Exercise Sheet 1: Relational Algebra

David Carral, Markus Krötzsch

Database Theory, 17 April, Summer Term 2018

Exercise 1.1. Consider a cinema database with tables of the following form (adapted from a similar example in the textbook of Abiteboul, Hull and Vianu):

Films

Title	*Director*	*Actor*
The Imitation Game	Tyldum	Cumberbatch
The Imitation Game	Tyldum	Knightley
...	...	...
The Internet's Own Boy	Knappenberger	Swartz
The Internet's Own Boy	Knappenberger	Lessig
The Internet's Own Boy	Knappenberger	Berners-Lee
...	...	...
Dogma	Smith	Damon
Dogma	Smith	Affleck
Dogma	Smith	Morissette
Dogma	Smith	Smith

Venues

Cinema	*Address*	*Phone*
UFA	St. Petersburger Str. 24	4825825
Schauburg	Königsbräcker Str. 55	8032185
CinemaxX	Hüblerstr. 8	3158910
...	...	...

Program

Cinema	*Title*	*Time*
Schauburg	The Imitation Game	19:30
Schauburg	Dogma	20:45
UFA	The Imitation Game	22:45
CinemaxX	The Imitation Game	19:30

Express the following queries in relational algebra:

1. Who is the director of "The Imitation Game"?
2. Which cinemas feature "The Imitation Game"?
3. What are the address and phone number of "Schauburg"?
4. Is a film directed by "Smith" playing in some cinema?
5. List the pairs of persons such that the first directed the second in a film and vice versa.
6. List the names of directors who have acted in a film they directed.
7. Always return $\{\mathit{Title} \mapsto \text{"Apocalypse Now"}, \mathit{Director} \mapsto \text{"Coppola"}\}$ as the answer.
8. Find the actors cast in at least one film by "Smith."
9. Find the actors cast in every film by "Smith."
10. Find the actors cast only in films by "Smith."
11. Find all pairs of actors who act together in at least one film.
12. Find all pairs of actors cast in exactly the same films.
13. Find the directors such that every actor is cast in one of his or her films.

Solution.

1. Who is the director of "The Imitation Game"?

$$\pi_{\mathit{Director}}(\sigma_{\mathit{Title}=\text{“The Imitation Game”}}(\mathit{Films}))$$

2. Which cinemas feature "The Imitation Game"?

$$\pi_{\mathit{Cinema}}(\sigma_{\mathit{Title}=\text{“The Imitation Game”}}(\mathit{Program}))$$

3. What are the address and phone number of “Schauburg”?

$$\pi_{Address,Phone}(\sigma_{Cinema=\text{“Schauburg”}}(Venues))$$

4. Is a film directed by “Smith” playing in some cinema?

$$\pi_{\emptyset}(\sigma_{Director=\text{“Smith”}}(Films) \bowtie Program)$$

   Possible answers: $\{\}$ (if there are no such films) and $\{\varepsilon\}$ (if there are).

5. List the pairs of persons such that the first directed the second in a film and vice versa. $\pi_{Director,RD}(\sigma_{Director=RA}(\sigma_{Actor=RD}(\delta_{Title,Director,Actor\to RT,RD,RA}(Films) \bowtie Films)))$

6. List the names of directors who have acted in a film they directed.

$$\pi_{Director}(\sigma_{Actor=Director}(Films))$$

7. Always return $\{Title \mapsto \text{"Apocalypse Now"}, Director \mapsto \text{"Coppola"}\}$ as the answer.

$$\{\{Title \mapsto \text{"Apocalypse Now"}\}\} \bowtie \{\{Director \mapsto \text{"Coppola"}\}\}$$

8. Find the actors cast in at least one film by “Smith.”

$$\pi_{Actor}(\sigma_{Director=\text{“Smith”}}(Films))$$

9. Find the actors cast in every film by “Smith.”

$$\begin{aligned}&\pi_{Actor}(Films) - \\ &\quad \pi_{Actor}[\pi_{Actor}(Films) \bowtie \pi_{Title}(\sigma_{Director=\text{“Smith”}}(Films)) - \pi_{Actor,Title}(\sigma_{Director=\text{“Smith”}}(Films))]\end{aligned}$$

10. Find the actors cast only in films by “Smith.”

$$\pi_{Actor}(Films) - \pi_{Actor}[Films - \sigma_{Director=\text{“Smith”}}(Films)]$$

11. Find all pairs of actors who act together in at least one film.

$$\pi_{RA,Actor}[\delta_{Actor\to RA}(Films) \bowtie Films - \sigma_{Actor=RA}(\delta_{Actor\to RA}(Films) \bowtie Films)]$$

