0 Álgebra relacional

0.1 Introducción

El álgebra relacional es un conjunto de operaciones sobre relaciones. Cada operación toma una o más relaciones como su(s) operando(s) y produce otra relación como resultado; por lo tanto las expresiones del álgebra relacional pueden anidarse a cualquier profundidad.

Codd definió originalmente un conjunto de operaciones del álgebra relacional y demostró que cualquier resultado del álgebra relacional puede obtenerse a partir de dichas operaciones. Matemáticamente se dice que este conjunto es completo y correcto. Este conjunto está formado por las operaciones: selección, proyección, producto cartesiano, unión y diferencia.

Existe otro conjunto de operaciones que pueden derivarse a partir de las operaciones básicas. Se incluyen como parte del álgebra relacional con el fin de simplificar las consultas que se hacen a una base de datos. Este conjunto está compuesto por las operaciones: intersección, junta, junta natural y división.

El conjunto de operaciones del álgebra relacional define un lenguaje que sirve para hacer consultas a una base de datos. Sin embargo, en la práctica comercial este lenguaje no se usa por la dificultad que presenta para la especificación de sus operaciones.

Entonces, se estudia con el propósito de presentar los aspectos teóricos formales relacionados con el lenguaje de manipulación de datos del modelo relacional, lo cual sirve de base para el estudio posterior de SQL (Structured Query Language) que es el lenguaje utilizado en el mundo comercial para la explotación de bases de datos relacionales.

0.2 Operaciones básicas

OPERACIÓN DE SELECCIÓN (σ)

Esta operación produce un subconjunto "horizontal" de una relación dada; es decir, el subconjunto de tuplas (t), tal que cada tupla t satisface una condición de selección (o sea, t hace que la condición sea verdadera). La operación se representa como:

$$\sigma_{\text{<condición de selección>}} \text{(Relación)}$$

Ejemplo:

Cursos

Nombre_c	Créditos	Nombre_Prof
Computación I	9	Luisa
Economía II	9	Roberto
Matemáticas I	8	Juan
Econometría	7	Roberto
Algorítmica	8	Luisa

$$\sigma_{\text{Nombre\_Prof = "Luisa"}} \text{(Cursos)}$$

Nombre_c	Créditos	Nombre_Prof
Computación I	9	Luisa
Algorítmica	8	Luisa

La relación resultante tiene *los mismos atributos* que la relación original. La <condición de selección> está compuesta por expresiones de comparación de la forma:

<nombre atributo> <op_comp> <constante> , o
<nombre atributo> <op_comp> <nombre atributo>

donde <op_comp> es cualquier operador de comparación {=, <, etc.}. Estas expresiones de comparación pueden estar conectadas arbitrariamente por los operadores lógicos AND, OR y NOT.

Características y propiedades de σ

- El operador es unario; por lo tanto no se pueden seleccionar tuplas de más de una relación.
- La operación se aplica individualmente a cada tupla de la relación.
- El conjunto de tuplas seleccionado se conoce como la **selectividad** de la condición de selección.

– La operación es **conmutativa**, es decir:

$$\sigma_{<cond1>} (\sigma_{<cond2>} (R)) = \sigma_{<cond2>} (\sigma_{<cond1>} (R))$$

– Operaciones Select en **cascada** se pueden sustituir por una condición de selección conjuntiva:

$$\sigma_{<cond1>} (\sigma_{<cond2>} (...(\sigma_{<condn>} (R))...)) =$$
$$\sigma_{<cond1>\ AND\ <cond2>\ AND\ ---\ AND\ <condn>} (R)$$

OPERACIÓN DE PROYECCIÓN (π)

La operación de proyección produce un subconjunto "vertical" de una relación dada, es decir, el subconjunto obtenido al tomar los atributos especificados, de izquierda a derecha, y eliminando luego las tuplas duplicadas en los atributos seleccionados. La operación se representa como:

$$\pi_{<lista\ de\ atributos>} (Relación)$$

Ejemplo:

Cursos

Nombre_c	Créditos	Nombre_Prof
Computación I	9	Luisa
Economía II	9	Roberto
Matemáticas I	8	Juan
Econometría	7	Roberto
Algorítmica	8	Luisa

$$\pi_{Nombre\_Prof} (Cursos)$$

Nombre_Prof
Luisa
Roberto
Juan
~~Roberto~~
~~Luisa~~

(Se eliminan duplicados)

Características y propiedades de π

– La expresión

$\pi_{\text{<lista1>}} (\pi_{\text{<lista2>}} (R)) = \pi_{\text{<lista1>}} (R)$

es válida, si <lista2> contiene a los atributos que están en <lista1>. En caso, contrario, no se sostiene.

– La operación **no** es conmutativa.

