15-415 Databases
Project #3 (Trivial XML Data Processing)

Assigned: Thursday, October 26, 2010
Due: Tuesday, November 9, 2010

Overview

This assignment asks you to augment your choice of your Project #1 or Project #2 to accept the bulk load of data in XML format, rather than CSV format. In particular, it asks you to consider XML technologies including DOM, SAX, XPath, XQuery, XSLT, and others, to parse XML data files. You may choose any of these technologies, or any other well-established technology, whether a W3C standard, or not. But, you may not write your own data-specific parser.

You should reject any malformed records. In order to do this, you should define and make use of an XML Schema (or a DTD, if you really insist) for your data format, e.g.
        http://download.oracle.com/javaee/1.4/tutorial/doc/JAXPDOM8.html

Teams

You may work in any subset of your Project #1 or Project #2 team. Other formulations are possible – please see a member of the course staff.

Sources of Data

The easiest thing to do is probably to write a program to reformat your CSV data from one of your earlier projects into XML. But, you are most welcome to find new data, already in XML format.

Deliverables

We'll describe the electronic submission process to you a little closer to the deadline. But, for now please know that we will ask you to submit the following –electronically--:

- A modified version of the source for the project
- A one page report describing what you did and how it works
- Some data with which to test
- Source code and documentation for any program you wrote to generate or reform data into XML
- The XML schema or DTD\