Guía de Álgebra Relacional

Prof. Claudio Gutiérrez, Aux. Mauricio Monsalve

Primavera de 2007

1. Problemas conceptuales

1. ¿Son la división y el producto operaciones inversas?
2. ¿Cuántas proyecciones existen, cuyo resultado es no vacío, sobre una relación no vacía de n atributos?
3. ¿Qué operadores relacionales son conmutativos?
4. Sea $R = (Persona, Idioma)$ y $S = (Idioma)$. ¿Quiénes hablan todos los idiomas? (¿Qué se debe cumplir para prescindir de S?)
5. Defina *intersección* y *join natural* utilizando sólo los operadores *diferencia*, *producto*, *proyección* y *selección*.
6. Suponga que las relaciones A, B, C y D tienen 100, 1000, 10000 y 50 tuplas, respectivamente. Ordene la operación $A \bowtie B \bowtie C \bowtie D$ para que sea más eficiente.
7. Sean R y S dos relaciones sin atributos comunes (sus esquemas son disjuntos). ¿Por qué la operación $R \bowtie S$ (join natural) es equivalente a la operación $R \times S$? ¿Qué ocurre en el caso opuesto, cuando R y S tienen el mismo esquema?
8. Sean R y S dos relaciones con el mismo esquema y que tienen n y m tuplas, respectivamente. Dé cotas para la cardinalidad de las siguientes operaciones:

   a) $R \cup S$, $R - S$ y $R \cap S$

   b) $R \bowtie S$, $R \bowtie_c S$ y $R \div S$

9. ¿Cuál es la filosofía de la división? ¿Cuándo ocuparía esta relación? Indique qué ocurre con $(A \times B) \div B$ y con $(A \bowtie B) \div B$.
10. Sean los esquemas de relación $R(A, B, C)$ y $T(B, D)$. La relación R tiene n tuplas y T tiene m tuplas, cumpliéndose que $n>m>0$. Indicar el tamaño mínimo y máximo de las siguientes expresiones relacionales (cotas para la cardinalidad). Indicar, además, si cambian los resultados al saber que B es llave de R y de T, y cuáles son los cambios.

    a) $R \bowtie T$

    b) $\sigma_c(R)$

c) $R - (T - R)$
d) $R \div T$
e) $R * (R \div T)$

11. Sea $R \subset S$. Simplifique al máximo las siguientes expresiones:

    a) $R \cup S$
    b) $R \bowtie S$
    c) $\sigma_{A=a}(R) - S$
    d) $\Pi_X(R) \cap \Pi_X(S)$

2. Problemas tipo

2.1. Álgebra alcohólica

Se tienen las siguientes relaciones en una base de datos:

```
Frecuenta(bebedor,bar,desde)
Sirve(bar,cerveza)
Gusta(bebedor,cerveza)
```

Donde "frecuenta" indica si un bebedor frecuenta un bar, "sirve" indica si un bar sirve un tipo de cerveza y "gusta" indica si un bebedor gusta de un tipo de cerveza. Suponga, además, que la base de datos modela toda una comunidad. Responda:

1. Qué cervezas sirve el bar "Don Lucho".
2. Los bebedores que no frecuenten bares con cervezas de su agrado.
3. Sabiendo que hay bares que sirven todos los tipos de cerveza, indicar cuáles son.
4. Aquellos bebedores que frecuentan todos los bares.
5. Aquellos bebedores que gustan todos los tipos de cerveza.
6. Esos bebedores que encuentran la cerveza "Shield" muy suave (no les gusta).
7. Los bares que van a quebrar, pues ningún bebedor los frecuenta.
8. Desde cuándo Petronila frecuenta el bar "Don Lucho".
9. Bebedores que beben lo mismo que Petronila.

