

## Operaciones del Modelo Relacional

El modelo relacional define las siguientes operaciones, aplicables a una relación  $R$ .

- **Inserción:** Cuando se efectúan inserciones de tuplas en una instancia de una relación, se genera una nueva instancia. La inserción de una tupla implica las siguientes operaciones

- Verificar la restricción de integridad de identidad de la relación. La restricción se viola si existe una tupla en la relación con el mismo valor para los atributos de la clave primaria que la tupla que se inserta.
- Verificar las posibles restricciones de dominio en los valores de los atributos de la tupla. Una restricción de dominio se viola cuando el valor del atributo no pertenece al dominio indicado para el mismo.
- Verificar restricciones de nulidad de los atributos de la tupla. Si un atributo de la relación  $R$  se define como no nulo, la tupla que se inserta no podrá contener el valor *null* para este atributo.
- Verificar las restricciones de integridad referencial de la relación. Si  $R$  posee claves foráneas hacia otras relaciones  $R_1, \dots, R_n$ , el valor de los atributos que constituyen una clave foránea de  $R$  hacia  $R_i$  deben coincidir con el valor de los atributos que son clave primaria para alguna tupla de  $R_i$ . Esta condición se aplica si y sólo si los valores de los atributos a inserta no son todos nulos.

En caso de violarse alguna de las restricciones definidas sobre la relación, ésta debe ser rechazada o debe indicarse al usuario que indique nuevos valores para garantizar que la restricción si se satisfaga.

- **Eliminación:** Cuando se elimina una o más tuplas de una instancia de una relación  $R$  se genera una nueva instancia. La eliminación de tuplas tiene implícita la verificación de la restricción de integridad referencial de todas aquellas relaciones que poseen claves foráneas a la relación de cuya instancia se elimina(n) la(s) tupla(s). Si todas las restricciones se cumplen, la eliminación es realizada, sino pueden tenerse varias alternativas:

- **Eliminación en cascada:** La eliminación de la tupla es realizada pero tiene como efecto eliminar todas las tuplas de las instancias de relaciones con clave foránea hacia la relación de donde se elimina y cuyo valor de clave foránea sea igual al valor de la clave primaria de la(s) tupla(s) a eliminar. Por ejemplo, cuando se elimina un departamento, se eliminan todos los empleados que en trabajan en éste.
- **Eliminación nula:** La eliminación de la tupla es realiza pero tiene como efecto colocar en *null* el valor de los atributos que son claves foráneas en las tuplas de instancias de relaciones con clave foránea hacia la relación

donde se efectúa la eliminación. Por ejemplo, cuando se elimina un departamento, se coloca en null el valor del departamento en el que trabaja cada empleado que laboraba en el departamento eliminado.

- **Eliminación restringida:** La eliminación se rechaza si existen tuplas en las instancias de una relación con clave foránea a la relación donde se efectúa la eliminación, que tuviesen como valor de los atributos que son clave foránea el valor de los atributos que son clave primaria en la(s) tupla(s) a eliminar. Por ejemplo, un departamento no puede ser eliminado si hay empleados que trabajan en él.

En base a estas características de eliminación, es conveniente indicar como parte del esquema que tipo de restricción de eliminación (cascada, restringida o nula) aplicar cuando se desee eliminar una tupla de una instancia de cada relación del esquema.

- **Modificación:** La modificación del valor de un atributo en una tupla, sólo trae implícita la verificación de las posibles restricciones de dominio y nulidad definidas para el atributo. Si el atributo a modificar es una clave primaria, adicionalmente deben aplicarse las reglas de verificación de restricciones de integridad por identidad e integridad referencial descritas en la operación de inserción. Si se modifica un atributo que constituya una clave foránea deben verificarse las restricciones de integridad referencial.
- **Consulta:** La operación de consulta debe permitir expresar cualquier requerimiento sobre la base de datos. En tal sentido debe proveerse de un lenguaje de especificación. Para el modelo relacional se han definido dos grandes tipos de lenguajes:
  - Orientados a tuplas, donde la especificación del "query" se da en base a condiciones a ser satisfechas por tuplas y el resultado son tuplas
  - Orientados a conjuntos, donde la especificación del "query" se da en base a condiciones a ser satisfechas por un subconjunto de una relación (recordemos que una relación es un conjunto) y el resultados son conjuntos.

A nuestro nivel, estudiaremos dos formalismos para representar este tipo de operación: un formalismo orientado a conjuntos llamado Álgebra Relacional y un formalismo orientado a tuplas llamado Cálculo Relacional de Tuplas.

## **Álgebra Relacional**

El álgebra relacional provee de una serie de operadores basados en teoría de conjuntos que permiten obtener nuevas relaciones a partir de las existentes en una base de datos relacional. Estas nuevas relaciones constituyen la respuesta a una interrogación (consulta o "query") realizado sobre la base de datos.

**Convención:** A partir de este momento usaremos indistintamente el término relación para indicar una relación propiamente dicha o una instancia de la misma.

