Álgebra relacional

El álgebra relacional es un lenguaje de consulta *procedimental*. Consta de un conjunto de operaciones que toma una o dos relaciones de entrada y genera una nueva relación como resultado.

Definiciones

Tupla
En terminología matemática, Es una secuencia (a lista) de valores. Una relación entre n valores se representa matemáticamente como un a n-tupla, es decir, una tupla con n valores, que se corresponde con una fila de una tabla.

Relación
Una tupla representa una relación entre un conjunto de valores, entonces la relación es una colección de tuplas, el término relación es equivalente a la tabla de una base de datos. De este término toma su nombre el modelo de datos relacional.

Cardinalidad: En matemáticas y bases de datos, la cardinalidad se refiere al número de elementos que contiene un conjunto. En este contexto se refiere a la cantidad de atributos.

Dominio: es el conjunto de valores atómicos que puede tomar un atributo. Es decir, es el conjunto de valores válidos que puede tener una columna específica en una tabla relacional. Este concepto es equivalente al tipo de datos del atributo.

Operaciones

Las operaciones fundamentales del álgebra relacional son:

1. Selección
2. Proyección
3. Renombramiento
4. Unión
5. Diferencia de conjuntos
6. Producto cartesiano

Las operaciones de selección, proyección y renombramiento, se denominan **operaciones unarias** por que operan sobre una sola relación. Por otra parte, las operaciones Unión, diferencia y producto cartesiano, se denominan **operaciones binarias** porque operan sobre pares de relaciones.

Selección σ:

La operación selección selecciona tuplas que satisfacen un predicado dado, se usa la letra griega sigma minúscula (σ) para denotar la selección. El predicado aparece como subíndice de σ. La relación de argumentos se da entre paréntesis a continuación del predicado.

El término selección en álgebra relacional tiene un significado diferente al utilizado en SQL, la operación selección en álgebra relacional corresponde en SQL a la cláusula WHERE.

Tomemos cómo ejemplo la relación profesores:

Profesores			
ID	**Nombre**	**Departamento**	**Sueldo**
10101	Srinivasan	Informática	65000
12121	Wu	Finanzas	90000
15151	Mozart	Música	40000
22222	Einstein	Física	95000
32343	El Said	Historia	60000
33456	Gold	Física	87000
45565	Katz	Informática	75000
58583	Califieri	Historia	62000
76543	Singh	Finanzas	80000
76766	Crick	Biología	72000
83821	Brandt	Informática	92000
98345	Kim	Electrónica	80000

Para seleccionar las tuplas de la relación profesores en las que el profesor pertenece al departamento de física se escribe:

$$\sigma_{\text{departamento = Física}}(\text{profesores})$$

El resultado sería:

$\sigma_{\text{departamento = Física}}(\text{profesores})$			
ID	**Nombre**	**Departamento**	**Sueldo**
22222	Einstein	Física	95000
33456	Gold	Física	87000

Buscar todos los profesores cuyos sueldos sea superior a 90000

$$\sigma_{\text{sueldo} > 90000}(\text{profesores})$$

En el predicado se permiten comparaciones =, ≠, <, <=, >, >=, combinaciones de varios predicados con AND (∧), OR (V), Negación NOT (¬)

Para encontrar las tuplas correspondientes a los profesores de física con un sueldo mayor a 90000 se escribe:

$$\sigma_{\text{departamento} = \text{Física} \wedge \text{sueldo} > 90000}(\text{profesores})$$

El resultado sería

$\sigma_{\text{departamento} = \text{Física} \wedge \text{sueldo} > 90000}(\text{profesores})$			
ID	**Nombre**	**Departamento**	**Sueldo**
22222	Einstein	Física	95000

Proyección π:

La operación de proyección devuelve la relación mostrando sólo los atributos listados, se indica con la letra griega pi (π), como subíndices de π se crea una lista de los atributos separados por coma y la relación se escribe a continuación entre paréntesis. Debido a que las relaciones son conjuntos, **se eliminan todas las tuplas duplicadas**.

