

Database and Information Systems

Homework 4

October 27, 2005; Due November 3 at 1:30 PM

For this homework, you should test your answers using Galax, an XQuery processor. See course home page for information about where to download the Galax system and Galax manual. Alternatively, you can ssh to *eniac-l.seas.upenn.edu* and run `~zives/galax/bin/galax-run` on your query source file.

Problem 1 [70 points]: The XML data files used for this problem are **proc.xml** and **inproc.xml** in `~zives/galax/` on *eniac-l*. Based on the two XML files, write the following queries in XQuery with the output delimited by the tags `<answer>...</answer>`:

1. Find the distinct **titles** of proceedings in **proc** that have paper information in **inproc**. Result should be nested in **title** tag.
2. For every **author** of a paper, output an **author** tag with a subelement representing the **name**. Find the **title** all proceedings they have published in during the 1998 year, and nest these within a **published** tag underneath the **author**.
3. Find the **authors** whose papers were most cited. The result should be nested in an **author** tag.
4. Find the conference(s) with the fewest papers in 1998.
5. Write a query that *maps* information for the year 1996 from **inproc.xml** and **proc.xml** into a schema that follows the pattern (where * represents repetition):

```
<publications>
  <author> *
    <name> .. </name>
    <in> *
      <forum> .. </forum>
      <article> .. </article>
    </in>
  </author>
</publications>
```

6. Find the authors who had papers published in two continuous years. Results should be in the following format:

```
<answer>
  <publication>
    <author>name1</name>
    <paper>title1</paper>
    <paper>title2</paper>
  </publication>
  <publication>
    ....
  <publication>
</answer>
```

7. Sort papers in **inproc** in ascending order of **year** (use tag `<year>`). When two papers were published in the same year, further sort them in ascending order of **booktitle** (use tag `<booktitle>`). Don't lose any data in `inproc` after sorting. Results should look like

```
<answer>
  <dblp>
    sorted inproceedings here
  </dblp>
</answer>
```

Problem 2 [30 points]: Design an XML Schema for an auction system. Recall that the PBAY system included the following entity sets:

Sellers(*sellerID*: int, *rating*: char, *email*: string)

Items(*itemID*: int, *type*: string)

Buyers(*buyerID*: int, *email*: string, *address*: string)

Start with these, add at least 5 total additional attributes that you think belong there, and determine what the appropriate relationship sets should be. Use nesting of XML structure to encode at least one of the relationship sets. Specify keys and keyrefs.