Los operadores del álgebra relacional se clasifican en dos tipos:

- **Operadores de la teoría de conjuntos:** Son aquellos utilizados para representar propiedades de conjuntos. Ellos son la unión ( $\cup$ ), la intersección ( $\cap$ ), la diferencia ( $-$ ) y el producto cartesiano ( $\times$ ).

Estos operadores permiten globalizar resultados a partir de resultados intermedios, de la siguiente forma:

- Tuplas de una relación  $R$  que satisfacen dos condiciones  $C_1$  y  $C_2$  (conjunción de condiciones). Si  $R_1$  es el subconjunto de  $R$  que satisface  $C_1$  y  $R_2$  es el subconjunto de  $R$  que satisface  $C_2$ , entonces el subconjunto de  $R$  que satisface  $C_1$  y  $C_2$  vendrá dado por  $R_1 \cap R_2$ .
- Tuplas de una relación  $R$  que satisfacen alguna de las condiciones  $C_1$  y  $C_2$  (disyunción inclusiva de condiciones). Si  $R_1$  es el subconjunto de  $R$  que satisface  $C_1$  y  $R_2$  es el subconjunto de  $R$  que satisface  $C_2$ , entonces el subconjunto de  $R$  que satisface  $C_1$  o  $C_2$  inclusivamente vendrá dado por  $R_1 \cup R_2$ .
- Tuplas de una relación que no satisface una condición  $C$ . Si  $R_1$  es el subconjunto de  $R$  que satisface  $C$ , entonces el subconjunto de  $R$  que no satisface  $C$  vendrá dado por  $R - R_1$ .
- Combinaciones posibles de tuplas de dos relaciones  $R_1$  y  $R_2$ . El conjunto de todas las posibles combinaciones de dos tuplas, una de  $R_1$  y una de  $R_2$  puede obtenerse a partir de la expresión  $R_1 \times R_2$ .

### **Ejemplos:**

Sean las relaciones

PERSONA(CI, NombreP)  
CANCION(CodigoC, NombreC)

EMISORA(CodigoE, NombreE)  
SINTONIZA(CI, CodEm)  
CI es clave foránea hacia PERSONA  
CodEm es clave foránea hacia EMISORA

GUSTA(CI,CodCan)

CI es clave foránea hacia PERSONA

CodCan es clave foránea hacia CANCION

COLOCA(CodEm,CodCan)

CodEm es clave foránea hacia EMISORA

CodCan es clave foránea hacia CANCION

- El requerimiento "Dar las emisoras que no colocan canciones que no le gustan a nadie" podría llevarse a cabo como:

- $EMI_1$  es el conjunto de las emisoras que colocan canciones que no le gustan a nadie.
- $EMI_2$  es el conjunto de las emisoras no colocan canciones que no le gustan a nadie.
- $EMI_2 \leftarrow EMISORA - EMI_1$

- El requerimiento "Dar las personas que sintonizan al menos la emisora X y la emisora Y, podría llevarse a cabo como:

- $SINTONIZA_X$  es el conjunto de personas que sintonizan la emisora X.
- $SINTONIZA_Y$  es el conjunto de personas que sintonizan la emisora Y.
- $SINTONIZA_{XY}$  es el conjunto de personas que sintonizan la emisora X y la emisora Y.
- $SINTONIZA_{XY} \leftarrow SINTONIZA_X \cap SINTONIZA_Y$

- El requerimiento "Dar las personas que sólo sintonizan la emisora X y la emisora Y", podría llevarse a cabo como:

- $SINTONIZA_X$  es el conjunto de personas que sintonizan la emisora X.
- $SINTONIZA_Y$  es el conjunto de personas que sintonizan la emisora Y.
- $SINTONIZA_{XY}$  es el conjunto de personas que sólo sintonizan la emisora X y la emisora Y.
- $SINTONIZA_{-XY}$  es el conjunto de personas que sintonizan emisoras que no son X e Y.
- $SINTONIZA_{XY} \leftarrow (SINTONIZA_X \cap SINTONIZA_Y) - SINTONIZA_{-XY}$

- El requerimiento "Dar las emisoras donde no se colocan la canciones Z y W" podría llevarse a cabo como:

- $COLOCA_Z$  es el conjunto de emisoras que colocan la canción Z.
- $COLOCA_W$  es el conjunto de emisoras que colocan la canción W.
- $COLOCA_{-ZW}$  es el conjunto de emisoras que no colocan la canción Z ni W.
- $COLOCA_{-ZW} \leftarrow COLOCA - (COLOCA_Z \cup COLOCA_W)$

Es importante destacar que para poder efectuar estas operaciones los conjuntos deben tener las mismas características. A nivel del modelo relacional esta implicación se traduce en el hecho de que las relaciones a unir, intersectar o restar tengan el mismo número de atributos y posicionalmente tengan los mismos dominios (el primer atributo del primer conjunto tenga el mismo dominio que el primer atributo del segundo conjunto y así sucesivamente).