Para listar el id, nombre y sueldo de la relación profesor tendíamos:

$$\pi_{\text{id, nombre, sueldo}}(\text{profesores})$$

El resultado sería

ID	Nombre	Sueldo
10101	Srinivasan	65000
12121	Wu	90000
15151	Mozart	40000
22222	Einstein	95000
32343	El Said	60000

33456	Gold	87000
45565	Katz	75000
58583	Califieri	62000
76543	Singh	80000
76766	Crick	72000
83821	Brandt	92000
98345	Kim	80000

Renombramiento ρ:

Resulta útil poder ponerles nombre a los resultados de las expresiones del álgebra relacional; la operación renombramiento, que se indica con la letra griega *ro* minúscula (ρ), permite hacerlo.

Dada una expresión E del álgebra relacional. La expresión $\rho_x(E)$ Devuelve el resultado de la expresión E con el nombre X.

Las relaciones r, por sí mismas, se consideran expresiones (triviales) del álgebra relacional. Por tanto, también se puede aplicar la operación renombramiento a una relación r para obtener la misma relación con un nombre nuevo.

Una segunda forma de la operación renombramiento es la siguiente. Suponga que una expresión del álgebra relacional E tiene cardinalidad n. Entonces, la expresión:

$$\rho_{x(A1', A2', \ldots, An)}(E)$$

Devuelve el resultado de la expresión E con el nombre de x y con los atributos con el nombre cambiado a A1, A2, An.

Composición de operaciones relacionales

Es importante el hecho de que el resultado de una operación relacional sea también una relación. Considere la consulta más compleja «Buscar el nombre de todos los profesores del departamento de Física. Hay que escribir:

$$\pi_{name}(\sigma_{nombre\_dept\ =\ Física}(profesores))$$

Téngase en cuenta que, en vez de dar el nombre de una relación como argumento de la operación proyección, se da una expresión cuya evaluación es una relación.

En general, dado que el resultado de las operaciones del álgebra relacional es del mismo tipo (relación) que los datos de entrada, las operaciones del álgebra relacional pueden componerse para formar una expresión del álgebra relacional.

Componer operaciones del álgebra relacional para formar expresiones del álgebra relacional es igual que componer operaciones aritméticas (como +, −, * y /) para formar expresiones aritméticas.

Unión ∪:

Considere una consulta para buscar el conjunto de todas las asignaturas que se enseñaron en el semestre 2 de 2009, el semestre 1 de 2010 o en ambos. La información se encuentra en la relación secciones.

Secciones						
asignatura_id	**secc_id**	**semestre**	**año**	**edificio**	**número_aula**	**franja_horaria_id**
BIO-101	1	1	2009	Painter	514	B
BIO-301	1	1	2010	Painter	514	A
CS-101	1	2	2009	Packard	101	H
CS-101	1	1	2010	Packard	101	F
CS-190	1	1	2009	Taylor	3128	E
CS-190	2	1	2009	Taylor	3128	A
CS-315	1	1	2010	Watson	120	D
CS-319	1	1	2010	Watson	100	B
CS-319	2	1	2010	Taylor	3128	C
CS-347	1	2	2009	Taylor	3128	A
EE-181	1	1	2009	Taylor	3128	C
FIN-201	1	1	2010	Packard	101	B
HIS-351	1	1	2010	Painter	514	C
MU-199	1	1	2010	Packard	101	D
PHY-101	1	2	2009	Watson	100	A

Para encontrar el conjunto de todas las asignaturas que se enseñaron en el semestre 2 de 2009 se escribe:

$$\pi_{\text{asignatura\_id}}(\sigma_{\text{semestre} = 2 \wedge \text{año} = 2009}(\text{secciones}))$$

Para encontrar el conjunto de todas las asignaturas que se dictaron en el semestre 1 de 2010 se escribe:

$$\pi_{\text{asignatura\_id}}(\sigma_{\text{semestre} = 1 \wedge \text{año} = 2010}(\text{secciones}))$$

