Carnegie Mellon University
Department of Computer Science
15-415/615 - Database Applications
C. Faloutsos & A. Pavlo, Fall 2016

Homework 1 (by Prashanth Menon) - Solutions

Due: **hard copy, in class** at **3:00pm, on Monday, Sep. 19**

VERY IMPORTANT: Deposit **hard copy** of your answers, in class. For ease of grading, please

1. **Separate** your answers, on different page(s) for each question (staple additional pages, if needed).
2. **Type** the full info on **each** page: your **name**, **Andrew ID**, **course#**, **Homework#**, **Question#** on each of the 5 pages.

Reminders:

- *Plagiarism*: Homework is to be completed *individually*.
- *Typeset* all of your answers whenever possible. Illegible handwriting may get zero points, at the discretion of the graders.
- *Late homeworks*: in that case, please email it
  - to all TAs
  - with the subject line exactly `15-415 Homework Submission (HW 1)`
  - and the count of slip-days you are using.

For your information:

- Graded out of **100** points; **5** questions total
- Rough time estimate: *approx. 6 hours* - 1 to 2 hours per question

Revision : 2016/10/01 15:16

Question	Points	Score
Entity-Relationship Diagram	25	
SQL Tables from the ER Model	15	
Relational Algebra	30	
Relational Tuple Calculus (RTC)	10	
Relational Domain Calculus (RDC)	20	
Total:	100	

Question 1: Entity-Relationship Diagram . . . . . [25 points]

On separate page, with '[course-id] [hw#] [question#] [andrew-id] [your-name]'

GRADED BY: Jiexi LIN

Consider a database to store information for a social networking website. The database has the following properties:

- Every user has a unique user ID (integer) along with a full name, age and phone number.
- Every group has a unique group ID (integer) and a name. Every group must have at least one user that serves as moderator of the group.
- A user may be a member of zero or more groups; groups may contain zero or more members (and one or more moderators).
- Users are allowed to create zero or more albums. An album has a unique album ID (integer), a creation date, and a name. An album is owned by exactly one user: the user that created it.
- An album can contain zero or more media files. For every media file, we record its unique URL , the date the file was added to the album, and a caption (if one exists).
- Users can zero or more photos to albums. Photos are a type of media file, but we also track the encoding (e.g., JPEG, PNG, etc.) and the size of the photo (in bytes).
- Users may add zero or more videos to albums. Videos are a type of media file, and we track the codec used to encode the video (e.g., MPEG-4), the length of the video (in seconds), and the video's bitrate.
- A media file may belong to at most one album.

Given this description of the database and its constraints, we have created a mostly correct Entity-Relationship Diagram, shown in Figure 1. This diagram is at this hyperlink - feel free to use it as a starting point.