- **Operadores sobre relaciones:** Si bien una relación matemáticamente es un conjunto, existen diferencias en cuanto al manejo ya que se habla de un conjunto cuyos elementos poseen estructura. Para tal fin, el álgebra relacional provee operaciones que permiten extraer información de cada tupla de una relación o tuplas de una relación que satisfacen una propiedad dada. Estas operaciones a su vez, pueden clasificarse en dos tipos:
  - Operaciones Primitivas: no pueden expresarse con ninguna construcción de las antes mencionadas.
  - Operaciones Derivadas: pueden construirse a partir de operaciones primitivas y operaciones de conjuntos.

Las operaciones primitivas son:

- **Selección:** Permite obtener las tuplas de una relación que cumplen con determinada condición. La condición utilizada es una expresión booleana de la forma:
  - $\langle \text{atributo}_1 \rangle \langle \text{operador} \rangle \langle \text{atributo}_2 \rangle$ , donde  $\langle \text{operador} \rangle$  puede ser  $\neq, =, <, >, \geq$  o  $\leq$ .
  - $\langle \text{atributo}_1 \rangle \langle \text{operador} \rangle \langle \text{valor}_1 \rangle$ , donde  $\langle \text{operador} \rangle$  puede ser  $\neq, =, <, >, \geq$  o  $\leq$ .
  - $\langle \text{condicion}_1 \rangle \text{ AND } \langle \text{condicion}_2 \rangle$ , donde  $\langle \text{condicion}_1 \rangle$  y  $\langle \text{condicion}_2 \rangle$  pueden tener cualquiera de las formas de construcción aquí dadas.
  - $\langle \text{condicion}_1 \rangle \text{ OR } \langle \text{condicion}_2 \rangle$ , donde  $\langle \text{condicion}_1 \rangle$  y  $\langle \text{condicion}_2 \rangle$  pueden tener cualquiera de las formas de construcción aquí dadas.

El operador de selección se denota como

$$\sigma_{\langle c \rangle}(\langle R \rangle)$$

donde  $\langle R \rangle$  es la relación de la cual se quieren seleccionar tuplas, y  $\langle c \rangle$  es la condición que deben cumplir las tuplas de  $\langle R \rangle$  para poder ser seleccionadas.

**Ejemplos:** Con esta nueva operación, podremos acercarnos más a la expresión que resuelve la consulta.

- "Dar las personas que sintonizan al menos la emisora X y la emisora Y"

- $\text{SINTONIZA}_X \leftarrow \sigma_{\langle \text{CodEm} = X \rangle}(\text{SINTONIZA})$
- $\text{SINTONIZA}_Y \leftarrow \sigma_{\langle \text{CodEm} = Y \rangle}(\text{SINTONIZA})$
- $\text{SINTONIZA}_{XY} \leftarrow \text{SINTONIZA}_X \cap \text{SINTONIZA}_Y$

En este caso, recordemos que los elementos de  $\text{SINTONIZA}_X$  deben ser cédulas de identidad y no pares  $\langle \text{CI}, \text{CodEm} \rangle$ . En tal sentido, el conjunto  $\text{SINTONIZA}_{XY}$  aún no devuelve el resultado esperado (devuelve vacío).

- "Dar las personas que sólo sintonizan la emisora X y la emisora Y":

- $SINTONIZA_X \leftarrow \sigma_{\langle CodEm = X \rangle}(SINTONIZA)$
- $SINTONIZA_Y \leftarrow \sigma_{\langle CodEm = Y \rangle}(SINTONIZA)$
- $SINTONIZA_{-XY} \leftarrow \sigma_{\langle CodEm \neq X \text{ OR } CodEm \neq Y \rangle}(SINTONIZA)$
- $SINTONIZA_{XY} \leftarrow (SINTONIZA_X \cap SINTONIZA_Y) - SINTONIZA_{-XY}$

$SINTONIZA_{XY}$  realmente no devuelve el resultado ya que  $SINTONIZA_X$  tiene tuplas de la forma  $\langle CI, 'X' \rangle$  y  $SINTONIZA_Y$  tuplas de la forma  $\langle CI, 'Y' \rangle$ . Luego, la intersección de estos conjuntos es vacía y la diferencia con  $SINTONIZA_{-XY}$  es vacía. Por lo tanto el resultado es vacío.

- "Dar las emisoras donde no se colocan ni la canción Z ni la canción W"

- $COLOCA_Z \leftarrow \sigma_{\langle CodCan = Z \rangle}(COLOCA)$
- $COLOCA_W \leftarrow \sigma_{\langle CodCan = W \rangle}(COLOCA)$
- $COLOCA_{-ZW} \leftarrow COLOCA - (COLOCA_Z \cap COLOCA_W)$

En este requerimiento no hay problemas, sin embargo se da información superflua, ya que el resultado no son sólo las emisoras que no colocan las canciones Z y W sino también las otras canciones que ellas colocan.

Nótese que cada uno de estos requerimientos pedían personas o emisoras, sin embargo la operación de selección devuelve toda la tupla y no parte de ella. Esto traía un conveniente ya que hacía que o el query no devolviese la información o por el contrario diese información de más. Para resolver estos problemas, se define el otro operador primitivo del álgebra relacional: la proyección.

