UNIVERSITY OF TORONTO
Faculty of Arts and Science

MIDTERM TEST FALL 2014 (Evening Section - TUE)

CSC343H – Introduction to Databases
Instructor - Manos Papagelis

Duration – 50min
No aids allowed

This test is worth 15% of your final mark. Please answer all questions in the space provided. You may use the blank pages dispersed throughout the exam for rough work. In your answers try to be concise. **Good luck!**

<table>
<thead>
<tr>
<th>Marks</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>20</td>
<td>20</td>
<td>100</td>
</tr>
</tbody>
</table>

Last Name
First Name
Student Number

Page 1 of 10
(Left blank – use as scratch paper)
[30 Marks; 10 marks each query] Question 1 (Relational Algebra)
Assume the following relational schema that represents information about world countries and cities. The schema also models information about world spoken languages (e.g., English, French, etc.), along with information about what percentage of a country’s population that is knowledgeable of a language:

country (country_id, country_name)
city (city_id, city_name, population, country_id)
language (lang_id, lang_name)
countrylanguage (country_id, lang_id, percentage)

The following foreign key constraints hold:

\[ \pi_{\text{country_id}}(\text{city}) \subseteq \pi_{\text{country_id}}(\text{country}) \]
\[ \pi_{\text{country_id}}(\text{countrylanguage}) \subseteq \pi_{\text{country_id}}(\text{country}) \]
\[ \pi_{\text{lang_id}}(\text{countrylanguage}) \subseteq \pi_{\text{lang_id}}(\text{language}) \]

Write the following queries in Relational Algebra using only the basic operators (\( \pi \), \( \sigma \), \( \rho \), \( \bowtie \), \( \cup \), \( \cap \), \( - \), :=). Assume the set semantics (not bag-semantics) for Relational Algebra.

**Q1:** Find the name of the largest city (i.e., largest population) in “Brazil”.

\[
R1 := \sigma_{\text{country.country_name}=\text{"Brazil"}} (\text{city} \bowtie \text{country}) \\
R2 := \sigma_{\text{country.country_name}=\text{"Brazil"}} (\text{city} \bowtie \text{country}) \\
\text{NotMaxPopulation} := \pi_{\text{population}}(\sigma_{\text{R1.population} < \text{R2.population}} (\text{R1} \times \text{R2})) \\
\text{MaxPopulation} := \pi_{\text{population}}(\text{R1}) - \text{NotMaxPopulation} \\
\text{Result} := \pi_{\text{city_name}}(\text{MaxPopulation} \bowtie \text{city})
\]
Q2: Find the names of countries where exactly one language is spoken.

\[ \text{AtLeastOneLang} := \pi_{\text{country}_\text{id}}(\text{countrylanguage}) \]
\[ \text{R1} := \text{countrylanguage} \]
\[ \text{R2} := \text{countrylanguage} \]
\[ \text{AtLeastTwoLang} := \pi_{\text{R1.country}_\text{id}}(\sigma_{\text{R1.country}_\text{id} \neq \text{R2.country}_\text{id} \land \text{R1.lang}_\text{id} = \text{R2.lang}_\text{id}}(\text{R1} \times \text{R2})) \]
\[ \text{ExactlyOneLang} := \text{AtLeastOneLang} - \text{AtLeastTwoLang} \]
\[ \text{Result} := \pi_{\text{country}_\text{name}}(\text{ExactlyOneLang} \bowtie \text{country}) \]

Q3: Find the name of the languages that are not spoken in any country (i.e., extinct languages).

\[ \text{AllLanguages} := \pi_{\text{lang}_\text{id}}(\text{language}) \]
\[ \text{SpokenLanguages} := \pi_{\text{lang}_\text{id}}(\text{countrylanguage}) \]
\[ \text{NotSpokenLanguages} := \text{AllLanguages} - \text{SpokenLanguages} \]
\[ \text{Result} := \pi_{\text{lang}_\text{name}}(\text{NotSpokenLanguages} \bowtie \text{language}) \]
[30 Marks; 10 marks each query] Question 2 (SQL Queries)

Assume the following relational schema that represents information about world countries and cities. The schema also models information about world spoken languages (e.g., English, French, etc.), along with information about what percentage of a country’s population that is knowledgeable of a language:

- `country (country_id, country_name)`
- `city (city_id, city_name, population, country_id)`
- `language (lang_id, lang_name)`
- `countrylanguage (country_id, lang_id, percentage)`

Foreign key constraints:
- `city.country_id` is a foreign key that refers to `country.country_id`
- `countrylanguage.country_id` is a foreign key that refers to `country.country_id`
- `countrylanguage.lang_id` is a foreign key that refers to `language.lang_id`

Answer the following queries in SQL.

**Q1.** Find the percentage of the population in "Canada" that is knowledgeable of the "French" language. The schema of the result should be (frenchincanada).

```sql
SELECT countrylanguage.percentage AS frenchincanada
FROM country, language, countrylanguage
WHERE country.country_id=countrylanguage.country_id
  AND language.lang_id=countrylanguage.lang_id
  AND country.country_name='Canada'
  AND language.lang_name='French';
```
Q2. Find all the countries that have more than 500 cities. Report the name of the country along with the number of cities it has, in a descending order of that number. The schema of the result should be (countryname, numofcities).