2.2. Campeonato de fútbol

Sean los siguientes esquemas sobre un campeonato de fútbol:

```
Equipo(cod_equipo, nombre, RUT_DT)
DT(RUT_DT,nombre,apellido)
Jugador(RUT_J,nombre,apellido,cod_equipo,suspendido)
Partido(cod_local, cod_visita, goles_local, goles_visita)
```

Sin usar el operador de agregación, use álgebra relacional para entregar:

1. La lista de los equipos y sus directores técnicos (con nombre y primer apellido).
2. Los directores técnicos retirados o cesantes.
3. Los nombres de los equipos que han ganado al menos una vez.
4. Los directores técnicos de los equipos que tienen 3 ó más jugadores suspendidos.
5. Los jugadores suspendidos cuyos equipos han jugado sólo 3 veces.
6. El equipo que ha jugado contra todos los demás.
7. El partido en el que ha habido la mayor diferencia de goles.

2.3. Trabajando con instancias

Sean las relaciones R y S definidas a continuación:

R	(A,	B,	C,	D)
	Pipo	15	10 %	2.71
	Pipo	25	16 %	3.14
	Ñata	21	11 %	1.67

S	(A,	C,	E)
	Pipo	21 %	I
	Pipo	9 %	II
	Ñata	33 %	III
	Pipo	25 %	IV

Obtenga -o indique si no se puede- el resultado de las siguientes expresiones:

1. $a \leftarrow R \cup S$
2. $b \leftarrow \Pi_{A,C}(R)$
3. $c \leftarrow \Pi_{A,C}(S)$
4. $d \leftarrow b \cup c$
5. $e \leftarrow b \cap c$
6. $f \leftarrow b - c$
7. $g \leftarrow c - b$
8. $h \leftarrow \sigma_{C>15\,\%}(S)$
9. $i \leftarrow \Pi_A(h)$
10. $j \leftarrow R * S$

2.4. Arriendo de propiedades

Considere los siguientes esquemas:

```
Arrendatario(RUT_A,nombre,apellido)
Arrienda(RUT_A,Id_casa,deuda)
Telefonos(RUT,fono)
Dueño(RUT_D,nombre,apellido)
Casa(Id_casa,RUT_D,numero,calle,comuna)
```

Universo del discurso: Un arrendatario arrienda una casa al dueño de ella. Puede o no tener deuda. Además, tanto el arrendatario como el dueño pueden tener varios teléfonos. Se sabe que (numero, calle, comuna) es llave de Casa. Además, una deuda igual a cero equivale a un arriendo sin deuda.

Con la información anterior, responda:

1. Los arrendatarios de la casa que queda en Carrera 1024, Santiago.
2. La deuda total que tienen los arrendatarios con María Pérez.
3. La deuda total para cada dueño.
4. El número de casas de cada dueño.
5. El total de teléfonos de cada arrendador.
6. El promedio de arrendatarios por casa.
7. El promedio de arrendatarios por dueño.
8. Los arrendatarios que tienen al menos 3 casas.
9. Los dueños que tienen deudores en todas sus casas.
10. El arrendatario que posee más casas.

2.5. Cursos y alumnos

Sea el siguiente modelo relacional:

```
Alumno(RUT_A, nombre_a, carrera, nivel, edad)
Curso(nombre_c, sala, cupo, RUT_P)
Inscrito(RUT, nombre_c)
Profesor(RUT_P, nombre_p, deptid)
```

Conteste:

1. Obtenga todos los RUTs de la base de datos.
2. Encuentre los nombre de todos los alumnos de 4to año (level=04) que están tomando un curso con J. Perez.

3. Encuentre la edad del estudiante de mayor edad que está en la especialidad de computación (major=cs) o esta tomando un curso con J. Perez.

4. Encuentre los nombres de todas las clases que usan la sala CC204 (meets at=cc204) y que tienen más de 15 alumnos.

5. Encuentre los nombres de los profesores que hacen clases en todas las salas que se ocupan con tal propósito.

3. Problemas avanzados

3.1. Problema del transporte

Una empresa de transporte rural de pasajeros tiene en su base de datos información acerca de los recorridos de sus buses. Un recorrido es una secuencia de paradas que comienzan en un terminal. El número de paradas varía de un recorrido a otro, pero hay al menos una parada.

