Pauta Ayudantía 1

Álgebra Relacional y Consultas de Primer Orden

Benjamín Saavedra — IIC2413 — Bases de Datos

Notación (recordatorio)

- Para una relación R, $\mathsf{att}(R)$ denota su conjunto de atributos.
- **Operadores base del álgebra relacional** (para relaciones R, S, atributos A, B y constante $a \in \mathbf{Data}$):
  - $R \cup S,\ R \cap S,\ R \setminus S$ (definidos si $\mathsf{att}(R) = \mathsf{att}(S)$)
  - $\sigma_{A=a}(R),\ \sigma_{A=B}(R)$, y su generalización $\sigma_\theta(R)$ con θ fórmula booleana de igualdades $(\wedge, \vee, \neg)$, además de $\sigma_{A\sim a}, \sigma_{A\sim B}$ con $\sim \in \{<, >, \leq, \geq\}$.
  - $\rho_{A\to B}(R)$ (definido si $A \in \mathsf{att}(R)$, $B \notin \mathsf{att}(R)$)
  - $\pi_{B_1,\ldots,B_k}(R)$
  - $R \bowtie S$ (join natural: combina tuplas *compatibles*, i.e. que coinciden en $\mathsf{att}(R) \cap \mathsf{att}(S)$; se cumple $\mathsf{att}(R \bowtie S) = \mathsf{att}(R) \cup \mathsf{att}(S)$)
  - $R \times S$ (caso particular de $\bowtie$ cuando $\mathsf{att}(R) \cap \mathsf{att}(S) = \varnothing$)
- **Consultas de primer orden**: $Q(x_1, \ldots, x_n) \leftarrow \varphi$, donde φ es una fórmula con $\mathsf{FV}(\varphi) \subseteq \{x_1, \ldots, x_n\}$ construida a partir de $x = a$, $x = y$, $R(t_1, \ldots, t_k)$ mediante $\wedge, \vee, \neg, \exists x, \forall x$.
- Recuerde que $\forall x.\, \varphi \equiv \neg\exists x.\, \neg\varphi$: esto es clave para traducir consultas del tipo *"para todos"* / *"ninguno"*.

Consejos generales

- Antes de escribir la fórmula/expresión, identifique *qué relaciones necesita combinar* y con qué atributo se conectan (llave que comparten).
- Para consultas del tipo *"que tengan dos o más ..."* necesitará comparar **dos testigos distintos** de una misma relación (dos variables existenciales con $\neg(\cdot = \cdot)$ en lógica; dos copias renombradas de la relación con ρ en álgebra).
- Para consultas del tipo *"que NO tengan ningún ..."* necesitará una negación de un existencial en lógica, o una diferencia $\setminus$ en álgebra.
- Para consultas del tipo *"para TODOS los ..."* piense en la equivalencia $\forall x.\varphi \equiv \neg\exists x.\neg\varphi$; en álgebra esto se traduce típicamente en dos diferencias anidadas sobre un "universo de candidatos" y una relación de evidencia (vea el Ejercicio A.1(h)).
- En álgebra, cuando necesite comparar dos tuplas de una misma relación, **renombre una copia con** ρ antes de aplicar $\bowtie$ o $\times$; de lo contrario los atributos compartidos forzarán una igualdad no deseada (o el operador quedará indefinido).

1 Parte A — Consultas en contexto

Para cada consulta se entrega (i) una consulta de primer orden $Q(x_1, \ldots, x_n) \leftarrow \varphi$, y (ii) una expresión de álgebra relacional equivalente.

Nota: Las consultas de esta parte no tienen una única respuesta correcta. Tanto en lógica de primer orden como en álgebra relacional casi siempre existen varias formas equivalentes de expresar la misma consulta (por ejemplo, cambiando el orden de los joins, usando $\cap$ en vez de un doble $\exists$, proyectando en un punto distinto, etc.). Las soluciones de esta pauta son una forma válida de resolver cada consulta, no la única.