12. Find all pairs of actors cast in exactly the same films. Consider first the query q returning a table with all pairs of actors a and a' such that a acts in every movie where a' acts.

$$\begin{aligned}q :=& \pi_{Actor}(Films) \bowtie \delta_{Actor\to RA}(\pi_{Actor}(Films)) - \\ &\quad \pi_{Actor,RA}[\pi_{Actor}(Films) \bowtie \delta_{Actor\to RA}(Films) - Films \bowtie \delta_{Actor\to RA}(\pi_{Actor}(Films))]\end{aligned}$$

    The query is then $q \bowtie \delta_{Actor,RA\to RA,Actor}(q)$

13. Find the directors such that every actor is cast in one of his or her films.

$$\begin{aligned}&\pi_{Actor}(Films) - \\ &\quad \pi_{Actor}[\pi_{Actor}(Films) \bowtie \pi_{Title}(\sigma_{Director=\text{“Smith”}}(Films))) - \pi_{Actor,Title}(\sigma_{Director=\text{“Smith”}}(Films))\end{aligned}$$

Exercise 1.2. We use ε to denote the *empty function*, i.e., the function with the empty domain, which is defined for no value. We use $\emptyset$ to denote the empty table with no rows and no columns.
Now for a table R, what are the results of the following expressions?

$$R \bowtie R \qquad\qquad R \bowtie \emptyset \qquad\qquad R \bowtie \{\varepsilon\}$$

Solution.

$$R \bowtie R = R \qquad\qquad R \bowtie \emptyset = \emptyset \qquad\qquad R \bowtie \{\varepsilon\} = R$$

To show this, we recall the definition of the natural join from the lecture:

$$R \bowtie S = \{f \cup g \mid f \in R \text{ and } g \in S \textit{ such that } f(a) = g(a) \text{ for all } a \in \mathit{Att}(R) \cap \mathit{Att}(S)\}$$

In the second case, $g \in S$ cannot be satisfied, since S is empty. In the third case, we have $S = \{\varepsilon\}$. Then $g \in S$ implies that $g = \varepsilon$, the empty function. By definition of the empty function, $\mathit{Att}(S)$ is empty.

Exercise 1.3. Express the following operations using other operations presented in the lecture:

- Intersection $R \cap S$.
- Cross product (Cartesian product) $R \times S$.
- Selection $\sigma_{n=a}(R)$ with a a constant.
- Arbitrary constant tables in queries (the constants in the lecture only had one single column and one single row; generalise this to any number of constants and rows)

Solution.

- $R \cap S = R \bowtie S$.
- Let U be the attributes of R and V be the attributes of S. Let W be such that $|W| = |V|$ and $W \cap U = \emptyset$. Then $R \times S = R \bowtie \delta_{\vec{V} \to \vec{W}}(S)$.
- $\sigma_{n=a}(R) = R \bowtie \{\{n \mapsto a\}\}$
- To create a constant table with a single row and many attribute-value pairs simply join several single attribute-value pair constant tables. Then use union to create a table with several rows.

Exercise 1.4. Consider the following identities and decide for each whether it is true or false. If true, prove your answer using the definitions from the lecture; if false, give a counterexample.

1. $R \bowtie S = S \bowtie R$
2. $R \bowtie (S \bowtie T) = (R \bowtie S) \bowtie T$.
3. $\pi_X(R \circ S) = \pi_X(R) \circ \pi_X(S)$ for all $\circ \in \{\cup, \cap, -, \bowtie\}$.
4. $\sigma_{n=m}(R \circ S) = \sigma_{n=m}(R) \circ \sigma_{n=m}(S)$ for all $\circ \in \{\cup, \cap, -\}$.
5. $\sigma_{n=m}(R \bowtie S) = \sigma_{n=m}(R) \bowtie S$, for n and m attributes of R only.

Why are these identities of interest?

Solution. Be quite formal in proving these – learning goal is also to understand how to derive "obvious" things properly.

1. Yes. By definition, $R \bowtie S = \{f \cup g \mid f \in R$ and $g \in S$ *such that* $f(a) = g(a)$ for all $a \in$ *Att*$(R) \cap$ *Att*$(S)\}$. Clearly, $R \bowtie S = S \bowtie R$.
2. Yes. Similar to the one above.
3. $\pi_X(R \cup S) = \pi_X(R) \cup \pi_X(S)$
   Let $f \in \pi_X(R \cup S)$, then there exists $f' \in R \cup S$ such that f' restricted to X is equal to f. Then f' belongs to R or to S (or to both). Without loss of generality, assume that f' belongs to R. Then $f \in \pi_X(R)$. The other direction is the same.