- **Proyección:** Como su nombre lo indica, permite proyectar una relación sobre un conjunto de atributos de esta. Como hemos utilizado el enfoque posicional, la operación toma una lista de atributos (obviamente los atributos deben existir en la relación a proyectar) y devuelve una serie de tuplas donde el orden de atributos en ellas es el orden de la lista de atributos.

La operación de proyección no necesariamente se aplica sobre un superconjunto de los atributos que son clave primaria. En tal sentido, puede pensarse que la proyección puede traer como consecuencia la aparición de tuplas repetidas. Esto es totalmente falso: en el resultado no se incluyen duplicados. Esto es debido al hecho de que en conjuntos la distinguibilidad de elementos se realiza por el valor del mismo y no por el nombre asignado.

Este operador se denota como

$$\pi_{\langle L \rangle}(\langle R \rangle)$$

donde  $\langle R \rangle$  es la relación a proyectar, y  $\langle L \rangle$  es una secuencia de nombres de atributos seguidos por coma y constituye la lista de atributos sobre los cuales proyectar  $\langle R \rangle$ .

**Ejemplos:** Ahora si podemos responder correctamente a algunos de los "queries" de la base de datos de emisoras y gustos musicales.

- "Dar las personas que sintonizan al menos la emisora X y la emisora Y"
  - $SINTONIZA_X \leftarrow \pi_{CI}(\sigma_{CodEm = X}(SINTONIZA))$
  - $SINTONIZA_Y \leftarrow \pi_{CI}(\sigma_{CodEm = Y}(SINTONIZA))$
  - $SINTONIZA_{XY} \leftarrow SINTONIZA_X \cap SINTONIZA_Y$
- "Dar las personas que sólo sintonizan la emisora X y la emisora Y":
  - $SINTONIZA_X \leftarrow \pi_{CI}(\sigma_{CodEm = X}(SINTONIZA))$
  - $SINTONIZA_Y \leftarrow \pi_{CI}(\sigma_{CodEm = Y}(SINTONIZA))$
  - $SINTONIZA_{-XY} \leftarrow \pi_{CI}(\sigma_{CodEm \neq X \text{ OR } CodEm \neq Y}(SINTONIZA))$
  - $SINTONIZA_{XY} \leftarrow (SINTONIZA_X \cap SINTONIZA_Y) - SINTONIZA_{-XY}$
- "Dar las emisoras donde no se coloca ni la canción Z ni la canción W"
  - $COLOCA_Z \leftarrow \pi_{CodEm}(\sigma_{CodCan = Z}(COLOCA))$
  - $COLOCA_W \leftarrow \pi_{CodEm}(\sigma_{CodCan = W}(COLOCA))$
  - $COLOCA_{-ZW} \leftarrow \pi_{CodEm}(COLOCA) - (COLOCA_Z \cup COLOCA_W)$

Las operaciones derivadas son:

- **Join (Enlace):** Permite enlazar relaciones a través de atributos comunes para reconstruir información. Con esta operación se permite implantar la noción de navegación en la base de datos que hemos estado mencionando a lo largo del curso.

Sean  $R(A_1, \dots, A_n)$  y  $S(B_1, \dots, B_m)$  dos relaciones. El operador de join de R y S respecto a una condición <Condición> se expresa como:

$$R \bowtie_{\langle \text{Condición} \rangle} S$$

El resultado de la operación es una relación Q con atributos  $Q(A_1, \dots, A_n, B_1, \dots, B_m)$  donde existirá una tupla por cada una de las combinaciones posibles de tuplas de R y S tales que satisfagan la condición dada.

Así, el operador de join puede verse como un operador reproducible a partir de los operadores de producto cartesiano y selección antes descritos, como:

$$R \bowtie_{\langle \text{Condición} \rangle} S = \sigma_{\langle \text{Condición} \rangle} (R \times S)$$

La condición del join sigue las reglas de formación descritas para el operador de selección.

### Ejemplos:

- "Dar el nombre de las canciones que le gustan a José Pérez". Esta consulta es directa. En la relación GUSTA tenemos la cédula de identidad y el código de la canción. Luego si seleccionamos la cédula de José Perez de la relación PERSONA y hacemos join con la relación GUSTA tendríamos el código de todas las canciones que le gustan a "José Pérez". Finalmente si hacemos join con la relación CANCION podremos saber el nombre de todas estas canciones. Así:

- Definamos JOSE como la relación que indica la cédula de identidad de José Pérez.

$$JOSE \leftarrow \pi_{CI}(\sigma_{NombreP='José Pérez'}(COLOCA))$$

- Definamos GS<sub>JP</sub> como la relación formada por los códigos de las canciones que le gustan a José Pérez.

$$GS_{JP} \leftarrow \pi_{GUSTA.CodCan}(JOSE \bowtie_{Jose.CI = Gusta.CI} GUSTA)$$