OPERACIONES DE CONJUNTOS

PRODUCTO CARTESIANO (X)

El producto cartesiano de dos relaciones, $R(A_1, A_2, \ldots, A_n)$ X $S(B_1, B_2, \ldots, B_m)$, es una relación $Q(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$, en ese orden. La relación resultante Q tiene una tupla por cada combinación de tuplas: una de R y una de S. De esta manera, si R tiene n_R tuplas y S tiene n_S tuplas, entonces R X S tendrá $n_R * n_S$ tuplas.

Ejemplo:

Profesores

Nombre	Categoría
Roberto	tc
Luisa	mt

Cursos

Nombre_c	Créditos	Nombre_Prof
Computación I	9	Luisa
Economía II	9	Roberto
Matemáticas I	8	Juan

Profesores X Cursos

Nombre	Categoría	Nombre_c	Créditos	Nombre_Prof
Roberto	tc	Computación I	9	Luisa
Roberto	tc	Economía II	9	Roberto
Roberto	tc	Matemáticas I	8	Juan
Luisa	mt	Computación I	9	Luisa
Luisa	mt	Economía II	9	Roberto
Luisa	mt	Matemáticas I	8	Juan

UNIÓN (U)

La unión de dos relaciones R y S (de tipos compatibles), es el conjunto de todas las tuplas (t), tal que cada tupla t pertenece a R o pertenece a S.

Ejemplo:

Hombres

Nombre	Fecha_nac
Juan	3/abr/80
Luis	25/feb/90
Armando	28/mar/75
José	30/dic/82

Mujeres

Nombre	Fecha_nac
Adriana	15/oct/72
María	30/sep/80
Oralia	8/jun/79
Ana	24/ene/62
Dolores	30/jul/70

Hombres U Mujeres

Nombre	Fecha_nac
Juan	3/abr/80
Luis	25/feb/90
Armando	28/mar/75
José	30/dic/82
Adriana	15/oct/72
María	30/sep/80
Oralia	8/jun/79
Ana	24/ene/62
Dolores	30/jul/70

DIFERENCIA (-)

La diferencia entre dos relaciones R y S (de tipos compatibles), en este orden R-S, es el conjunto de todas las tuplas (t), tal que cada tupla t pertenece a R y no pertenece a S.

Ejemplo:

Autos

Marca	País_origen
V.W.	Alemania
Ford	E.U.A.
Renault	Francia
Nissan	Japón

Autos_compactos

Marca	País_origen
V.W.	Alemania
Nissan	Japón
Renault	Francia

Autos - Autos_compactos

Marca	País_origen
Ford	E.U.A.

Características y propiedades de las operaciones de conjuntos

- Dos relaciones $R(A_1, A_2, \ldots, A_n)$ y $S(B_1, B_2, \ldots, B_n)$ son **de tipos compatibles** si tienen el mismo grado n (misma cantidad de atributos), y si $Dom(A_i) = Dom(B_i)$, para $1 \leq i \leq n$ (los atributos en la misma posición tienen el mismo dominio de valores).
- Se adopta la convención de que la relación resultante tiene los mismos nombres de atributos que la primera relación especificada con el operador.
- En la unión, los duplicados se eliminan (en términos generales, ninguna relación tiene tuplas duplicadas).
- La unión es conmutativa:

  $R \cup S = S \cup R$

- La unión puede aplicarse a cualquier número de relaciones. También, es una operación asociativa:

  $R \cup (S \cup T) = (R \cup S) \cup T$

- La operación de diferencia **no** es conmutativa, esto es:

  $R - S \neq S - R$

0.3 Operaciones derivadas

INTERSECCIÓN (∩)

La intersección entre dos relaciones R y S (de tipos compatibles), es el conjunto de todas las tuplas (t), tal que cada tupla t pertenece tanto a R como a S.
Ejemplo:

Autos

Marca	País_origen
V.W.	Alemania
Ford	E.U.A.
Renault	Francia
Nissan	Japón

Autos_compactos

Marca	País_origen
V.W.	Alemania
Nissan	Japón
Renault	Francia

Autos ∩ Autos_compactos

Marca	País_origen
V.W.	Alemania
Nissan	Japón
Renault	Francia

Características y propiedades de la intersección

- La operación de intersección $R \cap S$ es equivalente a R-(R-S)
- La intersección es conmutativa:

  $R \cap S = S \cap R$

- La intersección puede aplicarse a cualquier número de relaciones. También, es una operación asociativa:

  $R \cap (S \cap T) = (R \cap S) \cap T$

OPERACIÓN DE JUNTA (REUNIÓN O JOIN)

La operación de junta, denotada por $\bowtie$, es usada para combinar *tuplas relacionadas* de dos relaciones. Esta operación es muy importante para cualquier base de datos relacional ya que nos permite procesar vínculos entre relaciones.