El modelo relacional consiste en los siguientes esquemas:

```
Salida(NumRecorrido,IdLocalidad)
Parada(NumRecorrido,IdLocalidad,TiempoLleg)
```

Los atributos NumRecorrido e IdLocalidad identifican recorrido y localidad. El atributo TiempoLleg indica cuánto tarda, en promedio, en llegar un bus a la localidad desde que partió (la Salida).

Conteste:

1. ¿Cuál es el recorrido más largo según número de localidades? ¿Cuál es el recorrido más largo según tiempo tomado?

2. Indique los recorridos con ciclos, i.e. aquellos recorridos en que se visita una localidad más de una vez.

3. Indique la parada central de cada recorrido. (Si un recorrido tiene cinco paradas, interesa la tercera parada. Y si tiene seis, también interesa la tercera.)

3.2. Cálculo relacional

El cálculo relacional es un lenguaje de consulta que sólo indica los requerimientos que la respuesta debe satisfacer. En otras palabras, *no especifica la manera en la cual la consulta se llevará a cabo* (muy por el contrario del álgebra relacional).

Hay dos tipos de operaciones en el cálculo relacional:

Cálculo de Dominios Este tipo de operaciones altera el dominio (y el esquema) de acuerdo a su aceptación por una fórmula lógica. La forma de la consulta es: $\{X_1, ..., X_n : F(X_1, ..., X_n)\}$. Los términos $X_1, ..., X_n$ denotan *variables libres* (aunque se podrían incluir valores constantes). Por ejemplo, usando el esquema del problema "Álgebra bohemia", la consulta en cálculo relacional que devuelve los bares que sirven cerveza pero en los cuales no sirven lo que Tito gusta es:
$\{bar : (\exists cerveza)Sirve(bar, cerveza) \wedge (\forall cerveza)Gusta("Tito", cerveza) \Rightarrow \neg Sirve(bar, cerveza)\}$

Cálculo de Tuplas Este tipo de operaciones devuelve tuplas de acuerdo a su aceptación por una fórmula lógica. La forma de la consulta es: $\{t^{(n)} : F(t)\}$, donde t es una tupla y n es el grado de t (se escribe opcionalmente, pero sirve para evitar ambigüedades). Por ejemplo, usando el esquema del problema "Álgebra bohemia", la consulta en cálculo relacional que devuelve los bares que frecuenta Tito es:
$\{t^{(1)} : (\exists s) Frecuenta(s) \wedge s[bebedor] = "Tito" \wedge s[bar] = t[1]\}$

Las funciones lógicas de cálculo relacional, las *fórmulas*, están compuestas por:

- Comparaciones aritméticas entre variables libres, constantes o ambas: $=, \neq, \leq, ...$
- Operaciones lógicas entre fórmulas: si F y G son fórmulas, entonces $F \vee G$, $F \wedge G$, $\neg F$ también lo son.
- Cuantificadores: Si F es una fórmula, entonces $(\exists X)F$ y $(\forall X)F$ son fórmulas.

Estando en conocimiento de esto:

1. Demuestre que el cálculo de tuplas es equivalente al cálculo de dominios.
2. Demuestre que el cuantificador $\forall$ es prescindible (escriba $\forall$ en función de otros elementos que compongan fórmulas de cálculo relacional).
3. Pruebe que el cálculo relacional es más expresivo que el álgebra relacional. Para esto, escriba los operadores de álgebra relacional como consultas de cálculo relacional.

4. Soluciones selectas

1.1 No. Si bien se cumple $A = (A \times B) \div B$, no se cumple $A = (A \div B) \times B$.

1.2 Son $2^n - 1$ proyecciones distintas.

1.4 Quienes hablan todos los idiomas: $R \div S$. Si el conjunto anterior es no vacío, entonces es equivalente a: $R \div \pi_{Idiomas}(R)$ (prescindiendo de S).