   R

A	**B**
1	2

   S

A	**B**
1	3

   $\pi_A(R \cap S) = \{\} \subsetneq \pi_A(R) \cap \pi_A(S) = \{\{A \mapsto 1\}\}$
   $\pi_A(R \bowtie S) \subsetneq \pi_A(R) \bowtie \pi_A(S)$ – for the example, $\bowtie$ behaves as $\cap$
   $\pi_A(R - S) = \{\{A \mapsto 1\}\} \supsetneq \pi_X(R) - \pi_X(S) = \{\}$
4. yes
5. yes

These identities can be used to optimize queries, e.g., by pushing selection inwards using distributivity. The join then handles smaller tables.

Exercise 1.5. Let $R^{\mathcal{I}}$ and $S^{\mathcal{I}}$ be tables of schema $R[U]$ and $S[V]$, respectively. The *division* of $R^{\mathcal{I}}$ by $S^{\mathcal{I}}$, written as $(R^{\mathcal{I}} \div S^{\mathcal{I}})$, is defined to be the maximal table over the attributes $U \setminus V$ that satisfies $(R^{\mathcal{I}} \div S^{\mathcal{I}}) \bowtie S^{\mathcal{I}} \subseteq R^{\mathcal{I}}$. Note that the joined tables here do not have any attributes in common, so the natural join works as a cross product.

Consider the following table and use the division operator to express a query for the cities that have been visited by all people.

Visited

Person	**City**
Tomas	Berlin
Markus	Santiago
Markus	Berlin
Fred	New York
Fred	Berlin

Express division using the relational algebra operations introduced in the lecture. **Solution.**

- $\mathit{Visited} \div \pi_{\mathit{Person}}(\mathit{Visited})$
- Let X be the attributes unique to R, then $R \div S = \pi_X(R) - \pi_X[\pi_X(R) \times S - R]$

Exercise 1.6. Suggest how to write the relational algebra operations for using the unnamed perspective. What changes?

Solution. Key aspects: cross product $\times$ instead of natural join, no renaming, order matters in projections; $\{\sigma, \pi, \cup, -, \times\}$.

Exercise 1.7. We have seen above that $\cap$ can be expressed in terms of the other standard operators of relational algebra. Indeed, the set of operations $\{\sigma, \pi, \cup, -, \bowtie, \delta\}$ can express all queries of relational algebra: it is complete. Try to show that it is not possible to reduce this set any further:

For each operator, first try to find an example query that cannot be expressed when using only the other operators. Then try to find a general argument that shows that the operator is really needed.

Solution.

1. Natural join: The only operation that may produce tables with bigger arity than any of its inputs.

2. Renaming: The only operation that can rename attributes.

3. Projection: The only operation whose output may have smaller arity.

4. Difference: The only anti-monotonic operation.

5. Select: Note that, $\sigma_{x=c}(R)$ with c a constant can actually be expressed using the join operator algebra and an atomic constant table. Namely, $\sigma_{x=c}(R) = R \bowtie \{\{x \to c\}\}$. Nevertheless, we cannot use this technique to express $\sigma_{x=y}(R)$ with y an attribute. Even though $\sigma_{x=y}(R) = R \bowtie \{\{x \mapsto c_1, y \mapsto c_1\}, \ldots, \{x \mapsto c_n, y \mapsto c_n\}\}$ if $c_1, \ldots, c_n$ are all the constants occurring in R, this solution is not valid as it is dependent on a particular table/database. I.e., this transformation may not work if we consider a different table S containing some constant $c \notin \{c_1, \ldots, c_n\}$.

   Argument:

   - Let $\mathcal{I}$ a database containing a single table $R = \{\{x \mapsto 1\}, \{x \mapsto 2\}\}$.
   - Then, $\sigma_{x=y}(R \bowtie \delta_{x \to y} R) = \{\{x \mapsto 1, y \mapsto 1\}, \{x \mapsto 2, y \mapsto 2\}\}$. We show that such a table cannot be produced if we do not use the select σ operator.
   - Namely, we show that, in the previous case, we may only a table R that satisfies the following restrictions: if R has arity n, then R contains 2^n rows containing every single combination of the symbols 1 and 2.
   - The above claim can be shown via induction.

6. Union: The only operation that maps two columns to a single one. Consider the database $\{\{A \mapsto 1\}\}, \{\{A \mapsto 2\}\}$.