Relational Algebra In-class Exercise (Part II)

R & G, Chapter 4

Announcement

- Final exam date
  - Stevens official date: 12/21/22
  - My proposal: last week of class
  - Tentative course schedule for the proposed exam date is published in Course Syllabus
  - Finish the survey of final exam date (under "Quiz" in Canvas)
- Practice exercise for relational algebra expressions (under "Quiz" in Canvas)
- Assignment 3 Relational Algebra is available in Canvas (due at 11:59pm, Oct 19)

Join + Set Operations (III)

Q10 (in Last class): Find the name of sailors who are older than 20 and have not reserved a red boat

- Find sids of sailors with age over 20 as set S1
- Find sids of sailors who have reserved a red boat as set S2
- Take the set difference of S1 and S2
- Join with sailors, and return name of the sailors from join result

Join + Set Operations

Q10: Find the name of sailors who are older than 20 and have not reserved a red boat

- Find sids of sailors with age over 20 as set S1
- Find sids of sailors who have reserved a red boat as set S2
- Take the set difference of S1 and S2 as S3
- Join S3 with sailors, and return name of the sailors from join result

– Answer:

$$\rho(S1,\pi_{sid}(\sigma_{age>20}(Sailors)))$$

$$\rho(S2,\pi_{sid}((\sigma_{color='red'}(Boats))\bowtie \mathrm{Reserves})$$

$$\pi_{sname}(Sailors\bowtie (S1-S2))$$

Find the name of sailors with age over 20 who have not reserved a red boat

- **Answer 1 (join before set difference)**

$$\rho(T1,\pi_{sid}(\sigma_{age>20}(Sailors)\bowtie \text{Reserves}\bowtie \text{Boats})$$

$$\rho(T2,\pi_{sid}(\sigma_{color='red'}(Boats)\bowtie \text{Reserves}\bowtie Sailors)$$

$$\pi_{sname}(Sailors\bowtie (T1-T2))$$

  - Is Answer 1 correct? What will it return on the instances below?
  - Answer 1 is wrong because T1 does not include those sailors (e.g., Lubber) who have not reserved any boat

Sid	Bid	day
22	101	10/10/96
58	102	11/12/96

Reserves

Bid	Bname	Color
101	Interlake	Blue
102	Interlate	Red
103	Clipper	Green
104	Marine	red

Boats

Sid	Sname	Rating	Age
22	Dustin	7	45
31	Lubber	8	55
58	Rusty	10	35

Sailors

Find the name of sailors with age over 20 who have not reserved a red boat

- **Answer 2 (without using set-difference)**

$$\pi_{sname}(\sigma_{age>20}(Sailors) \bowtie Reserves \bowtie \sigma_{color \neq 'red'}(Boats))$$

  - Is Answer 2 correct? What will it return on the instances below?
  - Answer 2 is wrong because it may return the sailors (e.g., Dustin) who have reserved a red boat and a non-red boat.

Sid	Bid	day
22	101	10/10/96
58	103	11/12/96
22	102	12/10/96

Reserves

Bid	Bname	Color
101	Interlake	Blue
102	Interlate	Red
103	Clipper	Green
104	Marine	red

Boats

Sid	Sname	Rating	Age
22	Dustin	7	45
58	Rusty	10	35

Sailors

Division Operator (/)

- Useful for expressing "for all" queries like:
  Find sids of sailors who have reserved all boats.
- Division A/B
  - Put the table that includes "ALL XX" as the divisor B
  - Attributes of B MUST be a subset of attrs of A.

Division (I)

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q1: find the names of sailors who've reserved all boats

Division (I)

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q1. Find the names of sailors who've reserved all the boats

– The way of thinking:
  - Use Division operator (since it asked "all boats")
  - **Question**: is the following expression correct?

$$\rho(Tempsids,(\mathrm{Re}serves/Boats))$$

$$\pi_{sname}(Tempsids \bowtie Sailors)$$

Division (I) Cont.

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q1. Find the name of sailors who've reserved all boats

– The way of thinking:
  - Find the set A: (sid, bid) of sailors who have reserved some boats.
  - Find the set B: (bid) of all boats.
  - Take A / B to get *sid of* sailors who have reserved all boats
  - Join with Sailor table to get their names

$$\rho(A, \pi_{sid,\, bid}\text{Reserves})$$
$$\rho(B, \pi_{bid}\text{Boats})$$
$$\pi_{sname}((A/B) \bowtie \text{Sailors})$$

Division (II)

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q2. find the name of the sailors who've reserved all red boats

Division (II)

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q2. Find the names of sailors who've reserved all red boats

– The way of thinking:
  - Find the set A: (sid, bid) of all sailors who have reserved some boats.
  - Find the set B: (bid) of all red boats.
  - Take A / B to get sid of sailors who have reserved all red boats
  - Join with Sailor table to get their names

$$\rho(A, \pi_{sid,\, bid}(Reserves))$$
$$\rho(B, \pi_{bid}\sigma_{color='red'}(Boats))$$
$$\pi_{sname}((A/B) \bowtie Sailors)$$

Division (III)

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q3. Find the name of the sailors who have reserved boats of all colors

Division (III)

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q3. Find the name of sailors who have reserved boats of all colors

- The way of thinking:
  - Find the set A: (sid, color) pairs of sailors who have reserved some boats.
  - Find the set B: all colors of boats.
  - Take A / B to get sid of sailors who have reserved boats of all colors
  - Join with Sailor table to get sname

$$\rho(A, \pi_{sid,\, color}(Reserves \bowtie Boats))$$
$$\rho(B, \pi_{color}(Boats))$$
$$\pi_{sname}((A/B) \bowtie Sailors)$$

Can we do join with sailors to get sname before division? How?