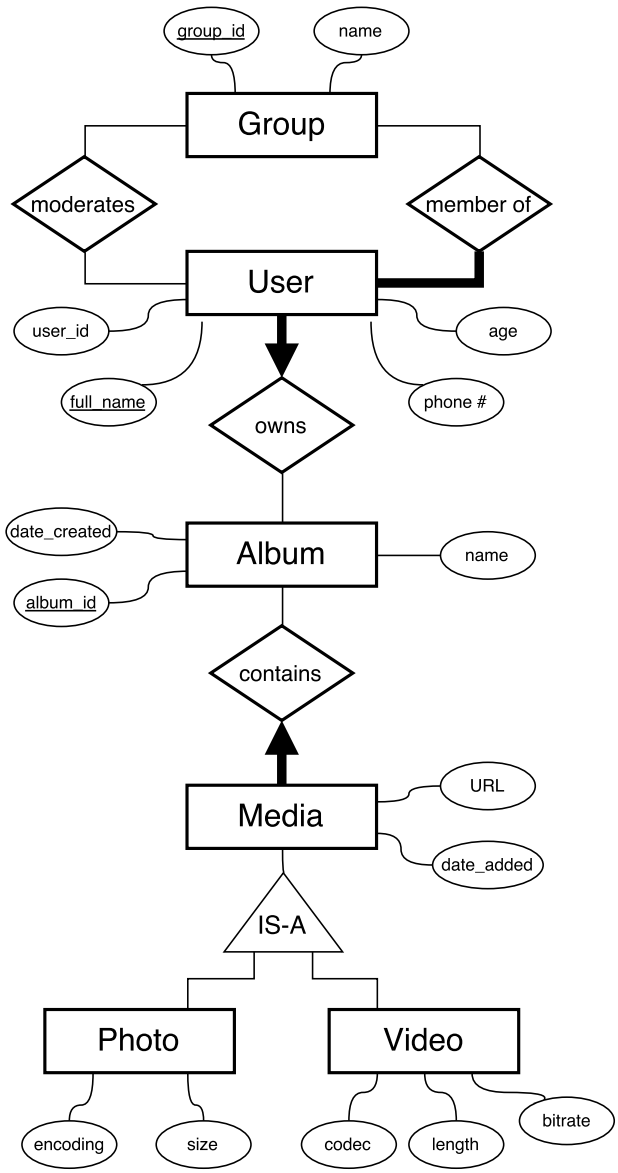


Figure 1: Almost correct ER diagram

(a) **[10 points]** Find and correct any mistakes in the given ER diagram. For this sub-question, list all the modifications (in the next one, we will ask you to list your additions/deletions). Specifically, here, number and list the modifications, like, e.g.

1. delete: arrow, from x to y
2. change to bold line: thin line, from z to w
3. change to bold box: entity e
4. underline: attribute a of entity e

Solution:

1. Change the bold line from `User` to `Group` through the `member of` relationship to a thin line.

2. Change the thin line from `Group` to `User` through the `moderates` relationship to a bold line.
3. Change the bold arrow from `User` to `Album` to a thin line.
4. Change the thin line from `Album` to `User` to a bold arrow.
5. Change the bold arrow from `Media` to `Album` to a thin arrow.
6. Delete the underline from the `full_name` attribute for entity `User`.
7. Underline the `user_id` attribute for the `User` entity.
8. Underline the `URL` attribute for the `Media` entity.

(b) **[5 points]** There may also be some missing element(s). If none, say 'none' - otherwise, add them to the picture, **and** list them, numbered. E.g.

1. add: attribute a, to entity e
2. add: bold line, arrow, from c to d.
3. add: weak entity, f, with attributes ....

Solution:

1. Add attribute `caption` to entity `Media`.

(c) **[10 points]** List and number all the bold lines and all the arrows that are in the final, corrected version of the diagram. E.g.

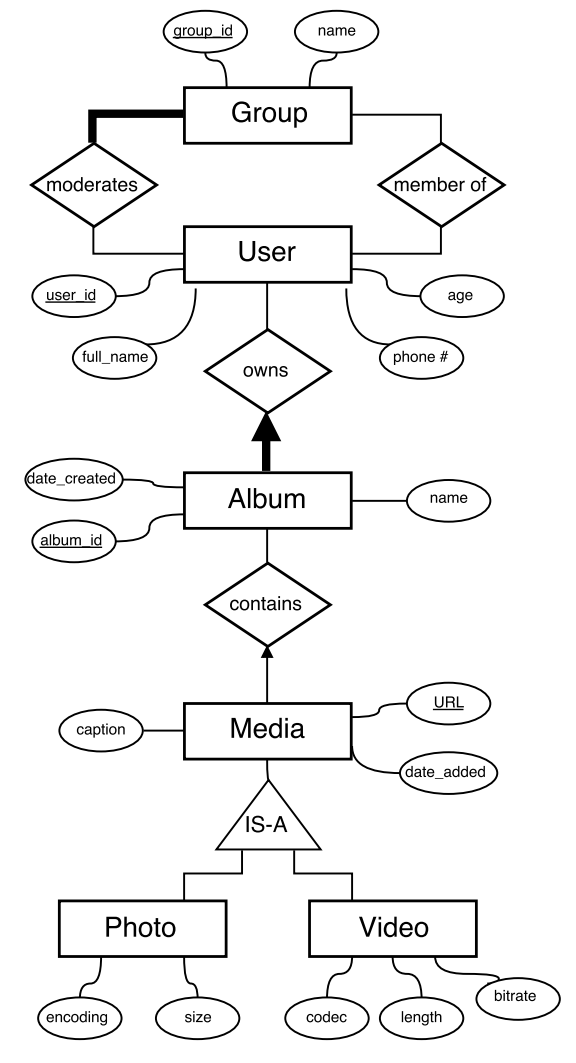
1. **bold**, line, from `Department` to `Employee`
2. thin, **arrow**, from x to y