Para contestar a la consulta es necesaria la unión de estos dos conjuntos; es decir, hacen falta todos los ID de sección que aparecen en alguna de las dos relaciones o en ambas. Estos datos se pueden obtener mediante la operación binaria unión, que se indica como en la teoría de conjuntos, por U. Por tanto, la expresión buscada es:

$$\pi_{\text{asignatura\_id}}(\sigma_{\text{semestre} = 2 \wedge \text{año} = 2009}(\text{secciones})) \cup \pi_{\text{asignatura\_id}}(\sigma_{\text{semestre} = 1 \wedge \text{año} = 2010}(\text{secciones}))$$

El resultado sería:

asignatura_id
CS-101
CS-315
CS-319
CS-347
FIN-201
HIS-351
MU-199
PHY-101

Tenga en cuenta que en el resultado hay ocho tuplas, aunque haya tres asignaturas distintas ofertadas en el semestre 2 de 2009 y seis asignaturas distintas ofertadas en el semestre 1 de 2010. Como las relaciones son un conjunto, los valores duplicados como CS-IOI, que se ofrece en ambos semestres, aparecen una sola vez.

Observe que en este ejemplo se toma la unión de dos conjuntos, ambos consistentes en valores de *asignatura_id*. En general, se debe asegurar que las uniones se realicen entre relaciones compatibles. Por ejemplo, no tendría sentido realizar la unión de las relaciones profesor y estudiante. Aunque ambas tengan cuatro atributos, difieren en los dominios de *sueldo* y *tot_créditos*. La unión de estos dos atributos no tendría sentido en la mayor parte de los casos.

Por tanto, para que la operación unión r U s sea válida hay que exigir que se cumplan dos condiciones:

1. Las relaciones r y s deben ser de la misma cardinalidad; es decir, deben tener el mismo número de atributos.
2. Los dominios de los atributos i-ésimos de r y de s deben ser iguales para todo i.

Tenga en cuenta que r y s pueden ser, en general, relaciones de la base de datos o relaciones temporales resultado de expresiones del álgebra relacional.

Diferencia de conjuntos −:

La operación diferencia de conjuntos, indicada por −, permite encontrar las tuplas que están en una relación, pero no en la otra. La expresión r − s da como resultado una relación que contiene las tuplas que están en r pero no en s.

Se pueden buscar todas las asignaturas que se enseñaron en el semestre del otoño de 2009 pero no en el semestre de la primavera de 2010, escribiendo:

$\pi_{\text{asignatura\_id}}(\sigma_{\text{semestre} = 2 \wedge \text{año} = 2009}(\text{secciones})) - \pi_{\text{asignatura\_id}}(\sigma_{\text{semestre} = 1 \wedge \text{año} = 2010}(\text{secciones}))$

El resultado sería

asignatura_id
CS-347
PHY-101

Como en el caso de la operación unión, hay que asegurarse de que las diferencias de conjuntos se realicen entre relaciones compatibles, por tanto, para que una operación diferencia de conjuntos r - s sea válida se exige que las relaciones r y s sean de la misma cardinalidad y que los dominios de los atributos i-ésimos de r y de s sean iguales, para todo i.

Producto cartesiano X:

La operación producto cartesiano, que se indica mediante un aspa X, permite combinar información de dos relaciones cualesquiera. El producto cartesiano de las relaciones r_1 y r_2 se escribe r_1 x r_2.

Recuerde que las relaciones se definen como subconjuntos del producto cartesiano de un conjunto de dominios. A partir de esa definición ya debería tener una idea intuitiva sobre la definición de la operación producto cartesiano. Sin embargo, dado que el mismo nombre de atributo puede aparecer tanto en r_1 como en r_2, es necesario crear un convenio de denominación para distinguir unos atributos de otros. En este caso se realiza adjuntando al atributo el nombre de la relación de la que proviene originalmente.