1.5 La intersección se puede escribir como $A - (A - B)$ o como $B - (B - A)$.

1.7 En la ausencia de atributos comunes, no dónde hacer comparación. Ahora, si todos los atributos son compartidos, se tiene que $A \bowtie B = A \cap B$.

1.8 Cotas:

(a) $Max(n, m) \leq |R \cup S| \leq n + m$. $Max(0, n - m) \leq |R - S| \leq n$. $0 \leq |R \cap S| \leq Min(n, m)$.

(b) $0 \leq |R \bowtie S| \leq Min(n, m)$. $0 \leq |R \bowtie_c S| \leq n \cdot m$. $0 \leq |R \div S| \leq \lfloor \frac{n}{m} \rfloor$.

1.9 Digamos que a se relaciona con algunos b. ¿Se relaciona a con todo b en B? Si la relación $R = (a, b)$ encapsula la relación de a con b, entonces la división contesta la pregunta: $R \div B$ son todos los a que se relacionan con todos los $b \in B$.

($(A \times B) \div B$) Se tiene que $A = (A \times B) \div B$.

($(A \bowtie B) \div B$) No se puede aseverar mucho del join natural, salvo que $(A \bowtie B) \div B \subseteq \pi_{Esq(A)-Esq(B)} A$.

1.11 Queda:

(a) $R \cup S = S$

(b) $R \bowtie S = R$

(c) $\sigma_A(R) - S = \emptyset$

(d) $\pi_X(R) \cap \pi_X(S) = \pi_X(S)$

2.1 (Álgebra alcohólica)

(1) Una consulta trivial: $\pi_{\{cerveza\}}[\sigma_{cerveza='Don\ Lucho'}(Sirve)]$

(2) $\langle \pi_{\{bebedor\}} Frecuenta \rangle - \langle \pi_{\{bebedor\}} Frecuenta \bowtie (Sirve \bowtie Gusta) \rangle$. Las negaciones se pueden tratar como complementos (si es que su complemento es más sencillo). Aparte, el orden de los paréntesis del trio de reuniones naturales no altera el resultado (en este caso).

(3) Algo que se relaciona con todo un grupo es una división: $Sirve \div \pi_{\{cerveza\}}(Sirve)$.

(4) Otra división: $\langle \pi_{\{bebedor,bar\}} Frecuenta \rangle \div \langle \pi_{\{bar\}} Sirve \rangle$.

(5) $Gusta \div \langle \pi_{\{cerveza\}} Sirve \rangle$.

(6) Otra resta: $\langle \pi_{\{bebedor\}} Gusta \rangle - \langle \pi_{\{bebedor\}} \sigma_{cerveza='Shield'} Gusta \rangle$. Nótese que una consulta de la forma $\pi_{\{bebedor\}} \sigma_{cerveza \neq 'Shield'}(Gusta)$ es incorrecta. Por ejemplo, para las tuplas $\{(Abdul, Shield), (Abdul, Crystal)\}$, la selección ($\sigma$) entregaría $(Abdul, Crystal)$, y luego la proyección entregaría $Abdul$, cuando $Abdul$ gusta de la cerveza $Shield$.

(7) Otra resta: $\langle \pi_{\{bar\}} Sirve \rangle - \langle \pi_{\{bar\}} Frecuenta \rangle$. Se ha supuesto que todo bar sirve cerveza.

(8) Una consulta trivial: $\pi_{\{desde\}} \langle \sigma_{bar='Don\,Lucho' \wedge bebedor='Petronila'}(Frecuenta) \rangle$

(9) Otra división, pero hay que suponer que un bebedor consumirá sólo la cerveza disponible en los bares que frecuenta, y sólo la que gusta, lo que corresponde a: $A \leftarrow \pi_{\{bebedor,cerveza\}}(Frecuenta \bowtie (Sirve \bowtie Gusta))$. Ahora, si bebe lo mismo que Petronila: $A \div \langle \pi_{\{cerveza\}}(\sigma_{bebedor='Petronila'} A) \rangle$.