Solution:

1. Bold line from `Group` to `User` through the `moderates` relationship.
2. Bold arrow from `Album` to `User`.
3. Thin arrow from `Media` to `Album`.

Clarifications/Hints:

- List your assumptions, if any. We will accept all reasonable assumptions.
- OPTIONALLY, to help graders save time, you may give your corrected ER diagram.



Solution:

Grading info:

- *(a)*
  - *-3 points for forgetting to delete the underline from the full_name attribute for entity User*
  - *-2 points for changing the number attribute to a partial key*
  - *-1 for incorrectly changing or forgetting to change the incorrect cardinalities*
  - *-1 points for making extra changes to the ER diagram that are incorrect, but this is a one-time penalty*
- *(b)*
  - *-5 points for forgetting to add caption attribute to* `Media` *or adding it to the wrong entity*
  - *-1 point for every incorrect change.*
- *(c)*
  - *-1 for each missing or incorrect arrow or bold line*
  - *Full points for either thick, or thin arrow from* `Media` *to* `Album` *(apologies for our confusing instructions).*

Question 2: SQL Tables from the ER Model . . . . . [15 points]

On separate page, with '[course-id] [hw#] [question#] [andrew-id] [your-name]'

GRADED BY: Kai KANG

Consider a database for an apartment-tracking application. It records information about apartment buildings, units and tenants. The constraints are exactly as shown in Figure 2. Apartment buildings, units and tenantes all have unique identifiers as shown in the figure.

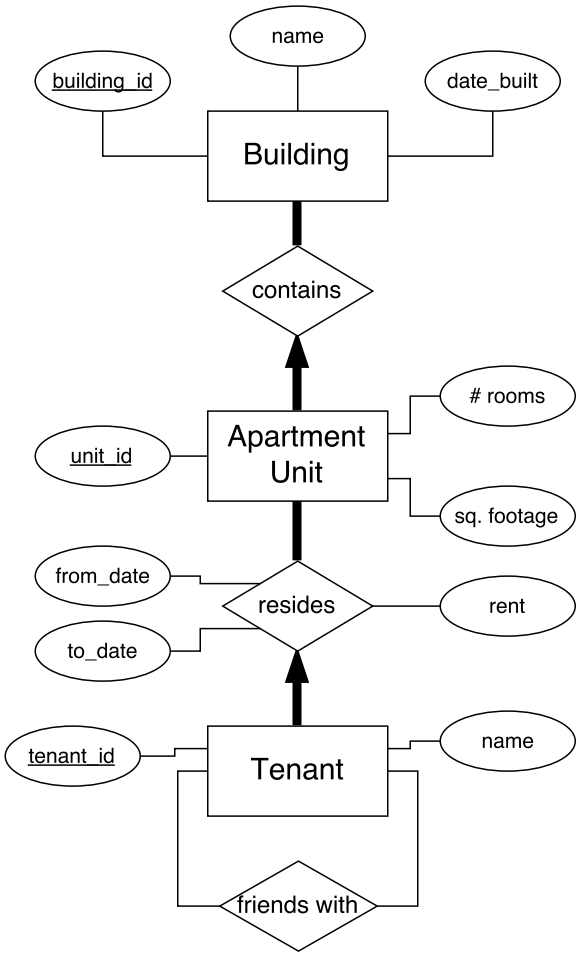


Figure 2: ER diagram for apartment tracking: turn to SQL tables

(a) **[15 points]** Give the DDL statements, that correspond to the above ER diagram.