1.1 Contexto: Cine

Un cine registra información sobre sus películas, salas y funciones en el siguiente esquema:

$$\begin{aligned}
\textbf{Pelicula}(\mathsf{pid}, \mathsf{titulo}, \mathsf{duracion}) \quad & \mathsf{ar} = 3 \\
\textbf{Genero}(\mathsf{pid}, \mathsf{genero}) \quad & \mathsf{ar} = 2 \\
\textbf{Sala}(\mathsf{sid}, \mathsf{tipo}, \mathsf{capacidad}) \quad & \mathsf{ar} = 3 \\
\textbf{Funcion}(\mathsf{pid}, \mathsf{sid}, \mathsf{dia}, \mathsf{hora}) \quad & \mathsf{ar} = 4 \\
\textbf{Confiteria}(\mathsf{sid}, \mathsf{item}, \mathsf{precio}) \quad & \mathsf{ar} = 3
\end{aligned}$$

donde **Genero** es multivaluada (una película puede tener varios géneros), **Funcion**($pid, sid, dia, hora$) indica que la película *pid* se exhibe en la sala *sid* el día *dia* a la hora *hora*, y **Confiteria**($sid, item, precio$) indica que en la sala *sid* se vende el ítem de comida *item* a *precio* (puede variar entre salas).

Instancia de ejemplo (opcional, para verificar sus respuestas):

Pelicula

pid	titulo	duracion
1	Matrix	136
2	Coco	105
3	Nosferatu	132
4	Intensamente	95

Genero

pid	genero
1	Accion
1	CienciaFiccion
2	Animacion
2	Comedia
3	Terror
4	Animacion

Sala

sid	tipo	capacidad
1	IMAX	200
2	2D	100
3	3D	150

Funcion

pid	sid	dia	hora
1	1	Viernes	20:00
1	2	Sabado	18:00
2	2	Viernes	16:00
3	1	Sabado	22:00
4	3	Viernes	15:00

Confiteria

sid	item	precio
1	Palomitas	3500
1	Bebida	2000
2	Palomitas	2800
2	Nachos	3000
3	Bebida	2200

Ejercicio A.1. Para cada consulta, entregue (i) una consulta de primer orden y (ii) una expresión de álgebra relacional equivalente.

(a) Obtenga el título y la duración de todas las películas del género "Terror".

(b) Obtenga los títulos de las películas que tienen alguna función en una sala de tipo "IMAX".

(c) Obtenga el identificador y el tipo de las salas donde se vende el ítem "Palomitas".

(d) Obtenga los títulos de las películas que pertenecen a dos o más géneros distintos.

(e) Obtenga los identificadores de las salas que no tienen ninguna función programada.

(f) Obtenga los títulos de las películas que nunca se han exhibido (no tienen ninguna función).

(g) Obtenga los títulos de las películas que se exhiben tanto en la sala 1 como en la sala 2.

(h) *(Desafío)* Obtenga los identificadores de las salas donde se exhiben **todas** las películas del género "Terror".

Solución A.1

(a)

$$Q(y, z) \leftarrow \exists x.\ \mathbf{Pelicula}(x, y, z) \wedge \mathbf{Genero}(x, \text{Terror})$$

$$Q \ = \ \pi_{\mathsf{titulo,duracion}}\big(\mathbf{Pelicula} \bowtie \sigma_{\mathsf{genero}=\text{Terror}}(\mathbf{Genero})\big)$$

(b)

$$Q(y) \leftarrow \exists x.\exists z.\exists s.\exists d.\exists h.\exists c.\ \mathbf{Pelicula}(x, y, z) \wedge \mathbf{Funcion}(x, s, d, h) \wedge \mathbf{Sala}(s, \text{IMAX}, c)$$

$$Q \ = \ \pi_{\mathsf{titulo}}\big(\mathbf{Pelicula} \bowtie \mathbf{Funcion} \bowtie \sigma_{\mathsf{tipo}=\text{IMAX}}(\mathbf{Sala})\big)$$

(**Pelicula** y **Funcion** comparten pid; el resultado comparte sid con **Sala**, por lo que ambos joins naturales conectan por la llave correcta.)

(c)

$$Q(s, t) \leftarrow \exists c.\exists p.\ \mathbf{Sala}(s, t, c) \wedge \mathbf{Confiteria}(s, \text{Palomitas}, p)$$

$$Q \ = \ \pi_{\mathsf{sid,tipo}}\big(\mathbf{Sala} \bowtie \sigma_{\mathsf{item}=\text{Palomitas}}(\mathbf{Confiteria})\big)$$

(d) Pedimos dos testigos de género distintos para la misma película.

$$Q(y) \leftarrow \exists x.\exists z.\exists g_1.\exists g_2.\ \mathbf{Pelicula}(x, y, z) \wedge \mathbf{Genero}(x, g_1) \wedge \mathbf{Genero}(x, g_2) \wedge \neg(g_1 = g_2)$$

$$Q \ = \ \pi_{\mathsf{titulo}}\Big(\mathbf{Pelicula} \bowtie \pi_{\mathsf{pid}}\big(\sigma_{\mathsf{g}_1 \neq \mathsf{g}_2}(\rho_{\mathsf{genero}\to\mathsf{g}_1}(\mathbf{Genero}) \bowtie \rho_{\mathsf{genero}\to\mathsf{g}_2}(\mathbf{Genero}))\big)\Big)$$