- El resultado (RES) será:

$$RES \leftarrow \pi_{NombreP}(GS_{JP} \bowtie_{GS_{JP}.CodCan = Cancion.CodCan} CANCION)$$

- "Dar el nombre de los radioescuchas que sintonizan sólo emisoras que colocan la canción W". Si conocemos los radioescuchas que sintonizan emisoras que coloquen la canción W y los radioescuchas que sintonizan emisoras que no coloquen la canción W, la diferencia entre estas dos relaciones nos dará los radioescuchas que sólo sintonizan emisoras que colocan la canción W. Así:

- Definamos EM<sub>W</sub> como las emisoras que colocan al menos la canción W.

$$EM_W \leftarrow \pi_{CodEm}(\sigma_{CodCan=W}(COLOCA))$$

- EM<sub>-W</sub> como las emisoras que no colocan la canción W.

$$EM_{-W} \leftarrow \pi_{CodEm}(EMISORA) - EM_W$$

- PERS<sub>W</sub> como las personas que sintonizan emisoras donde se coloca la canción W.

$$PERS_W \leftarrow \pi_{CI}(SINTONIZA \bowtie_{SINTONIZA.CodEm = EM_W.CodEm} EM_W)$$



- PERSONAS<sub>-W</sub> como las personas que sintonizan emisoras donde no se coloca la canción W.

$$PERS_{-W} \leftarrow \pi_{CI}(SINTONIZA \triangleright \triangleleft SINTONIZA.CodEm = EM_{-W}.CodEm \text{ } EM_{-W})$$

- Finalmente, el resultado del query sería la relación RES definida como

$$RES \leftarrow \pi_{NombreP}((PERS_W - PERS_{-W}) \triangleright \triangleleft CI = Persona.CI \text{ } PERSONA)$$

Hasta ahora hemos hablado en general de la operación de JOIN. Sin embargo, esta operación suele denominarse de dos formas diferentes dependiendo de los operadores de comparación usados en la expresión. Si la expresión incluye cualquier operador de comparación se habla de un **Theta Join**. Si la expresión sólo utiliza el operador de igualdad se habla de **Equijoin**.

Si observamos la definición de join, notamos que en caso de equijoins, se cumple que en la relación resultante siempre existirán uno o más pares de atributos que tendrán el mismo valor en cada tupla. Para eliminar esta redundancia de información en la relación resultante de un equijoin se introduce un nuevo operador: el **Natural Join**.

Sean  $R(A_1, A_2, \dots, A_n)$  y  $S(B_1, \dots, B_2)$  dos relaciones. El natural join de R y S sobre los atributos  $(A_{i1}, \dots, A_{is})$  de R y  $(B_{j1}, \dots, B_{js})$  de S se denota como

$$R *_{(A_{i1}, \dots, A_{is}), (B_{j1}, \dots, B_{js})} S$$

El resultado de esta operación será igual a:

$$R *_{(A_{i1}, \dots, A_{is}), (B_{j1}, \dots, B_{js})} S = \pi_L(R \mid \times \mid_C S)$$

donde:

- L es la lista formada por los atributos de R y los atributos de S menos  $\{B_{j1}, \dots, B_{js}\}$ .
- $C = \bigwedge_{n=1..s} (A_{in} = B_{jn})$  (conjunción de la igualdad de todos los atributos implícitos en la condición).

**Ejemplo:** Si consideramos las relaciones ESTUDIANTE y CARRERA dadas a continuación,

Carnet	CI	Nombre	Indice	Carrera
91-23218	2.345.333	José Gaspar	3.2764	0300
90-19888	6.786.669	Luis Alonso	3.8766	0800
92-24418	10.234.221	Diana Gómez	4.2034	0800

Codigo	Descripcion
0300	Ingeniería Mecánica

0800	Ingeniería de la Computación
------	------------------------------

Entonces ESTUDIANTE  $\bowtie$   $\langle \text{Carrera} = \text{Codigo} \rangle$  CARRERA sería)

Carnet	CI	Nombre	Indice	Carrera	Codigo	Descripción
91-23218	2.345.333	José Gaspar	3.2764	0300	0300	Ingeniería Mecánica
90-19888	6.786.669	Luis Alonso	3.8766	0800	0800	Ingeniería de la Computación
92-24418	10.234.221	Diana Gómez	4.2034	0800	0800	Ingeniería de la Computación

Y ESTUDIANTE  $\ast_{(\text{Carrera})(\text{Codigo})}$  CARRERA daría como resultado:

Carnet	CI	Nombre	Indice	Carrera	Descripción
91-23218	2.345.333	José Gaspar	3.2764	0300	Ingeniería Mecánica
90-19888	6.786.669	Luis Alonso	3.8766	0800	Ingeniería de la Computación
92-24418	10.234.221	Diana Gómez	4.2034	0800	Ingeniería de la Computación

- **Operador de División:** Con los operadores anteriores obteníamos tuplas que satisfacían una condición pero la verificación se hacía sobre una sola tupla. ¿Qué sucede si la condición debe ser compartida por muchas tuplas y no por sólo una? Por ejemplo supongamos el siguiente requerimiento:
- "Una persona se llama radioescucha incondicional si le gustan al menos todas las canciones que coloca una emisora. Dar el nombre de los radioescuchas incondicionales de la emisora X".