Division (III)

Q3. Find the name of sailors who have reserved boats of all colors

$$\rho(A, \pi_{sid,\ color}(Reserves \bowtie Boats))$$
$$\rho(B, \pi_{color}(Boats))$$
$$\pi_{sname}\ (A/B \bowtie Sailors)$$

Can we do join with sailors before division? How?

$$\rho(A, \pi_{sname,\ color}(Sailors \bowtie Reserves \bowtie Boats))$$
$$\rho(B, \pi_{color}(Boats))$$
$$A/B$$

Aggregate Queries by Relational Algebra

- Relational algebra does not provide any aggregate function in general.

- Relational algebra handles aggregate queries in a complicated way

Aggregate Queries

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q4. Find the names of sailors who have reserved at least two different boats

Aggregate Queries

- **Schema**
  - Boats (bid, bname, color)
  - Sailors(sid, sname, rating, age)
  - Reserves(sid, bid, day)

Q4. Find the names of sailors who have reserved at least two different boats

- The way of thinking:
  - *Find sailors S1 who reserved at least one boat*
  - *From S1, find sailors S2 who reserved at least two boats*
  - *From S2, select sailors with two different boats reserved*

S1: Sailors who have reserved at least ONE boat →

$$\rho(R,(\pi_{sid,sname,bid}(Reserves \bowtie Sailors))$$

S2: Sailors who have reserved at least TWO boats →

$$\rho(RPairs,(1\text{->}sid1,2\text{->}sname1,3\text{->}bid1,4\text{->}sid2,5\text{->}sname2,6\text{->}bid2),R \bowtie_{Sid1=Sid2} R)$$

Renaming to eliminate duplicate attributes

Condition join to make sure the same sailors

$$\pi_{sname1}(\sigma_{bid1\neq bid2} RPairs)$$

Final output: same sailor but two different boats

Running Example (1/3)

Reserves

Sid	Bid	day
22	101	10/10/96
58	103	11/12/96
22	102	12/10/96

Boats

Bid	Bname	Color
101	Interlake	Blue
102	Interlate	Red
103	Clipper	Green
104	Marine	red

Sailors

Sid	Sname	Rating	Age
22	Dustin	7	45.0
31	Lubber	8	55.5
58	Rusty	10	35.5

S1: Sailors who have reserved at least ONE boat

$$\rho(R,(\pi_{sid,sname,bid}(Reserves \bowtie Sailors))$$

Sid	Sname	bid
22	Dustin	101
22	Dustin	102
58	Rusty	103

R

Running Example (2/3)

S2: Sailors who have reserved at least TWO boats

R

Sid	Sname	bid
22	Dustin	101
22	Dustin	102
58	Rusty	103

$$\rho\,(RPairs,(1{->}sid1,2{->}sname1,3{->}bid1,4{->}sid2,\ 5{->}sname2,6{->}bid2),R \bowtie_{Sid1=Sid2} R)$$

Sid1	Sname1	bid1	Sid2	Sname2	bid2
22	Dustin	101	22	Dustin	101
22	Dustin	101	22	Dustin	102
22	Dustin	102	22	Dustin	101
22	Dustin	102	22	Dustin	102
58	Rusty	103	58	Rusty	103

RParis

Running Example (3/3)

Sid1	Sname1	bid1	Sid2	Sname2	bid2
22	Dustin	101	22	Dustin	101
22	Dustin	101	22	Dustin	102
22	Dustin	102	22	Dustin	101
22	Dustin	102	22	Dustin	102
58	Rusty	103	58	Rusty	103

RParis

Final output: same sailor but two different boats

$$\pi_{\text{sname1}} (\sigma_{bid1 \neq bid2} RPairs)$$

- **Return "Dustin" (twice)**

Further Thinking

- **Find the names of sailors who have reserved at least two different boats**

S1: Sailors who have reserved at least ONE boat →

$$\rho\,(R,(\pi_{sid,sname,bid}(\text{Re}serves \bowtie Sailors)$$

S2: Sailors who have reserved at least TWO boats →

$$\rho\,(RPairs,(1\text{–}>sid1,2\text{–}>sname1,3\text{–}>bid1,4\text{–}>sid2,\ 5\text{–}>sname2,6\text{–}>bid2),R \rhd\lhd_{Sid1=Sid2} R)$$

Renaming to eliminate duplicate attributes

$$\pi_{sname1}\ (\sigma_{bid1\neq bid2}\ RPairs))$$

Final output: Same sailor but two different boats

- Questions:

(1) If we remove 'sname' from R, which expression(s) will be changed?

(2) If we change equal-join in RPairs to be cross-product, do we need to change any other expression?

(3) How can we write the query: Find the names of sailors who have reserved at least three different boats?