(e)

$$Q(s) \leftarrow \exists t.\exists c.\ \mathbf{Sala}(s, t, c) \wedge \neg\big(\exists x.\exists d.\exists h.\ \mathbf{Funcion}(x, s, d, h)\big)$$

$$Q \ = \ \pi_{\mathsf{sid}}(\mathbf{Sala}) \ \setminus \ \pi_{\mathsf{sid}}(\mathbf{Sala} \bowtie \mathbf{Funcion})$$

(f)

$$Q(y) \leftarrow \exists x.\exists z.\ \mathbf{Pelicula}(x, y, z) \wedge \neg\big(\exists s.\exists d.\exists h.\ \mathbf{Funcion}(x, s, d, h)\big)$$

Para el álgebra, **cuidado**: no basta con $\pi_{\mathsf{titulo}}(\mathbf{Pelicula}) \setminus \pi_{\mathsf{titulo}}(\mathbf{Pelicula} \bowtie \mathbf{Funcion})$ —si dos películas *distintas* compartieran título, esa resta podría eliminar por error el título de la que nunca se exhibió (vea la Propiedad C.3). Lo correcto es restar por la llave pid y recién al final proyectar titulo:

$$Q \ = \ \pi_{\mathsf{titulo}}\Big(\mathbf{Pelicula} \bowtie \big(\pi_{\mathsf{pid}}(\mathbf{Pelicula}) \setminus \pi_{\mathsf{pid}}(\mathbf{Pelicula} \bowtie \mathbf{Funcion})\big)\Big)$$

(g)

$$Q(y) \leftarrow \exists x.\exists z.\, \textbf{Pelicula}(x,y,z) \wedge \big(\exists d_1.\exists h_1.\, \textbf{Funcion}(x,1,d_1,h_1)\big) \wedge \big(\exists d_2.\exists h_2.\, \textbf{Funcion}(x,2,d_2,h_2)\big)$$

$$Q \;=\; \pi_{\mathsf{titulo}}\Big(\textbf{Pelicula} \bowtie \big(\pi_{\mathsf{pid}}(\sigma_{\mathsf{sid}=1}(\textbf{Funcion})) \cap \pi_{\mathsf{pid}}(\sigma_{\mathsf{sid}=2}(\textbf{Funcion}))\big)\Big)$$

(aquí no hay riesgo de proyectar antes de tiempo, ya que no usamos diferencia: la intersección de dos conjuntos de pid es "segura".)

(h) *(Desafío.)*
Es una consulta de tipo "para todo": para cada sala s, toda película p que sea de género "Terror" debe tener una función en s.

$$Q(s) \leftarrow \exists t.\exists c.\, \textbf{Sala}(s,t,c) \wedge \forall p.\big(\neg \textbf{Genero}(p, \text{Terror}) \vee \exists d.\exists h.\, \textbf{Funcion}(p,s,d,h)\big)$$

(usamos $\neg A \vee B$ para expresar $A \rightarrow B$, ya que $\rightarrow$ no es parte de la sintaxis básica). Note que si **no** hay ninguna película de género "Terror", el $\forall p$ se cumple vacuamente para cualquier sala, así que Q debería devolver **todas** las salas (incluida una sin ninguna función programada).

Para el álgebra, primero obtenemos las películas de género Terror:

$$T := \pi_{\mathsf{pid}}\left(\sigma_{\mathsf{genero}=\text{“Terror”}}(\textbf{Genero})\right).$$

Luego obtenemos todas las salas:

$$S := \pi_{\mathsf{sid}}(\textbf{Sala}).$$

El producto cartesiano $S \times T$ genera todos los pares (sid, pid) que deberían existir si cada sala exhibiera todas las películas de Terror.

Por otro lado, los pares (sid, pid) que realmente tienen una función programada son:

$$F := \pi_{\mathsf{sid},\mathsf{pid}}(\textbf{Funcion}).$$

Entonces,

$$(S \times T) \setminus F$$

contiene exactamente los pares (sid, pid) donde una película de Terror no se exhibe en una determinada sala.

Por lo tanto,

$$\pi_{\mathsf{sid}}\left((S \times T) \setminus F\right)$$

corresponde a las salas en las que falta al menos una película de Terror.