- Use proper data types (we'll accept all reasonable choices).
- Avoid syntax errors (we'll forgive missing semicolons).
- Specify your decisions with respect to `CASCADE` deletions.
- **Without** using `CHECK` statements, enforce as many as possible of the implied integrity constraints as you can.

Solution:

```
CREATE TABLE Building (
    building_id INTEGER,
    name CHAR(20),
    date_built DATE,
    PRIMARY KEY (building_id)
);
CREATE TABLE ApartmentUnit (
    unit_id INTEGER,
    building_id INTEGER,
    num_rooms INTEGER,
    sq_footage INTEGER,
    PRIMARY KEY (unit_id),
    FOREIGN KEY (building_id)
        REFERENCES Building ON DELETE CASCADE
);
CREATE TABLE Tenant (
    tenant_id INTEGER,
    name CHAR(20),
    PRIMARY KEY (tenant_id)
);
CREATE TABLE Resides (
    tenant_id INTEGER,
    unit_id INTEGER,
    from_date DATE,
    to_date DATE,
    rent INTEGER,
    PRIMARY KEY (tenant_id),
    FOREIGN KEY (tenant_id) REFERENCES Tenant ON DELETE CASCADE,
    FOREIGN KEY (unit_id)
        REFERENCES ApartmentUnit ON DELETE CASCADE
);

CREATE TABLE Friends (
    tenant_1 INTEGER,
    tenant_2 INTEGER,
    PRIMARY KEY (tenant_1, tenant_2),
    FOREIGN KEY (tenant_1) REFERENCES Tenant ON DELETE CASCADE,
    FOREIGN KEY (tenant_2) REFERENCES Tenant ON DELETE CASCADE
);
```

Grading info:

- *(a) -5 points for missing a table*
- *(b) -1 point for missing primary key*
- *(c) -1 point for missing foreign key*
- *(d) -1 point for wrong ON DELETE actions*
- *(e) -1 point for wrong ON DELETE grammar*
- *(f) -1 point for wrong dateAdded type (we accept DATE, DATETIME, TIMESTAMP. But char and integer are not reasonable)*

Question 3: Relational Algebra . . . . . [30 points]

On separate page, with '[course-id] [hw#] [question#] [andrew-id] [your-name]'

GRADED BY: Prashanth MENON

Consider the relations of a database for the `2016 Olympics` as shown in Table 1. These relations record the athletes, events, and outcomes/results of the 2016 Olympic games. FYI, the *optional*, zero-point questions, will be used for partial credit, if needed

athlete_id	country	name	age
A1	U.S.A.	Michael Phelps	31
A2	U.S.A.	Justin Gatlin	34
A3	U.S.A.	Ryan Lochte	32
A4	Canada	Andre De Grasse	21
A5	Jamaica	Usain Bolt	30
A6	France	Christophe Lemaitre	26
A7	Japan	Masato Sakai	24
A8	Japan	Naito Ehara	60
A9	GBR	Duncan Scott	35
A10	GBR	James Guy	32

(a) `Athletes` Table

event_id	name
E1	100m Sprint
E2	200m Sprint
E3	200m Butterfly
E4	4x200 Freestyle Relay

(b) `Events` Table

event_id	athlete_id	result
E1	A5	Gold
E1	A2	Silver
E1	A4	Bronze
E2	A5	Gold
E2	A4	Silver
E3	A1	Gold
E3	A7	Silver
E3	A9	Bronze
E4	A1	Gold
E4	A3	Gold
E4	A7	Silver
E4	A8	Silver
E4	A9	Bronze
E4	A10	Bronze

(c) `Event_Results` Table: the outcome of every event

Table 1: Relations of the `Olympic` database.

