

Write an **XSLT stylesheet** that takes an XML file as input and generates an XHTML file listing all the metadata. This XHTML file should be viewable in a standard Web browser.

Submit your stylesheet (q3.xsl) and the XHTML output files for each XML file (e.g., 254.html). Store these in the subdirectory “q3”.

### 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. (Talk to Dom 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%)