Finalmente, eliminamos esas salas del conjunto de todas las salas:

$$Q = S \setminus \pi_{\mathsf{sid}}\left((S \times T) \setminus F\right)$$

La idea es construir primero todos los pares *(sala, película de Terror)* que deberían existir. Luego se restan los pares que sí aparecen en **Funcion**. De esta forma se identifican las salas donde falta al menos una película de Terror, y finalmente estas salas se eliminan del conjunto de todas las salas.

1.2 Contexto: Universidad

Una universidad registra la siguiente información:

Estudiante(rut, nombre, carrera)	ar = 3
Curso(codigo, nombre, creditos)	ar = 3
Inscribe(rut, codigo, semestre)	ar = 3
Dicta(codigo, profesor)	ar = 2
Requisito(codigo, prerrequisito)	ar = 2

donde **Inscribe**($rut, codigo, semestre$) indica que el estudiante rut se inscribió en el curso $codigo$ en el semestre $semestre$; **Dicta**($codigo, profesor$) es multivaluada (un curso puede tener más de un profesor); y **Requisito**($codigo, prerrequisito$) indica que para inscribir $codigo$ se necesita haber aprobado $prerrequisito$ (ambos son códigos de curso).

Instancia de ejemplo (opcional):

Estudiante

rut	nombre	carrera
1	Ana	IngCivil
2	Ben	Computacion
3	Zoe	Computacion

Curso

codigo	nombre	creditos
IIC2413	BasesDeDatos	10
IIC2233	ProgAvanzada	10
IIC1103	IntroProg	10

Inscribe

rut	codigo	semestre
1	IIC1103	2025-1
2	IIC2233	2025-1
2	IIC2413	2025-2
3	IIC2413	2025-2
3	IIC2233	2024-2

Dicta

codigo	profesor
IIC2413	Riveros
IIC2413	Soto
IIC2233	Riveros
IIC1103	Perez

Requisito

codigo	prerrequisito
IIC2413	IIC2233
IIC2233	IIC1103

Ejercicio A.2. Para cada consulta, entregue (i) una consulta de primer orden y (ii) una expresión de álgebra relacional equivalente.

(a) Obtenga los nombres de los estudiantes de la carrera "Computacion".

(b) Obtenga los nombres de los cursos dictados por el profesor "Riveros".

(c) Obtenga los nombres de los estudiantes inscritos en el curso "IIC2413".

(d) Obtenga los nombres de los estudiantes que están inscritos tanto en "IIC2413" como en "IIC2233" (no necesariamente en el mismo semestre).

(e) Obtenga los nombres de los cursos que no tienen ningún prerrequisito.

(f) Obtenga los nombres de los estudiantes que no están inscritos en ningún curso.

(g) Obtenga los nombres de los cursos dictados por más de un profesor.

(h) *(Desafío)* Obtenga los nombres de los estudiantes que están inscritos en **todos** los cursos que dicta el profesor "Riveros".

Solución A.2

(a)

$$Q(y) \leftarrow \exists x.\ \mathbf{Estudiante}(x, y, \text{Computacion})$$

$$Q \ = \ \pi_{\mathsf{nombre}}\big(\sigma_{\mathsf{carrera}=\text{Computacion}}(\mathbf{Estudiante})\big)$$

(b)

$$Q(y) \leftarrow \exists x.\exists c.\ \mathbf{Curso}(x, y, c) \wedge \mathbf{Dicta}(x, \text{Riveros})$$

$$Q \ = \ \pi_{\mathsf{nombre}}\big(\mathbf{Curso} \bowtie \sigma_{\mathsf{profesor}=\text{Riveros}}(\mathbf{Dicta})\big)$$

(c)

$$Q(y) \leftarrow \exists x.\exists w.\exists s.\ \mathbf{Estudiante}(x, y, w) \wedge \mathbf{Inscribe}(x, \text{IIC2413}, s)$$

$$Q \ = \ \pi_{\mathsf{nombre}}\big(\mathbf{Estudiante} \bowtie \sigma_{\mathsf{codigo}=\text{IIC2413}}(\mathbf{Inscribe})\big)$$

(d)

$$Q(y) \leftarrow \exists x.\exists w.\ \mathbf{Estudiante}(x, y, w) \wedge$$

$$\big(\exists s_1.\ \mathbf{Inscribe}(x, \text{IIC2413}, s_1)\big) \wedge \big(\exists s_2.\ \mathbf{Inscribe}(x, \text{IIC2233}, s_2)\big)$$

$$Q \ = \ \pi_{\mathsf{nombre}}\Big(\mathbf{Estudiante} \bowtie \big(\pi_{\mathsf{rut}}(\sigma_{\mathsf{codigo}=\text{IIC2413}}(\mathbf{Inscribe})) \cap \pi_{\mathsf{rut}}(\sigma_{\mathsf{codigo}=\text{IIC2233}}(\mathbf{Inscribe}))\big)\Big)$$