En este caso queremos obtener un conjunto cuyas tuplas deben satisfacer una propiedad sobre un conjunto de tuplas. El operador de división permite expresar este tipo de consultas.

El operador de división puede expresarse de la siguiente forma:

Sea  $r_1$  una instancia de una relación  $R_1(A_1)$  definida sobre el conjunto de atributos  $A_1$  y  $r_2$  una instancia de una relación  $R_2(A_2)$  definida sobre el conjunto de atributos  $A_2$ , tales que se cumple que  $A_2 \subseteq A_1$ . Sea  $D = A_1 - A_2$  el conjunto de atributos de  $R_1$  que no están en  $R_2$ . La división de  $r_1$  respecto a  $r_2$  (la denotaremos como  $r_1 / r_2$ ) permite obtener una instancia  $r_{r_1/r_2}$  de una relación  $\text{Res}(D)$  definida sobre el conjunto de atributos  $D$  tal que se cumple:

$$t \in r_{r_1/r_2} \leftrightarrow \forall t_{r_2} \exists t_{r_1} (t_{r_2} \in r_2 \wedge t_{r_1} \in r_1 \wedge t_{r_1}[A_2] = t_{r_2} \wedge t_{r_1}[D] = t)$$

¿ Qué quiere decir esto? Básicamente con este operador se pueden obtener todas las sub tuplas de  $r_1$  sobre los atributos del conjunto D que cumplen con que al ser proyectadas sobre  $A_2$  generan todas las tuplas de  $r_2$ .

Cómo puede describirse este operador utilizando operadores primitivos?

- Para poder obtener todas las sub tuplas de  $R_1$  sobre el conjunto de atributos D debemos hacer:

$$TUPLAS_D \leftarrow \pi_D(R_1)$$

Este conjunto es un superconjunto del conjunto resultado del query ya que pueden incluir sub tuplas correspondientes a tuplas de  $R_1$  que no cumplen con la condición de que al ser proyectadas sobre  $A_2$  generen las tuplas de  $R_2$ . Así necesitamos las tuplas de  $R_1$  que no cumplen con esta condición y restarlas de  $TUPLAS_D$ .

- El conjunto de tuplas de  $R_1$  que no satisfacen la condición de que al ser proyectadas sobre  $A_2$  generen todas las tuplas de  $R_2$  puede obtenerse así:
  - Se generan todas las posibles tuplas resultantes de combinar las tuplas en  $TUPLAS_D$  con las tuplas de  $R_2$ .

$$TUPLAS_1 \leftarrow R_2 \times TUPLAS_D$$

Este conjunto es compatible con el dominio de  $R_1$  ya que  $D \cup A_2 = A_1$ .

- Si se resta  $R_1$  de este conjunto y el resultado se proyecta sobre D obtenemos las tuplas de  $R_1$  que no satisfacen la condición deseada y por lo tanto es lo que necesitamos restar de  $TUPLAS_D$ .

$$TUPLAS_2 \leftarrow \pi_D(TUPLAS_1 - R_1)$$

- Finalmente, el resultado deseado es:

$$TUPLAS_{R_1/R_2} \leftarrow TUPLAS_D - TUPLAS_2$$

Regresemos al query que habíamos dejando a un lado: Una forma sencilla de hacerlo utilizaría el operador de división así:

- La propiedad sobre más de una tupla es el hecho de que halla gusto por todas las canciones que coloca la emisora X. Así, requerimos las canciones que coloca la emisora X.

$$CANCIONES_X \leftarrow \pi_{CodCan}(\sigma_{CodEm}(COLOCA))$$

- La relación *GUSTA* posee un superconjunto de los atributos de la relación *CANCIONES<sub>x</sub>*. En tal sentido, si dividimos *GUSTA* entre *CANCIONES<sub>x</sub>* obtendremos las cédulas de identidad de las personas que tienen gusto por las canciones de *CANCIONES<sub>x</sub>*. Así:

$$\text{INCONDICIONAL}_x \leftarrow \text{GUSTA} / \text{CANCIONES}_x$$

- Por último, para poder saber el nombre de esas personas basta con hacer join con la relación *PERSONA* y proyectar sobre su nombre.

$$\text{NOMBRE}_{\text{INCONDICIONALX}} \leftarrow \pi_{\text{NombreP}}(\text{PERSONA}) \bowtie \langle \text{Persona.CI} = \text{INCONDICIONALX.CI} \text{ INCONDICIONALX} \rangle$$

A este operador se le denomina división porque permite expresar la noción de una resta sucesiva y a través de restas sucesivas podemos visualizar la operación de división.