2.4 (Arriendo de propiedades)

(2) $\aleph_{<sum(deuda)>} \sigma_{nombre='Maria' \wedge apellido='Perez'} Due\overline{n}o \bowtie Casa \bowtie Arrienda$.

(4) ${}_{<rut\_d,nombre,apellido>}\aleph_{<rut\_d,nombre,apellido,count(id\_casa)>} Casa \bowtie Dueno$.

(6) $\aleph_{<avg(K)>} \langle {}_{<id\_casa>}\aleph_{<id\_casa,count(rut\_a)\,as\,K>} Arrienda \rangle$.

(8) ${}_{<rut\_a>}\aleph_{<rut\_a,count(id\_casa)>} \langle \sigma_{rut\_a=rut\_d}(Arrienda \times Casa) \rangle$.

(10) Una respuesta no agregada sería $Dueno \bowtie \pi_{rut\_d} \langle A - \pi_{A.rut\_d,A.K} \sigma_{A.K<B.K} A \times B \rangle$, donde $A \leftarrow {}_{<rut\_d>}\aleph_{<rut\_d,count(id\_casa)\,as\,K>} Casa \bowtie Dueno$, $B \leftarrow A$. Usando agregación, $Dueno \bowtie \pi_{rut\_d} \langle \sigma_{K=L} A \times (\aleph_{<Max(K)\,as\,L>} A) \rangle$.

3.1 (Problema del transporte)

(1) Recorridos más largos:

(1.1) Según número de localidades. Primero se debe conocer el largo de cada recorrido: $N \leftarrow {}_{<NumRecorrido>}\aleph_{<NumRecorrido, count(IdLocalidad)\, K>} M$, con $M \leftarrow \langle Salida \cup \pi_{\{NumRecorrido, IdLocalidad\}} Parada \rangle$. Con esto, el recorrido más largo será:
$\pi_{\{NumRecorrido\}} \langle N \bowtie_{K=L} (\aleph_{<max(K)\, L>} N) \rangle$

(1.2) En términos de tiempo, interesa el recorrido cuya última parada ocurra la mayor cantidad de tiempo posible después de la salida. Esto es:
$\pi_{\{NumRecorrido\}} \{ Parada \bowtie_{TiempoLleg=K} \langle \aleph_{<max(TiempoLleg)\, K>} Parada \rangle \}$

(2) Por cada recorrido, si se encuentran dos paradas con igual localidad pero distinto momento de llegada, se ha encontrado un ciclo. Luego la respuesta es:
$\pi_{\{NumRec\}} \sigma_{A.NumRec=B.NumRec \wedge A.IdLoc=B.IdLoc \wedge A.TiempoLleg \neq B.TiempoLleg} (A \times B)$,
en donde $A, B \leftarrow Parada$.

(3) Se supone que no es posible que dos paradas distintas de un mismo recorrido ocurran al mismo momento de llegada (es algo imposible físicamente). Luego el resultado se reduce a exigir que, dado un recorrido, las paradas previas sean tantas como las paradas posteriores, o las paradas posteriores sean una más. Sean $A, B, C \leftarrow Parada$. Entonces:
$R_1 \leftarrow \sigma_{A.NumRecorrido=B.NumRecorrido \wedge B.NumRecorrido=C.NumRecorrido} A \times B \times C$
$R_2 \leftarrow \sigma_{A.TiempoLleg>B.TiempoLleg \wedge A.TiempoLleg<C.TiempoLleg} R_1$
$R_3 \leftarrow_{<A.NumRecorrido>} \aleph_{<A.NumRecorrido\, N, count(B.IdLocalidad)\, U, count(C.IdLocalidad)\, V>} R_2$
Respuesta: $\sigma_{U=V \wedge U=V-1}(R_3)$.