(e)

$$Q(y) \leftarrow \exists x.\exists c.\ \mathbf{Curso}(x, y, c) \wedge \neg\big(\exists v.\ \mathbf{Requisito}(x, v)\big)$$

Igual que en A.1(f): restamos por la llave codigo y proyectamos nombre recién al final, para no perder cursos cuyo nombre coincida por casualidad con el de otro curso (Propiedad C.3):

$$Q \ = \ \pi_{\mathsf{nombre}}\Big(\mathbf{Curso} \bowtie \big(\pi_{\mathsf{codigo}}(\mathbf{Curso}) \setminus \pi_{\mathsf{codigo}}(\mathbf{Curso} \bowtie \mathbf{Requisito})\big)\Big)$$

(f)

$$Q(y) \leftarrow \exists x.\exists w.\ \mathbf{Estudiante}(x, y, w) \wedge \neg\big(\exists v_1.\exists v_2.\ \mathbf{Inscribe}(x, v_1, v_2)\big)$$

Misma precaución: restamos por rut, no por nombre:

$$Q \ = \ \pi_{\mathsf{nombre}}\Big(\mathbf{Estudiante} \bowtie \big(\pi_{\mathsf{rut}}(\mathbf{Estudiante}) \setminus \pi_{\mathsf{rut}}(\mathbf{Estudiante} \bowtie \mathbf{Inscribe})\big)\Big)$$

(g) Mismo patrón de dos testigos que en A.1(d):

$$Q(y) \leftarrow \exists x.\exists c.\exists p_1.\exists p_2.\ \mathbf{Curso}(x, y, c) \wedge \mathbf{Dicta}(x, p_1) \wedge \mathbf{Dicta}(x, p_2) \wedge \neg(p_1 = p_2)$$

$$Q \ = \ \pi_{\mathsf{nombre}}\Big(\mathbf{Curso} \bowtie \pi_{\mathsf{codigo}}\big(\sigma_{\mathsf{p_1} \neq \mathsf{p_2}}(\rho_{\mathsf{profesor} \to \mathsf{p_1}}(\mathbf{Dicta}) \bowtie \rho_{\mathsf{profesor} \to \mathsf{p_2}}(\mathbf{Dicta}))\big)\Big)$$

(h) *(Desafío.)*

$$Q(y) \leftarrow \exists x.\exists w.\ \mathbf{Estudiante}(x, y, w) \wedge \forall v.\big(\neg\mathbf{Dicta}(v, \text{Riveros}) \vee \exists s.\ \mathbf{Inscribe}(x, v, s)\big)$$

Igual que en A.1(h): si Riveros no dicta **ningún** curso, el $\forall v$ se cumple vacuamente y Q debería devolver **todos** los estudiantes (incluso uno sin ninguna inscripción).

Para el álgebra, sean

$$E := \pi_{\mathsf{rut}}(\mathbf{Estudiante}) \quad \text{(todos los estudiantes)}$$

$$D := \pi_{\mathsf{codigo}}(\sigma_{\mathsf{profesor}=\mathrm{Riveros}}(\mathbf{Dicta})) \qquad I := \pi_{\mathsf{rut},\mathsf{codigo}}(\mathbf{Inscribe})$$

Separando el universo de candidatos (E) de la evidencia (I), con el mismo argumento verificado en A.1(h) (cambiando S, T, F por E, D, I):

$$Q \;=\; \pi_{\mathsf{nombre}}\Big(\mathbf{Estudiante} \bowtie \big(E \setminus \pi_{\mathsf{rut}}((E \times D) \setminus I)\big)\Big)$$

Primero se consideran todos los pares posibles entre estudiantes y cursos dictados por Riveros mediante $E \times D$. Al restar las inscripciones reales I, obtenemos los pares que faltan, es decir, estudiantes que no están inscritos en algún curso de Riveros. Proyectando por rut identificamos esos estudiantes y los eliminamos de E. Así quedan exactamente los estudiantes inscritos en todos los cursos dictados por Riveros. Además, si Riveros no dicta ningún curso, $E \times D = \varnothing$, por lo que ningún estudiante es eliminado y el resultado contiene a todos los estudiantes, tal como exige la cuantificación universal.