La forma general de la operación de junta sobre dos relaciones $R(A_1, A_2, ..., A_n)$ y $S(B_1, B_2, ..., B_m)$ es

$$R \bowtie_{<cond\text{-}junta>} S$$

El resultado de la junta es una relación Q con n+m atributos $Q(A_1, A_2, ..., A_n, B_1, B_2, ..., B_m)$ en ese orden. Q tiene una tupla por cada combinación de tuplas -una de R y una de S- *que satisface la condición de junta.*

La condición de junta es de la forma:

<cond> AND <cond> AND ... AND <cond>

donde cada *cond* es de la forma $A_i \ \theta \ B_j$, con A_i como atributo de R, B_j atributo de S y θ un operador de comparación (<, <=, etc.).

La operación de junta es una abreviación para $\sigma_{<cond\text{-}junta>}$ (R X S).

Ejemplo: R(A,B,C)={<1,2,3>, <4,5,6>, <7,8,9>} y S(D,E)={<3,1>, <6,2>}.
entonces: $R \bowtie_{B<D} S$ (A,B,C,D,E)= {<1,2,3,3,1>, <1,2,3,6,2>, <4,5,6,6,2>}.

Tuplas cuyos atributos de junta son nulos *no aparecen* en el resultado.

Una operación de junta donde sólo aparece el operador = en la condición, se conoce como **equijunta (equijoin)**.

OPERACIÓN DE JUNTA NATURAL

Es una operación de equijunta la cual se denota por $\bowtie$ o por *. La definición estándar de junta natural requiere que cada par de atributos de junta (o sea, aquellos que se van a comparar para ver si satisfacen la condición de junta) tengan el mismo nombre. La junta natural se ejecuta *igualando todos* los pares de atributos que tienen el mismo nombre en ambas relaciones, desechando los atributos repetidos de la segunda relación.

La forma general de la operación de junta natural sobre dos relaciones R y S es (flexibilizando la notación del álgebra relacional):

$$Q \leftarrow R * S = R \bowtie_{R.atr1=S.atr1 \text{ AND } \ldots \text{ AND } R.atrj=S.atrj} S$$

con al menos un atributo en cada relación teniendo exactamente el mismo nombre.

Ejemplo: R(A,B,C) ={<1,2,3>, <4,5,6>, <7,8,9>} y S(C,E)={<3,1>, <6,2>}.
entonces: Q(A,B,C,E) = R * S = {<1,2,3,1>, <4,5,6,2>}.

Características de la operación:

- Los atributos que se comparan entre sí deben tener el mismo nombre en ambas relaciones (por lo tanto, no se debe hacer el cambio del nombre de atributo como sucede con la junta normal o el producto cartesiano).
- En la relación resultante sólo se conserva una ocurrencia de los atributos duplicados (además de los atributos que tienen nombres distintos).
- La condición de junta puede tener varias expresiones de comparación unidas *exclusivamente* por el operador AND. También, en cada comparación los operandos sólo pueden compararse por igualdad.

0.4 EXTENSIÓN DEL ÁLGEBRA RELACIONAL

FUNCIONES DE TOTALES (O DE AGREGADOS O DE ESTADÍSTICAS)

El primer tipo de solicitud que no puede ser expresado en el álgebra relacional es el de especificar **funciones de totales** sobre colecciones de valores de la base de datos. Un ejemplo podría ser calcular promedios o sumatorias.

Otro tipo de solicitud que no se puede hacer es el de agrupar las tuplas de una relación por el valor de alguno de sus atributos y aplicarles alguna de estas funciones.

Para manejar este tipo de solicitudes se extiende el álgebra relacional incluyendo una operación especial para poder especificar estas funciones. La operación *Función* se usará para tal fin y se denotará con el símbolo $\mathcal{F}$, aunque la notación no es estándar. Esta extensión se toma en cuenta debido a que la mayoría de los lenguajes relaciones de consulta comerciales incluyen recursos para ejecutar este tipo de solicitudes.

La sintaxis de esta operación es:

<atribs. de agrupación> $\mathcal{F}$ <lista de funciones> (<nombre de relación>)

donde:

<atribs. de agrupación> - es una lista de atributos de <nombre de relación>.
<lista de funciones> - es una lista de pares: <función> <atributo>, siendo <función> una de las funciones permitidas, como: Sum, Avg, Max, Min, Count; y <atributo> uno de los atributos de <nombre de relación>.

La relación resultante tiene a los atributos de agrupación más un atributo por cada elemento en la lista de funciones.

Si no se especifican atributos de agrupación, las funciones se aplican a los atributos de *todas las tuplas* en la relación, tal que la relación resultante *tiene una sola tupla.*

Hay que enfatizar que el resultado de aplicar una función de totales es una relación, no un escalar aunque el resultado sea un solo valor.