We have the following tables:

- `Athletes`: For every athlete, we record a unique `athlete_id`, the country they represent, their name, and their age.
- `Events`: This table lists all the events that are part of the 2016 Olympic games. Every event has a unique integer `event_id` and a name.
- `Event_Results`: Lists the outcomes of all events. Every outcome records the `event_id` of the event, the `athlete_id` of the athlete that won a medal in the event, (or was part of the team that won), and the standing of the athlete (i.e., gold, silver or bronze).

Given this database instance, answer the following questions:

(a) **[2 points]** Which of the following is the meaning of the expression:

$\sigma_{\texttt{age}\geq 25}(\texttt{Athletes})$

1. It lists the `athlete_id` and `name` of all athletes that are at least 25 years old.
2. It lists all attributes of all athletes whose `age` is greater than or equal to 25.
3. It lists all athlete `age`s that are greater than or equal to 25 .
4. None of the above. The real answer is .....

Solution: #2

(b) **[5 points]** We want to list the `name`s of the athletes that have won at-least one gold medal (eliminating duplicate names) . Which, if any, of the following expressions achieve that ? Mark **all** valid expressions.

1. $\pi_{\texttt{name}}\big(\sigma_{\texttt{result}='Gold'}(\texttt{Athletes} \bowtie \texttt{Event\_Results})\big)$
2. $\pi_{\texttt{name}}\big(\texttt{Athletes} \bowtie \sigma_{\texttt{result}='Gold'}\big(\texttt{Event\_Results}\big)\big)$
3. $\pi_{\texttt{name}}\big(\sigma_{\texttt{result}='Gold'}\big(\texttt{Athletes} \bowtie \pi_{\texttt{athlete\_id,result}}\big(\texttt{Event\_Results}\big)\big)\big)$
4. $\pi_{\texttt{name}}\big(\texttt{Athletes}\big) \bowtie \sigma_{\texttt{result}='Gold'}\big(\texttt{Event\_Results}\big)$
5. $\pi_{\texttt{name}}\big(\texttt{Athletes}\big) - \pi_{\texttt{name}}\big(\texttt{Athletes} \bowtie \sigma_{\texttt{result}\neq'Gold'}\big(\texttt{Event\_Results}\big)\big)$

Solution: #1, #2 , and #3
Grading info:

- *+1 point for each expression correctly listed (or not listed)*

(c) For the following expression:

$$\sigma_{\texttt{age}<25}(\texttt{Athletes} \bowtie \texttt{Event\_Results})$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List `athlete_id`, `country`, `name`, `age`, `event_id`, and `result` of athletes younger than 25 that have won medals.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: Six columns: `athlete_id`, `event_id`, `country`, `name`, `age`, and `result`.

iii. **[3 points]** How many tuples are in the answer?

Solution: 4

iv. **[3 points]** List all the tuples in the answer, as a table.

Solution:

athlete_id	country	name	age	event_id	result
A4	Canada	Andre De Grasse	21	E1	Bronze
A4	Canada	Andre De Grasse	21	E2	Silver
A7	Japan	Masato Sakai	24	E3	Silver
A7	Japan	Masato Sakai	24	E4	Silver

Grading info:

- *ii. -1 for missing columns or having extra columns*
- *iii. -1 for being off by one on number of tuples; -3 for getting the wrong number of tuples by more than one*
- *iv. -1 for missing one tuple; -2 for performing wrong operation (e.g. $\times$ instead of $\bowtie$) or getting some columns and tuples wrong; -3 for all tuples and columns wrong*

(d) For the following expression:

$$\pi_{\texttt{athlete\_id},\texttt{event\_id}}(\texttt{Event\_Results}) \div \pi_{\texttt{event\_id}}(\sigma_{\texttt{athlete\_id='A5'}}(\texttt{Event\_Results}))$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List the `athlete_id`s of all athletes that have won medals in every event that athlete with ID"A5" (i.e., Usain Bolt) has also won a medal in.

ii. **[2 points]** How many, and which are the columns (= attributes) in the answer?