2 Parte B — Nuevos operadores del álgebra relacional

El álgebra relacional que vimos en clases se construye a partir de un conjunto *base* de operadores: $\cup, \cap, \setminus, \sigma, \rho, \pi, \bowtie$. En la práctica es común definir operadores adicionales que faciliten escribir consultas. Para que un operador nuevo sea parte del álgebra relacional (y no la extienda a un lenguaje más expresivo), debemos demostrar que puede *simularse* usando sólo los operadores base: es decir, dar una expresión equivalente y demostrar la igualdad de las relaciones resultantes (típicamente por doble contención, o por "si y sólo si"). En cada ejercicio: (i) proponga una expresión en álgebra relacional para el operador definido, y (ii) demuestre formalmente la igualdad.

Ejercicio B.1 (semijoin). Para relaciones R, S se define el *semijoin* de R con S, denotado $R \ltimes S$, como la relación con $\mathsf{att}(R \ltimes S) = \mathsf{att}(R)$ tal que

$$R \ltimes S \;:=\; \{\bar{\tau} \in R \mid \exists \bar{\sigma} \in S.\ \bar{\tau} \sim \bar{\sigma}\}$$

(recuerde que $\bar{\tau} \sim \bar{\sigma}$ significa que $\bar{\tau}$ y $\bar{\sigma}$ son *compatibles*: coinciden en los atributos que comparten). Es decir, $R \ltimes S$ son las tuplas de R que tienen al menos un "compañero" compatible en S. Proponga una expresión equivalente en términos de $\pi, \bowtie$, y demuéstrela.

Solución. Proponemos

$$R \ltimes S \;=\; \pi_{\mathsf{att}(R)}(R \bowtie S).$$

Demostración.
($\subseteq$) Sea $\bar{\tau} \in R \ltimes S$. Por definición de semijoin, $\bar{\tau} \in R$ y existe $\bar{\sigma} \in S$ compatible con $\bar{\tau}$. Como $\bar{\tau}$ y $\bar{\sigma}$ son compatibles, por definición de $\bowtie$ se tiene

$$\bar{\tau} \bowtie \bar{\sigma} \in R \bowtie S.$$

Además, al proyectar $\bar{\tau} \bowtie \bar{\sigma}$ sobre los atributos de R, se recupera $\bar{\tau}$, es decir,

$$\pi_{\mathsf{att}(R)}(\bar{\tau} \bowtie \bar{\sigma}) = \bar{\tau}.$$

Por lo tanto,

$$\bar{\tau} \in \pi_{\mathsf{att}(R)}(R \bowtie S).$$

($\supseteq$) Sea $\bar{\tau}' \in \pi_{\mathsf{att}(R)}(R \bowtie S)$. Entonces existe $\bar{\upsilon} \in R \bowtie S$ con $\pi_{\mathsf{att}(R)}(\bar{\upsilon}) = \bar{\tau}'$. Por definición de $\bowtie$, existen $\bar{\tau} \in R, \bar{\sigma} \in S$ compatibles con $\bar{\upsilon} = \bar{\tau} \bowtie \bar{\sigma}$. Por lo tanto,

$$\bar{\tau}' = \pi_{\mathsf{att}(R)}(\bar{\upsilon}) = \pi_{\mathsf{att}(R)}(\bar{\tau} \bowtie \bar{\sigma}) = \bar{\tau}.$$

Así, como $\bar{\tau} \in R$ y $\bar{\tau} = \bar{\tau}'$ luego $\bar{\tau}' \in R$. Además, como $\bar{\tau}' = \bar{\tau}$ es compatible con $\bar{\sigma} \in S$, existe una tupla de S compatible con $\bar{\tau}'$. Por definición de semijoin,

$$\bar{\tau}' \in R \ltimes S.$$

Ejercicio B.2 (anti-join). Se define el *anti-join* de R con S, denotado $R \rhd S$, como la relación con $\mathsf{att}(R \rhd S) = \mathsf{att}(R)$ tal que

$$R \rhd S \;:=\; \{\bar{\tau} \in R \mid \nexists \bar{\sigma} \in S.\ \bar{\tau} \sim \bar{\sigma}\}$$

es decir, las tuplas de R que **no** tienen ningún compañero compatible en S.

(a) Demuestre que $R \rhd S = R \setminus (R \ltimes S)$.

(b) Use $\rhd$ para reescribir de forma más compacta su respuesta a la consulta A.1(e) ("salas sin función").

Solución.