### Otros operadores relacionales

- **Renombramiento de Atributos/Relaciones:** El operador de renombramiento permite asignar nombres diferentes a una relación, a sus atributos o a la relación y sus atributos simultáneamente. El operador de renombre se representa a través de la letra  $\rho$ . Este operador tiene las siguientes formas:
  - $\rho_S(B_1, B_2, \dots, B_n)(R)$  que renombra *R* como *S* y sus atributos a *B<sub>1</sub>, B<sub>2</sub>, ..., B<sub>n</sub>*
  - $\rho_S(R)$  que renombra *R* a *S*.
  - $\rho_{(B_1, B_2, \dots, B_n)}(R)$  que renombra los atributos de *R* a *B<sub>1</sub>, B<sub>2</sub>, ..., B<sub>n</sub>*.
- **Funciones de Agregación.** Los operadores del álgebra relacional no siempre permiten expresar cierto tipo de consultas que trabajan resumiendo los valores almacenados en una relación. Por ejemplo, en la base de datos de emisoras y gustos musicales no es posible satisfacer la consultar "Dar la emisora que posee el mayor número de radioescuchas". A este tipo de funciones que resumen ciertas propiedades de las tuplas de una relación se denominan funciones de agregación. El operador de agregación general se denotará como

$$\langle \text{atributos de agrupación} \rangle \mathcal{F} \langle \text{funciones de agregación} \rangle (R)$$

donde:

- $\langle \text{atributos de agrupación} \rangle$  es una lista de atributos de la relación *R* sobre los que se establecerá la agregación. La lista de atributos se utiliza para dividir la relación sobre la cual se aplica la agregación en grupos de tuplas similares. Existirán tantos grupos en una instancia de la relación como posibles valores para la lista de atributos de agrupación existan en la instancia. Si la lista de

atributos de agrupación es vacía se considera un único grupo constituido por la instancia de la relación.

**Ejemplo:** Si se considera instancia de la COLOCA

CodEm	CodCan	
1	1	$t_1$
2	3	$t_2$
3	2	$t_3$
2	2	$t_4$
2	4	$t_5$
3	5	$t_6$

Se tiene que:

- Si se considera como atributos de agrupación al atributo CodEm, existirán tres grupos, uno para cada uno de los tres posibles valores de CodEm, es decir  $G_1=\{t_1\}$ ,  $G_2=\{t_2, t_4, t_5\}$  y  $G_3=\{t_3, t_6\}$
  - Si se considera como atributos de agrupación al atributo CodCan, existirán cinco grupos, uno para cada uno de los tres posibles valores de CodCan, es decir  $G_1=\{t_1\}$ ,  $G_2=\{t_2\}$ ,  $G_3=\{t_5\}$ ,  $G_4=\{t_3, t_4\}$  y  $G_5=\{t_6\}$
  - Si se considera como atributos de agrupación al atributo CodCan y CodEm existirán tantos grupos como tuplas en la instancia, ya que CodCan y CodEm juntas constituyen la clave primaria de la relación COLOCA.
- *<funciones de agregación>* es una lista de operadores de agregación. Los operadores de agregación válidos son los siguientes:
- COUNT *<atributo>*, que determina el número de tuplas.
  - AVG *<atributo>*, que determina el promedio de los valores de *<atributo>* en la instancia de la relación.
  - MAXIMUM *<atributo>*, que determina el máximo de los valores de *<atributo>*. A pesar de que este operador se incluye como función de agregación puede representarse a través de operadores del álgebra relacional.
  - MINIMUM *<atributo>*, que determina el mínimo de los valores de *<atributo>*. A pesar de que este operador se incluye como función de agregación puede representarse a través de operadores del álgebra relacional.

**Ejemplo:**

- La consulta "Dar el nombre de las emisoras que poseen el mayor número de radioescuchas" puede representarse a través del siguiente programa en álgebra relacional.

SINTONIA(Emisora,NumRE)  $\leftarrow$   $\text{CodEm} \rightarrow \text{COUNT}_{CI}(\text{SINTONIZA})$   
 MAS\_ESCUCHADA  $\leftarrow$   $\text{MAXIMUM}_{NumRE}(\text{SINTONIA})$   
 RES  $\leftarrow \pi_{NombreE}(\text{EMISORA} \bowtie_{\text{Codigo} = \text{Emisora}} \text{SINTONIA})$

- **Outer Joins.** Los operadores de join vistos anteriormente producen como resultado las tuplas para las cuales se satisface la condición dada. Esto quiere decir que las tuplas para las que no existe valor asociado no se produce una fila resultado. Los operadores de outer join extienden esta noción añadiendo también las tuplas para las cuales no existe enlace. Estas tuplas tendrán valores nulos para los atributos correspondientes a la relación para la cual no existe enlace. Existen tres variantes del outer join:

- **Left Outer Join:** El "left outer join" de dos relaciones R y S sobre una condición C devuelve todas y cada de las posibles combinaciones de R y S que satisfacen la condición C. Adicionalmente, aquellas tuplas de R para las cuales no existe un enlace en S de acuerdo a la condición C se representarán como resultado del "left outer join" pero con valores nulos para las columnas de la relación S. El "left outer join" de las relaciones R y S se denota como

$$R \bowtie_C S$$

- **Right Outer Join:** El "right outer join" de dos relaciones R y S sobre una condición C devuelve todas y cada de las posibles combinaciones de R y S que satisfacen la condición C. Adicionalmente, aquellas tuplas de S para las cuales no existe un enlace en R de acuerdo a la condición C se representarán como resultado del "right outer join" pero con valores nulos para las columnas de la relación R. El "right outer join" de las relaciones R y S se denota como

$$R \bowtie_C S$$

- **Full Outer Join:** El "full outer join" de dos relaciones R y S sobre una condición C devuelve todas y cada de las posibles combinaciones de R y S que satisfacen la condición C. Adicionalmente, aquellas tuplas de R para las cuales no existe un enlace en S de acuerdo a la condición C se representarán como resultado del "full outer join" pero con valores nulos para las columnas de la relación S. De la misma forma, aquellas tuplas de S para las cuales no existe un enlace en R de acuerdo a la condición C se representarán como resultado del "right outer join" pero con valores nulos para las columnas de la relación R. Un "full outer join" de dos relaciones R y S puede verse como la unión de los resultados producidos por el "left outer join" y el "right outer join" de estas relaciones. El "right outer join" de las relaciones R y S se denota como

$$R \bowtie_C S$$

**Ejemplo:**

Si se consideran las siguientes instancias de las relaciones PERSONA, SSINTONIZA y EMISORA

**PERSONA**

CI	NombreP
1739108	José Pérez
6283353	Ricardo Díaz
10934828	Amalia Bustamante
6300336	Margarita Poveda


**EMISORA**

CodigoE	NombreE
10	Radio Merengue
20	Recuerdos 1500FM
30	Exitos 1900
40	Novedades Latinas 780


**SINTONIZA**

CI	CodEm
6283353	20
6283353	40
6300336	10
10934828	20
6300336	40

Los siguientes "outer joins" generan las relaciones que a continuación se muestran

PERSONA  PERSONA.CI = SINTONIZA.CI SINTONIZA

PERSONA-CI	NombreP	Sintoniza-CI	CodEm
1739108	José Pérez	NULL	NULL
6283353	Ricardo Díaz	6283353	20
6283353	Ricardo Díaz	6283353	40
6300336	Amalia Bustamante	6300336	10
6300336	Amalia Bustamante	6300336	40
10934828	Margarita Poveda	10934828	20

PERSONA  PERSONA.CI = SINTONIZA.CI SINTONIZA

PERSONA-CI	NombreP	Sintoniza-CI	CodEm
6283353	Ricardo Díaz	6283353	20
6283353	Ricardo Díaz	6283353	40
6300336	Amalia Bustamante	6300336	10
6300336	Amalia Bustamante	6300336	40
10934828	Margarita Poveda	10934828	20

PERSONA  $\bowtie$  PERSONA.CI = SINTONIZA.CI SINTONIZA

PERSONA - CI	NombreP	Sintoniza - CI	CodEm
1739108	José Pérez	NULL	NULL
6283353	Ricardo Díaz	6283353	20
6283353	Ricardo Díaz	6283353	40
6300336	Amalia Bustamante	6300336	10
6300336	Amalia Bustamante	6300336	40
10934828	Margarita Poveda	10934828	20

SINTONIZA  $\bowtie$  PERSONA.CI = SINTONIZA.CI EMISORA

CI	CodEm	CodigoE	NombreE
6283353	20	20	Recuerdos 1500FM
6283353	40	40	Novedades Latinas 780
6300336	10	10	Radio Merengue
6300336	40	40	Novedades Latinas 780
10934828	20	20	Recuerdos 1500FM

SINTONIZA  $\bowtie$  PERSONA.CI = SINTONIZA.CI EMISORA

CI	CodEm	CodigoE	NombreE
6283353	20	20	Recuerdos 1500FM
6283353	40	40	Novedades Latinas 780
6300336	10	10	Radio Merengue
6300336	40	40	Novedades Latinas 780
10934828	20	20	Recuerdos 1500FM
NULL	NULL	30	Exitos 1900

SINTONIZA  $\bowtie$  PERSONA.CI = SINTONIZA.CI EMISORA

CI	CodEm	CodigoE	NombreE
6283353	20	20	Recuerdos 1500FM
6283353	40	40	Novedades Latinas 780
6300336	10	10	Radio Merengue
6300336	40	40	Novedades Latinas 780
10934828	20	20	Recuerdos 1500FM
NULL	NULL	30	Exitos 1900



**TAREA:**

Expresar los siguientes requerimientos de la base de datos de gustos musicales y emisoras radiales en álgebra relacional:

- Dar los nombres de las personas que no les gusta ninguna canción que le guste a "Jorge Martínez".
- Dar las canciones que le gusta a los radioescuchas de la emisora Z.
- Dar las personas que no escuchan las emisoras que escucha "Luis Ramírez".
- Dar las emisoras que colocan sólo canciones que no le gustan a la persona P.
- Dar las canciones que le gustan a las persona que sólo escuchan emisoras que no colocan canciones que le gustan a los radioescuchas de la emisora W.