SELECT country.country_name AS countryname, count(city.city_id) AS numofcities
FROM country, city
WHERE country.country_id=city.country_id
GROUP BY country.country_name
HAVING count(city.city_id) > 500
ORDER BY numofcities DESC;

Q3. Find the names of cities in “Canada” with a population over 50000 that are also names of cities in “USA” (irrelevant of their population). The schema of the result should be (synonymcities).

(SELECT city.city_name  AS synonymcities
FROM city, country
WHERE city.country_id=country.country_id AND country.country_name='Canada' AND city.population > 50000)
INTERSECT
(SELECT city.city_name AS synonymcities
FROM city, country
WHERE city.country_id=country.country_id AND country.country_name='USA')
[20 Marks] Question 3 (JOIN OPERATOR)
Assume the following database instance that models information about students in departments:

<table>
<thead>
<tr>
<th>student</th>
<th>department</th>
<th>Full outer join</th>
</tr>
</thead>
<tbody>
<tr>
<td>student_id</td>
<td>name</td>
<td>dept</td>
</tr>
<tr>
<td>34</td>
<td>Tom</td>
<td>CS</td>
</tr>
<tr>
<td>12</td>
<td>John</td>
<td>Math</td>
</tr>
<tr>
<td>22</td>
<td>Peter</td>
<td>Bio</td>
</tr>
<tr>
<td>43</td>
<td>Mary</td>
<td>Phys</td>
</tr>
<tr>
<td>29</td>
<td>Xin</td>
<td>Math</td>
</tr>
<tr>
<td>20</td>
<td>Sofia</td>
<td>CS</td>
</tr>
<tr>
<td>NULL</td>
<td>NULL</td>
<td>Law</td>
</tr>
</tbody>
</table>

What is the output of the following SQL queries (write a single number besides “Answer:”)?

Q1. SELECT count(*) FROM student JOIN department ON student.dept=department.dept;

Answer: 3

Q2. SELECT count(department.chair) FROM student LEFT OUTER JOIN department ON student.dept=department.dept;

Answer: 3

Q3. SELECT max(student_id) FROM student RIGHT OUTER JOIN department ON student.dept=department.dept;

Answer: 34

Q4. SELECT count(DISTINCT department.chair) FROM student LEFT OUTER JOIN department ON student.dept=department.dept;

Answer: 2

Q5. SELECT count(*) FROM student FULL OUTER JOIN department ON student.dept=department.dept;

Answer: 7
[20 Marks] Question 4 (True/False Statements)

For each of the following statements, indicate whether they are true or false. Please clearly circle or underline the answer you think is correct. The marking scheme is as follows:

- +2: for correct answer
- -1: for incorrect answer
- 0: for not answering

The minimum mark for this question is 0 (never below).

<table>
<thead>
<tr>
<th>Answer</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>Assume a relations $R(X, Y, Z)$. In relational algebra, it holds that: $\sigma_{A/B}(R) = \sigma_A(R) \cup \sigma_B(R)$</td>
</tr>
<tr>
<td>TRUE</td>
<td>Assume relations $R(X, Y, Z)$ and $S(X, Y, Z)$. In relational algebra, it holds that: $\sigma_A(R \cup S) = \sigma_A(R) \cup \sigma_A(S)$</td>
</tr>
<tr>
<td>TRUE</td>
<td>In relational algebra (bag-semantics) the following holds: {a, a, b, b, c, c} \cap {b, b, c, d} = {c}</td>
</tr>
<tr>
<td>TRUE</td>
<td>In relational algebra (bag-semantics) the following holds: {a, a, b, b, c} \cup {b, b, c, d} = {a, a, b, b, c, d}</td>
</tr>
<tr>
<td>TRUE</td>
<td>In relational algebra (bag-semantics) the following holds: {a, a, b, b, c} - {b, b, c, d} = {a, a}</td>
</tr>
<tr>
<td>TRUE</td>
<td>Consider tables $R$, $S$, $T$ with $T=\emptyset$. Then $R \times S \times T = \emptyset$.</td>
</tr>
<tr>
<td>TRUE</td>
<td>In SQL, it holds that: TRUE AND NULL $\rightarrow$ NULL</td>
</tr>
<tr>
<td>TRUE</td>
<td>In relational algebra, projection ($\pi$) is distributive over set intersection: $\pi_A(R \cap S) = \pi_A(R) \cap \pi_A(S)$</td>
</tr>
<tr>
<td>TRUE</td>
<td>In relational algebra, projection ($\pi$) is distributive over set difference: $\pi_A(R - S) = \pi_A(R) - \pi_A(S)$</td>
</tr>
<tr>
<td>TRUE</td>
<td>If $</td>
</tr>
</tbody>
</table>
(Left blank – use as scratch paper)
(Left blank – use as scratch paper)