

## Database and Information Systems

### Homework 4

October 7, 2003; Due October 16 at 1:30 PM

#### Problem 1:

For the following problems, you have access to an XQuery processor called Galax. 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>.

Write the following queries in XQuery, with the output delimited by the tags `<result> ... </result>`.

1. Output the titles of all the papers (inproceedings in `db-inprocs.xml`) authored by Jeffrey D. Ullman.
2. For each `proceedings` published in year 1999, output the title of the proceedings and the title of each paper in the proceedings. (The output should look just very similar to that from a relational join between the two XML files.) Include a `<proceedings>` tag around the proceedings-title/paper-title pairs so we can see the associations.
3. For each proceedings published in year 1999, output an XML tree containing the title and year of the `proceedings`, plus (as subelements) the title of each paper published in that proceedings. (This should contain paper titles nested within `proceedings` entries.)
4. For each paper authored by Jeffrey D. Ullman, list the paper title (inside a `<paper>` tag), along with each of Ullman’s co-authors.
5. Find the total count of papers published in each conference held during the year 1996; output the conference title and the count. Remember the XQuery aggregation function for counting is called `fn:count`; see the web document <http://www.w3.org/TR/xpath-functions>.
6. List all the authors in VLDB for year 2002, presented in order of their names, removing any duplicates.

7. Find the title and *last* author (authors are listed according to their ordering, and recall functions `fn:position()`, `fn:last()`, `fn:first()`) for each paper in VLDB 1996.

**Problem 2:**

The database of the an online music store consists of the following entity sets:

- CD
- Distributor
- Artist

Design a DTD for the music store 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 “distributes” and “authors” relationship sets in XML. How does this differ from how we would do it in the relational model?