Fall, 2005  CIS 550

Database and Information Systems

Final Examination

Name: ____________________________________

Please sign the following statement: I agree not to discuss this examination with anyone until after the deadline of 2PM on December 16, 2005. I will only consult my textbook and course notes in working on the exam.

(Signature)

This examination is designed to be completed in 3 hours. Please answer the following questions.

1. Given an XML schema, `amazon-books.xml` of the form (where an annotation of * indicates zero or more occurrences, + means one or more, ? zero or one occurrence):

```
<amazon-books>
  <book>+</n
    <isbn>...</isbn>
    <title>...</title>
    <author>...</author>+
    <publisher>...</publisher>
    <price>...</price>
  </book>
</amazon-books>
```

And a target, mediated view, defined by the XQuery function `AllItems()`, of the form:

```
<for-sale>
  <item>*
    <item-id>...</item-id>
    <item-type>...</item-type>
    <item-title>...</item-title>?
    <mfr-or-publisher>...</mfr-or-publisher>
    <author>...</author>*
    <price>...</price>
  </item>
</for-sale>
```
Plus an existing concordance table, `concord.xml`, of the form:

```xml
<item-ids>
  <book>*
    <item-id>...</item-id>
    <isbn>...</isbn>
  </book>*
  <cd>*
    <item-id>...</item-id>
    <cddb-id>...</cddb-id>
  </cd>
</item-ids>
```

Write the following XQueries:

a. Over the mediated schema, retrieve all mfr-or-publisher names associated with at least 10 different item-ids.

b. Over the mediated schema, retrieve all items that cost less than the average price.
c. The **schema mapping** view from the amazon-book source to the mediated schema.

2. Using the Hybrid Inlining scheme of [Shanmugasundaram et al.], define a relational schema to store amazon-books. Use the form R(a,b,c): you do not need to specify attribute domains, nor do you need to use SQL.
3. Why would one want to use local-as-view mappings in data integration?

4. What is the relationship between serializability and the ACID properties?

5. Explain how the XML-XQuery model is similar to and different from Codd’s relational model.
6. What inputs does a query optimizer need in order to choose a good query plan?

7. Assume you are given the following B+ Tree over relation R(id,val):

```
     | 8 |
   /   |   |   |
(1,5) (4,1) (5,2) (6,3) (8,2) (10,1)
```

Redraw the B+ Tree after the tuple (3,4) is inserted.
8. Apply the *maximal push-down* heuristic to optimize these expressions over R(a,b), S(b,c), T(a,d):

   a. $\pi_{a,c}(\sigma_{b < 5}(R \bowtie S))$

   b. $\pi_{b,c}(\sigma_{(a < b) \land (a < 3)}(R \bowtie S))$

   c. $\pi_{d}(\sigma_{c < 5}(R \bowtie S \bowtie T))$

9. Given the relational instance R(pid, cid):

<table>
<thead>
<tr>
<th>pid</th>
<th>cid</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

   Show the answers returned by the recursive datalog query:

   q(y) :- r(x,y), x = 0
   q(z) :- q(y), r(y,z)
10. Given the schema R(ABCDEF) and functional dependencies: \( A \rightarrow B, B \rightarrow C, D \rightarrow EF \),
    decompose R into 3NF.