Fall, 2004 CIS 550

Database and Information Systems

Homework 4

October 14, 2004; Due October 21 at 1:30 PM

For this homework, you should test your answers using Galax, an XQuery processor. See http://www.seas.upenn.edu/~zives/assignments.htm for information about where to download the Galax system for Windows, Linux, or Solaris. Alternatively, you can ssh to **eniac-l.seas.upenn.edu** (*note the extra "-l": ordinary* ***eniac*** *will* ***not*** *work*) and run ~zives/galax/bin/galax on your query source file(s).

The XML data files for this problem set are in ~zives/galax. The files are ~zives/galax/db-proc.xml (list of proceedings) and ~zives/galax/db-inproc.xml (list of papers in proceedings). You may need to glance over the XML files (e.g., using *less*) to figure out the basic structure before you write your queries. Further information on Galax and its use can be found at its web site, http://db.bell-labs.com/galax. You can find further documentation on XQuery at its web site, http://www.w3.org/XML/Query.

Problem 1 [70 points]: Write the following queries in XQuery, with the output delimited by the tags ⟨answer⟩ . . . ⟨/answer⟩:

1. Output the years, titles, and conferences of all the papers (inproceedings in db-inprocs.xml) authored by Hector Garcia-Molina, nested inside a publication tag.
2. For each author of the paper entitled "Efficiently Publishing Relational Data as XML Documents.", list all conferences where they have published, grouped by author.
3. List the titles of all papers cited by "Efficiently Publishing Relational Data as XML Documents.".
4. Count the number of papers in db-inproc.xml.
5. Find the count of papers written by the author of the most papers.
6. List all conference proceedings titles in descending alphabetical order.
7. List the second author along with the paper title for each paper in SIGMOD 1999.

Problem 2 [30 points]:

The database of an online bookstore consists of the following entity sets:

- Book
- Publisher
- Author

Design an XML-Schema schema for the bookstore by choosing appropriate attributes or subelements for the entity sets. Each of the entities described above should have at least 7 attributes or subelements, including some sort of key. Also, explain how we might encode the "publisher-of" and "author-of" relationship sets in XML. How does this differ from how we would do it in the relational model?