Solution: One column: `athlete_id`.

iii. **[3 points]** How many tuples are in the answer?

Solution: 2

iv. **[3 points]** List all the tuples in the answer, as a table.

Solution:

athlete_id
A4
A5

Grading info:

- *ii. -1 for missing columns or having extra columns*
- *iii. -1 for being off by one on number of tuples; -3 for getting the wrong number of tuples by more than one*
- *iv. -1 for missing one tuple; -1 for missing column or having extra column; -2 for getting some columns and tuples wrong; -3 for all tuples and columns wrong.*

(e) For the following expression:

$$\pi_{\texttt{A.athlete\_id}}\big(\rho_A(\texttt{Athletes})\big) - \pi_{\texttt{ER1.athlete\_id}}\big(\rho_{ER1}(\texttt{Event\_Results}) \bowtie_{ER1.\texttt{athlete\_id}=ER2.\texttt{athlete\_id} \wedge ER1.\texttt{result}\neq ER2.\texttt{result}} \rho_{ER2}(\texttt{Event\_Results})\big)$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: Finds all the athletes (by athlete_id) that did not win two or more types of medals.

ii. **[2 points]** How many, and which are the columns (= attributes) in the answer?

Solution: One column: A.`athlete_id`.

iii. **[3 points]** How many tuples are in the answer?

Solution: 9

iv. **[3 points]** List all the tuples in the answer, as a table.

Solution:

A.athlete_id
A1
A2
A3
A5
A6
A7
A8
A9
A10

Grading info:

- *ii. -1 for missing columns or having extra columns*
- *iii. -1 for being off by one on number of tuples; -3 for getting the wrong number of tuples by more than one*
- *iv. -1 for missing one tuple; -2 for getting some columns and tuples wrong; -3 for all tuples and columns wrong*

Question 4: Relational Tuple Calculus (RTC) . . . . . [10 points]

On separate page, with '[course-id] [hw#] [question#] [andrew-id] [your-name]'

GRADED BY: Huanchen ZHANG

We will again use the `Olympic` database from the previous question (see Table 1).

(a) For the following RTC expression

$$\{t \mid \exists\, er \in \texttt{Event\_Results}\,(\, er.\texttt{athlete\_id} = \text{“}A4\text{”} \,\wedge\, t.\texttt{event\_id} = er.\texttt{event\_id}\,)\}$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List all the events (by event_id) that the athlete with ID "A4" won a medal in.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: There is one column: `event_id`.

iii. **[2 points]** How many tuples are in the answer?

Solution: 2

iv. **[2 points]** List all the tuples in the answer, as a table.

Solution:

event_id
E1
E2

Grading info:

- *ii. -1 for wrong number of columns; -1 for wrong attribute names*
- *iii. -1 if number of tuples is off by 1; -2 if off by 2 or more*
- *iv. -1 if contains any extra column or missing any column; -1 for each extra or missing tuple (up to -2); no penalty for missing or wrong attribute names*

(b) For the following RTC expression

$$\begin{aligned}\{t \mid \exists\, er1, er2 \in \texttt{Event\_Results}\\ (\, er1.\texttt{event\_id} \neq er2.\texttt{event\_id}\\ \wedge\, er1.\texttt{athlete\_id} = er2.\texttt{athlete\_id}\\ \wedge\, t.\texttt{athlete\_id} = er1.\texttt{athlete\_id})\}\end{aligned}$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List the `athlete_id`s of all athletes that have won at least two medals.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: One column: `athlete_id`.

iii. **[2 points]** How many tuples are in the answer?

Solution: 5

iv. **[2 points]** List all the tuples in the answer, as a table.