(a) Como $R \ltimes S \subseteq R$ (es inmediato de su definición), para $\bar{\tau} \in R$ se cumple $\bar{\tau} \in R \ltimes S \iff \exists \bar{\sigma} \in S . \bar{\tau} \sim \bar{\sigma}$ (el requisito "$\bar{\tau} \in R$" ya está garantizado). Entonces, para $\bar{\tau} \in R$:

$$\bar{\tau} \in R \setminus (R \ltimes S) \iff \bar{\tau} \notin R \ltimes S \iff \neg \exists \bar{\sigma} \in S . \bar{\tau} \sim \bar{\sigma} \iff \bar{\tau} \in R \rhd S.$$

Como ambos lados exigen además $\bar{\tau} \in R$, se concluye $R \rhd S = R \setminus (R \ltimes S)$. ■

Combinando con el Ejercicio B.1: $R \rhd S = R \setminus \pi_{\mathsf{att}(R)}(R \bowtie S)$, expresado enteramente en operadores base.

(b) **Sala** $\rhd$ **Funcion** son las salas (tupla completa) sin ninguna función; para obtener sólo el identificador:

$$Q = \pi_{\mathsf{sid}}(\textbf{Sala} \rhd \textbf{Funcion})$$

equivalente a la respuesta de A.1(e), en un solo operador en vez de una resta de dos proyecciones.

3 Parte C — Propiedades del álgebra relacional

Para cada propiedad: si es **verdadera**, demuéstrela (por extensión de la definición de cada operador, mostrando que ambos lados tienen los mismos atributos y las mismas tuplas); si es **falsa**, entregue relaciones concretas $R, S, \ldots$ (con sus atributos y algunas tuplas) donde ambos lados de la igualdad den resultados distintos.

C.1. (Conmutatividad de la selección) Para toda relación R y fórmulas θ_1, θ_2 (cuyos atributos pertenecen a $\mathsf{att}(R)$):

$$\sigma_{\theta_1}\big(\sigma_{\theta_2}(R)\big) \;=\; \sigma_{\theta_2}\big(\sigma_{\theta_1}(R)\big) \;=\; \sigma_{\theta_1 \wedge \theta_2}(R).$$

Solución. Verdadera. Basta el caso $\theta_1 := (A = a)$, $\theta_2 := (B = b)$ (el caso general es análogo). Para toda tupla $\bar{\tau}$:

$$\bar{\tau} \in \sigma_{A=a}(\sigma_{B=b}(R)) \iff \bar{\tau} \in R \wedge \bar{\tau}(B) = b \wedge \bar{\tau}(A) = a \iff \bar{\tau} \in \sigma_{B=b}(\sigma_{A=a}(R)),$$

donde la primera equivalencia se obtiene aplicando dos veces la definición de σ, y la segunda aplicando nuevamente dicha definición en el orden inverso, usando la conmutatividad de la conjunción. Por otro lado, por definición $\sigma_{(A=a)\wedge(B=b)}(R) := \sigma_{A=a}(R) \cap \sigma_{B=b}(R)$, y

$$\bar{\tau} \in \sigma_{A=a}(R) \cap \sigma_{B=b}(R) \iff \bar{\tau} \in R \wedge \bar{\tau}(A) = a \wedge \bar{\tau}(B) = b,$$

da la misma caracterización de arriba. Los tres conjuntos son iguales. ■

C.2. (Proyección distribuye sobre la unión) Sean R, S con $\mathsf{att}(R) = \mathsf{att}(S)$:

$$\pi_{B_1,\ldots,B_k}(R \cup S) \;=\; \pi_{B_1,\ldots,B_k}(R) \cup \pi_{B_1,\ldots,B_k}(S).$$

Solución. Verdadera. Con $L = \{B_1, \ldots, B_k\}$:

$$\begin{aligned}\pi_L(R \cup S) &= \{\pi_L(\bar{\tau}) \mid \bar{\tau} \in R \cup S\} = \{\pi_L(\bar{\tau}) \mid \bar{\tau} \in R \vee \bar{\tau} \in S\} \\ &= \{\pi_L(\bar{\tau}) \mid \bar{\tau} \in R\} \cup \{\pi_L(\bar{\tau}) \mid \bar{\tau} \in S\} = \pi_L(R) \cup \pi_L(S),\end{aligned}$$

usando el hecho general de que la imagen de una unión (bajo cualquier función, aquí π_L) es la unión de las imágenes. ■