Por ejemplo, si tenemos la relación cursos

cursos				
id	**asignatura_id**	**secc_id**	**semestre**	**año**
10101	CS-101	1	2	2009
10101	CS-315	1	1	2010
10101	CS-347	1	2	2009
12121	FIN-201	1	1	2010
15151	MU-199	1	1	2010
22222	PHY-101	1	2	2009
32343	HIS-351	1	1	2010
45565	CS-101	1	1	2010
45565	CS-319	1	1	2010
76766	BIO-101	1	1	2009
76766	BIO-301	1	1	2010
83821	CS-190	1	1	2009
83821	CS-190	2	1	2009
83821	CS-319	2	1	2010
98345	EE-181	1	1	2009

El esquema de relación r = profesores x cursos es:

```
(profesores.ID, profesores.nombre, profesores.nombre_dept,
profesores.sueldo cursos.ID, cursos.asignatura_id, cursos.secc_id,
cursos.semestre, cursos.año)
```

Con este esquema se puede distinguir entre profesores.ID y cursos.ID. Para los atributos que solo aparecen en uno de los dos esquemas se suele omitir el prefijo con el nombre de la relación. Esta simplificación no genera ambigüedad alguna. Por tanto, se puede escribir el esquema de la relación r como:

```
(profesores.ID, nombre, nombre_dept, sueldo, cursos.ID,
asignatura_id, secc_id, semestre, año)
```

Este convenio de denominaciones exige que las relaciones que sean argumentos de la operación producto cartesiano tengan nombres diferentes. Esta exigencia causa problemas en algunos casos, como cuando se desea calcular el producto cartesiano de una relación consigo misma. Se produce un problema parecido si se usa el resultado de una expresión del álgebra relacional en un producto cartesiano, dado que hará falta un nombre de relación para poder hacer referencia a sus atributos. Para estos casos, se utiliza la operación renombramiento.

Ahora que se conoce el esquema de relación de r profesor x cursos es necesario encontrar las tuplas que aparecerán en r.

Como es posible imaginar, se crea una tupla de r a partir de cada par de tuplas posible: una de la relación profesores y otra de la relación de cursos. Por tanto, r es una relación de gran tamaño, como se puede ver a continuación:

prof_ID	nombre	nombre_dept	sueldo	enseña_ID	asignatura_id	secc_id	semestre	año
10101	Srinivasan	Física	95000	10101	CS-101	1	Otoño	2009
10101	Srinivasan	Física	95000	10101	CS-315	1	Primavera	2010
10101	Srinivasan	Física	95000	10101	CS-347	1	Otoño	2009
10101	Srinivasan	Física	95000	10101	FIN-201	1	Primavera	2010
10101	Srinivasan	Física	95000	15151	MU-199	1	Primavera	2010
10101	Srinivasan	Física	95000	22222	PHY-101	1	Otoño	2009
12121	Wu	Física	95000	10101	CS-101	1	Otoño	2009
12121	Wu	Física	95000	10101	CS-315	1	Primavera	2010
12121	Wu	Física	95000	10101	CS-347	1	Otoño	2009
12121	Wu	Física	95000	10101	FIN-201	1	Primavera	2010
12121	Wu	Física	95000	15151	MU-199	1	Primavera	2010
12121	Wu	Física	95000	22222	PHY-101	1	Otoño	2009
15151	Mozart	Física	95000	10101	CS-101	1	Otoño	2009
15151	Mozart	Física	95000	10101	CS-315	1	Primavera	2010
15151	Mozart	Física	95000	10101	CS-347	1	Otoño	2009
15151	Mozart	Física	95000	10101	FIN-201	1	Primavera	2010
15151	Mozart	Física	95000	15151	MU-199	1	Primavera	2010
15151	Mozart	Física	95000	22222	PHY-101	1	Otoño	2009
22222	Einstein	Física	95000	10101	CS-101	1	Otoño	2009
22222	Einstein	Física	95000	10101	CS-315	1	Primavera	2010
22222	Einstein	Física	95000	10101	CS-347	1	Otoño	2009
22222	Einstein	Física	95000	10101	FIN-201	1	Primavera	2010
22222	Einstein	Física	95000	15151	MU-199	1	Primavera	2010
22222	Einstein	Física	95000	22222	PHY-101	1	Otoño	2009
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...

Por lo extensa de la relación solo se ha incluido una parte de las tuplas que constituyen r.

Referencias:

- SILBERSCHATZ, Abraham, KORTH, Henry, SUDARSHAN, S. Fundamentos de Bases de Datos. Sexta Edición. McGraw-Hill, 2014.