Solution:

athlete_id
A1
A4
A5
A7
A9

Grading info:

- *ii. -1 for wrong number of columns; -1 for wrong attribute names*
- *iii. -1 if number of tuples is off by 1; -2 if off by 2 or more*
- *iv. -1 if contains any extra column or missing any column; -1 for each extra or missing tuple (up to -2); no penalty for missing or wrong attribute names*

Question 5: Relational Domain Calculus (RDC) . . . . . [20 points]

On separate page, with '[course-id] [hw#] [question#] [andrew-id] [your-name]'

GRADED BY: Lu ZHANG

For this question, we will reuse the 2016 Olympic database from Question 3 (see Table 1).

(a) For the following RDC expression
$\{\langle n\rangle \mid \exists ai, \exists c, \exists n, \exists a(\langle ai, c, n, a\rangle \in \texttt{Athletes} \wedge a \geq 35)\}$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List the `name`s of athletes that are 35 years old or older.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: One column: `name`.

iii. **[2 points]** How many tuples are in the answer?

Solution: 2

iv. **[2 points]** List all the tuples in the answer, as a table.

Solution:

name
Naito Ehara
Duncan Scott

Grading info:

- *ii. -1 for wrong number of columns; -1 for wrong attribute names*
- *iii. -1 if number of tuples is off by 1; -2 if off by 2 or more*
- *iv. -1 if contains any extra column or missing any column; -1 for each extra or missing tuple (up to -2); no penalty for missing or wrong attribute names; -1 for redundant tuples*

(b) For the following RDC expression:

$$\{\langle en\rangle \mid \exists ai, \exists an, \exists aa, \exists ei, \exists en\ (\ \langle ai, \text{“}Canada\text{”}, an, aa\rangle \in \texttt{Athletes} \wedge \langle ei, en\rangle \in \texttt{Events} \wedge \langle ei, ai, \text{“}Silver\text{”}\rangle \in \texttt{Event\_Results})\}$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List all the `event_name`s of the events in which a Canadian athlete won a Silver medal.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: One column: `event_name`.

iii. **[2 points]** How many tuples are in the answer?

Solution: 1

iv. **[2 points]** List all the tuples in the answer, as a table.

Solution:

event_name
200m Sprint

Grading info:

- *ii. -1 for wrong number of columns; -1 for wrong attribute names*
- *iii. -1 if number of tuples is off by 1; -2 if off by 2 or more*
- *iv. -1 if contains any extra column or missing any column; -1 for each extra or missing tuple (up to -2); no penalty for missing or wrong attribute names*

(c) For the following RDC expression:

$$\left\{ \langle c, an1, aa1\rangle \;\middle|\; \exists ai1 \Big(\langle ai1, c, an1, aa1\rangle \in \texttt{Athletes} \right.$$
$$\left. \wedge\ \forall ai2 \left(\exists aa2, \exists an2 \big((\langle ai2, c, an2, aa2\rangle \in \texttt{Athletes}\ ) \ \Rightarrow\ (\ aa1 \geq aa2\ ) \big) \right) \Big) \right\}$$

i. **[0 points]** *Optional*: describe in English what the expression does

Solution: List the `country`, `athlete_name` and `age` of the oldest athlete for every country.

ii. **[1 point]** How many, and which are the columns (= attributes) in the answer?

Solution: Three columns: `country`, `athlete_name`, and `age`.

iii. **[4 points]** How many tuples are in the answer?

Solution: 6

iv. **[5 points]** List all the tuples in the answer, as a table.

Solution:

country	athlete_name	age
U.S.A.	Justin Gatlin	34
Canada	Andre De Grasse	21
Jamaica	Usain Bolt	30
Fance	Christophe Lemaitre	26
Japan	Naito Ehara	60
GBR	Duncan Scott	35

Grading info:

- *ii. -1 for wrong number of columns; -1 for wrong attribute names*
- *iii. -1 if number of tuples is off by 1; -2 if off by 2 or more*
- *iv. -1 if contains any extra column or missing any column; -1 for each extra or missing tuple (up to -2); no penalty for missing or wrong attribute names*