C.3. (Proyección y diferencia) Sean R, S con $\mathsf{att}(R) = \mathsf{att}(S)$. Demuestre o refute:

$$\pi_{B_1,\ldots,B_k}(R \setminus S) \;=\; \pi_{B_1,\ldots,B_k}(R) \setminus \pi_{B_1,\ldots,B_k}(S).$$

Solución. Falsa. Contraejemplo: sean $\mathsf{att}(R) = \mathsf{att}(S) = \{A, B\}$,

$$R = \{(A{:}1, B{:}1),\ (A{:}1, B{:}2)\}, \qquad S = \{(A{:}1, B{:}1)\}.$$

Entonces $R \setminus S = \{(A{:}1, B{:}2)\}$, luego $\pi_A(R \setminus S) = \{(A{:}1)\}$.

Por otro lado $\pi_A(R) = \{(A{:}1)\} = \pi_A(S)$, luego $\pi_A(R) \setminus \pi_A(S) = \varnothing$.

Como $\{(A{:}1)\} \neq \varnothing$, la igualdad es falsa. Intuitivamente: al proyectar *antes* de restar se pierde la información (el valor de B) que distinguía a la tupla que sobrevive de la que se resta.

C.4. (Conmutatividad y asociatividad del join) Para relaciones R, S, T cualesquiera:

$$R \bowtie S = S \bowtie R \qquad \text{y} \qquad (R \bowtie S) \bowtie T = R \bowtie (S \bowtie T).$$

Solución. Verdadera.

Conmutatividad. La compatibilidad es simétrica ($\bar{\tau} \sim \bar{\sigma} \iff \bar{\sigma} \sim \bar{\tau}$, pues sólo depende de $\mathsf{att}(\bar{\tau}) \cap \mathsf{att}(\bar{\sigma})$). Además, si $\bar{\tau} \sim \bar{\sigma}$, entonces $\bar{\tau} \bowtie \bar{\sigma} = \bar{\sigma} \bowtie \bar{\tau}$. Podemos ver esto analizando casos: todo atributo A cumple:

- si $A \in \mathsf{att}(\bar{\tau}) \cap \mathsf{att}(\bar{\sigma})$, ambas tuplas toman el mismo valor, pues $\bar{\tau}(A) = \bar{\sigma}(A)$ por compatibilidad.
- si $A \in \mathsf{att}(\bar{\tau}) \setminus \mathsf{att}(\bar{\sigma})$, ambas toman el valor $\bar{\tau}(A)$.
- si $A \in \mathsf{att}(\bar{\sigma}) \setminus \mathsf{att}(\bar{\tau})$, ambas toman el valor $\bar{\sigma}(A)$.

Por lo tanto, se cumple la igualdad de tuplas $\bar{\tau} \bowtie \bar{\sigma} = \bar{\sigma} \bowtie \bar{\tau}$, lo que nos permite concluir:

$$R \bowtie S = \{\bar{\tau} \bowtie \bar{\sigma} \mid \bar{\tau} \in R, \bar{\sigma} \in S, \bar{\tau} \sim \bar{\sigma}\} = \{\bar{\sigma} \bowtie \bar{\tau} \mid \bar{\sigma} \in S, \bar{\tau} \in R, \bar{\sigma} \sim \bar{\tau}\} = S \bowtie R.$$

Asociatividad. Queremos probar que

$$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T).$$

Una tupla de $(R \bowtie S) \bowtie T$ se obtiene a partir de $\bar{\tau}_R \in R$, $\bar{\tau}_S \in S$ y $\bar{\tau}_T \in T$ tales que

$$\bar{\tau}_R \sim \bar{\tau}_S \qquad \text{y} \qquad (\bar{\tau}_R \bowtie \bar{\tau}_S) \sim \bar{\tau}_T.$$

Esta última condición equivale a

$$\bar{\tau}_R \sim \bar{\tau}_T \wedge \bar{\tau}_S \sim \bar{\tau}_T,$$

pues en cada atributo compartido con T, el join $\bar{\tau}_R \bowtie \bar{\tau}_S$ toma el valor de $\bar{\tau}_R$ o de $\bar{\tau}_S$, según corresponda.

Por lo tanto, $(R \bowtie S) \bowtie T$ contiene exactamente las combinaciones de tres tuplas de R, S, T compatibles dos a dos. Análogamente, lo mismo ocurre con $R \bowtie (S \bowtie T)$.

Finalmente, si las tres tuplas son compatibles dos a dos, ambas formas de agrupar los joins producen la misma tupla, pues cada atributo recibe el mismo valor independientemente de la agrupación. Luego

$$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T).$$

■